

Universidad de Granada
Facultad de Filosofía y Letras
Departamento de Lengua Española

La evaluación oral:
una equivalencia entre las *guidelines* de ACTFL y
algunas escalas del MCER

Autor:

Alfonso Martínez Baztán

Director:

Dr. Miguel Calderón Campos

Granada, 2008

A Rosana.
Y a nuestra hija,
Índigo.

Amanecía y Scherezade guardó silencio
(*Las mil y una noches*. Anónimo).

Hace más de 26 años se produjeron, creo que azarosamente, dos elementos en principio diversos pero que han sido para mi vida importantes y han terminado por ir juntos, siendo el uno representante simbólico del otro. ACTFL (el Consejo americano para la enseñanza de lenguas extranjeras) promovió su test oral para evaluar el uso de la habilidad funcional de las lenguas, OPI (*Oral Proficiency Interview*), test que –aunque no exento de críticas- sería el modelo a seguir e imitar por casi todos los tests orales que desde entonces se crearon. Y al mismo tiempo, y probablemente muy cerca, el mismo año, 1982, Ridley Scott se decidía a llevar al cine la versión del libro *¿Sueñan los androides con ovejas eléctricas?* de Philip K. Dick, haciendo de él una magnífica película que rápidamente se convertiría en pieza de culto: *Blade Runner*.

La película, con la fascinante, envolvente y a veces melancólica música de Vangelis y esas inolvidables escenas en que Deckard (Harrison Ford) conoce a Rachel (Sean Young) o en la que esta se presenta en su casa para averiguar cuánta vida le queda, me llegó el mismo año de su estreno y no me ha abandonado ya. En ACTFL fui yo mismo el que entró, bastantes años más tarde, después de que en mi primera estancia en la Universidad de Utrecht en 1992 conociera su sistema de evaluar la lengua, la OPI, y quedara intrigado por su capacidad de hacerlo adaptándose a sus candidatos al mismo tiempo que con objetividad. A menudo, cuando evaluó a alguien, tengo el grato recuerdo de la escena en que Deckard lo hace con su candidata, Rachel, haciendo gala de su gran profesionalidad.

Desde entonces, y con el recuerdo del test y de la maquina que Deckard usa para ver el iris ampliado de los ojos de Rachel y analizar sus reacciones emocionales, no he dejado de plantearme si nuestros tests y sistemas de acercarnos a la lengua y evaluarla podrían ser tan objetivos y fiables como para estar seguros de que en nuestras notas y calificaciones no estábamos errando y condenando a los hablantes o estudiantes a una

situación inmerecida. O si a nuestra profesión se le podría exigir el mismo tipo de garantías y responsabilidades que a un médico, del que depende la vida de sus pacientes, a un arquitecto, que debe levantar edificios habitables que se sostengan, o a un ingeniero, que diseña puentes que aguantan movimiento, vibraciones y peso. Y siempre pensando que sí, o que al menos debemos ser conscientes de nuestras responsabilidades, posibilidades y limitaciones.

De ACTFL y la evaluación oral de la lengua española, así como de las formas de evaluar lenguas y las posibles equiparaciones entre el sistema de la OPI y el MCER (*Marco común europeo de referencia*) van a tratar las páginas de esta tesis. Sin embargo, la escena fundamental de la película *Blade Runner*, donde se despliegan las habilidades de evaluador de su protagonista, no va a tener -por razones obvias- cabida entre sus páginas. Pero quiero traerla ahora, aquí en esta presentación, con el deseo de que sirva de ayuda para entender la necesidad de plantear el tema de la evaluación oral de lenguas desde una postura seria que requiere honestidad e inteligencia para ser llevada a buen término.

Vamos a verla. Es la escena en la que Deckard conoce a Rachel. Aunque la cita es larga, es clarificadora. En ella, inesperadamente, averiguamos con Deckard que Rachel, de quien él se va a enamorar (y quizás nosotros mismos), no es humana. Ello la condena a muerte (qué paradaja) y a él a ser el juez de su destino (tragedia ante la que, humanamente, se revela).

Rachel: ¿Puedo hacerle una pregunta personal?

Deckard: Claro.

Rachel: ¿Nunca ha retirado a un humano por error?

Deckard: No.

Rachel: En su posición, eso es un riesgo...

(Aparece el doctor Tyrell cortando la conversación).

Tyrell: ¿Esto va a ser un test de empatía? Dilatación capilar por las así llamadas respuestas ruborizantes; fluctuación de la pupila; dilatación involuntaria del iris...

Deckard: Nosotros lo llamamos Voight Kampff, para abreviar.

Rachel: Señor Deckard, el doctor Eldon Tyrell.

Tyrell: Demuéstrelo. Quiero verlo trabajar.

Deckard: ¿Dónde está el sujeto?

Tyrell: Quiero verlo funcionar en una persona. Quiero ver una negativa antes de proporcionarle una positiva.

Deckard: ¿Y eso qué va a demostrar?

Tyrell: ¿Quiere... complacerme?

Deckard: ¿Con usted?

Tyrell: Intente con ella.

Deckard: Hay demasiada luz.

(Una gran ventana se oscurece reduciendo sensiblemente la entrada de luz).

Rachel: ¿Le importa si fumo?

Deckard: Eso no afectará al test. Bien. Le voy a hacer una serie de preguntas. Relájese y contéstelas lo más sencillamente que pueda. Es su cumpleaños y le regalan una cartera de piel...

Rachel: No la aceptaría. Y, además, denunciaría a la policía a la persona que me la regalara.

Deckard: Tiene un hijo. Éste le enseña su colección de mariposas y un frasco de arsénico.

Rachel: Lo llevaría al médico.

Deckard: Está viendo la televisión. De repente, se da cuenta de que una avispa le sube por el brazo.

Rachael: La mataría.

Deckard: Está leyendo una revista y se encuentra con la fotografía de una mujer desnuda.

Rachel: ¿Este test es para saber si soy una replicante o una lesbiana?

Deckard: Conteste a las preguntas, por favor. Se la enseña usted a su marido y a éste le gusta tanto que la pone en la pared de su dormitorio. ¿Se enfadaría?

Rachel: No le dejaría.

Deckard: ¿No le dejaría? ¿Por qué no?

Rachel: Debo ser suficiente para él.

(El tiempo discurre en el despacho de Tyrell).

Deckard: Una pregunta más. Está viendo una obra de teatro. Tiene lugar un banquete en el que los invitados se deleitan con un aperitivo de ostras vivas. El primer plato consiste en perro cocido.

(La máquina de Voight Kampff emite una señal visual que sólo Deckard parece advertir).

Tyrell: ¿Quiere usted salir un momento, Rachel? Gracias.

(Rachel sale del despacho)

Deckard: Ella es una replicante, ¿no es así?

Tyrell: Estoy impresionado. ¿Cuántas preguntas son las normales para detectar a uno?

Deckard: No le comprendo, Tyrell.

Tyrell: ¿Cuántas... preguntas?

Deckard: Veinte, treinta..., según el tipo.

Tyrell: Le ha costado más de cien con Rachel, ¿no es así?

Deckard: ¿Ella no lo sabe?

Tyrell: Está empezando a sospechar, creo.

Deckard: ¿Sospechar? ¿Cómo puede no saber lo que es?

Tyrell: El comercio es nuestro objetivo aquí, en la Tyrell. Y nuestro lema, «más humanos que los humanos». Rachel es un experimento, nada más.

(Blade Runner, Ridley Scott, 1982).

Como vemos, un magnífico ejemplo que plantea la posibilidad de realizar tests fiables, como el que Deckard utiliza para averiguar si su candidato, Rachel, es un robot, una “replicante”. La objetividad mostrada para averiguarlo es la garantía de poder eliminar al “replicante” sin problemas, pues si fallara la identificación, el hecho se convertiría en un acto de consecuencias terribles: el asesinato de un ser humano. Pero la asombrosa habilidad del evaluador no deja lugar a dudas, como se comprueba en la

escena, ante el necesario e inusual despliegue de más de cien preguntas para llegar a una conclusión, cuando lo normal hubieran sido unas veinte o treinta.

El peligro de la evaluación oral no es tan grande para nosotros los profesores de lenguas ni para las instituciones educativas que utilizan tests, pero la responsabilidad no es menor. Y muchas veces las consecuencias de una nota, para los candidatos, también pueden ser determinantes para sus vidas. Aunque no con la importancia tan evidente de los médicos, arquitectos o ingenieros, también sobre la práctica de la evaluación oral, como veremos más adelante, han recaído históricamente responsabilidades como las de decidir aspectos de la vida, nacionalización o trabajo de muchas personas inmigrantes. Y estos aspectos vuelven a estar hoy en día en acción cuando se plantean o determinan objetivos y niveles de cursos para acceder a puestos en universidades o trabajos en una Europa multicultural.

Esta tesis no va a abordar estos aspectos sociales de la evaluación, ni sus consecuencias, a las que cada vez, afortunadamente, se presta más atención, conscientes de que es labor que requiere de códigos de buenas prácticas que otorguen derechos a sus participantes y reglas deontológicas a sus evaluadores y administradores. Pero sí nos hemos planteado la evaluación desde una posición de respeto hacia los participantes y sus necesidades, respeto sin el que creemos que es imposible la buena práctica que se necesita no sólo para no herir a los candidatos ni malinterpretar sus actuaciones y discursos, sino para lograr que los discursos que se generen sean representativos y efectivamente evaluables.

No puedo terminar esta presentación sin mostrar mi agradecimiento a todas las personas que han hecho posible que lleve a término este trabajo que, gracias a su ayuda, me ha resultado apasionante. Ariadna tiene, por suerte para mí, muchos hilos. Todos ellos son importantes, algunos, además, imprescindibles.

En primer lugar a cuatro personas que se merecen todo mi afecto y amistad.

A Rosana Uriz, mi compañera durante 25 años, por haberme centrado durante todo este tiempo y sin cuya existencia estaría, sin duda, perdido. Y por su incansable ayuda y comentarios durante la realización de esta tesis, lo cual me ha desatascado en numerosas ocasiones, así como por las varias revisiones del documento. Ella sabe que se necesitarían poemas, y no estas breves palabras, para cuantificar algo de esto.

A nuestra hija Índigo Uriz, por su interés y aliento, por su comprensión y paciencia durante todas las horas que no ha podido usar el ordenador ni disponer de su padre, y por la alegría que le daba cada vez que terminaba el apartado de un capítulo o un capítulo. Dar la vida a un ser que sabe estar vivo es una inmensa alegría, Índigo.

A un buen amigo, además director de esta tesis, el Dr. Miguel Calderón, por su buena disposición y excelente guía que me han servido de aprendizaje y de inestimable ayuda, y por todo el trabajo que ha realizado desde que me convenció para que abordara este camino de investigación académica y que ha sido rico en consejos, análisis, comentarios y correcciones, siempre respetuosos, inteligentes y certeros.

Y a la Dra. Maite García Godoy, amiga antes que nada, sin cuyo interés y poder de convicción este trabajo no habría hecho ni empezar, así como por la buena compañía de su perspicaz humor.

Los cuatro saben que me han hecho sentirme mimado, quizás en exceso. Considero que he tenido la envidiable suerte de poder sentirlos a mi lado. Desde aquí les expreso todo mi cariño y gratitud, que no termina en estas líneas.

Asimismo, la suerte podría llamarse Ariadna, también debo y quiero agradecer la elaboración de este trabajo a muchas otras personas que voluntaria y amablemente han puesto parte de su tiempo e inteligencia a su servicio:

a los profesores del CLM que me ayudaron en la investigación con sus opiniones y datos, Aurora Biedma, Susana Díaz, María Estévez, Gracia Lozano, M^a Ángeles Martínez, Manolo Mayor, Luisa Molina, Encarna Morales, M^a José Pérez, Inmaculada Ramírez, Carmen Ruiz, Adolfo Sánchez, Rosana Uriz y Pilar Vela; a los estudiantes de mis cursos de lengua en el CLM, que fueron voluntarios participantes en la investigación; a los participantes directos en las entrevistas para evaluación oral realizadas en este trabajo, estudiantes de español que -con entusiasmo y alegría- se prestaron a esa evaluación oral mediante dos entrevistas, personas sin cuya ayuda ninguna de las conclusiones de esta tesis se habrían podido escribir; a los estudiantes del III Máster de ELE de la Universidad de Granada y del Máster de ELE de la UIMP, que

participaron en varios análisis y encuestas importantes para el desarrollo de esta tesis y de los que salió uno de los equipos de evaluación formado para calificar un grupo de entrevistas; a muchos otros participantes en los talleres de evaluación que imparto (en el Instituto Cervantes, en diversas Escuelas Oficiales de Idiomas, en la Universidad de Deusto o en la UIMP, entre otros) que siempre me han alentado y han colaborado en algunas de las mini-investigaciones que he tenido que llevar a cabo para comprobar y ajustar datos; a María Luisa Moreno y Antonio Orellana, por su ayuda con las traducciones de las *guidelines* de ACTFL; a Esther Uriz, por ayudarme con otras traducciones, en este caso al inglés; a María Dolores Delgado, miembro de ACTFL y evaluadora oficial de OPI, que se ha encargado de las calificaciones de las OPI; a Jesús López Mejías, profesor del Departamento de Psicología de la Universidad de Granada, que me facilitó enormemente las conclusiones del trabajo gracias a sus conocimientos de la magia y alquimia de los sistemas informáticos de análisis estadístico; a Kristi Jáuregui, por acogerme con tanto interés y cariño durante mi estancia en Holanda, en la Universidad de Utrecht; al servicio de bibliotecas de la Universidad de Granada por estar tan dispuesto y lleno de buena información, y al del Centro de Lenguas Modernas y sus bibliotecarias, especialmente a Maribel Díaz Guervós, por su ayuda siempre paciente e inestimable para obtener artículos y libros, y por localizarlos y conseguirlos tan pronta y cariñosamente, así como a Ángel Redondo, que también me guió en estas búsquedas; a Miguel Ángel Alegría por solucionarme, con sus dedos y mirar ubicuos, todos los problemas informáticos, allí cuando se planteaban, así como por la regrabación de todas las entrevistas al DVD que se adjunta a esta tesis; al Departamento de Lengua Española de la Universidad de Granada, por acogerme y ofrecerme su apellido, que espero saber llevar con dignidad y orgullo; y a mi familia entera, mis queridos madre, suegros, hermanos, tíos y cuñados, y a quienes me reconocen y tienen como amigo, ya que, además de darme o guardarme los momentos de vida que durante todo el tiempo que un trabajo de investigación como este no puedes disfrutar enteramente, han estado siempre al otro lado del hilo, por mail o teléfono, o más cerca cuando ha sido posible o necesario, con curiosidad e interés alentadores.

Ariadna, una vez más, que no tiene solo hilos, sino nombres. Todos ellos han contribuido a que esta tesis se pudiera realizar. A todos, mi agradecimiento más sincero.

Por supuesto ninguno de los fallos que se puedan encontrar en estas páginas se debe a otra persona sino a mí mismo.

ÍNDICE

RESUMEN	19
CAPÍTULO 1	
INTRODUCCIÓN	25
1.1 Explicación de cada capítulo	25
1.2 Introducción: algunas dificultades de la evaluación oral.....	29
1.3 Explicación de nuestro trabajo e hipótesis de investigación	35
1.4 Hipótesis de las equivalencias entre las escalas del MCER y las <i>guidelines</i> de ACTFL, para evaluación oral	38
CAPITULO 2	
ESTADO DE LA CUESTIÓN	41
2.1 Introducción a la evaluación de la lengua oral	42
2.2 Validación de tests a través del concepto de lengua utilizada en su constructo: el modelo de Bachman / Palmer.....	59
2.3 Técnicas cualitativas y cuantitativas de validación de tests	68
2.3.1 Validez de apariencia.....	76
2.3.2 Validez de respuesta	82
2.3.3 Validez de contenido	80
2.3.4 Validez de constructo	81
2.3.5 Validez concurrente.....	91
2.3.6 Validez predictiva.....	92
2.3.7 Uso práctico.....	93
2.3.8 Fiabilidad	94
2.3.9 Conclusión.....	97
2.4 Las características de una conversación natural y la evaluación oral.....	99
2.4.1 Planteamiento general.....	99
2.4.2 El papel del entrevistador en la génesis del discurso en las entrevistas orales	103
2.4.3. Otros elementos a considerar en la evaluación oral	107
2.4.3- 1) Elementos a reconsiderar	107
2.4.3- 2) Las intenciones del candidato	112
2.4.3- 3) El estrés comunicativo	114

2.4.3- 4) La formación de evaluadores	121
2.5 Sistemas de definir la capacidad lingüística: escalas.....	127
2.5.1 Tipos de escalas y sus características técnicas	127
2.5.2 La escala de evaluación oral de ACTFL	133
2.5.3 Las escalas del MCER.....	134
2.6 La entrevista oral de ACTFL.....	135
2.6.1 Introducción.....	137
2.6.2 Funcionamiento de la OPI	141
2.6.2- 1) Los principios de la escala de habilidades	141
2.6.2- 2) El constructo teórico de la lengua que se va a medir.....	147
2.6.2- 3) La estructura de la entrevista	159
2.6.2- 4) Los sistemas de elicitación de la lengua y el comportamiento del evaluador	170
2.6.2- 5) Las consideraciones necesarias para proceder a la calificación de las muestras obtenidas	181
2.6.3 Críticas al modelo de entrevista oral de ACTFL.....	185
2.6.3- A) Críticas sobre si la OPI es una conversación natural	186
2.6.3- B) Críticas sobre si la OPI es un test de competencia general o gramatical.....	188
2.6.3- C) Críticas sobre su valor como asesoramiento didáctico.....	191
2.6.3- D) Críticas sobre la validez y fiabilidad de la OPI.....	193
2.6.3- E) Críticas sobre si la escala se basa en criterios o en una norma.....	203
2.6.3- F) Conclusiones sobre las críticas al sistema de evaluar de AEI y a la OPI.	206
2.7 El Marco común europeo de referencia y la evaluación	207
2.7.1 Introducción.....	207
2.7.2 Contenidos del MCER.....	212
2.7.3 La evaluación dentro del MCER: el capítulo 9	223
2.7.4 Críticas y limitaciones del MCER.....	230
2.7.5 El documento para relacionar exámenes con los niveles del MCER	240
2.7.6 Conclusiones sobre el MCER.....	243
2.8 Equiparaciones propuestas entre los niveles del MCER y las <i>guidelines</i> de ACTFL.....	244

CAPÍTULO 3

TRABAJO DE CAMPO. ANÁLISIS DE LAS ESCALAS DEL MCER Y ACTFL Y DISEÑO, VALIDACIÓN Y USO DE UN TEST ORAL PARA EVALUAR SEGÚN EL MCER.....

3.1 Análisis de las escalas del MCER	250
3.1.1 Análisis crítico de los niveles establecidos por el MCER	250
3.1.2 Análisis general de los descriptores y de las escalas del MCER.....	253
3.1.3 Análisis crítico de la escala de <i>Autoevaluación: expresión oral</i>	258
3.1.4 Análisis crítico de la escala de <i>Expresión oral en general</i>	262
3.1.5 Análisis crítico de la escala de <i>Monólogo sostenido: argumentación</i>	267
3.1.6. Análisis crítico de la escala de <i>Niveles cualitativos de referencia</i>	272
3.1.7 Experimento sobre la interpretabilidad de los niveles del MCER.....	279
3.1.7- A) Pruebas realizadas con expertos. Grupo primero.....	281
3.1.7- B) Pruebas realizadas con expertos. Grupo segundo	284
3.1.7- C) Pruebas realizadas con expertos. Grupo tercero.....	288

3.1.8 Conclusiones del experimento sobre la ambigüedad en el MCER.....	291
3.2 Comparación entre los niveles de las <i>guidelines</i> de ACTFL y los establecidos por las escalas del MCER.....	293
3.2.1 Comparación con la escala de <i>Niveles comunes de referencia: escala global</i>	294
3.2.2 Comparación con la escala de <i>Niveles cualitativos de referencia</i>	301
3.2.3 Comparación con la escala de <i>Autoevaluación: Interacción oral</i>	304
3.2.4 Conclusión de las comparaciones entre las <i>guidelines</i> de ACTFL y las escalas del MCER.....	308
3.2.4- A) Contenidos de las escalas del MCER según los criterios de evaluación oral de ACTFL	309
3.2.4- B) Comparación de los contenidos en las escalas del MCER según los criterios de evaluación oral de ACTFL.....	311
3.2.5 Conclusión general de las comparaciones entre las escalas de habilidad oral del MCER y las <i>guidelines</i> de evaluación oral de ACTFL	313
3.3 Explicación del trabajo de campo realizado con el test y los candidatos voluntarios	318
3.3.1 Entrevistas utilizadas	319
3.3.2 Selección de candidatos.....	321
3.3.3 Preevaluación de los candidatos	323
3.3.4 Grabación y almacenamiento de las entrevistas	324
3.3.5 Calificación y evaluación de las entrevistas	325
3.3.6 Planteamiento sobre el tipo de test a utilizar	327
3.4 Descripción de las tareas de que se compone la segunda entrevista, diseñadas según el Marco común europeo de referencia	328
3.4.1 Tarea 1: “Trabajo de canguro”	328
3.4.2 Tarea 2: “Receta de comida”	329
3.4.3 Tarea 3: “Opinión sobre actividad cultural”	330
3.4.4 Tarea 4: “Explicar ciertos aspectos del trabajo”	331
3.4.5 Tarea 5: “Argumentar sobre la guerra...”	332
3.4.6 Tarea 6: “Explicar el viaje o experiencia...”	335
3.4.7 Tarea 7: “Explicar las ventajas de estudiar...”	335
3.4.8 Tarea 8, “Explicar un cómic”	336
3.5 Precauciones tomadas para que los resultados de la segunda entrevista sean comparables con los de una OPI de ACTFL	339
3.6 Validación del test	343
3.6.1 Validación de las tareas y su tipo de lenguaje con expertos.....	344
3.6.1- a) Validación de la tarea 1: “Trabajo de canguro”	348
3.6.1- b) Validación de la tarea 2: “Receta de comida”	349
3.6.1- c) Validación de la tarea 3: “Opinión sobre actividad cultural”	350
3.6.1- d) Validación de la tarea 4: “Explicar ciertos aspectos del trabajo”	351
3.6.1- e) Validación de la tarea 5: “Argumentar sobre la guerra...”	352
3.6.1- f) Validación de la tarea 6: “Explicar el viaje o experiencia...”	353
3.6.1- g) Validación de la tarea 7: “Explicar las ventajas de estudiar...”	353
3.6.1- h) Validación de la tarea 8: “Explicar un cómic”	354
3.6.1- i) Validación de la tarea 9: “Indicar el camino en un plano”	355
3.6.2 Validación de las tareas y su lenguaje con estudiantes	356
3.6.2- a) Validación de la tarea 1: “Trabajo de canguro”	358
3.6.2- b) Validación de la tarea 2: “Receta de comida”	359
3.6.2- c) Validación de la tarea 3: “Opinión sobre actividad cultural”	359
3.6.2- d) Validación de la tarea 4: “Explicar ciertos aspectos del trabajo”	360

3.6.2- e) Validación de la tarea 5: “Argumentar sobre la guerra...”	360
3.6.2- f) Validación de la tarea 6: “Explicar el viaje o experiencia...”	361
3.6.2- g) Validación de la tarea 7: “Explicar las ventajas de estudiar...”	361
3.6.2- h) Validación de la tarea 8: “Explicar un cómic”	361
3.6.2- i) Validación de la tarea 9: “Indicar el camino en un plano”	362
3.6.3 Validación del test en su conjunto y de su fiabilidad	362
3.6.3- A) Fiabilidad	363
3.6.3- B) Validez de constructo	364
3.6.3- C) Autenticidad	366
3.6.3- D) Interactividad	367
3.6.3- E) Impacto	368
3.6.3- F) Uso práctico	371
3. 7 Ficha del nuevo test del MCER utilizado en esta investigación.....	371

CAPÍTULO 4

ANÁLISIS DE LOS RESULTADOS DE LAS ENTREVISTAS ORALES..... 373

4.1 Corpus de entrevistas utilizadas en esta investigación	374
4.2 Carencias y problemas detectados al calificar utilizando las escalas del MCER... ..	377
4.2.1- A) Ambigüedades de interpretación.....	378
4.2.1- B) La preponderancia de criterios formalistas en ciertas calificaciones	380
4.2.1- C) Ponderación	386
4.3 Análisis de las entrevistas.....	392
4.3.1 Entrevistas de menor nivel que A1 en el MCER.....	393
4.3.1- A) Entrevista con el MCER del candidato 35 (Kayo): menos que A1 ..	395
4.3.1- B) OPI del candidato 35 (Kayo): inicial medio.....	398
4.3.1 C) Conclusión	399
4.3.2 Entrevistas de nivel A1	399
4.3.2- A) Entrevista del MCER del candidato número 3 (Johanes): A1	400
4.3.2- B OPI del candidato número 3 (Johanes): inicial alto.....	402
4.3.2- C) Conclusión.....	403
4.3.3 Entrevistas de nivel A2.....	404
4.3.3- A) Entrevista del MCER del candidato 7 (Carmen): A2	405
4.3.3- B) OPI del candidato 7 (Carmen): intermedio bajo	408
4.3.3- C) Conclusión.....	409
4.3.4 Entrevistas de nivel B1	409
4.3.4- A) Entrevista del MCER del candidato 8 (Juan): B1	411
4.3.4- B OPI del candidato 8 (Juan): intermedio medio.....	415
4.3.4- C) Conclusión.....	416
4.3.5 Entrevistas de nivel B2	416
4.3.5 -1) Entrevistas de candidatos de B2 que son intermedio alto en ACTFL	419
4.3.5 -1 A Entrevista del MCER del candidato 14 (Michelle): B2	419
4.3.5- 1 B) OPI del candidato 14 (Michelle): intermedio alto.....	422
4.3.5- 2) Entrevistas de candidatos de B2 que es avanzado bajo en ACTFL... ..	423
4.3.5- 2 A) Entrevista del MCER del candidato 20 (Nina)	424
4.3.5- 2 B) OPI del candidato 20 (Nina): avanzado bajo	427
4.3.5- 3) Conclusión	428
4.3.6 Entrevistas de nivel C1	428
4.3.7 Entrevistas de nivel C2	429
4.3.7- A) Entrevista del MCER del candidato 33 (Camila): C2.....	431

4.3.7- B) OPI del candidato 33 (Camila): superior.....	432
4.3.7- C) Conclusión.....	432
4.4 Casos raros.....	433
4.4.1 Entrevistas del candidato inicial medio en ACTFL, que es A1.....	434
4.4.1- A) Entrevista del MCER del candidato número 4 (María): A1	435
4.4.1- B) OPI del candidato número 4 (María): inicial medio	436
4.4.2 Entrevistas de candidatos de nivel B1 que son intermedio bajo en ACTFL ...	437
4.4.2- A) Entrevista del MCER del candidato 5 (Francesca): B1	438
4.4.2- B) OPI del candidato 5 (Francesca): intermedio bajo	439
4.4.3 Entrevistas de candidatos de B2 que son intermedio medio en ACTFL.....	441
4.4.3- 1 A) Entrevista del MCER del candidato 15 (María): B2.....	441
4.4.3- 1 B) OPI del candidato 15 (María): intermedio medio	443
4.4.3- 2 A) Entrevista del MCER del candidato 19 (Maya): B2	445
4.4.3- 2 B) OPI del candidato 19 (Maya): intermedio medio.....	448
4.4.4 Entrevistas del candidato de nivel C1 que es superior en ACTFL.....	449
4.4.4- A) Entrevista del MCER del candidato 29 (María): C1	450
4.4.4- B) OPI del candidato 29 (María): superior.....	453
4.5 Resultados estadísticos: explicaciones y comentarios.....	455
4.5.1 Análisis estadísticos.....	455
4.5.2 Equiparaciones entre los resultados obtenidos con las escalas del MCER y ACTFL	457
4.5.3 Conclusiones de la equiparación para una evaluación para certificación de dominio general	463
4.5.4 Conclusiones de la equiparación para una evaluación para usos pedagógicos	465
4.5.5 Comparación de estos resultados con otras equiparaciones existentes	467
4.5.5- A) Comparación con la equiparación de Tschirner.....	468
4.5.5- B) Comparación con la equiparación de Buitrago	469
4.5.5- C) Comparación con nuestra propia opinión cualitativa.....	470
CAPITULO 5	
CONCLUSIONES	477
5.1 Conclusiones.....	477
5.2 Discusión de las conclusiones	483
5.2.1 Las escalas de habilidad oral del MCER desde el punto de vista de la funcionalidad descrita en las <i>guidelines</i> orales de ACTFL	485
5.2.2 Equiparaciones entre nuestra evaluación oral según el MCER y las <i>guidelines</i> de ACTFL, desde el punto de vista pedagógico	486
5.2.3 Posibilidades del MCER para desarrollos académicos	488
5.2.4 Los niveles efectivos del MCER	491
5.2.5 Diferencias entre el MCER y la OPI	492
5.3 Limitaciones de nuestra investigación y propuestas de trabajos futuros.....	493
EPÍLOGO	497
APÉNDICES	503
APÉNDICE 1	
CARACTERÍSTICAS DEL TEST UTILIZADO EN ESTA INVESTIGACIÓN	503

APÉNDICE 2	
<i>GUIDELINES</i> DE HABILIDAD ORAL DE ACTFL.....	507
APÉNDICE 3	
EJEMPLO DE ENTREVISTA ORAL DE ACTFL (OPI).....	511
APÉNDICE 4	
FORMULARIOS PARA RELACIONAR EL TEST NUESTRO CON EL MCER ...	517
APÉNDICE 5	
FICHA USADA CON LOS EXPERTOS PARA EL EXPERIMENTO DE INTERPRETACIÓN DE NIVELES DE LA ESCALA DE <i>EXPRESIÓN ORAL</i> <i>EN GENERAL</i> DEL MCER	533
APÉNDICE 6	
ESCALAS DEL MCER CON LAS QUE SE HAN CORREGIDO LAS TAREAS DE NUESTRO NUEVO TEST	535
APÉNDICE 7	
FORMULARIO UTILIZADO PARA LA VALIDACIÓN DE LAS TAREAS Y SU TIPO DE LENGUAJE CON EXPERTOS.....	543
APÉNDICE 8	
DESCRIPCIÓN DE LA TAREA 9	557
APÉNDICE 9	
RESULTADOS DE LAS CALIFICACIONES EN LA OPI Y EN EL TEST DEL MCER.....	559
APÉNDICE 10	
CUADROS DE ANÁLISIS ESTADÍSTICOS DE CORRELACIÓN Y FIABILIDAD	565
APÉNDICE 11	
DVD CON LAS GRABACIONES DE LAS ENTREVISTAS.....(interior)	
BIBLIOGRAFÍA	571

RESUMEN

Esta tesis sobre evaluación oral consta de 5 capítulos y compara los sistemas de evaluar la lengua según la entrevista oral de lengua del *American Council on the Teaching of Foreign Languages* (ACTFL) y un test con el que evaluar la habilidad oral diseñado según unas escalas del *Marco común europeo de referencia* (MCER).

En ella (especialmente en el capítulo 2) hemos hecho una exhaustiva puesta al día de la evaluación oral: del sistema de ACTFL de evaluar oralmente las lenguas; de las posibilidades que el MCER presta a la evaluación oral del español; de las posibilidades de equiparar esos dos sistemas de evaluar el español; del papel del evaluador en las entrevistas orales de lengua; y de las técnicas de validación de exámenes.

Para ello hemos realizado varias investigaciones y hemos realizado los análisis y experimentos que se detallan:

- un análisis cualitativo de los descriptores de nivel del MCER (capítulo 3).
- un experimento con 50 profesores de ELE, para confirmar las hipótesis sobre los niveles de las escalas del MCER (capítulo 3).
- una comparación de los niveles y escalas del MCER con la escala utilizada en la OPI de ACTFL (capítulo 3).
- un nuevo test oral basado en 8 tareas, para calificar mediante algunas escalas del MCER (capítulo 3).
- la validación cualitativa y estadística de este nuevo test y sus tareas (capítulos 3 y 4 y Apéndices 7, 9 y 10).

- entrevistas de evaluación oral a 28 estudiantes, grabadas, en primer lugar las 28 OPI de ACTFL, y en segundo lugar las 28 del nuevo test diseñado según el MCER (capítulo 4 y Apéndice 11 con un DVD con las grabaciones de las entrevistas).

- hemos formado a un equipo de calificadores para evaluación oral según los criterios del MCER (capítulo 4).

- este equipo ha interpretado y calificado las 28 entrevistas del MCER y ha dado sus calificaciones a cada candidato (capítulo 4 y Apéndice 9)

- hemos hecho pruebas de fiabilidad entre los evaluadores (capítulo 4 y Apéndice 9).

- hemos hecho pruebas estadísticas para hallar los factores que evalúa el nuevo test del MCER (capítulo 4 y Apéndice 10).

- expertos de ACTFL han calificado las 28 entrevistas del modelo OPI de ACTFL y han dado sus resultados.

- hemos hecho pruebas para hallar la correlación entre ambos tests, la OPI y el del MCER, mediante el *índice de correlación de Spearman* (capítulo 4 y Apéndice 10).

- con todo ello, finalmente, hemos establecido dos fórmulas matemáticas para buscar la equivalencia entre una calificación en el test del MCER y otra en la OPI, y viceversa (capítulos 4 y 5 y Apéndice 10).

El capítulo 5, dedicado a las conclusiones, expone los resultados finales de las posibilidades de comparación entre ambos sistemas (el utilizado mediante las *guidelines* de evaluación oral de ACTFL y el utilizado mediante escalas de habilidad oral del MCER) y las limitaciones de esta investigación. Propone, también, nuevos trabajos necesarios para analizar las posibilidades de evaluación oral mediante el MCER.

La tesis se centra en investigar las posibilidades de equiparar dos tests de evaluación oral, explicando sus posibilidades, dificultades y limitaciones.

Un test es el sistema de evaluación utilizado por el Consejo americano para la enseñanza de lenguas extranjeras (ACTFL en sus siglas inglesas), la OPI (*Oral Proficiency Interview*); el otro, un test creado por nosotros a partir de ciertas escalas del *Marco común europeo de referencia para las lenguas* (MCER). Ambos tipos de test y escalas de calificación se explican en sus respectivos apartados (2.5 para sus tipos de escalas, 2.6 para ACTFL y 2.7 para el MCER) junto a sus objetivos, técnicas de elicitación del discurso y protocolos de evaluación, cuando existen. Las escalas que se

utilizan para calificar cada test son diferentes no sólo por su diverso sistema de compartimentar los niveles de la lengua, 10 niveles en ACTFL y 6 en el MCER, sino por su diferente forma de utilización al calificar las muestras de lengua y por la propia concepción de la lengua que suponen (de ello se trata en el apartado 2.5). Por ello nuestra tesis inicial es que la equiparación entre ambos sistemas no es sencilla o no es posible. La pregunta básica se podría resumir de la siguiente forma: ¿a qué nivel correspondería en el MCER un estudiante que obtuviera en una OPI de ACTFL el nivel de avanzado medio (por ejemplo)? O, al contrario, ¿qué nivel tendría en una OPI de ACTFL un estudiante que fuera B2, por ejemplo, en el MCER?

El MCER, como documento que abarca aspectos del extenso campo de la enseñanza de lenguas, el aprendizaje o la evaluación, tiene múltiples posibilidades de utilización. Entre estas figura su uso pedagógico, a través del llamado Portfolio europeo de lenguas (PEL), pero en el ámbito de la evaluación ha recibido la mayor parte de las críticas, debido a que su enorme flexibilidad, útil en contextos académicos, se torna indeterminación y ambigüedad cuando se trata de la evaluación. Establece 56 escalas de dominio de la lengua para diferentes habilidades y subhabilidades de expresión oral y escrita, comprensión oral y escrita, interacción y mediación. Pero carece de protocolos y técnicas de evaluación o de modelos estandarizados que ayuden a ponerla en práctica o de un fin específico de uso (el alcance del documento como modelo de evaluación, así como sus limitaciones y críticas se encuentra en el apartado 2.7).

ACTFL posee una sola escala de evaluación oral para el concreto y expreso uso de certificar el dominio lingüístico general sin importar los contextos de su aprendizaje, basada en cuatro elementos (funciones, contextos, temas y tipo de texto); y con ella se evalúa de forma holística el uso real de la lengua en situaciones determinadas, detectando los puntos fuertes del hablante y sus principales limitaciones (en el apartado 2.6 se explican con detalle sus objetivos, conceptos y mecanismos).

Dadas las iniciales diferencias en las posibilidades de uso de ambos sistemas, el MCER y ACTFL, y como el MCER no es un test ni da o propone ninguna forma de evaluación, para poder proceder a un intento de equiparación entre la evaluación hecha con él y la OPI hemos tenido que partir del diseño de un test oral mediante el MCER que se asemejara lo más posible a la OPI. Esto lo hemos hecho seleccionando un conjunto de escalas del MCER que fueran aptas para la evaluación de las destrezas orales con fines de certificación general, escogiéndolas pensando en el tipo de lengua que se produce mediante la OPI, sus funciones y tipo de texto, y utilizando las técnicas

de elicitación y protocolos fundamentales de este sistema (las precauciones para que el nuevo test produjera similares muestras de lengua que pudieran, por lo tanto, ser comparables a las de una OPI se encuentran en el apartado 3.5).

El diseño de este nuevo test se ha sometido a técnicas profesionales de validación de tests, de su constructo y fiabilidad, con análisis cualitativos con expertos, profesores y candidatos similares a la población meta (los datos cualitativos se encuentran en el apartado 3.6); también nos hemos fijado en las normas que el Consejo de Europa viene estableciendo para demostrar que un test está relacionado con los descriptores del MCER (esto se encuentra en el capítulo 2.7.5 y en el Apéndice 4). Al mismo tiempo, los resultados de las entrevistas se han sometido a diversos tratamientos estadísticos para averiguar la fiabilidad de sus calificadores, los factores que evalúa el test o su correlación con el test de ACTFL, la OPI (estos datos aparecen en el apartado 4.5 y, más detalladamente, en los Apéndices 9 y 10). Todo ello ha mostrado que este nuevo test puede ser considerado como un test paralelo a la OPI, con un índice de correlación de Spearman de 0.95, cuyos resultados, medidos con las escalas del MCER, pueden ser objetivamente comparados con los que se obtienen en una OPI.

Sin duda, todo el proceso hubiera sido mucho más sencillo si se pudieran calificar directamente las muestras de lengua obtenidas mediante la OPI con las escalas del MCER, además de con la de ACTFL, o viceversa, si los resultados obtenidos mediante un test elaborado según el MCER se pudieran calificar con las *guidelines* de evaluación oral de ACTFL, además de con las escalas del MCER. Así se podría decir fácilmente qué nivel tiene la muestra de la habilidad lingüística de un candidato, obtenida en un solo test, calificándola con ambos tipos de escalas. Pero no es posible (como apuntamos en la nota 12 en el capítulo 1) ya que, dadas las iniciales diferencias entre ambos sistemas, cualquiera de ambas cosas invalida los resultados o los convierte en inocuos. Por ejemplo, una OPI lleva a los candidatos hasta límites en los que se colapsa y desestructura su lengua y esto puede ser malinterpretado y penalizado con las escalas del MCER, que no consideran el asunto. Y, por otro lado, cualquier muestra de lengua obtenida con un test diferente a la OPI no puede ser calificada por la organización ya que carece del procedimiento y elementos necesarios que se establecen en ACTFL para que sea una “muestra aceptable” para su evaluación.

Nosotros no queremos simplemente saber en qué medida la escala de evaluación oral de ACTFL, de 10 niveles, se relaciona con los seis niveles del MCER en sus diferentes escalas (aunque este es un trabajo teórico de comparación cualitativa que

hemos intentado en el apartado 3.2 y del que nos servimos para establecer nuestras hipótesis de investigación) sino establecer en qué medida los resultados de un test y la calificación obtenida con sus escalas son equiparables con los de otro test y sus calificaciones; como hemos señalado: no solo qué sería un B2 en la escala de ACTFL, sino qué nivel se le asigna con una OPI de ACTFL a un candidato calificado según el MCER como B2; algo más concreto y práctico. Por esto hemos procedido a comparar resultados de evaluación oral y no sólo escalas. Su título sería más certero: La evaluación oral: una equivalencia entre la evaluación mediante las “guidelines” de ACTFL y mediante un test diseñado según algunas escalas del MCER.

Una vez planteada la investigación y diseñado el test para evaluar según el MCER, formamos un corpus de 28 candidatos de todos los niveles mediante las dos entrevistas (la OPI y nuestro test del MCER) con un total de 56 entrevistas. Las calificamos según sus respectivas escalas y comparamos los resultados obtenidos por cada candidato en ambos modelos. Siendo ambos tests dos tests casi paralelos, estos son, en principio, equiparables (los resultados están en el capítulo 4 y las conclusiones de las equiparaciones en el 5).

También a lo largo de toda esta tesis y en los capítulos del estado de la cuestión (capítulo 2), trabajo de campo, diseño y validación (capítulo 3) y análisis de resultados y conclusiones (capítulos 4 y 5), hemos puesto de manifiesto la situación actual de la evaluación oral con fines de certificación general y hemos señalado algunos de los puntos fuertes y las limitaciones de ambos sistemas de evaluación. Entre las limitaciones de la OPI señalamos (junto al aparato de críticas y defensa del modelo expuesto en el apartado 2.6.3) que esta no puede ser utilizada ni sus resultados extrapolados a campos ajenos a los que se definen en sus objetivos. Esta práctica, no obstante, es ajena a ACTFL, que deja claro que al realizar la evaluación no se pueden comparar resultados de candidatos entre sí ni juzgar al candidato en sí mismo sino a las muestras de lengua que de él se han conseguido.

En cuanto al MCER, tiene gran valor como herramienta pedagógica y de evaluación en ámbitos académicos o para tareas profesionales específicas, pero se le encuentran insuficiencias, entre las que se ha citado la falta de concreción en determinadas escalas, ambigüedades de interpretación y carencias teóricas en su concepción. Carencias provenientes de que las escalas han sido diseñadas basándose exclusivamente en opiniones de profesores de lengua. Para evaluación es dudosa su utilidad como sistema

de certificación del dominio lingüístico (se ha expuesto en el apartado 2.7.4 junto a las críticas al documento como modelo de evaluación), ya que no aporta elementos para realizar una adecuada evaluación, ni enseña a llevarla a cabo, ni tampoco en el documento que se ha estado desarrollando para relacionar exámenes con el MCER (expuesto en el apartado 2.7.5). Ello lleva a engaño respecto a las posibilidades de evaluar usando las escalas, respecto a la interpretación de los resultados obtenidos y respecto a la posibilidad real y objetiva de proceder a homologaciones entre todos los países europeos, sean cuales sean los métodos de enseñanza de lenguas, uno de los objetivos que se planteaba el MCER.

Los resultados de todo se exponen en los dos últimos capítulos (capítulos 4 y 5). El capítulo 4 muestra las calificaciones obtenidas por cada uno de los 28 candidatos en ambas entrevistas con ejemplos de algunos de sus discursos y con argumentos dados por sus calificadores. El capítulo 5 extrae las conclusiones oportunas, cinco, tras los análisis estadísticos a que se sometieron los resultados. En ellas damos dos fórmulas matemáticas para trasladar el resultado obtenido en un test al otro, es decir, el lugar de la escala que le correspondería a un candidato que tiene un determinado nivel según una escala, en la otra escala y viceversa. Podemos decir, no obstante, que este resultado matemático no lleva a una equiparación práctica efectiva, ya que no es ni unívoco ni conmutativo (véase el apartado 5.1). Es decir, a un avanzado bajo en ACTFL le pueden corresponder diferentes niveles del MCER; y si le corresponde B2, ello no significa que ser B2 en el MCER conlleve ser avanzado bajo en ACTFL. Entre las principales conclusiones podemos decir que NL (inicial bajo) es “menos que A1”; IL (intermedio bajo) es, como mínimo, A2; AL (avanzado bajo) es, como mínimo, B2; y S (superior) es C2; aunque no al contrario (C2 no tiene que ser superior y puede ser solo avanzado alto, o B2 no es siempre avanzado bajo, por ejemplo); y que el MCER, en cuestiones de evaluación oral de dominio general, según nuestro test, abarca un dominio de lengua más limitado que ACTFL, tanto en los niveles bajos como en los altos, ya que el nivel C2 del MCER no es el superior de ACTFL, sino simplemente avanzado alto, y porque el MCER carece de descriptores para los primeros niveles de ACTFL, inicial bajo y medio. En el apartado 5.2 se explican y aclaran estos resultados y sus implicaciones y en el 5.3 las limitaciones de la investigación. Remitimos al capítulo 5 para ver las conclusiones y su comentario.

CAPÍTULO 1

INTRODUCCIÓN

1.1 Explicación de cada capítulo

Este **capítulo 1** es una introducción general al hecho de hablar y evaluar una lengua extranjera y sus problemas fundamentales, que fluctúan entre el estatuto de la interlengua y el de la norma de la lengua 2, y se cruzan con problemas de adquisición, estilo de aprendizaje, estrategias pragmáticas de codificación y decodificación y motivaciones diferentes. Todo ello desdibuja el objeto en cuestión a medir -la lengua de extranjeros- y hace difícil una interpretación de dominio de habilidades única o útil.

En §1.3 y §1.4 explicamos nuestras hipótesis de investigación y ofrecemos las primeras equiparaciones entre las escalas de evaluación oral del MCER y las *guidelines* de evaluación oral de ACTFL que se intentarán demostrar o rebatir con esta investigación.

El **capítulo 2** presenta el actual estado de la cuestión de la evaluación oral de la lengua de extranjeros para certificación de dominio: qué es evaluar una lengua extranjera desde el punto de vista oral. Este capítulo se subdivide en 8 apartados que atienden a los distintos problemas que presentan las diversas metodologías y técnicas existentes en la práctica de la evaluación oral hoy en día.

2.1 se plantea la problemática de la evaluación oral y de la delimitación del objeto de estudio.

2.2 se centra en un modelo de evaluar extendido y aceptado y en su concepto de lengua y de validación: el modelo propuesto por Bachman / Palmer. En este apartado se

explican los problemas para delimitar el constructo a evaluar, la lengua oral, así como modos de hacerlo, limitaciones y sistemas para garantizar su validez.

2.3 trata de los aspectos técnicos, cualitativos y sicométricos, de la validación de tests¹ especialmente para evaluación oral, explicando las condiciones necesarias para realizar con éxito el diseño de una entrevista, y otros aspectos cualitativos para entender la dificultad de llevar a cabo una evaluación oral objetiva y fiable.

2.4 plantea las posibilidades de evaluación del discurso de extranjeros, considerando en qué medida evaluar la lengua oral es evaluar la conversación natural u otro producto diferente a definir; analiza si las características del lenguaje elicitado mediante entrevistas para evaluación poseen las características de una conversación o no, así como el papel e influencia del entrevistador en la producción del discurso del candidato. A la luz de estos datos, insiste en otros aspectos relacionados con la evaluación oral como son las intenciones del candidato y su interpretación por el entrevistador, el surgimiento de estrés comunicativo y su interpretación durante la entrevista o en la fase de corrección y calificación, o la necesidad de formación de evaluadores.

2.5 se centra en un elemento externo a las entrevistas pero importante porque delimita su concepto y ayuda a interpretar los resultados: las escalas de medición. En él se explican sus tipos y características y se describen las que se van a emplear en nuestra investigación, las del MCER y las *guidelines* de evaluación oral de ACTFL.

2.6 se dedica a explicar uno de los dos modelos que vamos a analizar y comparar en este trabajo, el de la OPI de ACTFL. El capítulo, además de una breve historia de este sistema y su importancia en el mundo de la evaluación oral, explica sus protocolos, sus métodos de realizar las entrevistas, sus objetivos, escala y tipología de preguntas, y termina con una selección de las principales críticas que se le han hecho y la respuesta dada a estas críticas por la propia organización o por trabajos independientes.

2.7 explica el otro modelo que se va a comparar en este trabajo, el que surge a partir del MCER. El capítulo explica el sentido y utilidad del MCER y sus posibilidades ante la evaluación oral, centrándose, por ello, en el capítulo 9, dedicado a la evaluación, así como en otros materiales editados para complementarlo, como el documento publicado

¹ Aunque el *Diccionario Panhispánico de dudas* (RAE, 2005) prefiere el plural invariable test, en esta tesis hemos decidido no seguir esa excepción y preferimos el plural tests, ya que opinamos que la diferencia entre el singular y el plural, imposible de pronunciar en español, ayuda a la lectura y clarifica el concepto.

para relacionar exámenes con los niveles del MCER. También se repasan las críticas que ha recibido el MCER y se señalan sus limitaciones.

2.8 termina el capítulo. En este apartado se explican brevemente algunas propuestas de comparación y equivalencia, no empíricas, que se han hecho respecto a los dos modelos de evaluación oral, el del ACTFL y el del MCER.

El **capítulo 3** se centra en la descripción del trabajo de campo realizado mediante el análisis de las diferentes escalas usadas en el MCER y la OPI para evaluar oralmente la lengua y el diseño de un nuevo test, según el MCER que se usó para entrevistar a estudiantes extranjeros que cursaron estudios de español en el CLM (Centro de Lenguas Modernas) durante el verano de 2006. Con el fin de llegar a conclusiones que fueran lo más firmes posible y que no fueran invalidadas este trabajo, ese test se diseñó teniendo en cuenta las principales críticas que se hacen a las entrevistas orales y a sus tareas.

3.1 es un primer acercamiento a las escalas del MCER a través del análisis crítico de varias de sus escalas con la intención de entender qué concepto de lengua establecen y en qué niveles lo definen. También se presenta un experimento realizado con 50 profesores y expertos con el fin de comprobar nuestras interpretaciones respecto a su ambigüedad.

3.2 compara de forma no empírica, sino cualitativa, varias de las escalas y niveles del MCER, con la escala y niveles de las *guidelines* de ACTFL. Estudia las diferencias halladas entre los modelos de evaluación oral del MCER y ACTFL y, tras ello, establece una propuesta de comparación y entre ambos modelos de evaluar y escalonar la lengua. Esta equiparación es la utilizada para hacer las hipótesis de investigación presentadas en el capítulo 1, en §1.4.

3.3 describe el trabajo el proceso de selección de estudiantes de diferentes niveles de dominio utilizados en la investigación y el sistema de calificación que se va a emplear con el nuevo test.

3.4 explica y describe las 8 tareas diseñadas para el nuevo test, elaborado según el MCER, que vamos a usar para llevar a cabo la investigación.

3.5 da cuenta de las consideraciones a tener en cuenta para que el test elaborado mediante el MCER y su forma de realización y calificación faciliten la comparación con la OPI de ACTFL.

3.6 explica el trabajo de análisis y validación de las tareas diseñadas para el nuevo test del MCER, con la explicación de sus objetivos y del lenguaje que pretenden elicit

y elicitan. Este apartado se divide en dos: primero presenta la validación hecha con expertos y profesores; después presenta la validación hecha con candidatos voluntarios.

3.7 explica el test final y muestra otros datos cualitativos sobre su validación y fiabilidad.

El **capítulo 4** se ocupa de los resultados de las entrevistas realizadas con el test del MCER y con la OPI. Comienza exponiendo las dificultades y problemas con que se encontraron los evaluadores para realizarlas y calificarlas. Continúa presentando los resultados de las entrevistas, con ejemplos transcritos de las más significativas y de los casos considerados como raros. Y termina con la presentación cuantitativa de los resultados, esto es, los análisis estadísticos que se realizaron con los resultados de los tests y de los calificadores. Estos datos estadísticos permiten comprobar si las hipótesis de investigación eran acertadas o deben desecharse. Los resultados se ofrecen en una serie de tablas explicativas de las comparaciones y equiparaciones entre las escalas del MCER y las *guidelines* de ACTFL. Termina con una comparación entre los resultados obtenidos por esta investigación y los que proponían otros investigadores, así como con las comparaciones no empíricas que se presentan en el capítulo 3, §3.2.

El **capítulo 5** presenta y comenta las conclusiones obtenidas, y hace una referencia a las limitaciones de esta investigación y a posibles trabajos posteriores y recomendaciones para realizarlos.

Se termina con la serie de **apéndices** que contienen documentos, tablas y otras cosas que se han citado durante el trabajo y se consideran necesarias para entender en detalle ciertos aspectos. Como se ve en el índice son 11. En estos 11 apéndices figura:

Apéndice 1. Ficha de las características de las tareas empleadas en el test del MCER, realizada según el cuadro de Bachman / Palmer (1996).

Apéndice 2. Escala original de habilidad y de dominio oral de ACTFL, en inglés.

Apéndice 3. Transcripción de una entrevista OPI como ejemplo de su estructura.

Apéndice 4. Formularios del Consejo de Europa para demostrar que un test se puede relacionar con el MCER, cumplimentados como si fuéramos a proponer la aceptación de la validación de nuestro test.

Apéndice 5. Ficha utilizada en nuestro experimento con expertos para ver las posibles interpretaciones de las escalas y niveles del MCER.

Apéndice 6. Todas las escalas del MCER utilizadas para corregir y calificar las entrevistas orales.

Apéndice 7. Documentos utilizados en la validación de las tareas y su lenguaje con expertos y la hoja de respuestas que se les dio para recabar sus opiniones.

Apéndice 8. Descripción de la tarea 9, desechada en el diseño final del test por falta de validez.

Apéndice 9. Cuadros con las calificaciones obtenidas por cada candidato en las entrevistas con nuestro test del MCER y la OPI de ACTFL así como las calificaciones dadas por cada evaluador, y el promedio de las calificaciones de cada tarea y cada candidato en el test del MCER.

Apéndice 10. Cuadros con todos los resultados estadísticos que se realizaron con las entrevistas y sus calificaciones.

Apéndice 11. DVD con los archivos de todas las entrevistas grabadas, 28 según la OPI de ACTFL y 28 con nuestro nuevo test basado en el MCER.

La **Bibliografía** va al final, contiene todas las referencias bibliográficas, de artículos y libros citados así como de documentos obtenidos en internet.

1.2 Introducción: algunas dificultades de la evaluación oral

Habla, por favor, ya por bien, ya por mal...

(Di que no es verdad.

Bolero. Alberto Domínguez).

Hay quien piensa que pocas cosas tienen el poder de callarnos como la pena. Efectivamente, es normal hablar. Hablamos, hablamos y hablamos. Está claro que los humanos hablamos. Hablan, hablamos, hablo... podemos seguir el paradigma, los paradigmas. Todos los estudiantes extranjeros lo intentan aunque unos con más éxito que otros: los usan, intentan combinarlos, se esfuerzan por hacerlo bien. ¿Hacerlo bien?

Aquí surge la norma, y con ella nuestro problema como evaluadores de lenguas, ¿cómo calificarla en un extranjero cuya interlengua es una lengua con estatuto propio y legítimo (Muñoz Licerias, 1992: 21) pero, claro, alejada de aquello a lo que se aspira en los tests, la adecuación al otro sistema?; ¿cómo medirla en bilingües que la usan con constantes cambios de código?; ¿cómo en aprendices, si las fases del desarrollo de su aprendizaje quedan a menudo ocultas y los tests no tienen en cuenta esos atisbos de elementos emergentes? La norma, por un lado. ¿Y el uso? Peor suerte corre este, ¿cómo aprehenderlo si la pragmática lo convierte en diversamente interpretable, y el estilo propio de cada individuo o la diversidad de estrategias de comunicación y compensación hacen de él algo terriblemente escurridizo? Son problemas a los que habrá que enfrentarse. En fin, en todos los humanos surge la lengua (Pinker, 2001: 26); por eso callar no es ausencia de lengua y ni siquiera de intención o de mensaje.

Pero no es difícil hablar, ni callar, cuando uno es el actor de su lengua, de sus lenguas, de sus interlenguas. Lo difícil es hacer hablar (¿o ya no tanto?: *reality shows*, *talk shows*, hay muchos métodos. Y los ha habido históricamente -incluidas la confesión y la tortura- que tienen ese mismo objetivo). Pero si parece difícil hacer hablar, más difícil aún es que quien habla diga lo que tú quieres oír (aquí ya se confunden el amante y el policía). Jueces, abogados, periodistas, profesores, amigos, familiares tienen -a veces- esta función en nuestra sociedad: hacer hablar. Y hacer preguntas es la técnica más básica y efectiva que de ello se ocupa.

Nuestro trabajo va a ir por este camino: hacer hablar y medir lo que se habla. Difícil camino si se tiene en cuenta su triple objetivo, pues hacer hablar para medir cómo se habla supone -muchas veces- hacer hablar para que el que habla diga lo que queremos oír, no tanto el mensaje como la función; y porque evaluar la lengua oral de hablantes extranjeros de español (de cualquier lengua extranjera) empieza por buscar que el candidato evaluado haga ciertas cosas con la lengua, cosas que logramos mediante determinadas tareas que lo motivan y determinan. Como decimos, una triple tarea: hablar de cómo otros hablan; de cómo otros hacen hablar; y de cómo estos escuchan e interpretan (y califican o evalúan) lo que hablan los que hablan.

De esto trata esta tesis. Concretamente de dos formas de hacer hablar y medir lo hablado²: la OPI (*Oral Proficiency Interview*) del Consejo americano para la enseñanza de lenguas extranjeras (*American Council on the Teaching of Foreign Languages*,

² Estos dos sistemas de evaluación se describen en el capítulo 2.

ACTFL)³, y un nuevo test, diseñado por nosotros a partir del *Marco común europeo de referencia* (MCER).

El español es una lengua importante y en crecimiento en todo el mundo y se estima que en las próximas décadas será una de las tres más extendidas en un mundo donde China, India, Rusia y Brasil tendrán un poder económico semejante al del G-8. Nuestro país -conjunción entre Europa y América- debe prestar atención a los sistemas de evaluar la lengua y asesorar sobre ella, y entre los métodos, a los dos más importantes y generales que existen hoy en día: uno basado en el MCER y el de la OPI de ACTFL -ambos de claro carácter panlingüístico e interesados en evaluar y homologar el conocimiento, dominio o habilidad de un conjunto de lenguas (europeas para el MCER, variadas en ACTFL).

Por ello este trabajo de analizar ambos sistemas y escalas de asesoramiento, evaluación y calificación del uso y aprendizaje de lenguas va a dedicar sus páginas, análisis y explicaciones a intentar encontrar los puntos en común, así como a describir sus limitaciones y diferencias; para establecer en qué medida y cómo, en qué contextos y situaciones, hay posibilidades de proceder a hacer homologaciones entre ambos y en cuáles no. Intentaremos clarificar los campos de actuación común evitando generalizaciones o desconocimientos que lleven a errores o a concepciones falsas sobre su utilidad pues, en principio, ambos tests son diferentes y tienen diferentes escalas y conviene conocer sus diferencias si queremos usarlos o manejarlos conjuntamente.

La OPI de ACTFL y los tests que se diseñen a partir del MCER, como el nuestro⁴, no son dos coches que se pueden coger indistintamente para, con mayor o menor comodidad, llegar al mismo lugar con unos minutos de diferencia: son dos medios que en sí mismos conducen a lugares diferentes, aunque se entrecrucen en numerosas ocasiones. Y es importante saber dónde están en cada momento y hacia dónde van, para poder usarlos de acuerdo a nuestras necesidades.

³ A partir de aquí los citaremos como MCER, OPI y ACTFL.

⁴ El MCER no es un test ni describe o define un test concreto. En nuestra investigación hemos diseñado, descrito y validado un test para evaluación oral según unas escalas del MCER, de forma que pudiéramos proceder a una comparación entre sus resultados y los de una OPI de ACTFL. Este test nuevo y diseñado por nosotros para esta investigación se explica detenidamente en el capítulo 3. Por comodidad, a partir de ahora lo nombraremos indistintamente como “el test del MCER” o “nuestro test del MCER”. Pero no hay “un test” del MCER ni nos referiremos a otros posibles tests diseñados mediante este documento. Siempre que hablemos del test del MCER nos estaremos refiriendo a nuestro test, el utilizado para esta investigación.

El campo de la enseñanza de lenguas sufrió avances notabilísimos en la segunda parte del siglo pasado ampliando y aclarando sus conceptos, revisando no pocos de ellos y ofreciendo muchas nuevas soluciones a sus necesidades. Para acercarnos a los más notables podemos citar los trabajos del Consejo de Europa⁵ desde la década de los años 60 y su interés por los enfoques comunicativos⁶ de enseñanza de lenguas y la introducción del concepto de sillabus funcional (Wilkins, 1976) basado en las nociones, situaciones, roles y funciones.

Para llevar a cabo estos trabajos que dieron lugar a la publicación del *Threshold Level* cuya versión española, *Un nivel umbral*, es de Slagter (1979)⁷, se necesitaron nuevos planteamientos respecto a la forma de organizar los niveles de dominio de la lengua (Trim, 2007: 19). Esto llevó, también, a la necesidad de plantearse otras formas de organizar el contenido lingüístico y de evaluar la lengua (Trim *et ál.*, 1980: 5), pues se vio la necesidad de evaluarla atendiendo a las necesidades comunicativas.

A ello contribuyeron desde el primer momento los cursos para fines específicos y sus escalas de evaluación basadas en necesidades contextuales concretas, en vez de los tests de escalas de gramática, vocabulario o pronunciación que provenían de los currículos y programas de enseñanza que no tenían en cuenta las necesidades comunicativas de los estudiantes (Carroll, 1983: 7).

Canale, que conocía estos intentos y los trabajos de Van Ek o Wilkins sobre nociones y funciones, hacía referencia a ellos cuando en 1980 diagnosticaba las nuevas demandas de la enseñanza de lenguas señalando la necesidad de nuevas formas de evaluarla con formatos que atendieran a situaciones comunicativas significativas y aseguraran que el conocimiento de la lengua era efectivo en contextos reales. Entre las nuevas formas de evaluar destacaba la interacción personal en situaciones auténticas y

⁵ Para ver la historia de este proceso que comenzó en los años 60 en el Consejo de Europa a través de su llamado “Major Project, Modern Languages” y dio lugar a los nuevos enfoques de enseñanza y aprendizaje de lenguas que se centraron en las necesidades del alumno, y que ha terminado, por ahora, con la publicación del MCER, véase el artículo de Trim (2007), así como otro anterior publicado en el *European Center for Modern Languages* (Trim, 2001, se puede ver en www.ecml.at/efsz/files/Trim.pdf) o el resumen de su historia, que el propio Consejo de Europa presenta en su página web: www.coe.int/t/dg4/linguistic/Historique_EN.asp (consulta realizada el 20 de enero de 2008).

⁶ Este interés centrado en las necesidades del hablante y en los enfoques de enseñanza comunicativa se ve ya en los documentos del Consejo de Europa de los años 50 y 60, así como en la realización de numerosos talleres de formación de profesores durante los años 80, una vez editadas las diferentes versiones del *Threshold level*. Concretamente entre 1984 -87 se impartieron 36 talleres internacionales (Trim, 2001: 4).

⁷ Hay una reedición en pdf, en línea, en *MarcoELE*, número 5. <www.marcoele.com/num/5/nivel_umbral.pdf>.

la necesidad de buscar sistemas de calificación justos y niveles de descripción de la competencia aceptables para diversas etapas del aprendizaje (Canale, 1980: 35).

Todas estas transformaciones en los conceptos de la lengua y la evaluación están bien expuestas en las dos siguientes citas de Bachman / Palmer (1996) y Bachman (1990), quienes consideran que se ha pasado de la interpretación de la lengua materna como principal factor de interferencia en el aprendizaje de una segunda lengua, a reconocer que hay otros factores determinantes, como los cognitivos, los afectivos o los contextuales. Valen la pena, aunque sean un poco largas⁸:

Dado lo que se sabía y desconocía respecto a la naturaleza del uso de la lengua, del aprendizaje y de la evaluación en esos momentos, cuestiones como estas simplemente eran impensables. La habilidad lingüística era concebida como un conjunto cerrado de elementos -gramática, vocabulario, pronunciación, ortografía- que se materializaban en las cuatro destrezas -comprensión oral, expresión oral, comprensión lectora y expresión escrita. Enseñando y evaluando estas cosas, se estaba enseñando y evaluando todo lo necesario. Los estudiantes se tomaban por individuos que aprendían la lengua, todos, fundamentalmente de la misma manera -mediante el proceso de estímulo y respuesta descrito por el conductismo-. Por último, se creía que el proceso de aprendizaje era casi más o menos el mismo para todas las personas, para todas las situaciones y para cualquier propósito (Bachman / Palmer, 1996: 4).

En los últimos veinticinco años se han producido cambios radicales en el campo de la evaluación de lenguas, debido a la interrelación de varias materias. Durante este periodo, en la lingüística hemos asistido a la aparición de la teoría de la sintaxis, a su desvanecimiento y a su reciente resurgimiento como paradigma dominante en la descripción de las características formales de los enunciados. También hemos sido testigos de indagaciones mucho más ampliamente fundadas sobre la naturaleza y uso de la lengua, capitaneadas por trabajos de sociolingüística, pragmática y etnografía de la comunicación. En la evaluación de lenguas hemos cambiado la concepción de la lengua como una estructura formal que se aprende mediante su enseñanza explícita, por otra que pone énfasis en la lengua como comunicación y reconoce el poder del uso comunicativo en su aprendizaje y adquisición. También se ha transformado nuestra visión de la adquisición de segundas lenguas, desde una en la que la interferencia de la primera lengua se consideraba como el principal determinante, a otra que reconoce gran variedad de influencias, cognitivas, afectivas y contextuales (Bachman, 1990: 297).

Todo esto no ha dejado de ocurrir y los trabajos e investigaciones, tanto en los campos de la enseñanza y adquisición, como de la pragmática o de la estadística aplicada a las ciencias humanas y al lenguaje, han producido nuevas teorías y diferentes aproximaciones a los problemas en el campo de la enseñanza, el aprendizaje, la adquisición y la evaluación.

⁸ Traduzco al español todas las citas de libros o artículos cuyo original es en inglés. Consúltese la bibliografía.

Uno de los primeros tests orales, de objetivos claramente orales, surgió en 1982 cuando ACTFL comenzó a promover un nuevo test que rápidamente se extendería por círculos de enseñanza e instituciones y sería enormemente influyente tanto en ámbitos académicos y educativos como de evaluación en general, además de muy decisivo en la nueva orientación que consideraba a la enseñanza de la lengua desde el punto de vista de algo que hay que aprender a usar y no solo saber.

Sus seguidores señalan su enorme popularidad, que lo convierte en el test más utilizado entre las instituciones americanas⁹; pero incluso sus mismos críticos reconocen esta poderosa influencia más allá incluso del campo de la evaluación (Salaberry, 2000: 292) y lo citan como uno de los primeros tests basados en criterios, evolucionado del primer test elaborado por el ILR (*Interagency Language Roundtable*) americano.

La llamada OPI de ACTFL¹⁰, cuyas *guidelines* de evaluación oral se revisaron en 1999 para adaptarlas a las críticas y comentarios que desde su aparición habían intentado explicarla, delimitarla o mejorarla, ha llegado hasta nosotros jalonada de constantes investigaciones y análisis y en nuestros días compite en el mercado mundial de certificaciones académicas, oficiales, estatales y universitarias junto a una nutrida cuenta de tests y sistemas de calificación que comparten el interés por delimitar espacios diversos de medición de las habilidades en el aprendizaje y uso de las lenguas. Pero se mantiene como la única específicamente diseñada para medir la habilidad oral con ambición de ser útil fundamentalmente fuera de los ámbitos académicos.

Y el último modelo, hasta el momento: el del MCER, surgido desde que en 2001 el Consejo de Europa promoviera una nueva política englobada dentro de un más amplio proyecto que es el MCER¹¹. Este documento, el MCER, consta de 9 capítulos, con uno

⁹ El §2.6 se ocupa de explicar este test. Él es sin duda “el sistema de medida más ampliamente usado por las instituciones americanas (...), utilizado como herramienta para analizar la fluidez en las muestras de la lengua obtenidas en los estudios sobre aprendizaje en el extranjero” (Magnan / Back, 2007: 48) y no solamente en el extranjero. Bachman, al hablar de él, lo considera “uno de los métodos de evaluar la habilidad lingüística más usados e imitados en todo el mundo” (Bachman, 1990: 302).

¹⁰ ACTFL es una organización que lleva a cabo múltiples actividades en relación con el aprendizaje y evaluación de lenguas, formación y asesoramiento. De su test oral llamado OPI, basado en sus propias *guidelines*, sus mecanismos de evaluación y protocolos de elicitación, se trata en §2.6. Esta tesis utiliza su modelo de test oral, realizado y calificado por miembros y evaluadores de la asociación. A partir de ahora, por comodidad, para referirnos a este modelo de evaluación oral mediante el modelo de OPI de ACTFL y sus *guidelines* de evaluación oral en 10 niveles diremos simplemente “el test de ACTFL”, ya que esta tesis trata de evaluación oral. Aunque ACTFL también dispone de otros modelos de tests, como por ejemplo un test para evaluar la habilidad de expresión escrita.

¹¹ El §2.7 explica el MCER. Entre los objetivos del Consejo de Europa están el interés creciente y la necesidad social, política, económica, también académica, de aunar esfuerzos sobre la evolución de los progresos de sus ciudadanos en la integración lingüística, en el aprendizaje y la enseñanza de lenguas, así como de proceder a homologaciones entre todos los países europeos sean cuales sean los métodos de

específico para evaluación. En él se ofrecen, en total, 56 escalas o escalas de habilidades y de dominio del uso de la lengua para calificar su aprendizaje, dominio y uso.

Ambos modelos, junto a otros prestigiosos tests como los de Cambridge, DELE o TOEFL, pretenden servir para calificar diversas habilidades en el uso de la lengua de forma auténtica, fiable, útil y aceptable, de acuerdo a los valores de honestidad y justicia que se han promovido en los últimos años (Bordón, 2006: 300). Todos ellos permiten día a día a programadores, educadores, empleadores, instituciones o estudiantes, realizar su labor, que además de ser compleja y tener varias y multifacéticas formas, necesita guías, criterios, objetos y mecanismos que provean de herramientas practicables (administrativamente), creíbles (académicamente), asumibles (económicamente) y aceptables (socialmente).

Pero ambos sistemas de evaluación y escalas de habilidad y dominio son diferentes y establecen niveles diferentes. Las *guidelines* de ACTFL determinan 10 niveles denominados: *Novice Low*, *Novice Mid*, *Novice High*, *Intermediate Low*, *Intermediate Mid*, *Intermediate High*, *Advanced Low*, *Advanced Mid*, *Advanced High* y *Superior*. El MCER describe 6 niveles denominados: A1, A2, B1, B2, C1 y C2. Por ello, equipararlos no es tarea sencilla. El mismo MCER hace referencia a su dificultad: “resulta difícil relacionar evaluaciones distintas, a pesar de la hechicería estadística de las técnicas tradicionales, ya que las evaluaciones generalmente valoran elementos radicalmente distintos, incluso cuando pueden abarcar los mismos ámbitos” (Consejo de Europa, 2002: 182).

Este va a ser nuestro trabajo.

1.3 Explicación de nuestro trabajo e hipótesis de investigación

Con este trabajo, vamos a intentar una equiparación entre los resultados que se obtengan al evaluar con la OPI y de evaluar con el nuevo test diseñado con algunas escalas de habilidad oral del MCER. Para ello he tenido que diseñar este nuevo test compatible con el formato de la OPI. El trabajo ha supuesto, al igual que a cualquier otro profesional que en su trabajo se sirva del MCER, determinar muchos aspectos del proceso de evaluación y tomar numerosas decisiones respecto al diseño de tareas o los

enseñanza de lenguas, sin que esto suponga introducir un sistema paneuropeo, ni rechazar la diversidad existente.

protocolos y técnicas de elicitación o realización del test. Pero además, con el fin de tener dos tests que puedan ofrecer resultados que acepten una comparación, se ha decidido eliminar el mayor número posible de variables que diferencien este test del otro (la OPI) y se han seleccionado las tareas y escalas de habilidad y dominio que más se acercan a lo que es el modelo de evaluación oral de ACTFL. El resultado se conforma un conjunto de 8 tareas elegidas para medir diferentes habilidades cercanas a las que se utilizan en la OPI, pues buscamos acercar los dos sistemas para que sean entre sí más fácilmente comparables.

Debe quedar claro que con ello hemos procedido a comparar un sistema existente en el mercado de forma estándar (la OPI de ACTFL), con otro modelo (el MCER) que no es en sí un sistema de evaluación sino un conjunto de escalas que hemos utilizado como referencia, y con el que hemos creado un nuevo test y hemos calificado a los candidatos. A ese test le llamaremos a partir de ahora el nuevo test del MCER, nuestro test del MCER o simplemente el test del MCER, aunque con el MCER se pueden elaborar muchos otros tests, tanto para otras muchas necesidades como para otros candidatos o situaciones concretas¹². El nuestro lo hemos diseñado intentando que sea general pero, como todo test, servirá solamente para lo que lo hayamos diseñado -si su diseño ha sido efectivo-. No pretendemos proponerlo como el único test de evaluación oral posible mediante el MCER con el que, a partir de ahora, haya que evaluar a todos los estudiantes.

Por lo tanto, lo que estamos comparando en realidad no es directamente el sistema de evaluación del MCER con el sistema de evaluación oral de ACTFL. Sino que, mediante el test que hemos diseñado basándonos en el MCER, estaremos analizando el grado de compatibilidad existente entre el sistema de calificar a un mismo candidato mediante ciertas escalas del MCER y mediante la OPI de ACTFL. Eso lo hemos hecho corrigiendo y calificando las muestras de lengua de cada entrevista según su propio sistema¹³ y comparando los resultados de ambos; es decir, viendo qué nivel logra un

¹² El Consejo de Europa da ejemplos de evaluación oral realizados mediante el MCER en inglés y francés, pero no en español, como muestras para ver el sistema de calificaciones. Véase la página web: www.coe.int/T/DG4/Portfolio/?L=E&M=/main_pages/illustrationse.html 8; una vez en la página ir a: Material illustrating the CEFR levels (consulta del 21 de enero de 2008). Estos ejemplos se basan en exámenes realizados por Cambridge y Eurocentres y contienen comentarios para justificar los niveles asignados. También hay ejemplos de evaluación de las habilidades de escritura, lectura y comprensión.

¹³ Por cuestiones metodológicas y de incompatibilidad entre sistemas y métodos de elicitación e interpretación de los datos, explicadas en sus aparatados correspondientes (el del ACTFL, §2.6, y el del Marco, §2.7), no es posible hacer otro tipo de comparaciones; por ejemplo, corregir y calificar las entrevistas OPI de ACTFL mediante las escalas del MCER, ni corregir y calificar las tareas del test del MCER mediante el sistema de ACTFL. Hacerlo lleva en sí a errores de medición graves que impedirían

candidato en cada escala y observando si todos los candidatos que obtienen el mismo nivel en un sistema se asignan al mismo nivel en el otro, o qué nivel se les asigna.

Detallamos a continuación las hipótesis y preguntas de investigación que creemos que son pertinentes y necesarias en nuestra investigación. Estas hipótesis de investigación parten de una primera pregunta de consideración general “¿Es posible evaluar de forma clara y demostrable la lengua oral: con qué limitaciones?” Tras la cual surgen las propias de esta investigación que son:

HIPÓTESIS 0 (a desechar): “Los niveles de las *guidelines* de evaluación oral de ACTFL y de las escalas del MCER no pueden compararse entre sí para la evaluación oral de la lengua”.

Es decir, partimos de que estos dos sistemas de evaluación oral de la lengua de extranjeros, la OPI de ACTFL -usado en Estados Unidos- y el que hemos diseñado según el MCER -en proceso actualmente en Europa- tienen escalas y sistemas de calificar la habilidad de uso de la lengua diferentes y no pueden equipararse o compararse entre sí.

Si se desecha esta hipótesis cero y se demuestra que sí pueden compararse y equipararse, entonces procede la hipótesis 1 (a demostrar):

HIPÓTESIS 1 (a demostrar): Se pueden establecer equivalencias entre los niveles de ambas escalas.

A pesar de distribuir a los estudiantes mediante diferentes definiciones de habilidad y distintas compartimentaciones de grado de dominio -ya que las *guidelines* de ACTFL establecen 10 grados de habilidad mientras que el MCER solamente 6- y a pesar de utilizar tareas y procesos diferentes para su medición, se puede establecer una

la ya de por sí difícil comparación. En su momento se ve (§2.6.2- 3 y §2.6.2 4), y lo explicamos al hablar del estrés comunicativo (§2.4.3- 3) que en el método de ACTFL se lleva a los candidatos hasta situaciones en las que se puede colapsar o desestructurar su forma de hablar, lo que, al no ser tenido en cuenta en el MCER en la forma o estructura de hacer las entrevistas, ni en el sistema de asignar una calificación final a las grabaciones, podría fácilmente -si evaluamos a candidatos entrevistados en una OPI según las escalas del MCER- ser interpretado como falta de habilidad y penalizado en la calificación final.

Y viceversa, no se puede calificar según las escalas de ACTFL a un candidato entrevistado con otras tareas, protocolos o estructuras diferentes a las de la OPI, ya que cualquier muestra obtenida sin seguir sus protocolos es una muestra que no contiene los suficientes elementos considerados en ACTFL y en la forma estandarizada que esta organización estipula, por lo que se considera una muestra no evaluable o no válida.

comparación, de la forma que se presentará en el apartado siguiente (§1.4) y en el cuadro 1.4 a.

Desechar la HIPÓTESIS 0 obliga a responder a la HIPÓTESIS 1 mediante la siguiente pregunta de investigación: ¿Cómo se establecen las equivalencias entre ambos sistemas?

Responder a esta pregunta lleva antes a clarificar las correspondencias y forma de estas correspondencias entre los diferentes elementos que componen el sistema de la evaluación oral: niveles y grados establecidos entre una y otra escala de dominio y habilidad, y sistemas de calificación. Finalmente, habrá que determinar las similitudes, diferencias y limitaciones existentes entre ambos sistemas de medir el dominio y habilidad lingüística¹⁴.

1.4 Hipótesis de las equivalencias entre las escalas del MCER y las *guidelines* de ACTFL, para evaluación oral

Las equivalencias que nosotros hemos encontrado, tras el análisis de los descriptores de las escalas de habilidad oral de cada sistema expuestas en §3.1 y §3.2, serían las que presentamos en el cuadro siguiente, 1.4 a¹⁵:

¹⁴ En realidad, ya se utilizan unas equivalencias entre el sistema americano (aunque no el de ACTFL) y el MCER. Son las utilizadas para homologaciones de cursos de lengua estudiados en España -a los que se da una calificación final utilizando el MCER como referencia- y los niveles de los cursos de lengua de las universidades americanas con los que se convalidan los cursos de lengua hechos en España. Estas equivalencias son ciertamente arbitrarias o no tienen fundamento por carecer de sostén las equivalencias entre los cursos de lenguas estudiados en España y la traducción de sus calificaciones de nivel al sistema de escalas del MCER. Ello no se ha hecho basándose en estudios ni investigaciones, sino en una traducción mecánica de los cursos tradicionales (inicial, intermedio, avanzado y superior) a las nuevas etiquetas del MCER. Y a esta mera traducción nominal se han sumado, demasiado a la ligera, centros, instituciones, profesores y editoriales, al calificar el nivel de sus programas, cursos o libros con la nueva nomenclatura, sin otros análisis ni estudios de dominio fiables. Véase Alderson (2007).

Estas equivalencias existentes serían algo así como:

A1 sería inicial bajo en cursos de lengua y 101 en programas americanos.

A2 sería falso inicial en cursos de lengua y 102 en programas americanos.

B1 sería intermedio en cursos de lengua y 201 en programas americanos.

B2 sería avanzado A y B en cursos de lengua y cursos de nivel 300 en programas americanos.

C1 sería superior A en cursos de lengua y 400 en programas americanos.

C2 sería superior B y perfeccionamiento en cursos de lengua.

En el Centro de Lenguas Modernas de la Universidad de Granada (CLM) la última propuesta, todavía en fase de prácticas, es: A1: inicial A; A2: inicial B; A2+: falso intermedio A; B1.1: intermedio A; B1.2: intermedio B; B2.1: avanzado A; B2.2 y B2+: avanzado B y falsos superiores; C1: superior A y superior B; y C2: perfeccionamiento.

¹⁵ Estas equivalencias están fundadas en los análisis comparativos hechos con varias escalas en el §3.2. Al final de él se explican las equivalencias y se comentan sus posibles consecuencias a la hora de calificar y evaluar. La demostración o refutación de ellas se encuentra en el capítulo 5.

Cuadro 1.4 a. Hipótesis, según nuestras interpretaciones, de las equivalencias posibles entre las escalas del MCER y las *guidelines* de ACTFL, en evaluación oral

ACTFL	NL	NM	NH	IL	IM	IH	AL	AM	AH	S
MCER	--	A1 A2	A1 A2	A1A2 B1	A2 B1	B1 B2	B2 C1 C2	B2 C1 C2	C1 C2	-

A1 iría desde inicial medio a intermedio bajo.

A2 comenzaría igualmente en inicial medio y alcanzaría hasta intermedio bajo o en algunas ocasiones y candidatos hasta intermedio medio.

B1 podría ser indistintamente intermedio bajo, medio o alto.

B2 (llamado avanzado en el MCER), C1 y C2 corresponderían a una horquilla que iría del intermedio alto al avanzado bajo o medio.

Pero un C1 y C2 también pueden ser avanzado bajo, medio o alto.

O sea, que un hablante calificado en A1 puede ser como uno calificado en A2. Un candidato calificado como A2 puede ser como otro de B1. Pero a su vez uno calificado como B1 puede ser como otro de B2. Y a su vez uno de B2 como otros de C1 o de C2. Y apenas se ve diferencia entre candidatos calificados como de nivel C1 y C2.

Creemos que en el MCER no hay suficiente claridad en la descripción de niveles de mayor habilidad, lo que en las *guidelines* de ACTFL sería el nivel superior.

Este análisis e interpretaciones es el que vamos a someter a investigación; en ella tendremos que demostrar o rechazar estas equivalencias y en su caso, establecer otras homologaciones o conclusiones.

El trabajo puede ayudar a aclarar cosas que hoy en día empiezan a estar en boca de profesionales (editores, profesores, instituciones, agentes laborales, etc.) y no están todavía claras. Por ejemplo, de qué se está hablando cuando se dice que una persona es de (o posee) un nivel B1 (según el MCER) o es intermedio medio (según ACTFL). Solamente si se establecen equivalencias entre los diversos métodos, pueden cobrar sentido sus diferentes nomenclaturas.

CAPITULO 2

ESTADO DE LA CUESTIÓN

Este capítulo, a través de sus 8 apartados, aborda la situación de la evaluación oral de lengua y la explicación de los dos modelos que utilizamos en nuestra investigación, el MCER y el de ACTFL.

En §2.1 se tantean las dificultades de la delimitación de la lengua como objeto de evaluación oral y los problemas de evaluar oralmente las lenguas.

En §2.2 se explican los conceptos del modelo de evaluación general propuesto por Bachman / Palmer, con la importancia de la definición del constructo de lengua a evaluar y las formas de garantizar la validez de los test, asegurándose de que la lengua a evaluar sea la que efectivamente se consigue a través de las entrevistas.

El §2.3, llamado “técnicas cualitativas y cuantitativas de validación de tests”, expone los principales conceptos para validación de tests, validez de apariencia, de respuesta, de contenido, de constructo, concurrente, predictiva, uso práctico y fiabilidad.

El §2.4 analiza algunos problemas de la evaluación del discurso de extranjeros, examinando las características del lenguaje elicitado mediante entrevistas y el papel e influencia del entrevistador en la producción del discurso.

El §2.5 se centra en los elementos utilizados para definir la lengua y calificarla: las escalas; explicando sus tipos y características y describiendo las escalas que se han

utilizado en nuestra investigación, las del MCER y las *guidelines* de evaluación oral de ACTFL.

En §2.6 se trata de ACTFL y su modelo de evaluación oral de la lengua y se revisan las principales críticas que se le han hecho y las respuestas dadas desde la propia organización o desde investigadores independientes.

El §2.7 presenta el otro modelo usado en este trabajo, el que surge a partir del MCER; analiza su utilidad, posibilidades y limitaciones como sistema la evaluación oral y repasa las principales críticas que ha recibido.

Y en §2.8 se muestran propuestas de otros autores respecto a la equivalencia entre el MCER y ACTFL.

2.1 Introducción a la evaluación de la lengua oral

Cuando pensamos en evaluación oral tenemos, sin duda, la imagen de un profesor sentado en una silla, hablando con un estudiante, y unas hojas con materiales como fotos, recortes de periódicos o cómics, para servirse de estímulo en determinados temas. También se piensa, sin duda por propia experiencia, en el estrés que ello puede provocar en el estudiante, nervioso por tener que hablar ante un desconocido, en una lengua extranjera. Y, visto desde fuera, escuchado más bien, el asunto puede parecer tranquilo, interesante e incluso sencillo. Sin embargo, esto que vemos es solo apariencia y, desde luego, es solamente la parte más externa de todo el proceso.

Evaluar la lengua oral es una de las cosas más difíciles de la evaluación de lenguas, puesto que los tests orales son los más difíciles de aplicar, los más costosos en medios humanos, los que precisan más formación para las personas encargadas de ponerlos en práctica, y los que más tiempo de realización llevan. Pueden parecer sencillos, ya que se trata simplemente de hablar con los candidatos, pero en realidad esta sencillez no es tal. Se hacen con el mismo tipo de planteamientos previos que el resto de pruebas de lengua, delimitando las características de los candidatos para los que el modelo de examen pueda ser útil y realista y estableciendo claramente los protocolos de evaluación que se vayan a utilizar y los criterios de corrección y calificación de los resultados.

Pero, además, requieren unos cuantos elementos específicos, tanto internos como externos; algunos de los cuales hacen de la evaluación oral algo específico y diferente a los demás tipos de tests, como la necesidad de dedicar gran cantidad de tiempo y trabajo a estos tres aspectos:

- La definición de la habilidad a evaluar, la competencia comunicativa o la habilidad lingüística.
- La formación de los evaluadores, ya que suelen condensar en una misma persona las figuras de entrevistador, corrector y calificador.
- La puesta en escena y administración de los tests, que necesita más tiempo y más evaluadores, puesto que a cada candidato hay que hacerle la entrevista personalmente.

Una de las principales características de un test oral, que lo diferencia del resto de tipos de tests, es que este debe ser capaz de generar en los candidatos el tipo de lengua que se desea medir. Un test oral trata de ser un mecanismo activo cuyo objetivo es conseguir que los candidatos produzcan un tipo de lengua con el que demuestren la existencia de determinadas habilidades lingüísticas. En ello, el evaluador y las tareas que propone llevan el peso principal. Pero como no se pueden provocar ni controlar mediante un test todos los factores que intervienen en la actuación oral de una persona, en su diseño hay que eliminar o minimizar los efectos de los factores que no son deseables y no interesen en la evaluación, como por ejemplo la cantidad de conocimiento que tengan sobre un tema o la educación o personalidad, y maximizar los efectos de la habilidad que se desea medir y hayan sido determinados previamente en los objetivos y escalas de evaluación. En este capítulo vamos a tratar de todos estos aspectos relacionados directamente con la evaluación oral.

Para evaluar oralmente la lengua es necesario partir de un concepto de lengua, ya que esta puede verse e interpretarse desde múltiples puntos de vista y parecer un objeto diferente a cada uno de ellos. Esta variabilidad de la lengua no es exclusiva de los extranjeros sino que pertenece igualmente a los sistemas nativos. Así, no es lo mismo para el periodista que para el médico, para el abogado que para el profesor de instituto, para el psicoanalista que para el profesor de ELE. Incluso entre los lingüistas mostrará aspectos distintos para el estudioso de la adquisición desde postulados innatistas (Muñoz Licerias, 1992), para el estudioso de su adquisición desde postulados

interaccionistas o para un cognitivista como VanPatten (VanPatten / Cadierno, 1993; VanPatten *et ál.*, 2000; VanPatten / Wong, 2003).

La lengua oral es rica y variable y una muestra tomada a un candidato de forma natural es, a todas luces, insuficiente y limitada; no es válida para calificar sus habilidades y lleva a interpretaciones erróneas. Menos aún si lo que pretendemos es evaluar a grandes poblaciones de candidatos con igual grado de objetividad y equidad.

No basta con formarse opiniones sobre los candidatos basadas en el conocimiento directo o indirecto que se tenga de ellos, por un trato personal, del aula, ni a través de una entrevista que no sabemos a dónde puede llevarnos, qué resultados nos puede ofrecer o cómo interpretarlos. Evaluar la lengua oralmente requiere de un examen específico que sea “una herramienta válida y fiable, que nos permita tomar decisiones acertadas (mediante) unos procedimientos de evaluación que sirvan para medir unos determinados conocimientos a través de las muestras que proporciona la actuación de los candidatos” (Bordón, 2006: 19). Y este “debe partir de una definición operativa del habla con la que se puedan desarrollar tests. Nunca el constructo va a captar la riqueza de la comunicación humana pero al menos el test sí debe ser capaz de medir todo lo que se haya definido en el constructo” (Fulcher, 2003: 19).

En fin, la evaluación oral es más costosa y difícil que todas las demás porque en todo momento, y aparte del diseño previo, necesita unos mecanismos en general más difíciles de aplicar. Como dice Bachman “la capacidad de hablar una lengua extranjera es tanto la más importante como la más problemática de las habilidades que intentamos medir” (Bachman, 1988: 149).

El problema de la definición de la habilidad ha sido tratado por diferentes teóricos de la lengua y la evaluación (Chalhoub-Deville, 1997) aportando diferentes modelos que intentan incluir todos los aspectos que intervienen en el uso de la lengua, la competencia lingüística, la estratégica y la pragmática o sociolingüística, pero que no ha dado resultados todo lo efectivos que se necesitan. La tensión entre la competencia lingüística y la habilidad comunicativa continúa y los sistemas desarrollados para medirlas siguen presentando problemas que necesitan más investigación. El último, el de Bachman / Palmer (1996), que se explicará enseguida.

Estamos de acuerdo con Ortega cuando dice “¿Qué metas, contextos y situaciones elegir? La respuesta a esta pregunta no ha recibido aún forma definitiva” (Ortega, 1996: 32). Tal es así que estamos todavía lejos de poder afirmar que se evalúa oralmente la lengua con la objetividad que se merece y la efectividad que la sociedad requiere, o de

poder hacerlo con total equidad y honestidad, igual que todavía estamos lejos de poder evaluar con suficiente fiabilidad otros aspectos como la pragmática (Walters, 2007: 176). Como dice Widdowson, “uno comienza a tener dudas sobre si *algún* modelo de la competencia comunicativa va a llegar a ser pedagógicamente útil como para servir de marco para la evaluación de la lengua”, (Widdowson, 2003: 170, la cursiva es del original).

La lengua de aprendices, la lengua 2, la lengua de no nativos, es en sí un continuum con pasos adelante y atrás, con aciertos por error y defectos por generalización de reglas, con elementos ocultos pero poderosos e influyentes en la determinación del discurso y de las reglas del discurso (Long, 1988; Muñoz Licerias, 1992, 1996; Ellis, 2001; Winford, 2003), con normas no aceptables para los nativos pero necesarias en el proceso, con fosilizaciones inapreciables y caminos cortados, a veces también con atajos, y en fin, con una multidimensionalidad y variabilidad inabarcable.

Medirla no es simplemente medir por separado su carácter formal, la gramática o el vocabulario, su coherencia discursiva o su adecuación sociocontextual. Y al diseñar las pruebas que lo hagan hay que plantearse cuestiones educativas, personales, culturales, académicas, sicolingüísticas, etc. para evitar que influyan subjetivamente y produzcan resultados indeseados. Todo dependerá, probablemente, de las necesidades de la situación y del sensato análisis de los resultados que se esperan, los medios de que se dispone y la responsabilidad de las decisiones que se quieran tomar (Ortega, 1996: 20).

El proceso debe definirse con cuidado desde el comienzo (el momento del planteamiento de los objetivos) hasta el final (la manera de hacer públicos sus resultados ponderando las calificaciones de que se compongan u ofreciéndolas de forma inteligible y útil). Las consecuencias de una prueba (lo que se llama “impacto” -en el ámbito social- o “efecto de rebote” -en el individual o académico) suelen pasar desapercibidas pero pueden llegar a tener importantes efectos sociales e incluso legales.

Y antes de usar sus resultados debemos estar seguros de que la forma en que los interpretamos y usamos es válida, pues de lo contrario corremos el riesgo de hacer interpretaciones incorrectas o tomar decisiones que pueden dañar a nuestros candidatos en vez de ayudarles o beneficiarles (Bachman, 2004: 3).

Debido a todas estas dificultades, las prácticas más usuales de la evaluación oral adolecen de grandes carencias a lo largo de todo su proceso, que comienza con el planteamiento de su diseño y llega hasta la formación de evaluadores, así como de un

gran desconocimiento tanto de las técnicas necesarias para ponerla en funcionamiento como de las limitaciones de cada formato.

Este tipo de evaluación, aunque es el método que más pruebas puede dar sobre el uso real de la lengua y el que más confianza da a los candidatos como prueba efectiva de que saben hablarla, una vez superado el nerviosismo que les provoca el tener que expresarse ante un desconocido que los juzga, es el método que se usa con mayor inconsciencia e ingenuidad respecto a sus dificultades y limitaciones. Cosa que queda muy clara cuando Alderson dedica su libro *Exámenes de idiomas* (1988) a intentar analizar las prácticas en el Reino Unido y explicar los procedimientos de una buena, válida, justa y fiable evaluación y termina advirtiendo de la necesidad de introducir mejoras en todos los campos, tanto de evaluación de inglés como de otros idiomas:

(...) esperamos que las cosas evolucionen, que haya mejoras y que progrese la profesionalidad de los evaluadores, de los tribunales de exámenes y de las compañías comerciales, de los ministerios y universidades (Alderson, 1998: 246).

Todo esto tiene su propia historia. Según Fulcher (2003: 5), el primer test oral se hizo en Estados Unidos en los años 20, consecuencia de la legislación de las normas sobre inmigración de 1924. También se debe mirar a los Estados Unidos para encontrar una de las escalas de evaluación más influyentes, la del FSI (*Foreign Service Institute*) de la que proceden muchos tests¹⁶ y en especial el de ACTFL (véase §2.6), así como para entender las causas de que las escalas de evaluación tengan el tipo de progresión aparentemente lineal que muestran, deudoras de las escalas lineales descritas para el ejército en los 50 (Fulcher, 2003: XIII).

Desde entonces, y en especial en los últimos 25 años, muchos han sido los temas sobre los que ha habido controversia y los problemas con los que ha habido que tratar.

Entre ellos (Alderson / Banerjee, 2002: 92-95):

- La capacidad de un test para obtener muestras suficientes que permitan elaborar un juicio sobre las habilidades que evalúa.
- El insuperable obstáculo de que los datos obtenidos lo son siempre de forma indirecta.

¹⁶ Hasta que en Europa se comenzaron a diseñar escalas como las de ALTE o el proyecto suizo que daría paso al MCER, la mayoría de escalas de evaluación estaban directa o indirectamente relacionadas con las del Instituto de Asuntos Exteriores norteamericano (FSI en sus siglas inglesas) (Parrondo, 2004: 968).

- La necesidad de continua investigación sobre cuál es la mejor forma de evaluar la lengua oral, cómo se define la competencia, habilidad y capacidad lingüística y cuáles son las mejores formas de elicitarlas.

- El análisis del efecto del evaluador y otros interlocutores en la conformación del discurso.

- La capacidad de las tareas para reflejar la habilidad del hablante.

- La posibilidad de que notas diferentes oculten capacidades similares y de que, al contrario, notas iguales oculten capacidades diferentes.

- La posibilidad de que la corrección gramatical tenga una poderosa influencia en la calificación final.

Alderson / Banerjee concluyen su larga lista diciendo que “los continuos trabajos que se están desarrollando en estas áreas indican que todavía estamos lejos de comprender bien el constructo lingüístico. (...) Cada vez hay más descubrimientos procedentes de áreas como la lingüística aplicada, el análisis del discurso y la adquisición de segundas lenguas” (2002: 95) y admiten que “el hecho de no poder atribuir con exactitud un significado a una nota es el agujero negro de la evaluación” (Alderson / Banerjee, 2002: 101).

A todos estos podríamos añadir otros como las dificultades para validar y obtener fiabilidad en los tests orales que se basan en escalas, criterios y constructos referidos a situaciones y contextos de la vida real, o problemas con los criterios de calificación y con la formación de evaluadores. Pensemos que cada uno de estos aspectos tiene diferentes implicaciones en el desarrollo de nuevos tests.

Uno de los últimos intentos para resolver algunas de estas dificultades ha sido el *Marco común europeo de referencia para el aprendizaje, la enseñanza y la evaluación de lenguas* (2002)¹⁷. Entre sus objetivos trata de establecer unos criterios comunes para la evaluación de lenguas en diferentes ámbitos y en diversas lenguas, que sean reconocidos internacionalmente de forma que sus calificaciones sean homologables e inteligibles y sus resultados interpretables. En él se hace una descripción de numerosos aspectos del uso de la lengua a través de 56 escalas escalonadas en seis niveles.

Su elaboración debe mucho a los deseos de establecer criterios y códigos de comportamiento comunes que comenzó la Asociación Internacional de Evaluación de

¹⁷ Del MCER se trata específicamente en §2.7 y a lo largo de este trabajo, pues es uno de los dos modelos utilizados en el trabajo de campo.

Lenguas (ILTA)¹⁸ y que en los años noventa llevó a ALTE (*Association of Language Testers in Europe*) a establecer un primer código de actuación, el *ALTE Code of Practice* (2001)¹⁹ pretendiendo entre otras cosas establecer esos criterios comunes en todo el proceso de evaluación de idiomas. Para ello, en 2001, creó unas listas de comprobación y análisis de tests con las que pretenden identificar los puntos débiles, las áreas necesitadas de una rápida mejoría y las que precisan de un desarrollo a largo plazo, atendiendo a todos los elementos del proceso: diseño y construcción, redacción, administración, corrección, puntuación, uso de los resultados, análisis de los mismos, redacción de las conclusiones y revisiones necesarias. Así mismo, presta especial atención a la calidad de los servicios que se da a los usuarios y candidatos.

En la evaluación de la lengua oral hay varios enfoques que se han ido desarrollando desde el último tercio del siglo pasado. Estos orientan el hecho de manera distinta según el concepto que tengan de los diferentes elementos que intervienen en la evaluación: la lengua, los sistemas de medir y escalonar su dominio, los objetivos, las formas de calificar, etc. En realidad, no son enfoques excluyentes (aunque puedan parecer a primera vista opuestos) ya que, aunque en las teorías se marquen distancias que aparenten ser insalvables, muchas de ellas se neutralizan en la práctica, a veces por falta de precisión en sus mecanismos de medición -lo que es debido a la dificultad de diseñar herramientas más precisas para evaluar un campo tan multidimensional-, otras veces porque se ponen al descubierto fallas en las teorías y limitaciones que las hacen menos diferentes de lo que parecen.

Sin menoscabo de un análisis más extenso en otros apartados, vamos a hacer una pequeña introducción que nos dé una visión general de cuatro de estas diferencias en los conceptos de evaluación oral, y una quinta en cuanto a los objetivos de corrección, a menudo gramatical, de los exámenes; introducción de 5 elementos que nos ayude a delimitarlos y a entender alguno de los problemas, a veces de interpretación, otras de uso práctico, en que se puede caer por mezclar o confundir conceptos:

- 1 - El concepto de lengua y sus escalas de medición.
- 2 - El objeto de la evaluación.
- 3 - El uso dado a la evaluación.

¹⁸ ILTA, finalmente, editó su *Code of Ethics* en 2000 (ILTA, 2000).

¹⁹ Su primera versión es de 1994.

4 - Los medios y mecanismos de hacer las entrevistas.

5 - Los objetivos de corrección de los exámenes.

1) El concepto de lengua y sus escalas de medición

Observando los conceptos de lengua a evaluar y las diferentes formas de definirla o de delimitar su dominio, podemos trazar dos posiciones o dos modos de evaluación:

a. - Una considera la lengua como un campo vasto cuyo uso se debe delimitar en áreas de dominio predefinidas, pues de lo contrario no es posible abarcarla, medirla y calificarla con objetividad. Esta compartimenta la lengua mediante puntos discretos ya sean desde un punto de vista estructural (gramática, vocabulario, pronunciación) o desde un punto de vista más pragmático (habilidad para iniciar intervenciones, tomar turnos de palabra, hacer uso de la interacción o compensar situaciones problemáticas);

b. - Otra, holística, no usa puntos discretos para compartimentar la lengua sino que intenta dar una calificación general del uso de la lengua sin predefinir sus componentes, porque estima que la existencia de estos no refleja el uso real de la lengua en ningún momento concreto ni ayuda a entender cómo va a ser la lengua que se use en la vida real. Normalmente la describe y delimita mediante definiciones de su uso concreto y contextualizado.

Liskin-Gasparro (1987: 33) manifiesta con un ejemplo práctico estas dos variedades cuando se pregunta cómo se califica en un test oral a los estudiantes a los que se presenta un dibujo de un chico hablando por teléfono en una cabina. El primer candidato -recordando lo estudiado en clase y los ejercicios hechos respecto al uso de los verbos en presente- dice: *El chico habla por teléfono*. Mientras que en el segundo candidato dice: *Juan hablo en teléfono con su amigo Jorge que es en el hospital porque es enfermo*.

Si los calificáramos en una escala de puntos discretos el primero obtendría mejores notas que el segundo por el uso correcto del léxico, de las formas verbales y de la corrección sintáctica. El otro quizás saliera perjudicado por no haber hecho bien su intervención, a pesar del uso correcto del léxico. La primera escala olvida que el lenguaje se aprende para satisfacer determinadas necesidades funcionales, en este caso

explicar un asunto, y que probablemente el segundo estudiante debería ser premiado por un uso creativo de la lengua para expresar sus propios pensamientos en vez de repetir ejemplos de lo aprendido en los ejercicios de clase. Lógicamente si los calificáramos según una escala holística, el segundo candidato saldría premiado.

Muchos han sido los intentos de establecer diferentes tipos de escalas de medición y -con ello- de delimitar la lengua a evaluar (véase §2.5). Las últimas escalas las ha escrito el MCER, que con sus 56 escalas pretende dar una completa visión de la lengua apta para todo tipo de usos y evaluaciones. De ello se tratará en los próximos apartados.

2) El objeto de la evaluación

Este segundo concepto también produce dos modelos de tests: los que se interesan por averiguar y predecir el conocimiento que el hablante tenga de la lengua entendida esta como sistema (en la definición clásica de la lingüística estructuralista o como competencia de uso de la lengua²⁰); y los que se interesan por demostrar el uso concreto de la lengua en contextos sociales o personales considerados útiles, bien para trabajar, para desenvolverse o para interactuar, independientemente de cual sea la competencia que se posea. Cada uno de ellos elabora diferentes mecanismos para evaluar la lengua.

a. - El primero pretende elaborar, mediante los datos obtenidos en el test, predicciones sobre la capacidad de usar la lengua en situaciones reales basándose en que el test es prueba de que esa competencia se posee y que se podrá poner en funcionamiento fuera del test, en la vida real. Trata, por lo tanto, de inferir el conocimiento de la lengua de forma indirecta a través del uso de la lengua que el candidato muestra durante el test. Y, para elicitarse el discurso que sea prueba de ese otro discurso ajeno al test, elabora unos mecanismos que demuestren que la lengua usada en el test es producida efectivamente por las habilidades que el candidato posee, que es lo que en última instancia se le quiere evaluar.

Este modelo se centra en la autenticidad de las tareas y de las inferencias que con ella se realizan; es decir, en la capacidad que muestren los sistemas de calificación para relacionar las pruebas obtenidas mediante el test, con el uso de la lengua en la vida real.

²⁰ Bachman aclara (1988: 161) que él se refiere a *competencia* en el sentido de Hymes, no limitado a la competencia lingüística definida por Chomsky, sino relacionada con todo lo que el hablante tiene en la mente para hacer uso de la lengua.

b. - El segundo se conforma con probar que el candidato posee unas habilidades en contextos concretos y que el uso directo de la lengua durante el test es prueba de ellas. Entonces, este modelo se interesa en buscar mecanismos que prueben que esa lengua es muestra de la mejor lengua posible de los candidatos, que es lo que se quiere evaluar en realidad.

Para este modelo la prueba del funcionamiento en la vida real se obtendría mediante el test mismo, considerando (sin entrar a definir las competencias existentes, las cuales también se acepta que existen), también de forma predictiva, que lo usado en el test se podrá volver a usar en otros momentos y situaciones similares. Estos tests también buscan una muestra de lengua apta para elaborar predicciones sobre su uso fuera del test, solo que utilizan mecanismos diferentes²¹.

Cabría plantearse si estas dos concepciones delimitan tests que son realmente excluyentes y en qué situaciones se solapan o hablan en realidad de lo mismo. Porque muchos de los tests existentes hoy en día basados en puntos discretos que intentan medir el conocimiento de elementos como el dominio del vocabulario, de la gramática, etc. (que son parte de ese sistema de la lengua al que pretenden acceder indirectamente), incluyen otros elementos pragmáticos y de compensación y adecuación sociolingüística que relativizan el conocimiento del sistema de la lengua y se fijan en el uso que de este se hace ante situaciones concretas. Ya que en las teorías pragmáticas actuales cada vez se tiene más en cuenta a los elementos contextuales como propios del sistema de la lengua y las condiciones de significación y sus elementos constitutivos pasan a ser estudiadas como parte de ese sistema general que es la lengua.

3) El uso dado a la evaluación

El tercer concepto se diferencia por el campo de actuación para el que se realiza la evaluación y atañe también al uso que se haga de sus resultados. En este sentido podemos diferenciar cuando se hace con fines académicos o cuando se hace con otros objetivos sociales.

²¹ Veremos en su momento, al explicar el modelo de la OPI (§2.6), cómo hay herramientas que ayudan a encontrar muestras de lengua aptas para este modelo de evaluación, evitando dar calificaciones erróneas por usar métodos de elicitación deficientes.

a. - Cuando se hace con fines académicos es para medir el estado de la lengua de un candidato (sea el que sea y desde el concepto que sea de lengua) con el fin de describir las actuaciones futuras que se puedan llevar a cabo para mejorarla. O con el de elaborar mecanismos que ayuden a mejorar la enseñanza, la calidad del aprendizaje o el aprendizaje en sí (tests de diagnóstico, tests de clasificación, tests de rendimiento, etc.).

b. - La evaluación con fines no académicos se desinteresa del contexto de aprendizaje, del diagnóstico de la lengua y del conocimiento pasivo que el candidato tenga. Se centra en medir la lengua para fines personales, sociales o laborales y en definir su uso en situaciones determinadas con el fin de demostrar que el candidato sabrá realizar fuera del test el mismo tipo de actividades que haya mostrado que sabe hacer durante el test.

Aunque estos dos modelos tampoco sean excluyentes²², suelen emplear conceptos o enfoques sobre la lengua diversos que llevan a resultados también diferentes.

4) Los medios y mecanismos de hacer las entrevistas

El cuarto elemento basa sus diferencias en la concepción material del test y sus mecanismos o herramientas y en los modos de operar para obtener la información; es decir, en los métodos de elicitar y obtener las muestras de lengua y en el papel que juega el evaluador como motivador del discurso.

Aquí encontramos la opción de no intervenir aun cuando esto pueda penalizar la actuación de un candidato; y la que propugna perseguir los objetivos de la evaluación mediante un protocolo de intervención que logre extraer de cada candidato las mejores muestras de su habilidad. Estas diferencias inciden directamente en la concepción de la entrevista oral como un diálogo natural o no y ofrecen técnicas diversas y excluyentes²³:

a. - El primer modelo de entrevista utiliza técnicas que pretendan acercarse a lo que es una conversación lo más natural posible, creyendo que ese es el mejor o más fiable

²² No son excluyentes porque se puede evaluar conjuntamente con fines académicos y laborales (para establecer un programa de formación de técnicos, de profesores, etc.) e, igualmente, una evaluación con fines sociales puede tener usos académicos, sirviendo como base para tomar decisiones de selección para cursos, o como dato homologable para cualquier otro fin.

²³ Véase §2.4, dedicado a la naturalidad de las entrevistas.

camino para demostrar la existencia de las habilidades. Este enfoque persigue la naturalidad absoluta del discurso y critica las entrevistas que no contemplan en sus objetivos la variedad de usos y aspectos que se han delimitado en los últimos 40 años en un buen número de estudios de pragmática como son la toma de turnos de palabra, el análisis de la interacción, el reconocimiento de intenciones, la negociación de significados o la diferencia entre conversación y actividades conversacionales.

b. - La segunda utiliza técnicas más planificadas y se aleja de la naturalidad de un diálogo real pretendiendo ganar -con esta pérdida de pragmática- eficiencia en sus resultados y sistemas de calificar. No presta tanta atención a la variedad de elementos sino que se centra en probar que lo que se logra elicitar realmente es prueba del mejor hacer de los candidatos.

5) Los objetivos de corrección de los exámenes

Y el quinto elemento marca diferencias en los objetivos de las mediciones de la lengua, en si estas pretenden averiguar si lo que el hablante sabe es parte del conocimiento explícito o implícito y qué cantidad de gramática conoce de forma implícita. La distinción la hacen Shohamy (1988: 170) o Ellis (2001) al hablar de estas dos formas de acceder a la gramática y al conocimiento de la lengua.

El conocimiento implícito es el que se produce en el discurso no planeado y el explícito el que surge en condiciones sin presión ante las que se puede recurrir a otras formas de conocimiento y reglas de gramática aprendidas.

El problema para Ellis está en saber qué tipo de acceso permiten los tests y qué conocimiento están midiendo (Ellis, 2001: 253). Para él, generalmente, si no siempre, se preocupan por medir el conocimiento explícito, ya que es difícil diseñar otros tests que logren inferir el conocimiento implícito. Y ello supone un problema de medición, ya que los estudios de adquisición de las lenguas manifiestan que se puede mostrar un determinado grado de corrección y, sin embargo, estar en un estadio inferior de adquisición (Ellis, 2001: 255-56).

La corrección no muestra el grado de conocimiento. Medir el conocimiento de la lengua por el grado de corrección al usar el conocimiento explícito puede llevar a errores de medición y de predicción sobre el verdadero dominio de la lengua. Así se puede penalizar a candidatos con un mayor desarrollo de sus interlenguas frente a otros

que sólo por una mayor corrección obtuvieran mejores notas (Martínez Baztán, 1995: 366).

A estos cinco enfoques sobre el concepto de la evaluación general, podemos añadir otros tres con diferentes perspectivas sobre el modo de construcción de los tests (Alderson / Banerjee, 2002; Fulcher, 2003).

Hasta ahora los principales modos de evaluar la lengua hablada han sido conductistas y se han basado en la premisa de autenticidad de que el modo de evaluar y el resultado obtenido se asemejen a la lengua usada en la vida real. Así son muchos de los que se utilizan desde los años 50, junto a los constructos lingüísticos derivados del modelo de lengua de Bachman (1990: 84) y Bachman / Palmer (1996: 68). Estos se caracterizan por describir una serie de tareas o ítems que provoquen un estímulo lingüístico, una actividad o comportamiento observable del cual se deduzca la existencia de una habilidad. Las tareas o ítems se diseñan con el objetivo de provocar una respuesta determinada y se valoran en la medida en que son capaces de provocarla de acuerdo con el constructo que se desee evaluar (Alderson / Banerjee, 2002: 99).

Pero esta no es la única perspectiva.

Fulcher (2003: 62) apunta tres: el nuevo conductismo, la teoría de los rasgos lingüísticos, y el interaccionismo.

- El nuevo conductismo estima que el constructo lingüístico viene definido principalmente por tareas (Fulcher, 2003: 62). Igual que el conductismo lingüístico, que cree que las tareas y sus contextos están indisolublemente ligadas a la habilidad que provocan.

- La teoría de los rasgos lingüísticos estima que los tipos de tareas utilizados son irrelevantes para el constructo a medir y que los rasgos lingüísticos son independientes de los contextos en que se generan. Así, las notas de un test se atribuirían fácilmente a las características del candidato, a su conocimiento y a procesos internos antes que a características del contexto o del lugar. Alderson / Banerjee explican esta perspectiva como una de las que más tradición tienen en el ámbito de la evaluación educativa, aunque la consideran simplista:

Pero esto es demasiado simplista. Sabemos por la lingüística aplicada y los estudios de adquisición de segundas lenguas, que el uso de la lengua no es independiente del contexto en que surge. La importancia del contexto se reconoce como un factor importante en la actuación así como en la valoración de esa lengua. Por ello por sí sola esta perspectiva parece inadecuada (Alderson / Banerjee, 2002: 100).

- El interaccionismo, por otra parte, cree que ambas perspectivas son erróneas y que hay que tomar elementos de todas partes. Algunas características de la habilidad lingüística proceden de los rasgos, otras del contexto y otras de la interacción.

La actuación lingüística es un signo de rasgos latentes que surgen en la interacción con características del contexto (...) Aunque lo difícil consiste en averiguar la medida en que se presentan, cómo los rasgos y el contexto interactúan y bajo qué circunstancias, para combinarlos en el diseño de un test (Alderson / Banerjee, 2002: 100).

Por otra parte, Fulcher (2003: 62), quien diferencia en la práctica los elementos (rasgos o tareas) y los tests en los que entran a formar parte, aclara que hay tests en los que las tareas son parte del constructo, como los tests para fines específicos, pero que también hay elementos de las tareas que no influyen en las posibilidades de generalizar las habilidades, esto es, que no impiden que se puedan generalizar a otras situaciones más allá de las directamente evaluadas en el test; por lo que también se pueden hacer tests en los que la influencia de la tareas quede minimizada.

Al diseñar un test se deben especificar adecuadamente y de la mejor forma posible, siempre dependiendo de las necesidades del momento y sus objetivos de evaluación, todos los elementos que se aspire a calificar; ya que dejarlos indeterminados, sea por no definir bien las habilidades a medir, por no aclarar cómo se interpretan las actuaciones de los candidatos en relación a la existencia de las habilidades, o por no controlar los efectos de factores que no se desean evaluar, hace que el resultado reste validez al test.

No obstante, y sin olvidar este problema de falta de definición, se debe tener en cuenta que los tests no son absolutos y siempre van a ser mejorables, bien porque se hallen nuevos datos en investigaciones sobre desarrollo o adquisición de la lengua, bien por el hecho de que su diseño complejo y costoso puede ser mejorado tomando en cuenta elementos que quizás no hayan sido bien considerados o se hayan minimizado en su diseño.

Esto supone una limitación de los resultados, que hay que tener clara y saber manejar para evitar hacer un mal uso de los datos y extraer de ellos conclusiones inadecuadas, exageradas, inexistentes o falsas.

Las limitaciones más importantes de las que hay que ser consciente son (Bachman, 1986: 380; Bachman, 1990: 30-40; Bachman, 2004: 14):

1- La falta de especificación de los contenidos del test, entre los que destacan las escalas de evaluación y gradación del dominio.

2- Las derivadas de una inadecuada forma de realización de entrevistas.

3- La dificultad de observación de las habilidades y -derivado de ello- la dificultad de interpretación de los datos obtenidos.

Del primero, la falta de especificación de los contenidos, hablamos también al explicar la importancia de diseñar tests con buenas tareas, que sean facilitadoras de las habilidades a evaluar; y volveremos a hablar al tratar de los sistemas de definir la habilidad lingüística mediante el uso de las escalas de medición y criterios de calificación de la lengua (§2.2). Al respecto, Bachman (2004: 28) dice que, además de la propia habilidad lingüística, es tal la variedad de factores que afectan a la actuación lingüística, como el estilo cognitivo, la ansiedad, la afectividad o el conocimiento del tema, que es imposible tenerlos todos en cuenta al diseñar el test o al interpretar los resultados, lo cual, de alguna manera, indetermina las interpretaciones.

Del segundo, las limitaciones derivadas de una inadecuada forma de realización de entrevistas, se hablará en los apartados dedicados a la naturaleza de las entrevistas y en los dedicados a los sistemas de elicitación del discurso (§2.6.2- 4) y el estrés comunicativo (§2.4.3- 3).

Respecto al tercer foco de limitaciones, la dificultad de observación de las habilidades, aunque se relacionan con las demás y a veces son difíciles de diferenciar, podemos decir con Bachman que proceden, en principio, de que las habilidades lingüísticas no se pueden observar de forma directa y ni siquiera la observación de la actuación viva de los hablantes es clara; ya que esta está sometida a variaciones derivadas del contexto, de las tareas que se estén realizando, de la hora del día, la alerta mental y otras características cognitivas y personales.

Bachman cita cinco elementos en relación a estas limitaciones (Bachman, 1990: 32-39; Bachman, 2004: 28-29):

1. Dependencia de un método de observación indirecto²⁴.
2. Deficiencias de la actuación y obtención de datos.
3. Imprecisión en la interpretación de los datos.
4. Subjetividad general de los tests y sus mecanismos²⁵.
5. Relatividad de las habilidades a evaluar²⁶.

El MCER, vamos a verlo en esta investigación, carece de un protocolo para realización de entrevistas y eso va a dificultar una adecuada y fiable evaluación y a obligar a los usuarios a llegar a acuerdos que no son fáciles de obtener. Igualmente, como se verá en §3.1 y en el capítulo 4, muestra bastantes dificultades para una adecuada interpretación de los datos. Algunas limitaciones se pueden disminuir mediante el diseño de otros mecanismos de los tests que contribuyen a dar objetividad. Trataremos de ello en §2.3, al explicar los procesos de validación de tests. Más adelante, (§2.6.2- 4), veremos diferentes sistemas de elicitación del discurso y sus efectos o problemas²⁷.

Todas las limitaciones de la evaluación de lenguas llevan a la necesidad de delimitar los numerosos aspectos que influyen en su realización y, en última instancia, a ser conscientes de qué tipo de herramientas, de entre todas las que se pueden utilizar,

²⁴ También el MCER dice claramente que “nunca se pueden evaluar las competencias directamente. Hay que basarse siempre en una serie de actuaciones a partir de las cuales se intenta generalizar una idea del dominio lingüístico (...) Todas las pruebas evalúan solo la actuación aunque partiendo de esa evidencia se puede intentar extraer inferencias respecto a las competencias que subyacen” (Consejo de Europa, 2002: 188).

²⁵ El MCER hace referencia a esta subjetividad en su apartado 9.3.8. No obstante, también dice que “el asunto de la subjetividad y la objetividad es bastante más complejo” (MCER, 2002: 9.3.8), ya que todo el proceso de diseño y selección de ítems y opciones de respuestas está basado en las opiniones de los redactores de forma igualmente subjetiva. Nosotros también consideramos que todos los elementos que intervienen en el diseño de un test son importantes fuentes de subjetividad y que -en los tests orales- la descripción de los protocolos de actuación de los evaluadores es importantísima, apoyada en su adecuada formación y en un análisis estadístico de sus resultados (véase §2.4).

²⁶ Esta relatividad de definir el objeto se hace más compleja al tratar de definir la lengua de las comunidades de residentes extranjeros. Esto también lo señala Winford argumentando que entre las comunidades de no nativos e inmigrantes las variedades se adaptan y adquieren su propia versión de la lengua extranjera (Winford, 2003: 237).

²⁷ Véase también el apartado dedicado al estrés comunicativo, §2.4.3- 3.

estamos manejando y qué consecuencias conllevan²⁸. Sin que se puedan hacer generalizaciones entre diferentes métodos ni comparaciones simples.

Todas estas precauciones que venimos señalando no deben apartarnos del intento de realizar una buena evaluación oral, tanto si se hace con fines académicos como con funciones laborales, sociales o jurídicas. Pues la evaluación oral es un tipo superior de evaluación que recoge a todas las demás, la única que tiene la capacidad de decir, si se hace adecuadamente, cómo es efectivamente la capacidad para comunicarse en una lengua extranjera desde cualquier punto de vista, la expresión y la comprensión en general o con fines específicos, la interacción y la toma de turnos de palabra, la compensación, la competencia gramatical y léxica, etc. Aunque para ello es necesario diseñar tests orales adecuados a cada situación y a los candidatos, ya que una conversación natural por sí misma, por muy satisfactoria que parezca, no es ni buena ni fiable (Bordón, 2006: 62; Bachman, 1990: 21, 34).

Como hemos visto a lo largo de este apartado, la evaluación se compone de productos pero es también, y sobre todo, un proceso que comienza con la detección de unas necesidades educativas o sociales a evaluar y termina con la publicación de las calificaciones o notas debidamente explicadas, con el objeto de que no lleven a errores de interpretación y uso. Este proceso tiene efectos a largo plazo sobre otros elementos generales como el sistema educativo o su metodología, ya que ayuda a definir lo que se espera de un alumno y con ello colabora a establecer los objetivos e incluso determina en gran medida los medios de lograrlos, o incluso sobre el uso social, laboral o legal que se le dé.

Y para que todo esto funcione bien y adquiera credibilidad, hay que asegurarse de su validez y fiabilidad. A esto vamos a dedicar los siguientes apartados.

Empezamos por Bachman / Palmer, (§2.2), por ser los autores más influyentes en el campo actual de la evaluación de lenguas, exponiendo su modelo, en el que dan a la lengua y a las tareas un papel primordial como prueba de que el test sirve efectivamente para determinar el dominio de los candidatos.

²⁸ En §2.3 se explican las técnicas principales que, fundamentalmente desde una perspectiva cualitativa y otra estadística, permiten llevar a los tests la objetividad necesaria para que estas limitaciones -si no se pueden reducir al máximo- tengan, al menos, poca influencia en los resultados o en la capacidad de medición de los tests.

En un apartado posterior (§2.3) entraremos en otra descripción, cualitativa o estadística, de demostrar que un test es efectivo, desde la óptica de Alderson (1998) y del propio Bachman (2004).

2.2 Validación de tests a través del concepto de lengua utilizada en su constructo: el modelo de Bachman / Palmer

En el apartado anterior terminamos hablando de la necesidad de validar un test siendo conscientes de sus limitaciones. Vamos a comenzar por explicar una forma de validación cualitativa, la de Bachman / Palmer (1996) antes que otros modelos, también cualitativos y estadísticos (§2.3).

Bachman / Palmer (1996) han conseguido que se considere el proceso de construcción de un test desde su comienzo, considerando en primer término la lengua que se quiere evaluar y haciendo de ella el centro de todo. Ellos han puesto “énfasis a la centralidad del constructo que se quiere medir (haciendo) que los evaluadores consideren lo que saben sobre la lengua y de su habilidad para usarla, con lo que los tests no se relacionan solamente con los elementos sicométricos que los analizan sino con los conocimientos sobre la lengua: los evaluadores deben saber lingüística aplicada, estar al tanto de las últimas teorías sobre la descripción lingüística, sobre la adquisición de lenguas y sobre su uso” (Alderson, 2002: 80).

La idea principal de la evaluación de lenguas de Bachman / Palmer parte de la asunción de que la lengua que se produce en un test podría no ser igual a la que se usa en la vida real, y de la necesidad de asegurarse de que los mecanismos del test intenten que eso suceda en la menor medida posible, garantizando que el discurso elicitado en las entrevistas orales sea una muestra de la lengua real del hablante que efectivamente use y domine en su vida y no solo una lengua artificial producida en el test, por el test o para el test. Por ello, para estos autores la autenticidad es un aspecto sustancial de la lengua a medir, y se garantiza seleccionando un área de lengua acotada mediante tareas y situaciones reales y demostrando que la actuación de un candidato al realizar las tareas del test es similar.

A este concepto se suma otro que Bachman interrelaciona hábil e inteligentemente con el fin de organizar un sistema fiable de medición de la lengua: la necesidad de

asegurarse de que los datos obtenidos mediante el test se pueden interpretar correctamente.

Tendemos a creer que una nota obtenida en un test preparado para evaluar una habilidad es directamente una indicación de esa habilidad. Pero esto solo será cierto si del test, del sistema determinado para obtener los datos (tareas, actividades, método, etc.) y de los sistemas de calificación determinados al efecto, se pueden deducir interpretaciones válidas sobre el uso externo de la lengua. Por lo tanto, hay que asegurarse de que la nota mide el uso real de la lengua y no una habilidad de hacer bien los tests²⁹.

La capacidad de un test para obtener y ofrecer datos fiables no es una consecuencia directa del test sino de su validez, fiabilidad y autenticidad³⁰. Es de suma importancia conseguir que todos los elementos que intervienen en el diseño y desarrollo de un test e interfieren en sus resultados lo hagan de forma positiva, sirviendo para facilitar la actuación de los candidatos y obtener la mejor muestra de sus habilidades lingüísticas. No se puede hacer un test y suponer sin más que sus resultados suponen un conocimiento o una actuación futura de los candidatos, sino que hay que demostrar que con esos resultados se pueden tomar esas decisiones. Y para ello hay que establecer una relación entre ambas cosas: el test y las interpretaciones de sus resultados (Bachman, 2004: 257).

En resumen, y para poder avanzar en las explicaciones profundizando en estas ideas que venimos desgranando, la autenticidad depende del control y definición de estos tres elementos:

1. El grado de definición del concepto de lengua (definir un concepto de lengua).
2. La necesidad de definir tareas para el test en las que se use ese tipo de lengua que se desea evaluar (definir las tareas apropiadamente).
3. Obtener evidencias de que el sistema de calificaciones usado tras el test refleja el área de lengua que se quiere medir (usar un buen sistema de calificaciones).

²⁹ Así, por un lado está la validación en sí del test y, por otro, la de las decisiones que se toman con sus resultados (Bachman, 2004: 257). Este planteamiento sobre la legitimidad del evaluador para tomar decisiones abre, según Leung / Lewkowicz (2000: 223) otras importantes cuestiones de ética profesional y uso democrático de test de las que no nos vamos a ocupar ahora.

³⁰ Estos conceptos de validez, fiabilidad y autenticidad se explican en §2.3. Por ahora los aclaramos diciendo que el primero es asegurarse de que un test mide lo que efectivamente pretende, el segundo asegurarse de que lo mide sistemáticamente y el tercero que lo que mide es efectivamente la lengua real y no una artificial producida por el test.

Vamos a explicarlos, brevemente, a continuación.

1. Definir un concepto de lengua

Cada test se diseña para medir unas características y habilidades específicas. Por ello la definición de su constructo, esto es, las habilidades y características de la lengua³¹ que se pretenden medir, debe hacerse de forma lo suficientemente clara como para que quede diferenciado de otras características que también pudieran surgir a la vez pero que no entran en lo que el test desea medir.

La descripción de la lengua que ha resultado ser más influyente es la de Bachman (Bachman / Savignon, 1986: 381; Bachman, 1990: 87; ampliado posteriormente, Bachman / Palmer, 1996: 66-78). Esta ha interrelacionado varios de los conceptos que al respecto se estaban manejando durante los años 80, especialmente los de competencia comunicativa y habilidad lingüística y ha proporcionado un modelo para numerosos exámenes.

El marco que proponen Bachman / Palmer (1996: 77) se compone de dos elementos (véase cuadro 2.2 A) y sirve de guía tanto para el proceso de diseño de tests como para analizar tests existentes. Estos componentes son, por un lado, los que dan forma al conocimiento organizativo y, por otro, los elementos del conocimiento pragmático.

Entre ellos la competencia sociolingüística se introduce como un subcomponente de la competencia pragmática, pero la competencia estratégica la consideran como una competencia ajena a las competencias lingüísticas cuya función es dar cuenta de cómo se relacionan los demás componentes para elaborar los mensajes adecuados a cada situación, mediante una serie de funciones de valoración, planificación y ejecución, que contribuyen a conformar la habilidad lingüística.

³¹ Como resumen Bordón (2006: 25-34) o Purpura (1999: 19-20) la definición de la habilidad lingüística ha dado importantes avances descriptivos durante el siglo XX desde el estructuralismo norteamericano, pasando por las teorías chomskianas sobre el innatismo del lenguaje humano o las aportaciones comunicativas de Hymes, la pragmática y Canale, hasta la propuesta de Bachman que aquí explicamos. Y añadimos, pasando por el empuje funcional de los trabajos del Consejo de Europa.

Cuadro 2.2 A. Componentes de la habilidad lingüística

CONOCIMIENTO ORGANIZATIVO	gramatical	vocabulario
		morfología y sintaxis
		fonología y ortografía
	textual	cohesión
		organización retórica y de la conversación
CONOCIMIENTO PRAGMÁTICO	funcional	funciones ideales
		funciones manipulativas (instrumentales, reguladoras e interpersonales)
		funciones heurísticas
		funciones imaginativas
	sociolingüístico	variedades y dialectos
		registros
		naturalidad y expresiones idiomáticas
		referencias culturales y figuras del discurso
COMPETENCIA ESTRATÉGICA		establecimiento de metas
		valoración
		planificación

(Bachman / Palmer, 1996: 77).

Leung / Lewkowicz (2000) explican que este modelo que considera la habilidad como una interacción con otros elementos que intervienen e influyen en el test, es el más ampliamente aceptado y usado en la evaluación de lenguas (Leung / Lewkowicz, 2000: 214). Pero no es el único sistema. Hay otras formas de definir el constructo lingüístico que se desee evaluar que no tienen por qué ser tan elaboradas. Alderson expresa, por ejemplo, que esto se puede hacer también a través de las especificaciones del test y la selección de materiales, las tareas, los conceptos que traslucen y las inferencias que se sacan de esos conceptos reflejadas en las notas (Alderson, 2000: 138).

Y tampoco está exento de críticas. Figueras (Figueras, 1997), al igual que Bordón (2006: 37), McNamara (1995) o Purpura (1999: 20), considera el modelo de Bachman como el más importante de los definidos hasta el momento (integrando a los de Hymes, Grice o Canale) pero, matizando estas opiniones, apunta que tal definición de la competencia:

(No aclara) el problema de identificar y definir -(ni) para el docente (ni) para el evaluador- la capacidad o habilidad que permite a un aprendiz de lengua combinar sus conocimientos de lengua para crear discurso. Bachman aclara el uso del término

competencia estratégica al definirlo como elemento que ayuda a caracterizar la capacidad mental de implementar los demás componentes de la competencia lingüística al usar la lengua en un contexto de comunicación. Pero esta aclaración es insuficiente y no facilita la tan necesaria definición (Figueras, 1997: 67).

También McNamara (1995) critica este modelo considerando que ignora la dimensión social de la habilidad lingüística y que, al centrarse sólo en aspectos psicológicos del lenguaje, no atiende a la influencia que tienen, en la formación del discurso del candidato, tanto su interlocutor como otros factores no cognitivos (McNamara, 1995: 173).

Finalmente, Widdowson encuentra muchas limitaciones a este modelo, así como a los anteriores de Hymes y Canale / Swain, respecto a la posibilidad de dar cuenta de la competencia comunicativa de una manera que sea efectiva tanto para enseñar como para evaluar (Widdowson, 2003: 170). Para él la competencia comunicativa, tal y como se define en estos modelos, incluido el de Bachman (1990) y Bachman / Palmer (1996), “permanece elusiva y no puede ser medida tal y como se propone en esos modelos” (Widdowson, 2003: 171). El sistema de Bachman / Palmer divide un sistema complejo (la competencia lingüística) en múltiples componentes estáticos y así no se pueden explicar las relaciones dinámicas que se dan en la comunicación.

Tampoco explica cómo se relacionan los diferentes elementos de la competencia general entre sí:

No hay manera directa de relacionar las habilidades de la actuación lingüística (lo comunicativo) con la competencia, aunque se diga, claro, que es la competencia estratégica la que hace eso y hace las conexiones contextuales necesarias para realizar la comunicación. Pero eso lleva a la necesidad de volver a definir esa competencia y cómo funciona (Widdowson, 2003: 169).

Para Widdowson “la competencia comunicativa no puede, como tal, ser medida. Quiero decir que los tests comunicativos son en principio imposibles, y eso es por lo que no sorprende que sean tan difíciles de diseñar. (...) No se puede evaluar la habilidad para comunicarse (...) Lo único que se puede enseñar, y evaluar, son aspectos de ella” (2001:18).

2. Describir las tareas apropiadamente

Para que las tareas sean efectivas hay que hacer que sean realistas y que provoquen el mismo tipo de lengua que se quiere evaluar fuera del test. El proceso debe contar con unas garantías que lo hagan viable; para lo cual, una vez descrito y delimitado el dominio y modelo de lengua usado en la vida real, se procedería mediante los siguientes pasos:

- Encontrar contextos, situaciones, tareas y actividades de la vida real en los que se utilice el dominio de la lengua que luego pretende provocar y calificar el test.
- Inventar o diseñar tareas para el test, de similares características a las de la vida real y que puedan producir su mismo tipo de lengua.
- Definir a los candidatos y sus características de forma que nos aseguremos de que ese tipo de lengua de la vida real que pretendemos provocar y medir en ellos les es habitual, y que el tipo de situaciones y tareas descritas les son igualmente adecuadas³².

Para proceder a la elaboración de tareas para tests o mejorar las ya existentes, Bachman / Palmer (1996: 49 y 108) han propuesto un cuadro de análisis de tareas que facilita la comparación de las tareas y su lenguaje con las tareas y lenguaje del test. El cuadro permite comparar las tareas de la vida real y las del test y analizar los puntos en que más se distancien, lo que aclara qué elementos deben rediseñarse o variarse. Se centra en cinco aspectos (véase cuadro 2.2 B): el lugar, el formato del test, las características del input de las tareas, las características de la respuesta esperada, y la relación entre las instrucciones del test y la respuesta dada.

³² Es fácil entender que estos autores consideran que las tareas son uno de los puntos fundamentales del proceso de evaluación oral. Ellas provocan la lengua del candidato, que es lo que se analiza, se califica y se utiliza para elaborar las predicciones sobre la existencia de las habilidades que son objeto de la evaluación. Esto implica dos cosas: que evaluamos la lengua real del candidato indirectamente, a través de la lengua producida en el test; y que no evaluamos a la persona que habla sino el discurso surgido mediante las tareas. Por ello hay que tener en cuenta que si deseáramos evaluar otras habilidades del candidato seleccionaríamos otras tareas que producirían otros datos y nos llevarían a otras conclusiones.

Cuadro 2.2 B. Cuadro de análisis de tareas³³

Elementos	Lengua real	Lengua del test
1. Características del escenario		
Características físicas		
Participantes		
Tiempo de la tarea		
2. Características del test		
<u>Instrucciones</u> Lengua usada (L1, L2) Canal (visual, oral, auditivo) Explicación de procedimientos		
<u>Estructura</u> Número de partes Diferenciación de las partes Secuenciación Importancia de cada parte Núm. de ítems/tareas por parte		
Duración (limitada o libre)		
<u>Sistema de puntuación</u> Sistema de corrección Procedimientos para cada ítem Explicación de criterios		
3. Características del input		
<u>Formato</u> Canal (oral, visual) Forma (lengua o no) Lengua usada (L1, L2) Extensión Tipo (ítem, instrucciones) Tiempo de procesamiento Fuente (en vivo, grabado)		
<u>Lengua del input</u> Características organizativas gramaticales y textuales.		

³³ Este cuadro lo usamos en el Apéndice 1 donde se muestra la ficha con las características del test diseñado para evaluar mediante el MCER.

Características pragmáticas funcionales y sociolingüísticas Características de contenido		
4. Características de la respuesta esperada <u>Formato</u> Canal (oral, visual) Forma (lengua o no) Lengua usada (L1, L2) Extensión Tipo (Multirrespuesta, cerrada, abierta) Tiempo de respuesta Fuente (en vivo, grabado) <u>Lengua del input</u> Características organizativas gramaticales y textuales Características pragmáticas funcionales y sociolingüísticas Características de contenido		
5. Relación input/respuesta Reacción Profundidad Dirección		

(Bachman / Palmer, 1996: 108).

Este cuadro es de enorme ayuda pues su objeto no es solo pretender naturalidad en el test y comprobar su validez, sino -imprescindible para que este tenga resultados efectivos- asegurarse de que con los resultados obtenidos se puedan establecer inferencias adecuadas respecto a los candidatos y su comportamiento lingüístico más allá del examen y lo que se defina un sistema de calificaciones válido que permita extraer razonablemente esas inferencias.

Para terminar este punto, diremos que estos autores definen las tareas como uno de los componentes responsables de la *autenticidad* conjuntamente con la interacción de otros dos elementos: el candidato (sus características) y el contexto del test. Bachman (1991: 687) considera que para que la evaluación avance es muy importante el desarrollo de una teoría que considere la competencia lingüística como algo formado por muchos componentes. Ello supone valorar la influencia del método del test y de las

características del candidato en su propia actuación, reconociendo la posibilidad de un conjunto de elementos que intervienen en la actuación (contexto, instrucciones, input, etc.) y la conforman.

Nosotros creemos -con Lazaraton (Lazaraton, 2002) y con el sistema de evaluar de las OPI- que estos tres elementos se deben ampliar con un cuarto elemento, la intervención del evaluador que, como se explicará más adelante (§2.4), consideramos decisiva, tanto en la modulación del test y sus tareas, como en la motivación del comportamiento del candidato y de la producción de su discurso mediante el manejo de los temas y preguntas.

Así, consideramos que un test oral es la interacción de varios factores internos y externos entre los que destacan:

- a. La habilidad en el uso de la lengua en situaciones determinadas.
- b. Las tareas que provocan el uso de la lengua en esas situaciones.
- c. Los conocimientos del candidato y otras características personales, los cuales se intenta que no influyan en la producción de la lengua, ni en su medición.
- d. El comportamiento del evaluador para elicitar la lengua mediante las tareas seleccionadas y el uso de sus preguntas.
- e. El método de realización del test, lo que abarca desde la elección de si es oral, escrito o de otro modo, hasta su estructura y constructo, y el valor que cada parte juega en la consecución de las muestras de lengua.
- f. El sistema de interpretar y calificar las muestras de lengua obtenidas.

3. Usar un buen sistema de calificaciones

Los criterios definidos en las escalas de un test deben ofrecer suficientes puntos de referencia como para que permitan interpretar la actuación dada en el test (Bachman / Savignon, 1986: 386). Esto es, los sistemas de calificar deben permitir extraer conclusiones adecuadas respecto al uso de la lengua³⁴. No vamos a explicar más de este punto. Es claro, aunque puede resultar complicado.

³⁴ ALTE en sus *Principles of Good Practice for ALTE Examinations* (ALTE, 2001: 11) considera que además de los criterios y escalas, los examinadores deben calificar con exactitud y consistencia, como medio de asegurar que las calificaciones sean fiables.

Hasta aquí el procedimiento que Bachman / Palmer utilizan para delimitar la lengua y asegurarse de que el test la evalúa efectivamente, garantizando autenticidad y evitando que elementos ajenos afecten al resultado del test. En el apartado siguiente se explican otras maneras, unas cualitativas, otras estadísticas o cuantitativas, de garantizar que un test mida y evalúe efectivamente lo que sus autores se hayan propuesto al principio. Para ello nos servimos fundamentalmente de lo que al respecto dicen los protocolos de ALTE (ALTE, 2001), Alderson (1995), Fulcher, (2003) y el propio Bachman (1990 y 2004).

2.3 Técnicas cualitativas y cuantitativas de validación de tests

Un test no es sólo el producto con el que se obtienen los datos a calificar sino que es, asimismo, la escala de medición utilizada para calificar los resultados, las pruebas de homogeneización de criterios de los evaluadores que los califican e incluso la forma en que se ofrece a los candidatos y a la sociedad en general los resultados y certificaciones. Su diseño es un proceso en el que se deben tener en cuenta muchos elementos: las características de la población a la que va destinado; el tipo de lenguaje que se pretende producir; las tareas facilitadoras de ese lenguaje; y los métodos y materiales que se utilicen, que deben ser aptos para todos los candidatos. Y todo ello debe someterse a pruebas que permitan demostrar que funciona y que es efectivo para los objetivos planteados.

Eso es la validación; y sus técnicas y procedimientos son variados. Vamos a verlos ahora para utilizarlos posteriormente en la validación de nuestro nuevo test, el test del MCER (véase §3.6). Pero como la validez y sus procedimientos de mostrarla es un asunto complejo, debemos aclarar, antes, los siguientes temas:

a. Las diferentes formas que hay para establecer y averiguar la validez de un test y los procedimientos empleados en ella:

- Validez externa, con preguntas a los candidatos sobre sus reacciones y actitudes ante una prueba.

- Validez interna, mediante opiniones de expertos que determinen qué se está midiendo con los ítems seleccionados.

- Validez de constructo, para determinar si las diferentes partes de un test miden cosas diferentes o no y, en última instancia, qué es lo que se mide a través del test³⁵.

b. El concepto de correlación, es decir, el método de análisis y comparación entre las diferentes pruebas de que se compone un test, que permite establecer y asegurar que cada una de ellas mide habilidades diferentes o las subhabilidades deseadas³⁶.

c. El concepto de fiabilidad, prueba de que un test sirve para medir efectivamente los conocimientos para los que estaba en principio planeado y de que sus resultados no dependen de otros factores, salvo el aumento de la habilidad o de la capacidad que se desea medir.

Comencemos por una situación concreta, para ayudarnos a entender los conceptos. Luego los explicaremos de forma teórica.

Consideremos un test de evaluación oral para alumnos extranjeros adolescentes que están estudiando un semestre en un país, aprendiendo la lengua en cursos intensivos con actividades extraescolares y conviviendo con familias. Si queremos saber en qué medida los estudiantes han alcanzado un grado de habilidad tras los cursos, podremos plantearnos varias cosas. Podríamos evaluar el rendimiento del curso suponiendo que al superarlo habrán adquirido las habilidades que en él se trataban y se trabajaban; o podríamos -por el contrario- evaluar la actuación de estos jóvenes en situaciones de la vida cotidiana como las que hayan tenido que vivir y a las que hayan tenido que enfrentarse durante el semestre, en las que aplicarán conocimientos aprendidos en el curso pero que no tratará de medir exclusivamente esos contenidos concretos. Y aún podríamos simplemente hacerles una prueba general de uso de la lengua, sin centrarnos en actividades cotidianas concretas, y describir sus habilidades generales al utilizar la L2.

³⁵ Algunos de estos conceptos (validez de respuesta y validez aparente, etc.) los veremos también explicados desde la óptica de Bachman / Palmer (1996, 2004), aunque nombrados de forma diferente, como “interactividad”, “autenticidad”, “impacto” y “uso práctico”, y de las ópticas de ALTE (2001) y de Fulcher (2003).

³⁶ También sirve para determinar si dos tests que se desea que midan las mismas habilidades lo logran y de qué manera o para comparar tests entre sí; y otras cosas como la fiabilidad entre evaluadores o la consistencia con sus calificaciones a lo largo del tiempo.

En los tres casos el test sería diferente y tanto los materiales como los contenidos se deberían adecuar a su objetivo. Pero en los tres casos habrá una población meta similar, los adolescentes extranjeros, con unas características de educación e interés diferentes a las que pueda tener un grupo de trabajadores de una empresa aeronáutica estudiando lengua en el extranjero, o de los jubilados extranjeros que viven en determinadas zonas del país.

Por lo tanto, habrá que diseñar las actividades, los contenidos y los materiales en que se plasmen los objetivos del test de acuerdo con unas características diferentes para cada caso. No podríamos pedir a un grupo de trabajadores ni de jubilados que charlaran sobre el horóscopo o se contaran un cómic, ya que esta actividad queda fuera de sus intereses normales, pero quizás sí sea una buena manera de motivar a los adolescentes³⁷.

Si, tras los análisis, se piensa finalmente que las tareas son demasiado artificiales o que los candidatos no las van a aceptar por estimarlas poco serias para evaluar sus habilidades, porque no producen el tipo de lenguaje esperado o por cualquier otra razón, se deberán rediseñar o eliminar para evitar que sus resultados sesguen la calificación del test en una dirección inapropiada. Si, por el contrario, no se aprecian elementos distorsionantes en ningún aspecto, puede aceptarse el test para este grupo concreto de personas; no para otros objetivos ni otros grupos de personas con otras características, ya que esto requeriría el planteamiento y diseño de nuevas tareas y el comienzo de todo el proceso de validación nuevamente.

Los pasos dados con este ejemplo de planteamiento de una tarea para incluir en un test para un grupo de personas determinado y para evaluar unas habilidades específicas, son, técnicamente, lo que vamos a detallar a continuación. Un tipo de validez se fundamenta en las opiniones y consecuencias lógicas de los elementos del test; se llama *validez de contenido* o, en otros autores, *validez interna*. Otro tipo de validez se basa en comparaciones empíricas, constatables, de la actuación de un candidato con relación a otras pruebas similares en situaciones que se conocen de antemano, autoevaluaciones, valoraciones de profesores, resultados de otras pruebas; se llama *validez empírica* o

³⁷ Hay que analizar cada actividad en todas sus facetas; desde el contenido y tipo de discurso que se espera que genere y el input que se ofrece (si se da escrito u oral y si en él hay elementos que ayudan a realizar la actividad o no) hasta si premia a determinado grupo de personas que tiene más creatividad o facilidad para realizarla, o si en el sistema de calificación hay elementos enmascarados que hacen de la evaluación otra cosa diferente a lo pretendido. Así lo hemos hecho nosotros con nuestro test del MCER y lo mostramos en el capítulo 3 y especialmente en §3.6.

externa, pues compara resultados en la prueba meta con otros resultados ya conocidos, externos al test.

En el ejemplo dado más arriba la *validez de contenido* consistiría en analizar el input que se ofrece a los adolescentes para realizar la tarea, así como el contenido que se busca que surja de esas tareas de charlar sobre el horóscopo o un cómic; el tipo de discurso generado al hacer las tareas. La *validez aparente* se obtendría mediante preguntas que les haríamos a ellos al preparar la prueba (a otros adolescentes similares que utilizaríamos en el proceso de pruebas) y tras realizar la prueba, respecto a qué les ha parecido y si creen que es una manera justa de evaluar sus habilidades. La *validez de respuesta* buscaría más datos concretos sobre qué han hecho ellos mientras se planteaban la tarea, qué habilidades han tenido que usar o poner en juego para contar el cómic o hablar sobre el horóscopo, en qué se han esforzado; no su opinión sobre qué tal lo han hecho, sino su introspección sobre cómo son capaces de hacerla.

Estos tres tipos de validez nos darían datos de la validez interna de la tarea pues se basan en datos sobre el propio test. Luego tendríamos que analizar el conjunto de las notas obtenidas por el grupo al realizar la tarea y ver en qué medida las actuaciones responden a lo que se supone que debería ser la actuación de los candidatos, por lo que ya se sabe de ellos; es decir, si los mejores son los que mejor lo han hecho y los peores los que lo han hecho peor. Si así fuera, sería un buen dato de que la prueba está funcionando, ya que se corresponde con otras medidas reales ya conocidas. Lo extraño y lo que nos haría tener que replantearnos el test o la tarea en concreto, sería el hecho de que los peores estudiantes hubieran obtenido los mejores resultados. Quizás habría que ver si la prueba mide lo que deseábamos o en realidad está midiendo otras cosas (en las que destacan los peores estudiantes) o analizar qué está pasando con estos resultados erróneos o inútiles. Esta sería la *validez concurrente*.

Si conservamos esa calificación para compararla con otras que en el futuro se realicen con la misma población o con un candidato concreto, para ver si efectivamente lo que se infería de esa prueba (que poseía tales habilidades en tal grado) se mantiene a lo largo del tiempo o sólo ha sido una mala o pasajera apreciación de un momento, estaríamos averiguando la *validez predictiva*.

Todavía hay un tipo de validez -que según Alderson (1998) es una validez superior a todas ellas- la *validez de constructo*, que trata de averiguar si toda la prueba, el conjunto de subpruebas y tareas de que se compone, poseen las características del conjunto deseadas. Alderson dice que esta validez es la más difícil de lograr, pero señala

cómo hacerla mediante correlaciones entre los resultados de las diversas partes del test entre sí.

Además, en los libros de análisis estadístico aplicado a lenguas y a tests, hay otros procedimientos que ayudan a entender más aspectos del funcionamiento del test, como si el número de tareas es suficiente para probar suficientemente la existencia de una habilidad determinada o si el número de calificadores es apropiado para dar una nota objetiva sin excesiva variación de criterios o resultados. Estos análisis requieren el uso de programas informáticos y expertos, ya que no se pueden hacer a mano por ser demasiado complejos. Alguno de ellos se citará también en los próximos apartados. Para tener una idea más clara de ellos se recomienda utilizar alguno de los libros que los explican (Bachman, 2004, Fulcher, 2003).

Entre los pasos a dar en el diseño de todo test fiable y válido³⁸, y que nosotros hemos seguido y mostramos en el capítulo 3, están (Alderson, 1998):

1. Contar con una serie de criterios iniciales y especificaciones de la prueba, de sus objetivos y de sus candidatos.
2. Discutir esos objetivos y habilidades a medir, prestando atención a las opiniones y planteamientos de profesores y expertos.
3. Plantear unas tareas, actividades y materiales mediante los que se pretende medir, que sean adecuadas y no premien ni penalicen a determinados candidatos por razones ajenas a la habilidad para realizarlas lingüísticamente. También atendiendo a las opiniones y planteamientos de profesores y expertos.
4. Hacer pruebas y ensayos con nativos.
5. Hacer pruebas con una población significativa de candidatos, de las mismas características que aquellos para los que se haya pensado el test (edad, conocimientos culturales, nivel de educación, sexo, nacionalidad, etc.).
6. Discutir, con candidatos voluntarios, el sentido y respuesta de cada ítem, actividad, tarea y test, para intentar comprobar si se responde de la manera que pretendía el diseñador del test.

³⁸ Luego procederemos a definir estos términos científicamente, ahora quedémonos con su sentido más coloquial.

7. Someter los datos y resultados de las pruebas a análisis estadísticos o analíticos cualitativos (en la medida que esto sea posible, dependiendo del tipo de tareas o ítemes seleccionados)

8. Establecer los criterios de corrección, inteligibles y practicables³⁹.

9. Rediseñar los nuevos ítemes, tareas, actividades, que haya sido necesario variar o reformar por no provocar los resultados necesarios.

10. Proceder a establecer comparaciones de los resultados del test con otras medidas conocidas de los estudiantes.

11. Determinar el alcance, limitaciones y uso del test, mediante el análisis de sus resultados y de su constructo.

12. Formar adecuada y sistemáticamente a los evaluadores para entender los criterios, objetivos y mecanismos del test.

Existen otras listas de protocolos, como las de ALTE (2001), Fulcher (2003)⁴⁰ o el Consejo de Europa (2003). Este, en su manual para demostrar que un test se ha hecho de acuerdo con los protocolos del MCER⁴¹, explica los pasos a seguir para validar correctamente un test (Consejo de Europa, 2003: 99- 123).

En la validación de tests ha habido dos concepciones: una que ve la validez como una cualidad de un test, como la prueba de que el test mide lo deseado; otra, más matizada (Fulcher, 2003; Bachman, 2004), solamente la ve como un argumento necesario para intentar probar que las inferencias hechas tras el uso del test son adecuadas.

Así, validar un test se ha llamado tradicionalmente al proceso por el cual se comprueba si el test sirve para medir lo que se pretende: “Se puede decir que una prueba

³⁹ Estos criterios ya están predeterminados en nuestro trabajo, pues se trata de lo que estamos comparando, los del MCER en este caso. Además de los puntos de esta lista, se deben incluir otros, no necesarios para nuestra investigación, como definir los criterios de presentación e información de las calificaciones; establecer métodos de reclamación, explicación y -en su caso- revisión de resultados y plantearse posibles mejoras futuras del test.

⁴⁰ ALTE dividen el proceso en las responsabilidades de los miembros de la asociación y las de los examinadores (ALTE, 2001: 3-7). Fulcher (2003: 161-69) explica el proceso de desarrollo de un test oral así como las necesidades económicas (dado en libras): *diseño del equipo de trabajo* (£147.010 anuales), *definición de objetivos*: 1 día (£869), *definición del constructo*: 5 días (£4156.34), *diseño de la tipología de tareas*: 20 días (£16.618), *pilotaje de los tipos de tareas*: 10 días (£8705.38), *descripción de escalas de calificación y sus descriptores*: 10 días (£8305.38), *redacción de documentos administrativos*: 2 días (£1294), *redacción de las especificaciones del test*: 10 días (£8305.38), *diseño de materiales para formación de los evaluadores*: 10 días (£8305.3), *formación de los evaluadores*: 2 días (£274.18), *investigaciones para sustentar las decisiones del test*: 10 días (£8231.58).

⁴¹ Des este manual se trata en §2.7.5.

o un procedimiento de evaluación tiene validez en la medida en que pueda demostrarse que lo que se evalúa realmente (el constructo) es lo que se debería evaluar y que la información obtenida es una representación exacta del dominio lingüístico que poseen los candidatos que realizan el examen” (Consejo de Europa, 2002: 177).

Pero en propiedad (Fulcher, 2003; Bachman, 2004) la validación no dice nada sobre una prueba, un test o unos ítems en sí, sino que nos habla sobre “el uso que se puede hacer de unos resultados, o de esos resultados como indicadores válidos de una determinada habilidad” (Bachman, 2004: 259). La validez, por lo tanto, “no es una característica de un test sino una cualidad de las interpretaciones que se hacen de sus resultados o de sus notas” (Alderson, 2002: 79). En palabras clarificadoras, la validez sirve para ver hasta qué punto la prueba representa lo observado, es decir, hasta qué punto se obtiene con ella información pertinente acerca de lo evaluado (Ortega, 1996: 20).

De esto se puede inferir que la validez no va a ser nunca ni definitiva, ni objetiva, sino un argumento elaborado mediante una serie de datos; argumento que necesitará más o menos elaboración según sean las necesidades y responsabilidades implicadas en los usos del test.

Aparte de estas dos visiones de la validez, el concepto de validación ha sufrido variaciones, pasando de ser considerado un hecho a posteriori para probar los usos de un test, a ser una parte importante a tener en cuenta desde el comienzo de su planteamiento. Así, ahora, el proceso de diseño de tests se relaciona con su validación desde el principio, desde el planteamiento del dominio de lengua que se quiere evaluar; y se mantiene a lo largo de todo el trabajo de diseño de tareas y actividades, para comprobar si estas son adecuadas para las personas que van a usarlo y si son buenas para que los candidatos produzcan el tipo de lengua que demuestre la existencia de las habilidades que el test quiere medir. Y llega hasta el punto de comprobar si el modo como se ha diseñado el test permite que sea practicable, útil y factible de administrar.

Todo este proceso y su validación no es sencillo. Pero debería ser la preocupación primordial de todos los evaluadores, puesto que si una prueba no es válida para el objetivo para el que se ha preparado, los resultados no significan lo que se cree que significan. En esta línea, matiza Henning (1987) que siempre que evaluamos es *para* algo (con la preposición marcada intencionalmente):

Un test es válido en la medida en que evalúa lo que se supone que evalúa. De ello se sigue que el término válido, cuando lo usamos para describir un test debería ir acompañado de la preposición para. Cualquier test puede ser válido para un propósito pero no para otros (Henning, 1987: 89).

Sólo que incluso profesores de idiomas y evaluadores expertos son a menudo incapaces de ponerse de acuerdo sobre lo que evalúa un ítem (Alderson, 1998: 74). La dificultad de llegar a acuerdos muestra que un test no es cosa fácil de diseñar y que su naturaleza es más complicada de lo que a menudo parece. Esta cita de Alderson hablando sobre la variabilidad de juicios entre expertos, puede ser clarificadora:

Es interesante para los investigadores del mundo de la evaluación constatar esta variedad en la emisión de los juicios: sugiere que los aspectos que hay que juzgar son mucho menos claros, más controvertidos, quizá más complejos de lo que se pensó en un principio. Si se analiza tal complejidad a fondo surgen todo tipo de nuevos matices. Sin embargo el responsable de una prueba se encuentra ante un dilema en tales circunstancias. Si los “expertos” no se ponen de acuerdo sobre lo que evalúa una prueba, ¿qué es en realidad lo que esta evalúa? (...) La respuesta a este dilema, creemos, se encuentra en el hecho de que la validez de una prueba es más relativa que absoluta. Si los procedimientos de validez del contenido revelan problemas relacionados con la validez de contenido de la prueba, se deben recoger otros tipos de evidencia de validez: validez externa, validez aparente, validez de respuesta, y así sucesivamente (...) se debe aceptar que en la evaluación de idiomas, como en otros muchos aspectos, hay todavía mucho que desconocemos (Alderson, 1998: 170)

No obstante las dificultades, diseñar procedimientos de control de calidad para garantizar a los candidatos de las pruebas que los resultados representan verdaderamente lo que se les está diciendo al darles una nota o un certificado, es un proceso obligado e importante del que debería responsabilizarse cada institución. Pero esto que está claro para los profesionales del test no lo está para el resto de profesionales de la enseñanza de lenguas ni para los administradores, que demasiado a menudo desatienden los objetivos y circunstancias que influyen en una buena evaluación y, desconociendo la importancia de atender a los aspectos cualitativos y cuantitativos del diseño de tests, aceptan prácticas y mecanismos que no han sido debidamente probados o copian y utilizan modelos de tests que han sido válidos para otras situaciones, sin plantearse si serán útiles para las nuevas. Inconscientes ante el verdadero trabajo de diseño de tests y sin apreciar la cantidad de elementos que hay que tener en cuenta para que los resultados no queden sesgados y sean útiles para la población a que la que van destinados.

Vamos a analizar cada una de estas formas de averiguar la validez, su definición y mecanismos, especialmente para los tests de evaluación oral.

2.3.1 Validez de apariencia

La validez aparente se refiere a la credibilidad que den los candidatos, los estudiantes o los usuarios institucionales y la sociedad en conjunto, a una actividad, tarea, prueba o test; es decir, a la aceptabilidad personal o pública que le den los agentes que tienen relación con ella.

La pregunta básica que debemos hacernos para averiguar la validez de apariencia de un test o de una actividad es si es realista y aceptable para el candidato y para el resto de integrantes del proceso de evaluación; pregunta que implica si la actividad es buena para evaluar las habilidades que pretende evaluar el test. Un test no logrará aceptabilidad ni personal ni social si los candidatos creen que no se les está midiendo adecuadamente.

Este tipo de validez es importante porque supone tener en cuenta a los candidatos y respetar sus intereses, pero también por sus consecuencias sobre otros aspectos del proceso de la evaluación y en concreto sobre la capacidad para provocar una respuesta adecuada, ya que si los candidatos no toman en serio la tarea tampoco responderán, no se esforzarán, no se comprometerán al hacerla, y no pondrán en juego los elementos cognitivos necesarios para que el resultado de la tarea sea efectivo⁴².

Por lo tanto, la validez de apariencia es importante por sí misma y porque influye y está relacionada con la validez de respuesta y en alguna medida -ya lo veremos- con la validez de contenido⁴³. Como dice Fulcher (2003: 217), “si el candidato no se toma el test en serio, si no entiende qué debe hacer, si su contenido no le es familiar, o algo en el

⁴² Estos elementos se explican en la validez de respuesta, que veremos en seguida, en §2.3.2.

⁴³ Desde que los enfoques comunicativos han cobrado importancia en la enseñanza y en la evaluación, se ha dado más importancia a la validez aparente y se propugna que un examen debe parecerse a la vida real lo más posible. Así, autores como Bachman / Palmer (1996) -como hemos visto en el punto anterior, §2.2- centran esta validez en la medida en que las tareas provocan un tipo de lenguaje que existe en la vida real. Alderson, en cambio, engloba estos aspectos y otros relacionados con el tipo de método utilizado para evaluar, dentro de otro tipo de validez, la validez de contenido.

Para ver cómo la validez de apariencia va unida a la de respuesta y de contenido, véase §3.6.1- i y §3.6.2- i, un ejemplo nuestro, que se tuvo que desechar para el test por carecer de validez de respuesta.

test o en su contexto le resulta ofensivo, es improbable que las notas reflejen su habilidad”.

Como también veremos en el capítulo 3, la validez de apariencia se obtiene haciendo preguntas a los candidatos y personas no expertas en evaluación sobre sus reacciones, opiniones, actitudes e impresiones al realizar una prueba. Los resultados de esas preguntas, si se desea, se pueden analizar estadísticamente para determinar la aceptabilidad de cada tarea.

Pero lograr tareas de mucha autenticidad conlleva también sus propios problemas, ya que cuanto más auténticas son, más se complican, tanto al diseñarlas como al calificarlas o al tener que realizarlas los candidatos; y al ser complicadas son susceptibles de quedar a merced de muchas influencias y se hace más difícil su calificación (Leung / Lewkowicz, 2000: 216).

Para Bachman / Palmer la validez de apariencia se corresponde, como acabamos de apuntar, con lo que entienden por autenticidad⁴⁴ y está relacionada con la validez de respuesta (§2.3.2), de contenido (§2.3.3) y de constructo (§2.3.4), elementos todos que contribuyen a definir la autenticidad. Esta, así entendida, abarca a todos los componentes de un test y su administración, ya que engloba aspectos como el diseño de las tareas, la definición del dominio lingüístico a medir, la relación del contenido del test con las implicaciones y generalizaciones que se puedan establecer con sus resultados, la diferente percepción que puedan tener diferentes candidatos sobre la lengua que se mide, los materiales con los que se mide y el modo de hacerlo. En definitiva, para ellos la autenticidad es una de las más importantes características de un test de lengua y debe ser analizada desde todas las perspectivas posibles⁴⁵.

⁴⁴ No vamos ahora a tratar este tema de la autenticidad en profundidad. Ya se ha hablado acerca de la autenticidad de las tareas y del lenguaje a evaluar y producir mediante ellas en §2.2 y más adelante habrá ocasión de discutir otras relaciones con otros elementos del proceso de evaluación, §2.4.

⁴⁵ Para averiguarla en toda su amplitud, proponen plantearse su presencia en situaciones y momentos del test que tradicionalmente no se analizan, como el lugar en que tiene lugar el test o la forma en que se ofrecen las instrucciones de cada tarea. Para esto se plantean “en qué medida la descripción de los objetivos de las tareas en su ámbito de actuación real incluye información sobre localización, input, respuesta esperada, y relación entre input y respuesta”. Esto nos ayuda a entender el amplio concepto de la autenticidad de estos autores llevado a todos los extremos del diseño y administración de un test. Según ellos, habría que considerar cosas como el espacio físico en que se hace el test y sus diferencias con el espacio en que suelen tener lugar las tareas, el canal por el que se propone, el tipo de información que se da y las diferencias entre esta y la realización de una tarea similar en la vida real.

2.3.2 Validez de respuesta

Para Alderson, la validez de respuesta se refiere a si los candidatos ponen en práctica las destrezas que pretendía el redactor de la tarea o si por el contrario usan otras habilidades y conocimientos que no se habían planeado o no interesan. Igualmente, se encarga de analizar qué hacen los candidatos durante el proceso de responder a la pregunta o tarea, examinando su comportamiento, su sistema de razonamiento y sus pensamientos⁴⁶.

La validez de respuesta pretende averiguar cosas como si en un test de comprensión lectora se responde correctamente a las preguntas de comprensión porque se ha entendido el texto, o si se hace por adivinar o por otros mecanismos que ayuden a sacar una buena nota a pesar de no haber comprendido. En estos casos, aunque el candidato haya sacado buenas notas no sabremos nada sobre su capacidad de comprensión lectora. Igualmente la validez de respuesta pretende averiguar si hay candidatos que, habiendo entendido el texto, responden mal a las preguntas y no pueden demostrar todo lo que saben. Y en un test oral se intenta averiguar otras cosas como si el discurso generado por la actividad propuesta es el tipo de discurso que se pretendía, o en qué medida el sistema establecido para dar las calificaciones sirve para medir con igualdad a diferentes hablantes, dada la variedad de discursos que una misma tarea puede producir.

Bachman matiza, aunque no la llama *validez de respuesta*, que los candidatos ponen en marcha diferentes procesos y estrategias al realizar una tarea y que es necesario averiguar qué procesos utilizan y en qué medida. Estos casos no se pueden observar directamente y se necesita un procedimiento de búsqueda de datos, preguntando a los propios candidatos durante la realización del test o al terminarlo (Bachman, 2004: 276). La forma que tienen estos autores de abordar este aspecto parte de una perspectiva diferente. Se centran no tanto en la respuesta que provocan las tareas o en el resultado de la tarea, sino en algo previo: tener en cuenta las características de los candidatos para que luego las tareas puedan ser efectivamente llevadas a cabo por ellos (Bachman / Palmer, 1996: 143-144). A esto lo llaman *interactividad*, queriendo referirse a la medida en que interactúa cada tarea con el candidato. Para ello plantean la necesidad de definir la medida y manera en que se supone que ciertas características del candidato existen y

⁴⁶ La respuesta no es solamente cosa del candidato, quien debería demostrar qué habilidades posee, sino que es responsabilidad de otros elementos que forman parte del test, como la capacidad de una tarea para provocarla. Como veremos más adelante, §2.4, esta responsabilidad también puede compartirse con la acción del entrevistador.

se pondrán en funcionamiento ante la correcta realización de una tarea. Responsabilizan, pues, al planificador del test y al diseñador de tareas, del efecto que estas puedan tener sobre el candidato cuando este quiera o tenga que realizarlas, ya que son ellos quienes se deben plantear toda la gama de cuestiones que permitirán o no que la tarea sea efectiva.

Los elementos que consideran implicados en esta interactividad entre la tarea y el candidato son:

- a. Las características de los candidatos.
- b. Las habilidades lingüísticas necesarias:
 - b1- El conocimiento que deban tener de la lengua necesaria para realizar la tarea.
 - b2- Las estrategias metacognitivas que haya que poner en funcionamiento (discurso planificado o improvisado, actividades independientes o no, etc.).
- c. El conocimiento previo que sea necesario tener sobre el tema en cuestión.
- d. Otros factores afectivos que intervengan premiando o limitando la actuación de los candidatos.

Mediante estos cuatro elementos se plantean averiguar cómo será el proceso de respuesta y reacción a una tarea cualquiera, pues se concentran con detalle en la medida en que los elementos planificados -relativos a todas las facetas que un candidato pone en práctica cuando responde- van a estar presentes en sus respuestas.

Como vemos, esta perspectiva de la validez de respuesta es diferente a la señalada por Alderson. Se centra en un paso anterior, externo a la administración del test, el del diseño. Alderson, en cambio, se centraba en un paso posterior, en el resultado del discurso y el análisis de por qué se produce. A ambas perspectivas vamos a añadir, más adelante, otra que nosotros creemos que también se debe considerar. La tomamos de Lazaraton (2002) y las técnicas de una OPI, y se produce en el proceso interno de realización y administración del test, de la entrevista oral en nuestro caso. Esta es la corresponsabilidad del entrevistador en la formación del discurso, y el manejo de las técnicas de elicitación. Lo veremos cuando expliquemos el papel del entrevistador y hablemos de su responsabilidad como generador o motivador del discurso de los candidatos (§2.4).

2.3.3 Validez de contenido

La validez de contenido se refiere a si el contenido y los temas utilizados en un test son representativos de lo que se pretende medir y si miden lo que se desearía. Se examina mediante un análisis sistemático del contenido de la prueba para ver si los materiales y las tareas de que se compone son una buena muestra de las destrezas lingüísticas que se quieren medir: si se corresponden con las definiciones de lo que se haya determinado que va a ser el test, con los currículos o las programaciones -cuando se ha diseñado partiendo de una situación académica concreta-, o si se corresponde con los criterios establecidos en las escalas -cuando el test se ha diseñado de acuerdo a un criterio-⁴⁷.

Este proceso de comparación entre objetivos y mecanismos utilizados para lograrlos se puede hacer (ALTE, 2001: 8) analizando la medida en que el contenido se adecua a la escala o constructo en que se base la medición de sus habilidades (MCER, ACTFL, ALTE, etc.) o pidiendo opiniones a expertos, colegas, profesores, otros evaluadores. Las preguntas que se hacen a los expertos suelen hacer referencia a los ítemes, a su contenido, a los textos y materiales utilizados, a la complejidad de la lengua requerida para entenderlos y para contestarlos, al tema, la organización retórica, elementos culturales supuestos, a otros elementos pragmáticos, etc., según cuales sean los contenidos que se desee medir con el test⁴⁸.

Un problema que podemos encontrar al intentar hacer este tipo de validación de contenido es, como dice Alderson, que las opiniones de los expertos no son siempre homogéneas, decisivas o coincidentes. Ya hemos dicho que en sí misma la validación es un proceso de muchas facetas y cada elemento que se intenta validar contribuye a dar validación al conjunto del test. Si por un mecanismo no se logran resultados habrá que utilizar otros tipos de validez que en conjunto den luz sobre su aspecto más importante,

⁴⁷ Bachman aclara que no se puede comparar el contenido de un test con lo definido en el test sino que debe demostrarse su equivalencia con lo obtenido mediante el test a través de las notas. Es decir, no se puede decir que el test mide lo que se haya establecido que va a medir, a través de sus explicaciones, declaraciones de objetivos, etc. sino que se debe demostrar que lo que se ha producido mediante él es lo que se había definido previamente. “La primera limitación de la validez de contenido, entonces, es que se centre en el test, en vez de en las notas del test” (Bachman, 1988: 152). Opinión que, a su vez, comenta con un ejemplo circense: no debemos confundir lo que el cartel de un circo muestre con lo que vamos a ver bajo la carpa, que no sólo puede ser diferente en tamaño y color sino en que, quizás, nada de lo que nos enseñen sea lo que se había anunciado en el cartel.

⁴⁸ Así lo hemos hecho nosotros, como se muestra en §3.6.

qué mide el test y si mide lo que efectivamente estamos intentando, y con suficiente justicia como para no penalizar a los candidatos.

Los siguientes apartados dedicados a la validez de constructo (§2.3.4), concurrente (§2.3.5) y predictiva (§2.3.6) son bastante importantes para llegar a una conclusión efectiva sobre qué se supone que mide un test, con independencia, incluso, de las opiniones de los expertos o de uno mismo.

2.3.4 Validez de constructo

Para Fulcher esta validez también se llama *validez estructural de un test* (Fulcher, 2003: 195). Nos parece una forma clara de explicarlo: la estructura interna de un test puede estar compuesta por diversas tareas, designadas cada una para evaluar diferentes rasgos o aspectos de las habilidades. Cuando el constructo es algo tan amplio como hablar una lengua, se utilizan normalmente escalas holísticas para su evaluación, pero también se pueden usar escalas de rasgos diferentes para cada tarea⁴⁹.

Este tipo de validez combina lo visto hasta ahora (validez aparente y de respuesta) y otros tipos de validez que veremos luego (validez de contenido, validez predictiva y validez concurrente). En realidad, como dice Bordón, “es un tipo superior de validez (...) al que contribuyen tanto la validez interna como la externa y se refiere a la comprobación de hasta qué punto el examen refleja la teoría (el constructo) que subyace a él” (Bordón, 2006: 62). Como también dice ALTE (2001: 8), “el diseñador de exámenes define rasgos de habilidad con el objeto de medirla; estas definiciones son su constructo”.

En este sentido, el modelo de habilidad comunicativa que subyace al test o se define con él representa el constructo del mismo. Como parte integrante de la validez de un test, la validez de constructo trata de identificar en qué medida el test en su conjunto se puede relacionar con la medida de algo concreto previamente definido; solo que ahora esta medida es genérica y de alguna forma -por decirlo así- total. Esto es, que no está centrada en un aspecto o faceta (su apariencia, su capacidad para producir respuestas, su contenido representativo) sino en su visión global, lo cual permite en última instancia asegurar qué significa el test.

⁴⁹ Este va a ser el caso al trabajar con el MCER, como se verá en §3.3.

Por ello, para Bachman / Palmer (1996: 44) esta validez consiste en encontrar suficiente y adecuada justificación para realizar las interpretaciones que con el test se pretenden hacer. Para estos autores, el sentido último de este tipo de validación reside en probar que se posee una forma justa y consistente de calificar a un candidato a través de su actuación en un test. “La validez de constructo se utiliza por lo tanto para referirse a la medida en que podemos *interpretar una nota como indicador de una de las habilidades* que queremos medir” (Bachman / Palmer, 1996: 21, la cursiva es nuestra) y para ello hay que poder establecer adecuadamente las relaciones entre sus tres elementos más importantes, la lengua real, la lengua de las tareas y el sistema de calificaciones.

Para Bachman / Palmer la obtención de los datos no se soluciona de forma estadística, sino en la medida en que se lograran suficientes evidencias como para justificar las calificaciones. También Bachman advierte de que “la validez de constructo suele ser malentendida y considerada importante solo en situaciones en las que el contenido no se puede averiguar porque el dominio de las habilidades a evaluar no ha sido bien definido” (Bachman, 1988: 151). Por ello, alerta sobre malinterpretaciones respecto a la validez de constructo: “La validez de constructo tiene que ver con la manera en que una actuación en un test es consistente con las predicciones que se hacen en base a una teoría de habilidades (lingüísticas), o constructo” (Bachman, 1988:151).

Estos autores también relacionan el concepto con otros elementos estadísticos como concurrencia (correlación con otras medias conocidas o que se realicen con el candidato en la misma línea -lo que llamamos aquí validez concurrente- §2.3.5) y capacidad de predicción o validez predictiva (§2.3.6) para los que las correlaciones, así como otras técnicas estadísticas, también van a ser imprescindibles (Bachman, 2004).

Quizás un ejemplo sea más ilustrativo: si queremos un test para clasificar estudiantes en un curso universitario que pretende mejorar la habilidad oral, hay que darse cuenta de que el conocimiento gramatical es solo uno de los aspectos incluidos en esta habilidad académica y que definir en términos gramaticales el constructo a evaluar es algo demasiado estrecho. Para averiguar si los candidatos pueden acceder a un programa universitario determinado, una vez hecho el test, se debe poder identificar qué tipo de candidatos no pueden acceder a los cursos, cuáles sí y cuáles necesitarían un apoyo formativo para poder acceder en el futuro. Y para ello, el sistema de calificaciones del test debe estar de acuerdo con las áreas de habilidad lingüística que se desean medir, con el tipo de tareas utilizadas y con el tipo de lenguaje usado.

Bachman / Palmer (1996) proponen plantearse 10 cuestiones prácticas para comprobar esta validez:

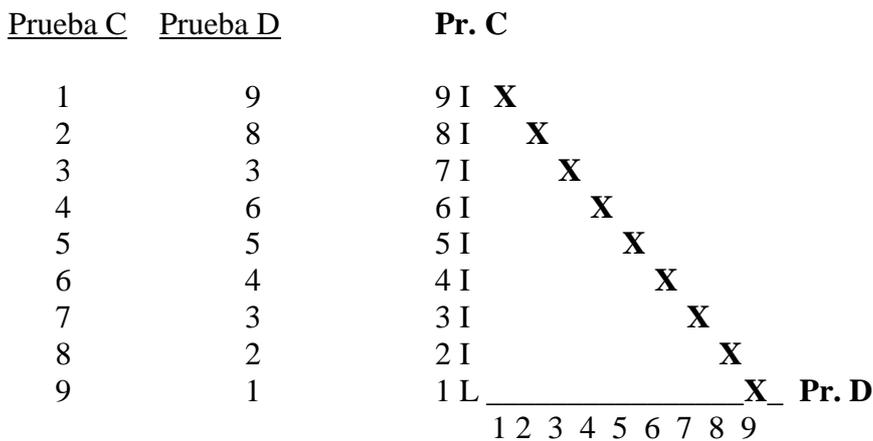
- 1 - *La habilidad lingüística, ¿está definida claramente y sin ambigüedades?*
- 2 - *La habilidad lingüística que produce el test, ¿es relevante y adecuada para su propósito?*
- 3 - *¿En qué medida reflejan las tareas la definición del constructo?*
- 4 - *¿En qué medida refleja el sistema de calificaciones la definición del constructo?*
- 5 - *El sistema de notas y calificaciones, ¿sirve para hacer las interpretaciones sobre la habilidad de los candidatos?*
- 6 - *¿Qué características del lugar de realización del test pueden llevar a los candidatos a actuar de forma diferente?*
- 7 - *¿Hay características en la forma del test que pueden llevar a los candidatos a actuar de formas diferentes?*
- 8 - *¿Qué elementos del input pueden llevar a los candidatos a actuar de formas diferentes?*
- 9 - *¿Qué elementos de la respuesta esperada pueden causar diferencias entre la actuación de los candidatos?*
- 10 - *¿Qué elementos de relación entre el input y la respuesta esperada pueden hacer variar la actuación de los candidatos? (Bachman / Palmer, 1996: 140-142).*

Las cinco últimas se relacionan de forma directa con la fiabilidad (§2.3.8), pero la 1, 2, 4 y 5 se refieren al concepto de constructo, junto a la 3, que incide en la validez de apariencia, la cual -evidentemente- es necesaria para lograr la validez de constructo.

Hasta aquí nos hemos centrado en la definición y concepto de la validez de constructo según Bachman / Palmer. Ahora vamos a intentar otro camino, un camino estadístico que nos parece bastante claro, a pesar de utilizar unos cuantos conceptos matemáticos que antes debemos entender, y lo suficientemente expresivo como para poder tenerlo en cuenta y extraer de él unas cuantas conclusiones cualitativas necesarias. Este lo explica Alderson mediante el uso de correlaciones entre los componentes de un test.

Este camino permite interpretar adecuadamente los datos y establecer qué mide cada parte de un test, cómo se relacionan unas con otras y cuál tiene más valor en el conjunto, con independencia de lo que se haya definido previamente. Lo que se plantea aquí, de forma cuantitativa y estadística, es averiguar en qué medida hay relación entre las diversas partes que componen una prueba. Y la solución a esta pregunta ofrece datos muy interesantes para diferenciar los elementos, arrojando luz respecto a qué puede estar midiendo cada uno de ellos, por encima de las apariencias y opiniones de los expertos o diseñadores, y eso permite entender su alcance e influencia en todo el test ayudando a establecer las formas de ponderación.

Gráfico 2. Correlación negativa



El gráfico 1 indica los resultados de la correlación en un grupo de 9 personas que realizan dos pruebas (Pr. A y Pr. B) y en las dos pruebas obtienen la misma posición según los resultados: el mejor en una prueba es igualmente el mejor en la segunda, el que queda en segundo lugar en la primera prueba queda igualmente en segunda posición en la segunda prueba, etc. El gráfico 2 indica lo contrario: el mejor en la primera prueba es el peor en la segunda (queda en la novena posición), el segundo mejor en la primera prueba es el penúltimo en la segunda (octava posición), y así sucesivamente.

Obviamente no se suelen encontrar estas correlaciones tan claras, pues siempre hay algunas diferencias entre las posiciones logradas por los mejores y los peores. Pero en general, las gráficas muestran una línea ascendente de izquierda a derecha cuando hay una correlación entre dos pruebas (gráfico 3) y una línea desfigurada o ausencia de línea cuando no hay correlaciones ente los elementos que se comparan (gráfico 4).

Gráfico 3. Correlación positiva

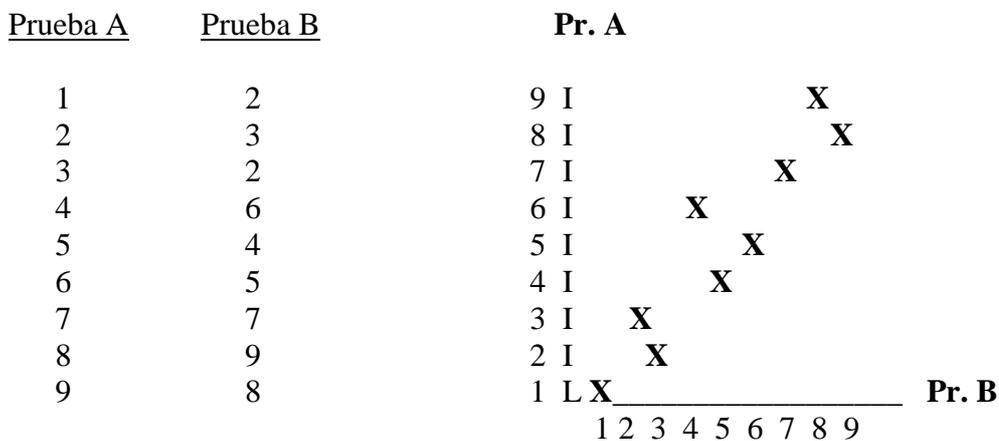


Gráfico 4. Falta de correlación

<u>Prueba C</u>	<u>Prueba D</u>	<u>Pr. C</u>
1	1	9 I X
2	7	8 I X
3	5	7 I X
4	8	6 I X
5	6	5 I X
6	4	4 I X
7	9	3 I X
8	3	2 I X
9	1	1 L _____ X Pr.
		1 2 3 4 5 6 7 8 9

Podemos imaginar que no hay ninguna correlación entre tener una habilidad lingüística determinada y nadar rápido. Sin embargo, podemos también entender que sí existirá cierta correlación entre hablar una lengua y ser capaz de escribir en ella (salvo el caso de no estar alfabetizado), o entre el dominio gramatical y la habilidad para organizar adecuadamente frases en un texto, aunque la naturaleza exacta de estas correlaciones también dependa de otros factores y no sea absoluta, como en el gráfico 1, sino relativa, como en el gráfico 3.

El concepto de correlación, es decir, el análisis de la relación de equivalencia, semejanza o desemejanza entre dos pruebas, es uno de los conceptos más interesantes y productivos que podemos utilizar en la evaluación. Él nos va a dar multitud de datos sobre la fiabilidad de una prueba (es decir sobre si repetida daría los mismos resultados), sobre el grado en que dos o más pruebas son iguales, semejantes o diferentes (es decir, si miden las mismas o distintas habilidades) o sobre el grado de homogeneidad y similitud con el que un equipo de evaluadores orales está actuando al proceder a recabar datos o interpretarlos y dar notas. Las correlaciones, por lo tanto, se usarán en todo el proceso de validación de pruebas y tests, desde validar su constructo hasta ver su fiabilidad o si el conjunto de evaluadores o entrevistadores está haciéndolo de la misma manera.

Estos gráficos de arriba son formas visuales de establecer las correlaciones sencillas cuando hay pocos elementos que correlacionar. No obstante, cuando el número de candidatos a una prueba es alto, y así debe ser si queremos obtener datos estadísticos fiables, se procede mediante otras técnicas y fórmulas, como el *coeficiente de*

correlación de Pearson, usado para averiguar correlaciones cuando los elementos tienen el mismo valor, y el *coeficiente de correlación de Spearman*, usado cuando los elementos que se correlacionan tienen diferente valor y solo se pueden agrupar de forma ordinal. Ambos, por suerte, se calculan automáticamente con programas estadísticos⁵¹.

Una correlación total como la de la gráfica 1 tiene un índice matemático de + 1.0 (lo que indica que las dos pruebas son equivalentes y una de ellas no dice nada nuevo respecto de la otra). Ya hemos dicho que no es normal que esto ocurra, pero puede ser necesario aproximarse a estos índices si lo deseado es tener dos pruebas que midan lo mismo pero se vayan a utilizar en momentos diferentes, por motivos diversos (una más breve que la otra, por falta de tiempo; o para no repetirla, si los estudiantes se mantienen en meses consecutivos y no se les puede pasar el mismo test dos veces seguidas).

Otro sistema es el llamado *correlación de acuerdo al orden de la escala*, el cual nos vendrá bien entender para comprender el proceso y los análisis de una matriz de correlaciones como las que se usan en la validación del constructo (o en la formación de evaluadores) y poder extraer datos sobre qué evalúan las diferentes actividades del test y qué implican sus calificaciones. La fórmula⁵² para obtener el coeficiente de correlación entre dos pruebas es la siguiente:

$$CC = 1 - \left(\frac{6 \times \text{SUMA } d^2}{n(n^2 - 1)} \right)$$

CC = coeficiente de correlación; d = diferencia de posición obtenida por el candidato en el ranking de notas de las pruebas; n = número de estudiantes.

Una forma de averiguar la validez de constructo de una prueba es haciendo las correlaciones entre las diferentes partes de la prueba. Como bien explica Alderson, lo

⁵¹ Ambos son sistemas que se hallan con el programa SPSS. Se puede ver una explicación sencilla en Bachman (2004: 84-91). Para una mejor comprensión se debe acudir a libros especializados. El último, la correlación de Spearman, es el que hemos usado en los análisis estadísticos realizados con los resultados obtenidos mediante nuestro test del MCER y entre estos y los de la OPI, junto a una regresión factorial que permite averiguar cuántos factores se están evaluando con el test. Véase capítulo 4 y Apéndice 10.

⁵² Se puede ver una explicación en Alderson (1998: 263).

normal es que una prueba contenga diversas partes para medir cosas diferentes que contribuyan a la visión de conjunto.

Una vez hechas las correlaciones entre las partes se considera que si los resultados se aproximan a 1.0 (0.9, 0.8) significa que ambas partes son demasiado equivalentes y semejantes; si quedan cercanas al 0.5 (desde 0.3 a 0.6) significa que las partes tiene cierta relación -como cabe esperar siendo todas ellas integrantes de una misma prueba de lengua y que lógicamente ponen en funcionamiento habilidades cognitivas semejantes- pero midiendo cosas diferentes dentro de un mismo campo; y si se acercan al 0.0 o resultan negativas, significa que su relación es nula, algo no deseado en nuestro caso (ya que aunque mediante pruebas diferentes, estamos evaluando un concepto lingüístico que debe tener alguna correlación entre una prueba y otra, entre una manera de medirlo y otra, y el resultado no puede ser cero).

Los datos obtenidos con todas las correlaciones se pasan a una matriz de correlaciones (véase un ejemplo en el cuadro 2.3.4) -necesarias para interpretar finalmente los datos y hallar la validez de constructo-. En estas se presentan diferentes datos:

- Los análisis hechos entre cada parte del test y cada una de las demás partes.
- Los análisis entre cada parte del test y el conjunto del test eliminando esa misma parte.
- Los análisis entre una parte y el conjunto de todo el test (la nota final total, en la que también se encuentra la parte con que se hace la correlación).

En este último caso los índices que indican que entre las partes hay una semejanza pero sin llegar a ser partes demasiado equivalentes (que acabamos de decir rondan entre el 0.3 a 0.6) pueden subir hasta 0.7, ya que es lógico que haya una mayor correlación entre una parte y otra si en la parte total se incluye la misma parte con la que se está hallando la correlación.

Copiamos y explicamos la matriz dada por Alderson como ejemplo, por considerarla ilustrativa y explicativa del análisis de datos que nos va a permitir entender si hay una buena validez del constructo en un test (Alderson, 1998: 178).

Cuadro 2.3.4. Matriz de correlaciones de una prueba de lengua⁵³

	Lectura	Competencia Lingüística	Expresión escrita	Expresión oral	<u>Total</u>	Total menos componente
Lectura	-	0.53	0.27	0.44	<u>0.73</u>	0.50
Competencia Lingüística	0.53	--	0.43	0.66	<u>0.84</u>	0.72
Expresión Escrita	0.27	0.43	--	0.45	0.66	0.46
Expresión Oral	0.44	0.66	0.45	--	<u>0.86</u>	0.66
<u>Total</u>	<u>0.73</u>	<u>0.84</u>	0.66	<u>0.86</u>	--	--

Esta prueba consta de 4 partes en principio diferentes, de Lectura, Competencia lingüística, Expresión escrita y Expresión oral. Copiamos la explicación de Alderson, por ser la forma más sencilla de entenderla:

(...) la correlación más alta es la existente entre los componentes de competencia y la prueba de expresión oral (0,66). La correlación está por encima de lo que podría esperarse si las pruebas en realidad evalúan distintas destrezas, pero la coincidencia no es lo suficientemente grande como para sugerir a los redactores la supresión de una de las dos partes. Todas las partes excepto la de expresión escrita tienen correlaciones con la prueba global por encima de un 0,7. El hecho de que la correlación entre la expresión escrita sea más bien baja (0,66) puede deberse al hecho de que esta prueba resultó ser poco fiable, y las correlaciones entre las pruebas no fiables muestran coeficientes bajos puesto que los resultados se deben al factor suerte. Las correlaciones de las partes de competencia lingüística y expresión oral con la nota global (total) están por encima de 0,8. Esto muestra que ambos componentes influyen de forma importante en la puntuación global final. Cuando cada una de estas partes se correlaciona con la puntuación global menos ella misma, las correlaciones se reducen a 0,72 y a 0,66. De todas formas, estas correlaciones son todavía las más altas entre las partes de la prueba y el total y muestran lo importantes que son estas dos secciones dentro de la serie (Alderson, 1998: 178-179).

Como se ve, este sistema permite averiguar la coherencia total de un test, el valor de cada una de sus partes e incluso si alguna está midiendo lo mismo que las otras, aunque no se pretendiera eso en su diseño inicial. Analizar de esta manera el comportamiento de las pruebas permite tomar decisiones importantes tanto para su diseño o rediseño como respecto al valor dado a cada parte, hacer ponderaciones dando notas más altas a una parte frente a otra o exigiendo obtener mayores puntuaciones en una prueba que en

⁵³ Hemos señalado diferentes índices: en negrita el 0.66, subrayados los datos por encima del 0.7, con cursiva los índices superiores al 0,8 y con color gris los índices de 0.72 y 0,66.

otras e interpretar qué está midiendo cada parte en efecto, y entender mejor el valor de la certificación que otorga la prueba.

¿Cómo hacer una valoración de la validez de constructo en pruebas abiertas u orales como las que vamos a considerar en este trabajo? Los tests abiertos, por su especial contenido y constructo, son más proclives a análisis abiertos, subjetivos, basados en opiniones cualitativas, como las aportadas por Bachman / Palmer (1996: 133 y stes.). No obstante, también se pueden someter a análisis estadísticos, comparándolos entre sí, y con otras pruebas abiertas o cerradas⁵⁴.

Además, un punto importante en un test abierto reside en la naturaleza de los juicios establecidos sobre la actuación de los candidatos, calificación que se hace siempre por un evaluador (o un equipo de evaluadores en base a una serie de criterios previamente establecidos). En las diversas o sucesivas administraciones de un test oral por parte de diferentes evaluadores o de uno mismo a lo largo del tiempo, la forma del test puede cambiar por la actuación diferente de los evaluadores (o del evaluador) tanto al elicitar el discurso como al calificarlo según las escalas establecidas, lo cual influye en su fiabilidad (§2.3.8) pero igualmente en su contenido y constructo.

Estos dos elementos, el sistema de elicitar el discurso y de calificarlo por parte de los evaluadores, deben incluirse por lo tanto en las consideraciones del diseño del test como elementos determinantes. Y al analizar la validez de constructo de una prueba oral se debe considerar al evaluador y su formación como un punto tan importante como la propia escala de calificaciones a utilizar o las tareas empleadas.

La mejor forma de hallar esta validez reside en incluir las correlaciones de los calificadores entre los datos a analizar; tanto las correlaciones internas y la medida en que las notas que den sean consistentes en un mismo evaluador a lo largo del tiempo, como las correlaciones externas, la medida en que las notas que den sean consistentes con las calificaciones del resto de evaluadores⁵⁵.

⁵⁴ Cuando la distancia entre el valor de una calificación y la siguiente no se establece con números que siempre miden intervalos iguales, para averiguar las correlaciones se utiliza el *coeficiente de Spearman*, que es el que hemos utilizado en los análisis de nuestros tests. Véase capítulo 4.

⁵⁵ No terminan aquí las técnicas de validación del constructo. Existen muchas otras que deben ser utilizadas según las circunstancias, como el análisis factorial, los análisis multirasgo-multimétodo o los análisis de varianza (ANOVA). Estos tipos de análisis requieren unos procedimientos especiales y la ayuda de especialistas y programas informáticos. Pero son muy útiles los datos que se extraen de ellos, ayudando a entender si las notas reflejan la existencia de una habilidad, si por el contrario se deben a otros factores que se deben desechar, si el número de tareas que se incluyen en un test es suficiente para demostrar que una habilidad se domina o si el número de evaluadores que califica a un candidato es suficiente para dar notas objetivas o muestra una variabilidad grande que invalide los resultados. Todos

2.3.5 Validez concurrente

La validez concurrente y la validez predictiva (§2.3.6) son dos tipos de validez también consideradas externas, ya que utilizan datos de otras pruebas, datos de cursos o profesores anteriores, etc. cuyos resultados se comparan con los de la prueba actual. Según ALTE (2001: 9) esto supone relacionar los resultados con criterios externos o escalas de criterios independientes como las usadas en la validez de contenido.

Ambas son generalizaciones de los resultados del test a otras situaciones de fuera del test. Ya que los resultados de un test pueden verse afectados por muchas variantes que no son solamente los cambios en la habilidad, generalizar las notas significa mostrar que una nota obtenida en un test en un contexto determinado significa lo mismo que otra en otro contexto, con otro test, otras tareas y otros examinadores (Fulcher, 2003: 196).

La validez concurrente consiste en la comparación entre los resultados de la prueba con cualquier otro dato conocido de los candidatos del mismo periodo de tiempo en que se realiza la prueba. Esto puede ser otra prueba semejante o paralela, otra prueba diferente, las opiniones de los profesores, de un experto o de otras personas, sobre puntos relevantes similares a los que pretende evaluar la prueba, o incluso mediante la comparación con una tarea de la vida real que integre los elementos de la misma habilidad que se quiere medir en el test (ALTE, 2001: 9).

La comparación entre ambas pruebas o datos se hace como una correlación y en la medida en que muestren resultados semejantes (los coeficientes no suelen superar más de 0.5/0.7 en estos casos, ya que las pruebas nunca son idénticas), se puede decir que nuestra prueba acierta a determinar esos conceptos.

Cabe preguntarse que si para validar una prueba estamos utilizando otra ya conocida, qué sentido tiene intentar validar la nueva prueba. Al diseñar una prueba se intenta hacerlo en base a determinados parámetros y necesidades de organización, atendiendo a una tipología de candidatos que van a ser sus destinatarios, a los tiempos de administración de que se disponga cuando se aplique, etc. Los datos extras que estamos utilizando para validar la prueba no siempre están disponibles, por ejemplo cuando queramos evaluar a un candidato del que no se conozca nada porque acabe de

esos datos pueden hacer replantearse numerosos elementos de la composición, contenido y forma de administración de un test, así como de su fiabilidad Para explicaciones más complejas al respecto se recomiendan libros como el de Bachman (2004) o Fulcher (2003), ambos con buenos ejemplos de cada una de estas técnicas.

llegar a un centro. La validación que estamos haciendo durante en el proceso de pilotaje de las pruebas se ayuda de estos datos extras a través de candidatos que se conocen y de los que se puede establecer una correlación con otros resultados u opiniones. Una vez validada la prueba, sabiendo que efectivamente sirve para nuestros propósitos, servirá para -sin necesidad de hacer esas otras averiguaciones ni correlaciones- estimar que los resultados ofrecidos ante la actuación de otros candidatos desconocidos son realistas y prueban unas características y habilidades que se desconocen pero que anteriormente - con otros candidatos- resultaron ciertas.

Este tipo de validez, la validez concurrente, también es necesaria para el diseño de dos pruebas paralelas o similares que se desea que ofrezcan idénticos resultados, por ejemplo si van a usarse en diferentes momentos, si existe una prueba que es cara o larga de administrar y se necesita en algún momento otra más breve o barata; o si se necesita evaluar a un grupo de personas varias veces consecutivas pero -lógicamente- no se les puede dar siempre el mismo modelo de examen. En estos casos se necesitan diversas pruebas que sean concurrentes, paralelas o similares y que obtengan un índice de correlación alto, de 0.9 al menos. No obstante, ya que es muy difícil conseguir dos pruebas iguales, lo normal es proceder a la ponderación de sus resultados y a establecer algún sistema de equiparación justo: si sabemos que una resulta más sencilla que la segunda, podemos hacer que el aprobado se obtenga con más puntuación en una que en otra.

2.3.6 Validez predictiva

Este es otro tipo de validación externa. Se diferencia de la validez concurrente en que los datos no se recogen en un tiempo posterior a la prueba, pues se trata de averiguar en qué medida la prueba predice que un candidato va a ser capaz de hacer determinadas cosas en adelante.

La recogida de datos en la fase de pilotaje del test trata de probar si efectivamente lo que predice la prueba es real, si el candidato será capaz de realizar lo que la prueba dice. Para ello se deben realizar pruebas al cabo del tiempo sobre la habilidad que predecía la primera prueba. Se suelen emplear en pruebas de dominio, que son las que intentan predecir lo bien que un candidato sabrá hacer determinadas cosas en su vida futura. Esto puede ser a la semana, al mes o al cabo de más tiempo, según lo que prediga la primera

prueba. Por ejemplo, si queremos hacer un test para clasificar estudiantes en un centro de lenguas, la prueba segunda con que se necesita comparar la prueba de clasificación para probar que efectivamente lo que se deducía de la prueba de clasificación era verdad pueden ser las opiniones de los profesores respecto a si sus estudiantes están bien agrupados en los diferentes niveles de los cursos. Si la prueba era para establecer diferencias entre conocimientos y habilidades para asignarlos a grupos diferentes, y al cabo de -por ejemplo- una semana, los profesores dicen que los estudiantes están muy mal organizados, la capacidad predictiva (y en última instancia la validez de la prueba de clasificación) para decir qué dominan y qué no dominan los candidatos y con ese dato asignarles a un curso determinado, habrá fracasado.

Bachman / Palmer entienden la validez predictiva en el sentido ya analizado de la relación entre un test y el lenguaje que elicitaba, comparado con el de las tareas de la vida real (§2.2)⁵⁶.

2.3.7 Uso práctico

Todavía hay un elemento que se debe considerar, aunque no se trata normalmente por todos los autores ya que es relativamente moderno en la bibliografía sobre evaluación de lenguas, y sin embargo es importante: la viabilidad o la posibilidad de uso práctico de los tests diseñados. A esto hacíamos referencia al principio de este capítulo cuando advertíamos de la dificultad de llevar a la práctica y poner en escena tests orales frente a los demás modelos de tests habituales en la evaluación de lenguas.

Este concepto se entiende como el grado en que un examen se puede poner en práctica en relación a la cantidad de medios que se necesitan para su uso. Según ALTE (2001: 14) el uso práctico de un examen depende de dos factores: por un lado, los medios que se necesiten para validarlo y probar que es operativo; por otro, la cantidad de medios que se necesitan para administrarlo y darlo a los candidatos. Si un test necesita excesivos medios, se pueden producir variaciones en la forma en que se administra o problemas en su uso; para evitarlo debe revisarse o variarse.

⁵⁶ También hablan de un concepto que no vamos a tratar porque no es necesario en nuestra investigación, al estar conectado con los efectos de un test. Es lo que ellos denominan “impacto” de un test (Bachman / Palmer, 1996: 29).

El tema es aparentemente sencillo de entender e incluso de controlar en la mayoría de los tests utilizados normalmente, aunque a veces los descuidos en la forma de administrarlo pueden llevar a pérdidas de fiabilidad (véase §2.3.8). No obstante, cuando las habilidades a medir son orales, su importancia adquiere una relevancia mayor y es muchas veces la causa de que las pruebas orales no se hagan de forma del todo efectiva e incluso de que su proceso de diseño y validación no sea el adecuado.

2.3.8 Fiabilidad

Donde dije digo, digo Diego
(Dicho popular).

Hasta aquí hemos tratado los conceptos básicos sobre validación de tests, pero no hemos terminado de explicar los pasos necesarios para probar que un test vale efectivamente para lo que se haya probado. Nos queda un último punto, que no es exactamente validación, pues no trata de averiguar qué mide efectivamente un test o si mide lo que se pretendía⁵⁷. Se trata de la fiabilidad.

La fiabilidad es la prueba de que el test mide la habilidad lingüística deseada y no otras cosas; y de que se puede confiar en sus resultados. Con ella se debe demostrar que el resultado del test es debido a este, o al incremento de la habilidad por parte del candidato, y no a otras variables aleatorias (Bachman, 2004: 166). En un test el resultado o nota se puede deber al conocimiento o habilidad del candidato, pero en cierta medida se debe también a otros elementos indeseables que le afectan (edad, constructo, método, suerte o conocimiento del tema, cultura, etc.).

Ya hemos visto que el diseño del test trata de minimizar todo esto reduciendo el valor de variables irrelevantes para los objetivos del test. No obstante, nunca se logra completamente, por lo que siempre diremos que hay un factor de error, suerte o azar. La fiabilidad de un test depende de que estos elementos estén verdaderamente minimizados y controlados. En palabras de Fulcher (2003: 212) “representa el grado en que una nota

⁵⁷ Aunque últimamente las fronteras entre ambos conceptos se están disolviendo, más aún que lo que ya Alderson explicaba en su libro (Alderson, 1998).

no depende de otros factores salvo la variación de la habilidad que se haya decidido que se desea medir”. O también, “hasta qué punto es uniforme la extracción de información y la medición de la información obtenida, es decir, hasta qué punto las propiedades del instrumento medidor o la aplicación del mismo provoca errores en la medida” (Ortega, 1996: 20).

Para Bachman, la fiabilidad puede definirse como la correlación entre dos grupos de notas paralelas (Bachman, 2004: 159). Es decir, “el grado en que se repiten el mismo orden de los candidatos en cuanto a las calificaciones obtenidas en dos convocatorias distintas de la misma prueba de evaluación” (Consejo de Europa, 2002:177). Si pudiéramos hacer un mismo test con las mismas personas varias veces sucesivas -sin que estos participantes hubieran aprendido nada entre una y otra vez, y sin que el hecho de haber contestado ya una vez a las preguntas les ayude a hacerlo mejor las siguientes veces- estas personas deberían obtener siempre los mismos resultados, todas las veces igual que la primera.

Es, por lo tanto, la capacidad para medir cambios sistemáticos -debido al aprendizaje o al aumento de la habilidad-, y la garantía de que otros cambios no sistemáticos -debidos a otras características no evaluadas mediante el test- no afectan a sus resultados, no se tienen en cuenta, no se contabilizan o no influyen en su realización.

ALTE (2001: 11), al igual que Bachman, aclara este punto aún más diciendo que es el grado en que el test está exento de errores de medida.

La fiabilidad, explica Bachman, puede depender de varios factores: (1) inconsistencia entre los ítems o las tareas del test; (2) inconsistencia debida al tiempo en que se administra el test (3) inconsistencias en las diferentes formas en que existe o se administra el test y (4) inconsistencia entre los evaluadores entre sí y con respecto a sí mismos (Bachman, 2004: 160). Es, pues, algo que busca la consistencia en los resultados por encima de factores como el formato del test, los calificadores y otras características de los candidatos o del contexto en que se realizan los tests. Si estas variaciones no pueden ser contabilizadas se llaman *error de medida* y reducen la fiabilidad de un test y de las notas que obtienen los candidatos.

Todos sabemos, cuando vamos a comprar un coche, que este tiene un comportamiento determinado que se va repetir mientras se mantenga en buenas condiciones, como una aceleración en un tiempo determinado, una inclinación al tomar

las curvas a cierta velocidad, un consumo cada 100 kilómetros, etc. Si estas características no estuvieran controladas y no pudiéramos asegurar que el coche se va a comportar de esa manera sistemática predecible, no estaríamos seguros de que al pisar el acelerador un poco conseguiríamos una cierta velocidad y que al apretar más a fondo esta aumentaría en una proporción conocida. Y, seguramente, no nos sentiríamos seguros al volante o no lo compraríamos. Lo mismo pasa con los tests.

Bachman / Palmer (1996), cuando explican el concepto, extienden su análisis a todos los elementos externos relacionados con el test, su lugar y proceso de realización, mediante la siguiente lista de planteamientos:

- *¿Cambian las características del lugar en que se realiza el test de una vez a otra?*
- *¿Pueden cambiar las características del test o de sus partes o hay diferentes formatos del test?*
- *¿En qué medida cambia el input dado de una tarea a otra, de una parte del test a otra o en diversas administraciones del test?*
- *¿En qué medida varían las características de la respuesta esperada de una parte a otra, de una tarea a otra o en diversas administraciones del test?*
- *¿En qué medida cambia la relación entre el input y la respuesta de una forma injustificada de una parte a otra o en diversas administraciones del test? (Bachman / Palmer, 1996:139).*

Estos elementos tratan de asegurarse de que las pruebas se administren y corrijan de forma consistente, de que las instrucciones estén claras y no haya tareas o ítems ambiguos.

En cuanto a las técnicas estadísticas, se centran en el análisis del contenido del test y de las respuestas dadas por los candidatos. Se puede hacer de diversas maneras. La menos aconsejable es la de repetir el test al mismo candidato, pues lógicamente este puede hacerlo de forma diferente esta segunda vez. Veremos más adelante que la fiabilidad en pruebas abiertas orales no tiene por qué cumplir esta premisa ya que si el test fuera enteramente fiable, los candidatos no deberían poder mostrar mayores habilidades la segunda vez, ni los protocolos de realización -responsabilidad del evaluador- deberían permitir que la primera vez el discurso generado sea de poca calidad, de manera que sea fácilmente mejorable en una segunda oportunidad, o viceversa.

Otra manera de obtener fiabilidad sería dando a los candidatos otra prueba paralela, similar o equivalente a la anterior (de igual validez concurrente -§2.3.5). Pero muy pocas veces existe esta otra prueba paralela, por lo que el sistema tampoco es operativo.

Otra manera consiste en lo que se llama la *consistencia interna del test*⁵⁸; esta radica en simular el sistema de las pruebas paralelas dividiendo el test en dos partes y calificándolas por separado para averiguar su correlación, la medida en que son equivalentes. Pero este procedimiento tiene el peligro de que una mitad resulte ser más fácil que la otra o diferente y la fiabilidad resulta poco satisfactoria.

El método más efectivo y utilizado es una extensión del anterior de las dos mitades pero suponiendo que se realizan todas las posibles divisiones del test en dos mitades, para lo que se necesita algo más que una simple correlación, un conjunto de correlaciones entre todas las subdivisiones posibles. La fórmula utilizada para estos casos es la obtenida por Kuder Richardson en dos versiones (KR-20 y KR-21⁵⁹) según se utilicen notas finales y varianza entre notas obtenidas por el conjunto de estudiantes presentados a la prueba o el análisis de cada uno de sus ítems.

Un aspecto importante a tener en cuenta en pruebas abiertas y orales es que la fiabilidad del test depende en gran medida de la consistencia de sus calificaciones, como en cualquier otro test, pero en este caso estas se producen por el juicio (y aplicación de todo tipo de protocolos) del evaluador; por ello la fiabilidad de una prueba abierta oral depende del evaluador⁶⁰ y de los protocolos de formación y de corrección establecidos. Para ello se utilizan correlaciones y otras estimaciones estadísticas para averiguar la fiabilidad de los evaluadores y por ende del test, con relación a las notas de un mismo evaluador y de este con relación al grupo de evaluadores.

2.3.9 Conclusión

Hemos citado al comienzo las palabras de Alderson sobre la imposibilidad efectiva de saber qué sentido dar a una calificación, y el agujero negro de la evaluación, así como el escepticismo de Widdowson sobre la definición de la competencia con fines de evaluación. Es cierto que “con lo que se sabe sobre la variación en la actuación lingüística es problemática cualquier conceptualización de la fiabilidad” (Alderson / Banerjee, 2002: 101).

⁵⁸ Remítimos a libros especializados para explicaciones sobre su sistema y posibilidades. Hay una explicación sencilla en Alderson, 1998: 265-66.

⁵⁹ Decimos lo mismo que en la nota anterior, ahora en Alderson, 1998: 267-68.

⁶⁰ Más adelante, en §2.4, comentamos estos procesos en los que el evaluador es el centro de la evaluación. También se trata el tema en el apartado dedicado a la formación de evaluadores, §2.4.3- 4.

Alderson dice que en sus análisis de los tribunales ingleses halló deficiencias grandes en todos los campos que acabamos de explicar, que deberían ser subsanadas. Entre sus críticas cita la ausencia de ensayos previos, la ausencia de ensayos en pruebas abiertas y la ausencia de evidencias empíricas sobre la validación de los exámenes, junto a falta de datos, dificultad de acceso a ellos y deficiencias en la comprensión respecto a lo que es la validación, la validez y la fiabilidad. Y cuando sí se afirma que hubo un proceso de validación se echan en falta las necesarias explicaciones respecto a cómo se desarrolló el proceso o se nota una falta de procedimientos que garanticen la consistencia de los resultados de las pruebas en los años consecutivos o en diferentes versiones de los mismos exámenes. “Se hacen afirmaciones infundadas sobre la calidad de los exámenes que deberían analizarse de forma crítica”, “debería medirse la fiabilidad, no solo afirmarla” (Alderson, 1998: 245).

En ese tiempo, 1995, cuando Alderson escribía su libro, afirmaba que no existían un conjunto unitario de criterios a seguir. Pero creemos necesario animar sobre las posibilidades de diseñar tests para ámbitos generales o académicos atajando al mismo tiempo los miedos y frustraciones que a todo profesor o profesional de las lenguas le pueden surgir ante el camino que hemos señalado. Evitando excesos, tanto por el lado de la dificultad técnica o estadística como por el lado del todo vale; para saber en cada momento a qué atenerse y qué recursos utilizar o desechar por innecesarios.

Nosotros estamos totalmente de acuerdo con la afirmación de que “los ensayos previos son importantes para corroborar o negar el valor de las opiniones de los revisores, moderadores o examinadores” (Alderson, 1998: 245) y con esta premisa hemos diseñado un conjunto de tareas para evaluar oralmente a candidatos que serían calificados de acuerdo con el MCER (véase capítulo 3) intentado hacer todos los ensayos de validación necesarios para desarrollar los argumentos que indiquen que el test puede servir para evaluar a los candidatos. Y hemos pretendido que este conjunto de tareas sirvan para comparar los resultados con las pruebas de otro test ya existente, la OPI de ACTFL. Es una tarea costosa pero no imposible.

En este sentido son balsámicas y clarificadoras las palabras de Bachman / Palmer en el comienzo de su libro *Language Testing in Practice* (1996: 7) sobre las falsas expectativas y frustraciones en las que los no expertos en la materia pueden caer al plantearse el diseño y uso de un test. Infundadas, a su juicio, si se tienen en cuenta unas pocas cosas de antemano que den lugar al optimismo y animen a los profesores a participar en el proceso de evaluación. Son las siguientes:

- No dejarse llevar por la idea de que solo hay un test (un buen test o el mejor) para una situación concreta.
- No infravalorar el proceso de diseño de test.
- No tener una fe ciega en los aspectos cuantitativos y tecnológicos del diseño.
- No tener falsas expectativas sobre lo que un test determinado hace o debería hacer.

Estas falsedades llevan a problemas y errores como el de usar un test que no sirve para los candidatos o no responde a las necesidades de la situación; a utilizar un test solo porque parezca bueno o sea famoso; a frustrarse al no encontrar o no poder diseñar el test perfecto; o a perder la confianza en la capacidad personal para diseñarlo. No existe un test que sea el mejor test y que sirva para medir la lengua, al contrario, hay muy diversos tests que la acotan y que hay que usar en virtud de la necesidad y objetivos de cada momento.

2.4 Las características de una conversación natural y la evaluación oral

2.4.1 Planteamiento general

¿Es la OPI una conversación natural? La pregunta intentaba descalificar a este modelo de entrevistas haciendo que al no serlo perdiera su naturalidad y con ello parte de su validez como método de evaluación. Para buscar una respuesta se hicieron muchas investigaciones sobre las entrevistas de lengua en general, las LPI⁶¹. Las principales objeciones a las entrevistas orales decían que lo que generaban no era una conversación porque la estructura y concepto de una entrevista oral no se asemejaba a lo que realmente es una conversación natural. De todas estas investigaciones se pueden extraer muchas conclusiones que ayudan a entender el proceso interno de una entrevista oral.

Son muchas las características que se han señalado como necesarias para hacer que “una serie de frases dichas por dos o varios interlocutores sean realmente un discurso o un diálogo” (Soler, 2000: 259). ¿Se producen en una entrevista oral esas características?; la estructura de un test oral, su tipología de preguntas y mecanismos para elicitar la información y la interacción ¿ayudan a producir muestras de lengua

⁶¹ A diferencia de las OPI (*Oral Proficiency Interview*), una LPI es cualquier otro tipo de entrevista oral, *Language Proficiency Interview*.

naturales?; ¿qué elementos de la conversación natural y de la competencia comunicativa surgen en una conversación de entrevista oral?; ¿cuáles de ellos se desean evaluar y, por lo tanto, se han considerado en las escalas utilizadas para hacer las calificaciones? En definitiva: ¿invalida a un test oral la falta de determinados elementos propios de una conversación natural?⁶²

Es obvio que una de las principales limitaciones de una entrevista para hacer surgir los elementos de una conversación está en la (carencia de) habilidad del candidato. Hablar una lengua extranjera es una cosa difícil y limitada enormemente por las situaciones, el vocabulario, la capacidad de entendimiento y muchos otros elementos lingüísticos, pragmáticos y culturales.

En una entrevista con un no nativo de poco nivel no pueden aparecer todos estos elementos, no solo debido al formato especial de la entrevista sino por la propia naturaleza del discurso a evaluar y sus carencias. Invalidar una entrevista por estas limitaciones es también invalidar casi toda la práctica pedagógica y de enseñanza de lenguas. En el aula, así como en la mayoría de ejemplos de diálogos de los libros de enseñanza de lenguas para extranjeros, tampoco se dan muchas de estas características pragmáticas. Falta la toma libre de turnos, las normales interrupciones en el discurso, el desarrollo progresivo de temas, la colaboración del oyente para construir el diálogo, implicaturas, etc.⁶³

Lo importante no es que se puedan identificar todas esas características sino que encontremos aquellas necesarias para evaluar y, sobre todo, que el hecho de que si aparecen otras ello no penalice o premie a la persona que las haya producido, frente a otras personas en las que no surjan, si estas otras características no estaban previamente descritas y consideradas en la evaluación y en las escalas de dominio utilizadas⁶⁴. Para lograr esto, los objetivos de evaluación deben ser homogéneos y claros para todas las

⁶² Podemos dar una respuesta con las palabras de Moder / Halleck (1998: 118), para quienes aunque la competencia comunicativa interviene en muchas situaciones, la interacción no es siempre lo más interesante en ellas. Aducen que un estudiante de inglés, por ejemplo, seguramente se encontrará ante muchas situaciones con asimetría de roles (hacer una petición, explicar un problema, solicitar un servicio) o ante relaciones muy variadas que disten de ser una conversación informal interactiva.

⁶³ Una conversación natural, incluso entre un nativo y un no nativo, es solo posible en condiciones naturales. Cualquier otro intento quedará lejos de producir los mismos elementos y características que hubieran surgido de forma natural. Las condiciones del discurso en evaluación cambian notablemente respecto a las condiciones en un discurso natural. Cualquier conversación se convierte en un tipo especial de conversación, con sus características especiales, como puede serlo cualquier otro modelo de entrevista, como la de un periodista, una cita médica o una entrevista de trabajo. En la evaluación oral, el diálogo no es natural ni realista porque el candidato y el evaluador modifican y desvirtúan los datos y las intenciones, las implicaturas y los elementos pragmáticos y contextuales. Pero esto no debe invalidarlas.

⁶⁴ En §4.2.1- A, al hablar de la ambigüedad de interpretación de las escalas y, en concreto, del candidato 33, se ve cómo un criterio nuevo, no considerado inicialmente, puede afectar a la evaluación.

personas y se deben mantener en todo momento, sin dejarse llevar por aspectos que ocasionalmente pueden aparecer y parecer interesantes⁶⁵.

Muchos trabajos se han esforzado por dilucidar si la OPI era una conversación o no⁶⁶ y han planteado nuevos problemas que han obligado a matizar el alcance de las descalificaciones. También han abierto nuevos campos de análisis con los que se ha pasado a investigar el papel que juega el propio entrevistador en una entrevista y su constructo (Lazaraton, 2002, Brown, 2003) y con ello en la consecución de que esta sea natural o no.

Por ello, las críticas a la naturalidad del discurso han sido sustituidas ahora por análisis respecto a si las técnicas de elicitación del discurso y el comportamiento del evaluador generan un discurso especial o influyen subordinando el discurso del candidato al comportamiento del evaluador y comprometiendo el resultado de la entrevista y sus calificaciones.

Así, se ha llegado a entender que los modos de los entrevistadores pueden variar mucho, haciendo que cada entrevista sea diferente a otra e impidiendo a los candidatos tener las mismas oportunidades (Lazaraton, 1996: 166) e imposibilitando que sean evaluados o calificados de la misma manera y que puedan demostrar sus habilidades en igualdad de condiciones. Esto puede invalidar gran parte de los tests orales que se están haciendo en la actualidad o, por lo menos, comprometer la fiabilidad y el constructo del test (Brown, 2003: 19).

El tema es importante porque compromete no solo al constructo del test y a su contenido sino los juicios que de ellos resulten. Y porque en ello se ven implicados no solo el evaluador y el candidato, sino la propia estructura del test elegido.

Young (1995) encontró que en las entrevistas hay diferencias provocadas por la propia estructura de estas y su influencia en el tipo de interacción, a menudo ocasionadas por el nivel del entrevistado⁶⁷. Brown dice que las diferencias que se

⁶⁵ En ACTFL, la toma libre de turnos de palabra y el desarrollo más conversacional se consideran en el nivel avanzado y allí aparecen dotados de sentido de dominio del lenguaje. Claro que en otros niveles pueden surgir, como muchos otros elementos, pero no se consideran como aspectos a tener en cuenta en la calificación hasta que el resto de elementos necesarios para dominar el nivel no hayan mostrado que se posee dicho nivel.

⁶⁶ De ello se trata en el tema relativo a la OPI en §2.6.3. Respecto a las entrevistas en general, sean LPI o OPI, la mayoría ha encontrado que contienen numerosas características de una conversación que se comparten con otros elementos más estructurados y propios de una entrevista. Young / Milanovic, (1991); Lazaraton (1992; 1996 y 2002); Young (1992 y 1995); He / Young, (1998); Ross (1992 y 1998); Moder / Halleck, (1998).

⁶⁷ Entre sus conclusiones resalta el hecho de que, en su opinión, la estructura de la entrevista del FCE (*First Certificate of English*) puede impedir que los candidatos muestren sus diferencias de nivel

pueden generar por la estructura del test o la actuación del entrevistador son tales que pueden llevar a variar el constructo del test (Brown, 2003: 20)⁶⁸.

He y Young (1998) señalan que el discurso es algo que se construye constantemente mediante la acción de todos los participantes (He / Young, 1998: 5)⁶⁹. También Lazaraton la señala como algo natural diciendo que “el evaluador es tan responsable de lo que esté sucediendo en una entrevista como el candidato” (Lazaraton, 2002: 44). Para esta autora, el resultado de una evaluación oral, como el de una conversación, es el resultado de la colaboración entre todas las partes. Los enunciados de las preguntas no son meros estímulos en forma de pregunta a los que el candidato responde, sino que vienen formulados dentro de una situación en el transcurso de la entrevista y responden a un contexto generado en ella, de significados establecidos, nuevos o por establecer, y pueden ser elaborados de muchas formas, reelaborados, o reducidos.

Otro elemento que se está viendo como determinante del resultado de una entrevista oral es el propósito que el candidato tenga al hablar, junto a su percepción de lo que se le pide que haga en cada momento (He / Young, 1998: 11), es decir, las intenciones de su discurso, con independencia de las que el evaluador haya planeado⁷⁰. Esto depende de un mutuo acuerdo entre los participantes, así como de la colaboración entre ambos para dar sentido las preguntas y situaciones que se generen en el diálogo.

(Young, 1995: 36), lo que le lleva a decir que este examen, el examen de UCLES/FCE no permite al candidato dar lo mejor de sí mismo (Young, 1995: 28).

⁶⁸ Los análisis realizados por el momento (Lazaraton, 2002, con los Certificados de Cambridge entre otros) no son alentadores. Aunque Lazaraton aclara que este hecho no es en sí malo. Y se pregunta qué hacer con ello: ¿tolerarlo?, ¿alentarlo?, ¿prohibirlo? Pero no da la respuesta pues “los efectos de esta ayuda sobre las calificaciones no están aún claros” (Lazaraton, 1996: 151). Ello la lleva a plantear la necesidad de más investigaciones y estudios que analicen cualitativamente, según los principios del análisis de la conversación, los tests. Para nosotros no es cuestión de tolerarlo, ya que lo dejaría al libre albedrío de cada evaluador y del nivel de empatía que surja con sus candidatos en el proceso interactivo de desarrollo de temas, haciendo cada test diferente y vulnerable a variables no consideradas en su descripción. Tampoco creemos que la solución sea prohibir una práctica que es buena en sí y una característica de la conversación y por lo tanto de las habilidades que se desean medir de alguna manera.

⁶⁹ Para estos autores la habilidad no preexiste a la entrevista sino que se construye durante ella, ellos (los entrevistadores) producen o fabrican las habilidades que se supone que se miden (He / Young, 1998: 8). En esta línea está el trabajo de Ross (1998) para quien en las entrevistas orales se crea un contexto específico que influye en el tipo de respuestas que los candidatos dan (1998: 336).

⁷⁰ Este aspecto de la interpretación de las intenciones tiene enorme importancia a la hora de realizar una entrevista para evaluación oral, como se ve en §2.6 al explicar la OPI, y como hemos pretendido en el diseño de nuestro test oral del MCER. Ello permite que la elicitación sea adecuada y no se cree estrés (véase §2.4.3- 3 y §2.6.2- 4). Junto a ello, es igualmente importante ser consciente de la estructura de la entrevista y del manejo adecuado de cada fase.

Diseñar un buen test oral y una adecuada formación de evaluadores que permita elicitar e interpretar los discursos de forma consistente⁷¹ requiere ser consciente del papel tan importante que el entrevistador tiene en la génesis del discurso, junto con la dificultad de la tarea, de la situación, y del contexto producido en la interacción a lo largo y dentro del test. Por ello nos parece importante entender el proceso de creación de sentido que se produce en una entrevista, proceso que deben conocer el entrevistador y el calificador, y al que está sujeto el candidato consciente o inconscientemente.

Pero todavía los resultados de los análisis e investigaciones no aportan suficientes datos. El problema está en dilucidar el tipo de interacción que se produce en las entrevistas y, lo que es de más calado, el papel del entrevistador en todo ello, en la realización del proceso mediante el cual se toman las decisiones finales sobre la habilidad y en la forma de tomar esas decisiones (Ross, 1992: 173). Lo cual exige más profundidad de análisis de la conversación y de la interacción, para analizar en qué medida todo ello influye en el discurso, en la estructura de la entrevista y en las calificaciones (Leung / Leukowicz, 2000: 217).

Todos estos factores dan al entrevistador un papel mucho más importante del que hasta ahora se había considerado y parece ser que ello puede tener consecuencias en las calificaciones, si no se le presta la debida atención. A partir de ahora será necesario analizar cómo actúa el candidato, cómo interpreta las preguntas en el transcurso de la reorganización del contexto y sentido dentro de la entrevista y, asimismo, cómo interpreta el entrevistador las respuestas, sobre la marcha.

Esto lo vamos a ver someramente en el siguiente apartado y en §2.4.3- 2 y §2.4.3-3.

2.4.2 El papel del entrevistador en la génesis del discurso en las entrevistas orales

Es bien sabido que un elemento que resta validez y fiabilidad a un test es el hecho de que un mismo candidato pudiera recibir diferentes calificaciones en la misma entrevista. El análisis de la conversación comentado en el apartado anterior ha traído

⁷¹ Esto lo hemos hecho en el capítulo 3 y lo veremos de forma práctica en el capítulo dedicado a exponer y analizar los resultados de las entrevistas y especialmente en §4.2.1- B.

nuevos elementos a considerar con tanta o más importancia que los anteriores y ahora un nuevo encargado de demostrar la validez del resultado es el análisis del discurso hecho desde dentro⁷².

En este sentido el entrevistador ha dejado de ser un mero actor secundario que ponía en práctica, con mayor o menor celo, todo el mecanismo de actividades, tareas y preguntas que un equipo de expertos había diseñado cuidadosa y estadísticamente. El evaluador ha pasado a considerarse como *cooperador necesario* en la construcción del discurso del candidato, mediante la selección de temas a tratar, mediante el uso de la entrevista y de las técnicas elicitoras, adaptándose al nivel y necesidades del candidato, interpretando su discurso o interpretando las escalas de calificación. La carga de la culpa recae ahora sobre él haciéndole cómplice de los resultados y va a ser necesario, para recuperar la objetividad de los tests, establecer primero cómo y en qué medida se produce esa complicidad, para luego conseguir que ello no suceda de forma penalizadora.

McNamara / Lumley (1997: 154) lo vieron al decir que “la enorme riqueza de la interacción que surge cara a cara en la calificación del habla tiene sus propios problemas; la nota del candidato es claramente el resultado de un conjunto de variables, de las que solo una es la habilidad del candidato”.

Los trabajos que han centrado las discusiones en este elemento ahora considerado trascendental en el proceso de evaluación oral y distintivo de los demás tipos de evaluación empezaron con el estudio ya citado de Young / Milanovic (1992) en el que analizaban el tipo de interacción que se producía en la entrevistas. Luego se comenzó a señalar problemas debidos a la estructura de las entrevistas, directamente relacionados con el tipo de actuación e interacción que posibilitan (McNamara / Lumley, 1997), Lazaraton, 2002). Entre los principales datos que se han ido señalando destacan los de la influencia del entrevistador, mediante la ayuda que le da al candidato al realizar las preguntas o modificarlas y acomodar su discurso mediante la selección que haga de las funciones y temas a tratar y la forma de desarrollarlos, mediante el método de

⁷² Si en los años 80 y 90 las preocupaciones estaban centradas en la construcción del mecanismo de evaluación, del test y su validez, a finales de los 90 se centraron en el papel del evaluador como constructor del discurso y en su influencia en las calificaciones finales. Ya no son solamente el test en sí y sus tareas los catalizadores de las diferencias de habilidad existente entre diferentes candidatos sino que el evaluador, conductor de la máquina, aparece ahora como el responsable de su uso y de los resultados que con ella se obtengan. La pregunta es ahora ¿en qué medida influye el evaluador en la obtención y calificación de las habilidades del candidato? O más crudamente, ¿en qué medida es causante de las diferencias de nivel que él mismo califica?

elicitación o debido al estilo de hacer las preguntas (Ross, 1992; Ross / Berwick, 1992, Lazaraton 1996; Ross, 1998).

Brown (2003: 1) apunta algunas de las maneras que usa el entrevistador para implicarse en la construcción del discurso del candidato mediante el manejo de diversos elementos de la entrevista:

- El manejo de la estructura y las secuencias de los temas de los que se habla.
- Las maneras y técnicas de elicitar la información.
- El tipo de feedback que se da a cada candidato.

También se ha señalado que el entrevistador adapta las preguntas modificándolas de acuerdo con el nivel captado en las respuestas y comprensión del candidato, a veces de la forma natural propia de una conversación normal con no nativos, otras de forma deliberada, de manera que se podía llevar a saber el nivel de estos según el grado de adaptaciones que se produzcan, pues a mayor nivel del candidato se producen menos adaptaciones de las preguntas y del discurso (Ross, 1992; Ross / Berwick, 1992). Aunque no sean concluyentes, hay datos de cómo los candidatos actúan de forma distinta según sus entrevistadores y las diferentes variables de las entrevistas (Young, 1995; Young / Halleck, 1998; Ross / Berwick, 1996; Brown, 2003; Ross 1998).

Entre los sistemas de ayudar al candidato se han marcado⁷³ varios tipos de adaptaciones (Ross, 1992: 177; Ross y Berwick, 1992: 165; Brown, 2003: 17):

- Haciendo preguntas binarias o preguntas que presentan diferentes opciones de respuesta en ellas mismas.
- Desplegando información antes de realizar una pregunta, con frases o informaciones que ayudan a centrar el tema.
 - Simplificando la gramática usada.
 - Reduciendo la velocidad.
 - Articulando de forma especial determinadas palabras que marcan lo especialmente importante.
- Simplificando el léxico.

⁷³ De algunas de estas formas se habla en §2.6.2- 4 al explicar los sistemas de elicitación que utiliza la OPI.

- Haciendo preguntas de comprensión tipo *¿qué significa eso?* que dan la oportunidad de reexplicarse.
- Haciendo preguntas de clarificación.
- Reexplicando las preguntas que no se han entendido bien.
- Reformulando las preguntas.

Lazaraton describe hasta 8 tipos de ayuda. Los enumeramos con una breve explicación (Lazaraton, 1996: 158-65):

1. Preparación del tema: previamente a las preguntas concretas se hace una introducción preparando el contexto para las siguientes preguntas. Por ejemplo: *¿Tiene un trabajo a tiempo parcial?* puede ser una pregunta preparatoria para hablar de planes.

2. Ayudar con el vocabulario que desconoce o completar los enunciados. Este hecho complica la calificación final del candidato en términos de habilidad léxica, pues impide tener certeza respecto a su dominio y conocimiento del vocabulario.

3. Dar feedback evaluativo o juzgar el discurso con expresiones tipo *¡bien!* Esto, dice, puede ser un problema si es malinterpretado por el candidato y en vez de un comentario sobre el contenido del discurso se toma por un comentario sobre su dominio que puede orientarle en una dirección lingüística determinada.

4. Repetir sus palabras, haciéndose eco de lo que dice el candidato o corrigiéndole el discurso⁷⁴.

5. Repetir las preguntas más lentamente: este, señala, es el primer tipo de acomodación al discurso de un no nativo.

6. Hacer preguntas cerradas que solo requieran una sencilla confirmación. Esto impide a los candidatos hablar y mostrar realmente sus habilidades.

7. Ayudarles a terminar las frases y a sacar las conclusiones. Esto, aunque es una actitud muy natural en el discurso nativo, donde las conclusiones se alcanzan por consenso entre los hablantes y demuestra que el evaluador está comprendiendo al candidato, sin embargo también impide al candidato expresarse y le relaja en su actitud de demostrar sus habilidades, permitiéndole que simplifique la amplitud y profundidad de su discurso.

⁷⁴ ACTFL tiene en sus advertencias sobre el comportamiento del evaluador una cuantas que intentan que este no sea fuente de inconsistencias y errores en la evaluación: no reduzca la velocidad al hablar, no haga eco ni corrija, etc. Véase §2.6.2- 4.

8. Reformular las preguntas con otras palabras, para facilitar la comprensión. Acción que, además, y dependiendo del nivel del candidato, puede dificultar la comprensión del objetivo de la conversación o desdibujar el contexto de las preguntas.

Como venimos viendo, el estatuto e importancia del trabajo del entrevistador para evaluación oral ha cambiado desde que se ha estudiado su influencia en la producción del discurso. Creemos que este asunto no sólo debe ser introducido en el diseño del test mediante la formación de evaluadores sino que obliga a replantear la necesidad de definir otros elementos como el marco en que tenga lugar la entrevista, su estructura, los objetivos de cada fase, el sistema de interacción que se va a producir, o las formas de elicitar el discurso e interpretarlo.

La explicación de algunos de ellos la vamos a tratar en el próximo apartado.

2.4.3. Otros elementos a considerar en la evaluación oral

Para terminar de exponer la situación y características de la evaluación oral, y antes de pasar a exponer y confrontar los dos modelos que vamos a utilizar en esta tesis (§2.6 y §2.7), vamos a tratar de exponer algunos de los elementos que deben ser nuevamente considerados a la luz de las nuevas concepciones del papel del entrevistador señaladas en el apartado anterior.

El apartado consta de cuatro puntos: los elementos a reconsiderar en un test (§2.4.3-1), las intenciones del candidato (§2.4.3-2), el estrés comunicativo (§2.4.3-3), y la formación de evaluadores (§2.4.3-4).

2.4.3-1) Elementos a reconsiderar

Al evaluar y calificar el discurso de un candidato hay dos procesos importantes: el de la administración del test y el de su calificación. El primero supone una interacción entre el evaluador y el candidato mediante una serie de tareas y preguntas que provocan el discurso que se va a calificar. El segundo supone una interacción entre el evaluador y el discurso obtenido, mediante el uso de las escalas de calificación.

A nuestro juicio, siguiendo a Lazaraton (2002) así como al sistema de realización de entrevistas de la OPI de ACTFL, en ambos tipos de interacción, tanto en la fase de administración como en la de calificación, se deben tener en cuenta varios elementos más: la consideración de la intención del hablante, de la complejidad y objetivos de la tarea a realizar, y el papel que el evaluador juega en la realización de la entrevista.

Así pues, en una entrevista oral se conjugan multitud de elementos: el evaluador, la tarea, las preguntas, el candidato, el discurso del candidato, la intención de cada pregunta y de cada respuesta, el calificador, las escalas y la complejidad de las tareas. Tanto para llevar a cabo una buena evaluación y elicitación del discurso como para hacer una buena calificación final, es necesario tener en cuenta todos estos elementos; cada uno en su momento oportuno, en el que adquiere su verdadero valor.

Algunos de ellos, tradicionales en los libros sobre evaluación, analizados desde la nueva óptica que considera el comportamiento del evaluador, las intenciones del candidato y la complejidad de las tareas (Young, 1995; Lazaraton, 1996; Ross, 1998; Brown, 2003) suponen un replanteamiento de la concepción de las entrevistas y su estructura. Teniendo en cuenta este nuevo actor, el entrevistador, que siempre había estado oculto y anónimo en las entrevistas, y ese otro aspecto de las intervenciones del candidato, sus intenciones, que pocas veces era atendido o respetado, debemos considerar de otra manera el discurso y proceder a elaborarlo y calificarlo con nuevos mecanismos. Así, hay que considerar al menos dos cosas nuevas: por un lado lo que Lazaraton (2002: 31) entiende como analizar “qué está pasando ahora en la entrevista” y, por otra, lo que llamamos (Hernández, 2006: 30) el análisis del *estrés comunicativo*, que en las entrevistas de ACTFL provoca lo que llaman *brake down* o ruptura lingüística.

Esta nueva óptica exige un replanteamiento de los siguientes conceptos:

- 1- El papel de las escalas de evaluación y de calificación.
- 2- Los métodos de elicitación del discurso y tipología de preguntas utilizados.
- 3- El comportamiento de los evaluadores.
- 4- El método y estructura de la entrevista oral y la grabación de las entrevistas.
- 5- Las pretensiones que tenga el candidato ante las tareas del test.
- 6- La interpretación que el evaluador haga del discurso obtenido, tanto durante la entrevista como durante la fase de calificación.

1. El problema de las escalas y cómo escalonar la competencia comunicativa se ha tratado tradicionalmente respecto a su diseño (véase §2.5), así como respecto a la debida formación de los evaluadores para entender, interpretar y usar las escalas con consistencia y homogeneidad.

Debemos considerar que un buen test oral debe reflexionar sobre este tema como algo más amplio, abarcando al comportamiento del evaluador y planteándose si este varía de unos candidatos a otros, influyendo en los resultados y en la objetividad de sus calificaciones, y analizando qué criterios utiliza para interpretar el discurso de los candidatos y calificarlo.

2. Los métodos de elicitación y la tipología de preguntas también son un tema clásico, pues muchos tests tienen establecidas las preguntas, que se dan escritas a sus evaluadores. A la vista del papel predominante que el entrevistador juega en la construcción e interpretación del discurso, el tema debe ser analizado con mucho más cuidado para lograr que las oportunidades sean las mismas para todos los candidatos. Así, los entrevistadores deben aprender unos protocolos de actuación que permitan utilizar las técnicas del discurso y la conversación con el fin de obtener de cada candidato lo mejor de sus habilidades no por suerte o azar, o por la buena práctica de un determinado entrevistador sino por profesionalidad. En §2.6 y §2.6.3 vamos a ver el modelo de ACTFL, con siete tipos de preguntas, y diferentes trabajos sobre análisis del discurso en la OPI.

3. El comportamiento de los evaluadores era un tema clásico pero tangencial: sólo se analizaba para formarlos con el fin de poner en funcionamiento la correcta administración del test, interpretar el discurso de los candidatos de acuerdo con las escalas y calificarlo de forma homogénea. El tema se ha hecho más extenso ahora, lo acabamos de ver en §2.4, y obliga a replantear buena parte de la naturaleza de los tests orales.

4. El método y estructura de la entrevista también era un tema clásico del que se citaban las posibles influencias que el método podía tener en los resultados, así como otros aspectos sobre la tipología de tareas y su complejidad cognitiva o si su dificultad se adecuaba a los niveles. A todos estos elementos sumamos la necesidad de grabar todas las entrevistas, pues, ante la complejidad de realización de una entrevista oral, es

la única garantía de validez y fiabilidad, o de posibles revisiones de calificaciones, como enseguida veremos.

Los tipos de tareas que se piden al estudiante pueden influir en los resultados de un test. No es lo mismo un test de descripción de imágenes que otro de narración de una historia. La precisión lingüística usada depende de la tarea seleccionada y de su complejidad cognitiva: si les obligamos a tareas muy arriesgadas pueden generalizar y fallar una regla adquirida hace poco y -por el contrario- si quieren ser más eficaces y comunicativos, pueden optar por la estrategia de simplificar su forma de hablar.

Pero, como acabamos de ver, si el evaluador no es solo una máquina que repite preguntas y propone tareas, queda abierta una nueva puerta respecto a la capacidad que tiene para modificar el test, incluso en los tests que tienen un marco cerrado. Lo que obliga a plantearse el proceso del test y entra de lleno en aspectos de su diseño desde el principio, en el planteamiento de sus objetivos, estructura y selección de tareas, así como en el del diseño de los sistemas de elicitación y del uso de las escalas de calificación.

Toda esta creciente complejidad en la concepción de la entrevista y el papel del entrevistador exige que esta se grabe para tener almacenado el trabajo de los dos interlocutores. Los elementos que intervienen en una entrevista son tantos y tan variados que, aunque no se puedan controlar y provocar todos de la misma manera con todos los candidatos, es necesario tenerlos grabados para observar en qué medida con el material obtenido se puede realizar una calificación suficiente y adecuada. ¿Cuáles son las garantías de que la entrevista que se ha generado entre el entrevistador y el candidato ha sido adecuada, posee todos los elementos necesarios para realizar las calificaciones y no ha sufrido colapsos, faltas de cooperación o situaciones de pérdida de fiabilidad? ¿Cuáles son las garantías de que un examen oral ha obtenido la nota que le correspondía, se ha corregido adecuadamente? Solo si se graba se pueden obtener garantías de objetividad, validez y fiabilidad porque es la única forma de que no solo el proceso de calificación sea claro y revisable sino de que el proceso de elicitación sea adecuado.

Para que sea una buena herramienta, una entrevista debe estar construida con una estructura que ayude a determinar el nivel del candidato, cuyo diseño fomente lo esencial, facilite el surgimiento del discurso y no penalice arbitrariamente sino en virtud de la falta de dominio o habilidad. Por esto, tanto estos aspectos tradicionales de los que estamos hablando, como las preguntas que haga el evaluador y su forma de hacerlas, así

como su comportamiento, influyen en la consecución del discurso que va a ser evaluado y del que se extraen las conclusiones sobre el nivel del candidato.

Es pues, crítico, haber realizado una buena entrevista que contenga todos los elementos deseados para la calificación. Y esto exige, además, su grabación, pues, entre otras cosas, impide que los evaluadores tomen las decisiones basándose en aproximaciones o en previsiones de actuaciones que no se dan efectivamente en el examen. Grabar tiene la ventaja de que garantiza el cumplimiento de todos los estándares, permite análisis posteriores, y permite revisar las calificaciones dando al candidato unos derechos que sin duda se merece, además de permitir la creación de un corpus a efectos de formación de evaluadores. Los actuales sistemas digitales evitan que estas grabaciones sean intrusivas e influyan en los resultados. Debidamente informados (y, si es necesario, evitando que aparezca el nombre del candidato) este hecho no debe ser un problema para los candidatos, sino una garantía de profesionalidad.

Así pues, es de gran importancia que se logren obtener muestras y pruebas del mejor saber hablar de los candidatos, y que la entrevista tenga una estructura adecuada donde se puedan entender y considerar los diferentes elementos que contribuyen al éxito de la evaluación (tareas seleccionadas, tiempos de cada fase, temas que se vayan a tratar, grado de profundidad exigido a los análisis, comparaciones, explicaciones, etc.).

El sistema de ACTFL estima que puede haber entrevistas mal hechas por causa del evaluador, que no han sido efectivas para lograr los elementos necesarios y que no son calificables. Se llaman *non ratable interview*. Este hecho, la consideración institucionalizada de que la entrevista puede ser fallida, no se considera en otras entrevistas que no se hacen grabadas, lo que hace pensar que su fiabilidad puede ser baja, ya que oculta, o la falta de conciencia respecto al hecho de que pueden darse circunstancias que hagan imperfecta una entrevista y la invaliden.

También el nivel de autoexigencia por parte del evaluador es mayor si la entrevista se graba y esto se convierte en la garantía de que la entrevista se ha realizado siguiendo las normas diseñadas⁷⁵.

⁷⁵ Este mismo evaluador ha sentido cómo al evaluar a estudiantes usando grabadora o sin ella cambiaba su autoexigencia, la formulación de preguntas y el grado de interés en perseguir un tema en profundidad ante el cansancio o la dificultad mostrada por un candidato difícil. En muchas ocasiones los evaluadores orales deben hacer varias entrevistas por hora y durante varias horas, lo que invalida parte del trabajo, que si no se graba no se puede demostrar que ha sido bien realizado, ni analizarse para calificar adecuadamente.

5. / 6. Los elementos más nuevos serían: (5) el análisis de las intenciones del candidato con el estrés que puede producirse al elaborar el discurso y (6) la necesidad de una adecuada interpretación del discurso por parte del evaluador, tanto durante el proceso de realización de la entrevista como durante la fase final en que se debe proceder a interpretar la fortaleza o debilidad del discurso y calificarlo.

El primero ayuda a centrar en todo momento el discurso de acuerdo con el momento de la entrevista, las características del discurso que va a provocar la tarea, y la personalidad del hablante. El segundo hace que al calificar no se tengan en cuenta realizaciones del discurso del candidato fallidas por factores ajenos al dominio lingüístico; ambos, junto al manejo adecuado de las preguntas elicitoras, logran que se consiga una buena muestra de lengua que sea evaluable sin penalizaciones.

Vamos a ocuparnos de estos dos últimos elementos en los próximos subapartados.

2.4.3- 2) Las intenciones del candidato

Un aspecto muy importante es analizar la intención y los procedimientos que usan los candidatos para cumplir sus propias expectativas (Lazaraton, 2002: 31); esto es, darse cuenta de que cuando el candidato entabla una interacción y ofrece una respuesta, pretende hacer “algo” concreto. Este “algo” no viene determinado sólo por la petición del evaluador al hacer una pregunta o por la tarea que utilice, sino que depende del manejo de las preguntas, de las propias intenciones del candidato, de su interacción con la tarea, o de su personalidad. Y en gran medida depende del momento en que surja, y por lo tanto, de la estructura de la entrevista. Lograr identificar qué está pasando en cada momento e interpretar su sentido de acuerdo con las intenciones del hablante, es algo fundamental para evaluar adecuadamente el discurso que se está produciendo.

Pongamos un ejemplo que volveremos a ver en las explicaciones del método de elicitación de ACTFL (§2.6.2- 4). En la entrevista se plantea al candidato, casi al principio, que nos cuente cómo es su ciudad, una tarea en principio sencilla y que a un nivel medio puede hacer cualquier estudiante (digamos intermedio bajo de ACTFL o B1 del MCER); conoce el tema, es familiar y no se necesita demasiado vocabulario ni estructuras complicadas para dar una pequeña información sobre la ciudad; ya que no se pide nada más complicado como explicar, aclarar o detallar ningún aspecto.

Esta es la respuesta que nos dio un estudiante ante la pregunta “¿Puedes hablarme de tu ciudad?”:

- Mi ciudad es pequeña, no es tan interesante como Chicago porque en Chicago hay más cosas y en mi ciudad es muy aburrida.

En su momento hemos señalado que, ante la brevedad de su respuesta, es importante evitar el comportamiento instintivo de presionar directamente para obtener más datos; hemos dicho que eso sería una mala práctica de elicitación. La solución al problema reside en darse cuenta de las intenciones del candidato, lo que se hace analizando sobre la marcha el desarrollo de su respuesta. Todos los seres humanos hacemos esto de forma intuitiva en una conversación natural y seleccionamos las preguntas siguientes en virtud de nuestros intereses comunicativos. En una prueba oral debemos poner en práctica esa misma perspicacia, pero dejando de lado nuestros intereses comunicativos para someternos a los de la evaluación.

¿Cuál es el objetivo de la pregunta sobre su ciudad? Si está dentro de la fase de calentamiento y toma de contacto con el candidato, es permitirle adaptarse a las circunstancias y ayudarle a hacerse con la situación, entender el acento del evaluador, tranquilizarse y tomar confianza, para que pueda comenzar a hablar con tranquilidad. Y si este objetivo no se está cumpliendo, deberemos hacer algo en la siguiente pregunta e intervención para que la fase funcione y cumpla su objetivo. Cosa que no se logra presionándole para obtener más información así como así.

Pero si el objeto de la pregunta era la búsqueda de la existencia de determinadas habilidades, deberemos retomar ese objetivo y perseguirlo mediante una pregunta nueva que facilite al candidato demostrar que puede hacerlo. La prueba de que no sepa hacer algo y carezca de habilidades lingüísticas para ello estará en el momento en que habiendo tenido las posibilidades reales de demostrarlo, el candidato no haya sido capaz de hacerlo; y no sólo por el mero hecho de que no haya sido capaz de hacerlo, quizás por nerviosismo u otras circunstancias.

En nuestro caso, la pregunta por su ciudad trata de averiguar la capacidad del candidato para elaborar frases breves dando una información familiar de forma creativa. La breve respuesta obtenida no es suficiente prueba de que no sepa responder, pero presionarle para conseguir más información puede colapsar su discurso. Hay que darse cuenta de cuál es el problema que impide al candidato hablar como esperamos con nuestra tarea y pregunta: sus intenciones, sea por lo que sea, por querer hacerlo como en

su lengua materna, por su personalidad, por no haber captado el sentido de la pregunta, etc., están lejos de las nuestras, al haber comenzado a responder haciendo una comparación entre dos ciudades, de calado cognitivo mucho más difícil de lo que pretendíamos y haberse dado cuenta de que no podía continuar, (“no es tan interesante como Chicago porque en Chicago hay más cosas y en mi ciudad...”).

La intención del hablante era darnos una explicación de su ciudad por medio de la comparación con otra. Pero eso choca con la nuestra. Nuestro deber es, pues, reorientar la comunicación del candidato y hacerle entender qué es lo que queremos, para que tenga la posibilidad efectiva de hacerlo y de probar que tiene tal habilidad o que veamos que efectivamente carece de ella.

La solución que dimos en su momento era reformular la pregunta haciéndole centrarse en su ciudad y olvidando Chicago. Para que nos hable con sencillez de algo que conoce y no plantea complicaciones, en vez de intentar comparaciones y discursos más complejos.

Para poder realizar esta labor de interpretación del sentido de las respuestas e intenciones del candidato es crucial que la entrevista tenga fases, objetivos claros en cada una de ellas y tareas con un análisis de su complejidad, que permita averiguar si la tarea ha surtido el efecto deseado o no. Trabajar sin más, planteando situaciones para ver cómo responden los candidatos ante ellas, es trabajar a ciegas, sin poder extraer conclusiones basadas en la habilidad real de los candidatos.

Por ello es fundamental el diseño de la entrevista y su estructura, así como el de las escalas de calificación y tareas, y la formación de los evaluadores para que sepan realizar el trabajo no solo de administrar sino de guiar y elicitar el discurso.

2.4.3- 3) El estrés comunicativo

Los fallos que se pueden provocar en una entrevista si el evaluador no maneja adecuadamente el discurso, los hemos denominado estrés comunicativo⁷⁶ (Martínez

⁷⁶ El estudio de la ansiedad y sus efectos negativos en el aprendizaje de lenguas ha sido un campo de investigación y análisis desde que a finales de los años 80 Horwitz / Cope publicaran su trabajo sobre la validez de una escala para medir la ansiedad ante una lengua extranjera (Horwitz *et ál.*, 1986). Otros estudios y otras herramientas y subescalas de ansiedad fueron definiéndose paulatinamente (McIntyre / Gardner, 1989) para áreas específicas como el efecto de la ansiedad en las actitudes y actuaciones de

Baztán, 2005: 118). Cuando surgen conviene identificarlos y buscarles una solución (Warren, 2008).

Para explicar este concepto tenemos que comenzar por analizar y darle la vuelta al argumento de que el nivel del discurso y de la nota obtenida en un test será mayor cuanto mejor se dominen las habilidades necesarias para el aspecto que se esté evaluando. El argumento, del revés, quedaría así: una muestra de discurso de poca habilidad es una prueba de falta de habilidad y un argumento para dar una nota baja. Esto, después de lo que venimos señalando en los apartados anteriores, especialmente en §2.4, es falso.

Evidentemente, aunque un discurso de más nivel sí sea prueba de nivel, no es cierto que un discurso de menos nivel sea prueba de menor habilidad. Porque, en el transcurso de una situación cualquiera (y en un test esto puede suceder muy a menudo) puede haber muchas causas para que el nivel de habilidad del candidato baje o no se demuestre en su totalidad (a veces el mismo candidato puede no esforzarse lo suficiente como para realizar adecuadamente la tarea) provocando que el candidato no muestre o no pueda mostrar sus habilidades en profundidad, aunque sí las puede poseer y dominar habitualmente.

Estos factores, externos o internos, que van a afectar al discurso del candidato son lo que vamos a considerar en una prueba oral como “estrés comunicativo”. Identificarlos adecuadamente es labor del evaluador, para no juzgar la lengua que provocan como una prueba definitiva del nivel. Y es tarea del evaluador modelar su actuación en dirección y manera tales que permitan al candidato demostrar que posee mayor nivel.

Por estrés comunicativo, para ser concretos, nos referimos a:

(...) cualquier situación ante la que un hablante se encuentra con barreras psicológicas, sociales o lingüísticas tales que sus mensajes se hacen ininteligibles, se desestructuran o se colapsan. Esto le puede suceder a una persona en su propia lengua bien porque debe hablar ante un público que le atemoriza, ante una persona que le ponga nervioso, al querer expresar ideas que no tiene claras o ante limitaciones de tiempo. El causante de los fallos en estos casos no es la falta de dominio lingüístico, sino estos otros casos citados. Igual ocurre -pero con mucha

estudiantes en tests orales (Phillips, 1992), ansiedad ante la lectura en una lengua extranjera (Saito *et ál.*, 1999), ansiedad y el nivel de lengua que se domina (Bailey *et ál.*, 2000), ansiedad de estudiantes universitarios de lenguas extranjeras (Casado / Dereshiwsky, 2001) o ante la escritura en una lengua extranjera (Cheng, 2002).

más frecuencia- con los extranjeros, provocándoles la imposibilidad para expresar ideas para las que conocen las palabras.

Este estrés puede ser difícil de captar para un nativo o para un evaluador no bien formado y llevar a malinterpretar las posibilidades reales del discurso de los extranjeros pudiendo parecer que un inmigrante (o un estudiante) no sepa hablar sin darnos cuenta de que pretende hacer una intervención de más nivel y profundidad de lo que realmente puede. Y debemos considerar también que el estrés comunicativo lo puede ocasionar el entrevistador mediante preguntas inapropiadas o de más nivel del que correspondería. Y que la inconsciencia de estos hechos puede llevar a interpretar incorrectamente las respuestas obtenidas (Martínez Baztán, 2005: 118-119).

La aparición espontánea de estrés en los candidatos puede lastrar el efecto de su discurso provocando colapsos en su lengua y en su formalización lingüística, tales que provoquen la apariencia de un nivel mucho menor del que efectivamente pueden ofrecer en sus discursos y llevar a errores de apreciación graves en los evaluadores. Asimismo estos pueden ser quienes provoquen el estrés en el estudiante ocasionando los efectos citados.

Este estrés del que venimos hablando no tiene por qué ocurrirle solamente a una persona que habla una lengua extranjera. De la misma manera un hablante nativo puede sufrir estrés comunicativo cuando se enfrenta a una situación que no controla o que por diversas razones provoca que no pueda comunicarse de forma apropiada⁷⁷.

En el lenguaje escrito nativo también podemos observar ejemplos de esto; por ejemplo, ante la prisa y falta de espacio para describir un accidente de coche, en algunas de las declaraciones hechas en los formularios de seguros de accidentes en las que los conductores trataban de resumir los detalles con un mínimo de palabras:

El tío estaba por toda la calle y tuve que hacer varias maniobras bruscas antes de atropellarle.

El otro coche chocó contra el mío sin previo aviso de sus intenciones (Calsamiglia, 2002: 213).

⁷⁷ Caplan (1996: 21) ejemplifica de manera muy concreta este tipo de ansiedad comunicativa cuando explica cómo una persona, hablante nativa de español y poseedora del código lingüístico de la lengua española, en determinadas ocasiones puede no ser capaz de utilizarlo. Recurre a la frase “me siento acalorada”. Esta expresión puede utilizarse tanto para pedir al conductor del vehículo en el que viaja que abra la ventanilla, como para indicar al doctor en una consulta médica los síntomas de su enfermedad. Según Caplan, la ansiedad producida en alguna de estas dos situaciones puede provocar que el hablante no logre emplear -o no se atreva- la expresión “me siento acalorada” (que es sumamente descriptiva), provocando así un fallo en la comunicación, bien por emplear otra frase que no haga referencia concreta a lo que realmente quiere comunicar, bien porque no sea capaz de expresarse y por consiguiente no sea capaz de cumplir con la función comunicativa del lenguaje.

Los efectos de esta ansiedad son aún más fuertes en un hablante de lengua extranjera y aún más si está en un examen, sabe que se está midiendo su nivel de lengua en tan solo unos minutos y siente la responsabilidad de querer mostrar en ese poco tiempo, a menudo sin objetivos claros, todo lo que es capaz de hacer o sabe.

También el diseño y la estructura del test y de la entrevista puede ser parte del problema. Young (1995: 36) concluye que en el examen de UCLES/FCE⁷⁸ no se permite al candidato dar lo mejor de sí mismo porque la estructura de la entrevista del FCE puede impedir que los candidatos muestren sus diferencias de nivel. También Brown (2003: 19), tras analizar la variación provocada en el discurso de un mismo candidato por dos evaluadores diferentes, concluye que uno de ellos provocó que el candidato no entendiera el sentido pragmático de sus preguntas y se limitó a dar respuestas breves y con constantes rupturas.

Veamos algún ejemplo de extranjeros en situación de estrés⁷⁹:

Discurso 1:

- *En Estados Unidos, ah... muy... ah... la familia de mí en Estados Unidos, am... trabaja mucho, mucho, am... nunca en la casa nunca porque hermanos están en escuela y padres están en ah... trabajo y ... am... cuando no tra, tienen trabajar or... am... escuela em... mi familia ir a... partidos de fútbol, fútbol americano or... partidos de... am... mi hermanos cuando... ah... tienen partidos de béisbol, fútbol o... no sé.*

Discurso 2:

- *La madres, mm... eh... no sé... comid... am... preparar comida y ... am... el padre trabajo en el campo y much... creo que muchas am... gentes en Granada, am, trabajan en el campos también sí, y ah....*

Discurso 3:

- *Porque... yo tengo... eh, yo tengo en mi escuela, yo tengo... dos amigos muy, muy amable. Un... una... una, una quedado, un ha quede en mi escuela secundario y ... un buen amigo, un buen hombre, em... y yo, yo quiero él ser... ¿you know?... yo quiero, yo quiero él ser... tener... la...el mismo cosas de mi.. Es, es... es un... em... un cosa muy, muy ... importa.*

⁷⁸ Siglas que corresponden al organismo encargado de los primeros tests de inglés de la Universidad de Cambridge: *University of Cambridge Language Examinations Syndicate/ First Certificate of English.*

⁷⁹ Estos ejemplos son nuestros. Hay otros en §4.2.1- B, §2.6.2- 3 (Tiffany y François) y §2.6.2- 4 (ejemplos de preguntas que crean estrés).

Discurso 4:

- *Cuando, en... em... yo fue, yo fui a Bérchules. Y cuando en Bérchules tiene, em... ha tuvo mucho nieve y... por eso, em... no podemos, em... regresar a Granada porque los autobuses no pueden ah ir a Bérchules. Por eso... em... em... nos, nos... que... em... nos, nos quedemos un familia de allí. Muy, muy amable, demasiado amable, es increíble... y... yo aprendí como jugar ah... una... un... un partido con tarjetas de España. Yo tengo las tarjetas en mi mochila ahora. Y, em... y ah... nosotros comemos mucho comida todo el tiempo con esta familia y, em... es en general es un buen tiempo. Es... em... un hombre. Un... em... él...él fue un gay.*

Los ejemplos 1 y 2 son de una persona (la llamaremos Marie) y los ejemplos 3 y 4 de otra (le llamaremos Mike). Cada uno muestra dos formas diferentes de expresarse, producidas por las diferentes tareas que están intentando desarrollar, una en la que logran comunicarse y otra en que no lo logran.

El primer discurso de cada persona (1 y 3) es la muestra del discurso en la que no logra comunicarse bien, y el segundo ejemplo es la muestra del discurso ante una tarea más sencilla (2 y 4). Por ello estos últimos ejemplos quizás parezcan mostrar mayor nivel o mejores posibilidades de expresión.

Del análisis de los dos tipos de discurso empleado se puede concluir que un mismo candidato puede mostrar diversos tipos de discurso con niveles de habilidad diferentes. Así, ciertas características de un discurso pueden desaparecer en determinadas situaciones, mostrando unos errores que en otras situaciones no aparecerían: se olvidan verbos y no se conectan claramente las ideas, aparecen problemas de vocabulario y las frases y sintagmas son menos amplios y más desestructurados.

Veamos las posibilidades expresivas de cada alumno separadamente, tomadas de los análisis de Hernández (2006: 56-60).

Primero los ejemplos 1 y 2 (de Marie), luego los 3 y 4 (de Mike).

(E = entrevistador, M = Marie).

(1)

E.- *¿Me puedes explicar las diferencias de la familia de Estados Unidos y aquí en España?*

M. - *En Estados Unidos, ah... muy... ah... la familia de mí en Estados Unidos, am... trabaja mucho, mucho, am... nunca en la casa nunca porque hermanos están en escuela y padres están en ah... trabajo y ... am... cuando no tra, tienen trabajar or... am... escuela em... mi familia ir a... partidos de fútbol, fútbol americano or... partidos de... am... mi hermanos cuando... ah.... tienen partidos de béisbol, fútbol o... no sé.*

(2)

E- Pero, *¿esa es una familia normal?, ¿tú qué crees?*

M- *Sí*

E.- *Sí, ¿por qué?*

M.- *Ah... porque la madres, mm... eh... no sé... comed... am... preparar comida y... am... el padre trabajo en el campo y much... creo que muchas am... gentes en Granada, am, trabajan en el campos también sí, y ah...*

En ambos discursos la estudiante habla de la familia pero intentando expresar y describir cosas diferentes. En un caso el lenguaje se desestructura, dando la apariencia de no saber hablar (1). Pero tenemos otras pruebas de que sí puede (2). Aunque en ambos el tema es la familia, las funciones son diferentes: en el primero compara las familias de América y de España (1); en el segundo sólo habla de su familia (2).

No saber detectar la diferente intención y las funciones del discurso del candidato conforme a lo explicado en el punto anterior (la intención del hablante) puede llevar a realizar una mala calificación de sus discursos y penalizar el resultado final. Habrá que considerar si los fallos son por falta de habilidad, que es la única causa que puede ser calificada en el test, o por culpa de una mala realización de las entrevistas, por un mal comportamiento del evaluador, una deficiente técnica de elicitación, creación de situaciones contextuales difíciles, o un uso de tareas complicadas. O, incluso, si es por la personalidad o intención del candidato, o la malinterpretación de la pregunta en un contexto complicado.

Lo mismo podríamos decir de los otros ejemplos, 3 y 4, de Mike:

A primera vista habríamos podido pensar que la persona del discurso 4 tiene más nivel que la del discurso 3. Pues bien, el discurso 3 está extraído de un momento de la entrevista en la que el profesor le pide su opinión sobre el matrimonio homosexual. La tarea está por encima de sus posibilidades y por ello el lenguaje se le desestructura.

(E = entrevistador, M = Mike).

(3)

E.- *Sí, ¿y a ti por qué te parece bien? ¿Tú crees que tienen el mismo derecho los gays que los heterosexuales? ¿Por qué te parece a ti bien?*

M.- *Porque... yo tengo... eh, yo tengo en mi escuela, yo tengo... dos amigos muy, muy amable. Un... una... una, una quedado, un ha quede en mi escuela secundario y ... un buen amigo, un buen hombre, em... y yo, yo quiero él ser... ¿you know?... yo quiero, yo quiero él ser... tener... la...el mismo cosas de mi.. Es, es... es un... em... un cosa muy, muy ... importa.*

En el discurso 4, sin embargo, el estudiante está narrando un hecho en pasado. Si bien es cierto que no narra del todo correctamente e incluso se equivoca con ciertas formas del pasado, el discurso se entiende y la comunicación se lleva a cabo, al contrario que en el ejemplo 3 donde es imposible saber de qué está hablando.

(4)

M.- Cuando, en... em... yo fue, yo fui a Bérchules. Y cuando en Bérchules tiene, em... ha tuvo mucho nieve y... por eso, em... no podemos, em... regresar a Granada porque los autobuses no pueden ah ir a Bérchules. Por eso... em... em... nos, nos... que... em... nos, nos quedemos un familia de allí. Muy, muy amable, demasiado amable, es increíble... y... yo aprendí como jugar ah... una... un... un partido con tarjetas de España. Yo tengo las tarjetas en mi mochila ahora. Y, em... y ah... nosotros comemos mucho comida todo el tiempo con esta familia y, em... es en general es un buen tiempo. Es...em.. .un hombre. Un... em... él...él fue un gay.

El análisis morfosintáctico que se realizó de estos discursos (Hernández, 2005: 59-60) muestra que Marie, en su discurso normal (2), utiliza frases casi enteras y de tres elementos: sujeto, verbo y complemento (“ayudan su madre en, en la cocina”, “El padre trabajo en el campo”, “Muchas gentes en Granada trabajan en el campos también”). Además, en cada respuesta usa una cantidad que va de dos hasta cinco frases, compuestas de verbo y complementos. Sin embargo, en su otro discurso, colapsado (1), las frases que formula no son del todo correctas en español, faltan verbos, a veces no los conjuga, recurre en más ocasiones al uso de infinitivos y repite más ideas hasta que logra enlazarlas.

Y en el análisis comunicativo también se hallaron diferencias entre los discursos (Hernández, 2005: 61-64). Marie tiene, normalmente, una comunicación fluida: no busca excesivamente palabras, ni repite ideas, ni reinicia frases. Cumple con la tarea, describir su entorno en frases cortas sobre un tema familiar. En cambio, luego, al intentar comparar los tipos de familias, el oyente tiene que hacer esfuerzos para entender lo que quiere decir y se hace prácticamente necesaria una interpretación de sus palabras. Tarda más en dar respuestas y recurre con demasiada frecuencia a las muletillas porque busca información a la que no accede, no puede conectar ideas, repite palabras o se pierde.

La conjunción de estos dos elementos que estamos explicando, las intenciones del candidato y el estrés producido durante la entrevista, provoca deficiencias en el sistema de evaluar y en las calificaciones. Es labor del diseño del test tenerlas en cuenta

permitiendo que el instrumento de medida sea eficaz. Así como es labor de la formación de evaluadores, de la que vamos a tratar en el siguiente apartado, que el test sea bien utilizado y sirva para obtener las muestras de lengua para las que se diseñó.

Para ello es importante que tanto las técnicas de elicitación como el diseño de la estructura faciliten que el candidato haga bien el trabajo que se propone al someterse a una entrevista. Igualmente es imprescindible plantearse aspectos internos como la actuación del evaluador al elicitar el discurso y la necesidad de interpretarlo en directo. La tarea de interpretación de datos (y de calificación) debe tener en cuenta que a menudo se premia, penaliza o compensa al candidato en virtud de una mala práctica del evaluador que ha provocado que el candidato no haya sido capaz de mostrar todo lo que realmente sabe hacer. Por ello habrá que establecer protocolos que eviten esta práctica inconsciente por parte del evaluador.

Para minimizar los riesgos hay que elaborar protocolos que eviten la ansiedad y ayuden a entender en todo momento el proceso en que está inmerso el candidato, más allá de lo que pretenda el evaluador con sus preguntas y su intención pragmática. Todo ello hace que otro de los principales problemas de la evaluación oral, aparte de los factores que estamos apuntando, esté en la posibilidad de uso práctico del procedimiento, su dificultad añadida de que por muy bien diseñado que esté, el test oral, salvo que sea en formato digital, por ordenador o en internet, siempre dependerá de la forma en que se administre en cada momento. Para administrar un test no es suficiente con tener una buena escala y unas buenas tareas. Hay que saber manejar los temas y la situación, las preguntas y su sentido.

2.4.3- 4) La formación de evaluadores

El campo del entrenamiento de los evaluadores recibe cada vez más atención y son muchas las instituciones, agencias y universidades que han desarrollado procesos de formación para sus evaluadores y protocolos de actuación. Esto es debido al desarrollo del interés por la objetividad de los tests y por la práctica justa, objetiva y fiable de los evaluadores, como consecuencia de los estudios y códigos publicados, como el código

de actuación de la Asociación de evaluadores de lengua europeos (ALTE, 2001)⁸⁰, o los de ética y de buenas prácticas⁸¹ de la Asociación internacional de evaluadores de lenguas (ILTA, 2000).

Pero la atención es aún más importante cuando el test es oral, ya que en un examen oral el examinador se convierte en el examen. Es decir, a diferencia de otros exámenes en los que hay un papel con preguntas y el examinador corrige las respuestas dadas y saca los resultados, en un examen oral el “papel”, el “formato”, es el propio entrevistador, quien hace las preguntas, busca la interacción, reacciona a las respuestas del candidato, guía el diálogo, cambia o introduce un tema, facilita o provoca la discusión, y, mediante su actuación, interviene en el producto tanto o más que el propio candidato.

Incluso en los exámenes orales que están preparados con una lista de preguntas y temas previstos, con juegos de roles o papeles con cómics, noticias de periódico o discusiones. El entrevistador no es un mero lector de esas preguntas, enunciados e instrucciones, sino que incluso con la forma en que dé las instrucciones está interviniendo en el resultado.

En §2.4 señalamos cómo los análisis del discurso generado en las entrevistas han hecho cambiar la visión del estatus del entrevistador. Nos toca ahora sacar conclusiones y llevarlas al terreno de la formación de evaluadores orales, con el fin de que la evaluación sea lo más objetiva, consistente y fiable posible. Tanto para el proceso externo de corrección como para el interno de elicitación.

En ACTFL el tema de la formación del evaluador se tiene en cuenta desde el principio, pues esta organización solo certifica como evaluadores a los miembros que hayan pasado un curso de formación que se compone de varias fases y otorga una validación apta solamente para cuatro años, tras los cuales se debe proceder a reacreditarse como evaluador. Sus talleres de formación provienen de los que ya comenzó el Foreign Service Institute en 1979 cuando realizó el primer taller de formación de evaluadores para aprender el modelo de evaluación, practicar la

⁸⁰ El Código de actuación de ALTE en su primera versión es de 1994.

⁸¹ En su versión 3, de junio de 2005, en su apartado 6 de derechos de los candidatos, decía que estos tienen derecho a que los tests les sean administrados e interpretados por personas adecuadamente entrenadas y que tengan un código profesional de ética (ILTA, 2000: 5). El *Code of Ethics for ILTA* ha sido finalmente aprobado en el *Language Testing Research Colloquium*, Barcelona en 2007, según explica Fulcher, presidente de ILTA en 2006 y miembro permanente del comité honorífico de la asociación, en una entrevista concedida a Newfields en *JALT Testing & Evaluation SIG Newsletter*, 2007, p. 12. No hemos tenido acceso a su versión final. Véase Newfields, 2007.

calificación de entrevistas, analizar y discutir sus calificaciones y aprender las técnicas de elicitación del discurso (Liskin-Gasparro, 1987: 23). Posteriormente, estos se han extendido y son práctica habitual de la organización.

El MCER, aunque considera que se debe obtener la adecuada formación para demostrar que se entienden los criterios de evaluación y se utilizan protocolos adecuados y consistentes, no considera la necesidad de establecer un protocolo de formación para utilizar el MCER, ni a nivel comunitario ni estatal. Este es, sin embargo, importante y necesario. Así lo considera Figueras, quien dice que “el entrenamiento (de los evaluadores para) la evaluación del aprendizaje de lenguas en relación con el MCER debería ser una parte regular de la formación de un profesor” (Figueras, 2005 b: 276), y así lo consideramos nosotros tras los análisis realizados con la interpretación de las escalas y calificación de entrevistas⁸².

No caben dudas sobre la necesidad de una buena formación de evaluadores para evaluación oral, y nadie discreparía de las palabras de Bordón cuando dice que “la estabilidad de los resultados no depende únicamente de la idoneidad de los criterios, sino también de su rigor en la aplicación. Para ello es indispensable contar con evaluadores entrenados en su utilización” (Bordón, 2006: 244), o de Alderson cuando hace hincapié en que “si la puntuación de una prueba no es válida y fiable, todo el trabajo (...) habrá sido una pérdida de tiempo (...) todos los esfuerzos serán vanos si los usuarios de la prueba no pueden confiar en los resultados que los examinadores dan a los candidatos” (Alderson, 1995: 106). No obstante, esta formación es más costosa y difícil que para el resto de las pruebas, ya que exige prácticas y estudios de mayor consideración, en el tiempo y económicamente, así como garantías extras sobre la fiabilidad de los juicios, entre las que se encuentra la grabación de entrevistas como único medio de garantizar tanto la fiabilidad y consistencia de evaluadores como el derecho de reclamaciones y revisión de calificaciones de todo candidato.

A continuación explicamos los principales elementos de la formación de evaluadores para certificación oral de lenguas. Comenzamos por los más evidentes y comunes a la formación de evaluadores e incluimos otros derivados del papel del entrevistador durante la entrevista, específicos de la evaluación oral.

⁸² Las ambigüedades de interpretación de las escalas, así como otros problemas surgidos al poner en práctica las escalas del MCER para evaluación oral se exponen en §3.1 y §4.2.

Los principales elementos sobre los que hay que formar a los evaluadores son para evitar tanto la inexactitud en la puntuación como la variedad en la interpretación de los datos. Por ello la formación debe ir encaminada a disminuir la incertidumbre respecto a esos elementos, lo que supone familiarizarse con los sistemas de puntuación, conocer y comprender las escalas que se van a usar y los principios que subyacen a ellas, interpretar sus descriptores de forma coherente y aprender a aplicar el sistema. Y esta formación debe ser periódica y sistemática, incluso para los viejos evaluadores o los ya formados. De nada servirá tener diseñado un buen test si los evaluadores no saben utilizarlo consistentemente⁸³.

Pero además, en las pruebas de expresión oral “la persona que dialoga con el candidato debe dominar las técnicas que ayudarán a cada candidato a sentirse cómodo, mientras presta atención a la vez a detalles como el tiempo y la forma de hacer las preguntas, para garantizar que todos los candidatos tengan las mismas oportunidades de mostrar sus habilidades” (Alderson, 1995: 115). Así, los equipos de evaluación oral deben estar formados por entrevistadores y calificadores que sean capaces de poner el test en práctica, interpretar bien sus tareas, perseguir el tipo de lenguaje y características requeridos y calificar con los mismos criterios y justicia a todos los candidatos. El proceso no es fácil de llevar a la práctica: hay que hacer que todos los evaluadores compartan el mismo concepto de lengua⁸⁴, los mismos criterios y objetivos de evaluación, y realicen el mismo tipo de análisis del discurso.

Un aspecto muy importante en estos equipos de evaluadores orales es el del conocimiento y uso de técnicas de elicitación del discurso, confiado habitualmente al albedrío o a la confección de una lista de temas y preguntas para el test. Esto no es suficiente. La formación en las técnicas de elicitación debe incluir elementos importantes e influyentes como la necesidad de ser conscientes de cómo la génesis del discurso de los candidatos puede verse influida por la actuación del entrevistador o de cómo el manejo de las preguntas y temas influye en los resultados obtenidos. Hay que conocer las causas y objetivos de la selección de temas y tareas del test y su uso debe producirse en el momento adecuado de la entrevista, en la fase adecuada en la que no se

⁸³ A la validación clásica que se da en todo test para averiguar si efectivamente mide lo que se espera, hay que añadir en los tests orales un nuevo tipo de validación, la validación de los evaluadores o su fiabilidad interna y externa.

⁸⁴ El concepto de lengua se materializa en el constructo a medir mediante el test. Los evaluadores deben comprenderlo y compartirlo, evitando que en sus calificaciones o técnicas de elicitación sea otro modelo el que influya y determine el resultado. Llegar a compartir el constructo no es fácil; supone dejar de lado las consideraciones que cada evaluador o profesor suele tener en su sistema de trabajo y que quizás no se consideren en el examen.

vaya a malinterpretar o a crear estrés. Igualmente es importante ser consciente de qué está pasando en cada momento de la entrevista y qué está haciendo el candidato en ese momento con su discurso o respuestas, como hemos señalado en §2.4.

Todo ello exige hacer ensayos previos con ejemplos y candidatos de todo tipo, entrevistando a voluntarios y corrigiendo muchas pruebas consideradas normales junto a otras raras, para prever todas las variantes y saber qué hacer ante imprevistos. Y, lógicamente, en pruebas orales estos requisitos conllevan que se hagan grabadas⁸⁵.

El proceso supone trabajar en grupos de evaluadores, escuchar y ver ejemplos de entrevistas grabados; dar una puntuación independiente; discutir las notas con el resto de evaluadores hasta alcanzar un consenso; y analizar los casos difíciles y raros. A estos elementos externos que se aplican una vez realizadas las entrevistas, se deberían sumar otros internos, que se aplican durante la realización de las entrevistas, como el aprender a llevar la conversación; a reaccionar a las respuestas de los candidatos; a entender las respuestas y analizarlas sobre la marcha, adecuando la siguiente pregunta al nivel requerido por la estructura del test en relación con la respuesta dada por el candidato; y a evitar el estrés que puede dificultar el discurso del candidato⁸⁶.

El examinador se convierte en el principal elemento de este proceso y la necesidad de hacer el trabajo de forma adecuada se conjuga con la de que todos los evaluadores lo hagan de la misma forma, logrando los mismos resultados. Esto es, siendo consistentes consigo mismos y con el resto. Bachman explica que las causas de inconsistencia en un mismo examinador pueden deberse a factores como cansancio u otros. Por ejemplo, dice, “si corrige un test después de varios de muy poco nivel puede tender a dar mejores notas que si hubiera corregido ese mismo test después de otros test de muy buen nivel. O puede que califique de diferente forma los últimos tests de un gran grupo que los primeros” (Bachman, 2004: 169). Igualmente, explica que la inconsistencia entre diferentes evaluadores se debe a la diferente forma en que califican los mismos tipos de respuestas. Por ejemplo, en un test oral, diferentes evaluadores pueden valorar

⁸⁵ La inicial resistencia de evaluadores y candidatos a utilizar cintas en las entrevistas es fácilmente reducible si se utilizan grabadoras digitales que, por su formato y sistema de uso, son poco molestas. La grabación, junto al uso adecuado de las técnicas de elicitación del discurso, ayuda a evitar el estrés comunicativo.

⁸⁶ La prueba del nueve de la formación de evaluadores está en entrevistar a nativos y calificar esas entrevistas hechas a nativos, que no van a cometer errores gramaticales. Ya que esto obliga a utilizar los criterios evitando que la gramática, o el hecho de que el discurso “parezca” el de un nativo, sin más explicaciones, sea el criterio fundamental de corrección.

diferentes características del habla del candidato y dar a su actuación diferentes calificaciones (Bachman, 2004: 169).

Para asegurarse de que todos los evaluadores van a realizar la tarea igual y consistentemente se utilizan técnicas de análisis estadísticos de correlación⁸⁷ con sus resultados, tanto de uno mismo a lo largo del tiempo, como de uno con relación al resto de evaluadores⁸⁸. Estos deben conjugarse con otra técnica sencilla como es hallar las medias de cada evaluador. Esto es porque no es suficiente con establecer que todos los evaluadores ordenen de mejor a peor a los mismos candidatos, sino que es preciso asegurarse de que les asignen el mismo nivel.

En §2.6 se explican los protocolos de una OPI para realizarse con objetividad y consistencia, así como estudios sobre su fiabilidad y validez. También se explican sus técnicas de elicitación y su tipología de preguntas (§2.6.2- 4). En la OPI, ya lo hemos dicho, la formación es tan fundamental y los protocolos son tales, que se considera como una posibilidad el hecho de que una entrevista se pueda considerar como no válida por defectos de forma en el entrevistador o por defectos de elicitación.

En §2.7 se explica el sistema del MCER y se hace referencia a lo que considera ese documento como importante para realizar una evaluación fiable conforme a sus postulados.

Antes de terminar queremos señalar una cosa importante y siempre olvidada respecto a la evaluación y al diseño de tests y de talleres o cursos de formación. No se debe olvidar que una evaluación es siempre algo limitado, y especialmente limitado a lo que se haya demostrado que el test puede medir. Por ello, es importante considerar en la formación de evaluadores el hecho de que un test tiene limitaciones. Hay que conocer las del test en cuestión y aceptarlas, para evitar extraer conclusiones falsas o tener expectativas erróneas o darle un valor predictivo inadecuado.

⁸⁷ La explicación de las correlaciones aplicada a la evaluación se hace en §2.3.4.

⁸⁸ En §4.5 y en el Apéndice 10, se dan estos datos referentes a los evaluadores utilizados en nuestra investigación con el examen del MCER.

2.5 Sistemas de definir la capacidad lingüística: escalas

*La palabra, la mitad es de quien la dice
y la otra mitad de quien la escucha
(Dicho popular).*

2.5.1 Tipos de escalas y sus características técnicas

La forma en que se definen las capacidades lingüísticas de un individuo y la forma en que se explican los resultados de su evaluación dependen de las escalas utilizadas. Una vez recogidos los datos mediante un instrumento o prueba de evaluación, la interpretación de sus resultados depende del patrón utilizado (Ortega, 1996: 28). Los mismos datos, interpretados con diferentes escalas, ofrecerán distintos resultados. Lo que es fácil de ver si, por ejemplo, interpretamos un discurso oral no planificado, mediante distintas escalas, gramatical, comunicativa, de dominio de léxico o de fluidez: en cada caso llegaremos a unas conclusiones, dependiendo de lo que queramos encontrar y comentar.

Sin duda, hay algo arbitrario en cualquiera de las divisiones que pudieran seleccionarse⁸⁹. La decisión depende de la teoría lingüística con que decidamos examinar o enfocar la lengua⁹⁰. Así, y según sea el concepto que se maneje, dos

⁸⁹ Este problema, y el escepticismo contra el que había que luchar para determinar niveles, también se lo plantearon en el Consejo de Europa los miembros del equipo dedicado a establecer lo que finalmente se llamó *Threshold Level*, como se explica en Trim (2007: 19).

⁹⁰ A comienzos de los 70 se discutió sobre la idoneidad de los tests existentes y la necesidad de nuevas escalas basadas en criterios de referencia. Junto a ese cambio hubo otros sustanciales en los contenidos de las escalas, apareciendo los modelos holísticos (Bordón, 2006: 210).

La historia de estos cambios no ha sido sencilla porque no es fácil compartimentar el continuum de una lengua en rasgos objetivos que sean aceptados por los profesionales. Según Bachman (1990: 6) las primeras investigaciones comenzaron en los años 40. B. J. Carroll, con sus estudios sobre validación de constructos, fue uno de los pioneros. Estos estudios partieron de trabajos de corte sicologista que intentaron establecer los parámetros en que se podía dividir la actuación lingüística atendiendo a las habilidades verbales y analizando las variables con análisis estadísticos y procedimientos analíticos. Pero había cierto escepticismo al respecto. Lantoff / Frowley (1988: 185) dudaban de las posibilidades de encapsular el comportamiento humano lingüístico, arguyendo que no había acuerdo entre los sicometristas sobre el número de niveles en los que dividir la competencia. Según Spolsky (2000: 542), uno de los primeros intentos de establecer una escala de habilidad oral fue es el de Sammartino, en 1938, que creó una escala para evaluar oralmente la lengua con 7 niveles y se debe reconocer a su escala como un claro precedente de las que fueron utilizadas hasta la aparición de los protocolos de ACTFL (1982, 1986).

candidatos que parecen poseer el mismo dominio lingüístico para un evaluador, podrían ser distintos para otro, calificándolos como de diferente nivel al utilizar otro sistema.

El objeto de este apartado es poder entender mejor las diferencias entre las escalas de los dos sistemas que vamos a comparar, las del MCER y la de ACTFL. Para ello vamos a explicar algunos de sus aspectos técnicos y sus características, y con ello nos acercaremos a la comprensión de sus posibilidades como sistemas de evaluación oral⁹¹. Ambos sistemas son diferentes en su forma de describir la lengua, en su manera de compartimentarla, en sus niveles y en su uso. Por lo que es posible que den resultados distintos y aporten datos diferentes respecto a las habilidades de los candidatos⁹². Para entenderlo necesitaremos explicar, si bien someramente, algunos de los principales conceptos técnicos sobre el diseño de escalas.

Las escalas que hasta ahora se han desarrollado para medir los tests las podemos explicar atendiendo a diferentes puntos de vista:

1. - Atendiendo a su contenido, hay escalas holísticas y escalas basadas en puntos discretos o analíticas. Ambos sistemas no suelen establecer intervalos equivalentes.

Los criterios analíticos discriminan entre atributos o rasgos de la lengua, de manera que se establecen niveles de actuación para cada componente y se distinguen planos como coherencia, cohesión, pronunciación, velocidad, fluidez, etc. Las escalas del MCER⁹³ (cooperación, intercambio de información, etc.) son de este tipo y se centran en cinco criterios (fluidez, interacción, coherencia, corrección y rango). Al usarlas, los evaluadores son libres de decidir si el candidato debe cumplir todo lo de la escala o solo una parte de sus descripciones funcionales.

Los otros criterios, los holísticos, definen estadios de actuación en los que el nivel se establece de manera general y no se puntúa separadamente. Suelen usarse en ámbitos laborales, diplomáticos, políticos, sociales o militares, donde se necesita averiguar el

⁹¹ Este apartado sirve, además, como introducción a ambos sistemas de evaluar, de los que se trata detalladamente en los siguientes apartados, §2.6 para el sistema de ACTFL y §2.7 para el del MCER.

⁹² Efectivamente, como se ve en el capítulo 4, el uso de una escala u otra ha dado resultados diferentes con los mismos candidatos. Estos podrían haber sido aún más diferentes si no hubiéramos tomado las precauciones para hacer dos tests similares que pudieran ser calificados con parecidos criterios (véase §3.5), o si se hubieran seleccionado otras escalas para la calificación del discurso. Nosotros hemos realizado la investigación acercando lo más posible los dos sistemas y hemos obtenido unos resultados objetivos (véase §4.5). Pero el MCER tiene muchas otras posibilidades y haber elegido cualquier otra, a pesar de que tiene una intención de servir como sistema de homologación fiable, habría dado otros resultados.

⁹³ La explicación detallada del MCER y sus posibilidades en la evaluación se tratan en el § 2.7.

uso general de la lengua y no un aspecto concreto de ella. La escala de ACTFL⁹⁴ es, en este sentido, holística.

2. - Respecto a la forma de agrupar los resultados, hay escalas basadas en un ranking o norma, y escalas basadas en un criterio.

Las notas de un test referido a una norma se interpretan con referencia al resto de resultados de los compañeros o con referencia a un grupo previamente dado. Esas eran las escalas usadas, principalmente, hasta que aparecieron los enfoques comunicativos. Ahora, en cambio, se tiende a pensar que también el peor del grupo podría cumplir y superar los objetivos del curso; por lo tanto, se tiende a aprobar a quienes cumplan con unos criterios de habilidad externos al grupo, descritos previamente en unos criterios.

Las escalas referidas a un criterio permiten interpretar la actuación de un candidato con relación a un nivel de habilidad definido en el criterio. La comparación entre los candidatos no tiene sentido ni hay forma de hacerla, ya que todos se ajustan al criterio o no (Bachman 1990: 211), y no es posible dispersarlos en una gráfica de mejor a peor⁹⁵. Tanto las del MCER como la de ACTFL son escalas referidas a un criterio⁹⁶.

3. Y dentro de las escalas de referencia a un criterio se pueden considerar diferencias basadas en la forma de interpretar sus resultados, si estos se hacen de forma ascendente (a.) o descendente (b.), o si se hacen con referencia a la maestría (c.) o a un continuum (d.); y, en la forma de organizar su dominio de lengua, si esta se hace de forma organizada (e.) o sin organizar (f.).

Explicamos estos seis conceptos someramente.

(a. / b.): Escalas ascendentes o descendentes

La teoría y la práctica han presentado las escalas en forma ascendente⁹⁷, organizando los niveles desde el más bajo al más alto, como una consecuencia de las

⁹⁴ La explicación detallada del sistema de evaluar de ACTFL se trata en el siguiente apartado, § 2.6.

⁹⁵ No obstante, el uso de criterios asegura que los resultados puedan ser homologables y comparables con los de cualquier otro candidato que haya realizado un examen similar aunque con tareas, contenidos o materiales diferentes (García Ramos, 1989: 45).

⁹⁶ El tema de si la escala de ACTFL era o no referida a criterios ha sido controvertido. De ello se trata en §2.6.3.

⁹⁷ Ver otra visión de este aspecto de la interpretación ascendente o descendente de las escalas en el apartado siguiente y en §2.6.3- C al explicar las críticas a la escala de AEI y la OPI, que son descendentes en vez de ascendentes.

concepciones pedagógicas de la enseñanza de lenguas y de los estudios sobre adquisición del lenguaje⁹⁸. La escala de ACTFL se utiliza de forma descendente, de arriba a abajo⁹⁹. El MCER, en cambio, no se decanta por ninguna manera de interpretar sus escalas. Su apartado 3.7 dice que “algunas personas prefieren leer una escala de descriptores desde los niveles inferiores a los superiores; otras personas prefieren hacer lo contrario” (Consejo de Europa, 2002: 39)¹⁰⁰.

(c. / d.): Con referencia a la maestría o a un continuum

La determinación de los puntos de corte de cada nivel puede hacerse bien con referencia a la maestría, bien con referencia a un continuum. El primero consiste en establecer un nivel mínimo de competencia que se debe satisfacer mínima pero obligatoriamente. Con ello se diferencia a los candidatos que lo han alcanzado de los que no. Este sistema no supone ningún tipo de gradación de calidad respecto a los logros. Es el que utiliza ACTFL entre sus macro niveles inicial, intermedio, avanzado o superior¹⁰¹.

El segundo, con referencia a un continuum, sitúa las capacidades en una línea que recoge todos los grados de menor a mayor o de peor a mejor y puede dispersar las actuaciones a lo largo de toda la línea. El MCER está abierto a utilizarse con el enfoque de maestría o con el de continuum (Consejo de Europa, 2001: 185; Bordón, 2006: 48), para ello el enfoque de continuum se puede equiparar con los niveles comunes de referencia (capítulo 3 del MCER), y el enfoque de maestría se puede describir en el cuadro conceptual de las categorías (capítulo 5: 52-53) y los niveles ofrecidos por el MCER (capítulos 5 y 6)¹⁰².

En esta variedad y apertura, el MCER se diferencia de la escala de la OPI radicalmente, ya que la OPI marca los niveles por referencia al nivel de maestría y evita

⁹⁸ Ello supone una relación con estas teorías sobre el aprendizaje. Sin embargo, esta relación no es directa porque estas teorías no se dedican a construir escalas y sus resultados no son suficientemente claros como para describir niveles detalladamente (Luoma, 2004: 50). En realidad, las escalas reflejan más bien lo que sus creadores entienden que es un adecuado o deficiente funcionamiento de la lengua (Leung / Lewkowicz, 2000: 220).

⁹⁹ Véase §2.6.3- D.

¹⁰⁰ Si hacemos caso a lo que dice Lowe (véase §2.6.3- D), esta libertad en la forma de usar las escalas del MCER puede tener consecuencias a la hora de evaluar.

¹⁰¹ Para una explicación de sus niveles véase §2.6.2- 1, y para una explicación de sus sistemas de calificaciones, §2.6.2- 5.

¹⁰² Incluso, en su sistema de diseño y escalonamiento, el MCER pretendió dar un sentido de progresión que fuera útil y de sencilla interpretación para los profesores, uno de los principales destinatarios del documento (North, 2002: 97).

expresamente seguir el sistema de continuum para dar una adecuada calificación. Aunque la OPI también recoge este aspecto de continuum en subniveles una vez que nos situamos dentro de cada macronivel (bajo, medio o alto), si bien ya dentro de cada uno de ellos tampoco se puede graduar el discurso de los candidatos. Todos ellos se sitúan en el mismo punto de la línea dentro de un nivel dado, por ejemplo, intermedio medio. El MCER, en cambio, permite hacer constantes subdivisiones en virtud de las necesidades pedagógicas o de evaluación que se requieran.

(e. / f.): Con un dominio organizado de lengua, o sin organizar

Cuando una escala organiza su contenido con criterios donde determinados elementos tienen más valor que otros o son de dificultad mayor, esta es de dominio organizado. Por ejemplo, los tests basados en las programaciones dadas en los cursos de lengua de acuerdo con criterios como dificultad, posibilidad de aprenderse, importancia comunicativa, etc. O las escalas que describen diferentes niveles de habilidad de los candidatos en competencia comunicativa general. Así son tanto la de ACTFL que estamos utilizando en nuestros análisis (Bachman, 2004: 301), como las del MCER.

Cuando no se pueden establecer diferencias en lo que se quiere medir, por ejemplo, si queremos evaluar el dominio del vocabulario que ha alcanzado un grupo de estudiantes, pues el léxico cuyo dominio se quiere comprobar no tiene una organización, ya que todos los ítems para medirlo son igual de importantes, la escala es de dominio sin organizar.

4. Otros elementos del diseño de escalas que hay que considerar son el número de niveles en que se dividen, la cantidad de criterios que se utilizan en cada nivel, la manera de presentar y escribir los descriptores y, por último, los propósitos para los que se diseñan.

Es central ver cuántos niveles describir, ya que esto es un elemento conceptual de importancia práctica: cuantos más niveles se diferencien, más claridad habrá en la descripción del progreso de las habilidades lingüísticas. Pero, al mismo tiempo, será más difícil su uso, ya que los calificadores deben poder entender los niveles y su

sistema de diferenciación y ponerlos en práctica coherentemente y esto se hace más complicado con muchos niveles¹⁰³.

Si al realizar una escala para un test se decide utilizar criterios analíticos, los diseñadores tienen que decidir cuántos criterios utilizar, sin que suponga, como en el caso anterior, un problema de uso conceptual y práctico. El MCER sugiere que este número sea variable, “la experiencia demuestra que más de cuatro o cinco categorías comienzan a provocar una sobrecarga cognitiva, y que siete categorías es psicológicamente un límite máximo” (Consejo de Europa, 2001: 193).

Otro elemento a tener en cuenta, importante para la interpretación de los descriptores, es la forma de escribirlos, que debe ser mediante frases concretas, prácticas, no muy largas, inteligibles por sí solas y no muy detalladas, ya que de lo contrario sería difícil encontrar ejemplos de discursos que contuvieran esos elementos. Las directrices del MCER aconsejan que los descriptores tengan, a ser posible, una formulación positiva del contenido; sean breves; sin jergas; describiendo tareas y sus grados de destreza mediante una formulación concreta y no solo estableciendo que se debe poder hacer lo de los niveles precedentes pero de mejor manera; con claridad y transparencia; estableciendo lo realmente característico de la actuación de los individuos; e independientes (Consejo de Europa, 2001: 198)¹⁰⁴. Las descripciones de la escala de ACTFL se tratan en 2.6.3- D, donde se señalan sus criterios de parquedad en las definiciones, y la necesidad de basarse en hechos y experiencias en vez de en teorías lingüísticas, entre otros (Lowe, 1986, 1998).

Para terminar estas explicaciones técnicas tenemos que prestar atención a la forma en que las escalas se muestran a los usuarios. Las escalas usadas en un test no tienen por qué explicarse de la misma manera en todo momento, pues sus características pueden variar de acuerdo con las necesidades de los usuarios. Así, y según los propósitos de sus diferentes usuarios, diseñadores, candidatos, administradores, agentes laborales, etc., pueden variar en la forma en que se presentan y explican. Las diferencias entre las

¹⁰³ Sean los que sean, en los talleres de formación de evaluadores hay que analizar sus contenidos y formas de diferenciar la actuación de los candidatos de acuerdo a esos criterios, y analizar cuantitativamente la forma y consistencia de aplicar los descriptores a un mismo modelo de discurso, tanto para cada evaluador consigo mismo, como para todo el grupo de evaluadores. El tema de la formación de evaluadores se ha presentado en §2.4.3- 4.

¹⁰⁴ Véase un análisis crítico sobre sus descriptores y su ambigüedad en §3.1.1.

versiones residen en la cantidad de información, detalle, ejemplos o terminología usada (Luoma, 2004: 60).

La forma de presentar las escalas puede ser más técnica, para los examinadores; con más especificaciones, para los entrevistadores; o más genéricas, para los candidatos. El contenido y la forma de escritura de estas escalas es, en lógica consecuencia, diferente: las que se hacen para los calificadores sirven para ayudarles a tomar decisiones consistentes; las redactadas para los examinadores les ayudan a entender y establecer niveles y definir sus características de habilidad en términos de elementos débiles o fuertes; las diseñadas para los administradores se centran en que estos conozcan toda la información necesaria de forma concisa.

El MCER explica que, entre sus escalas, unas se destinan especialmente a los candidatos o estudiantes, por ejemplo, para su autoevaluación, otras a los administradores o examinadores, y otras a los responsables de elaborarlas (Consejo de Europa, 2001: 40).

Concluyendo, estas son las principales diferencias entre ambos modelos de escalas: el MCER y ACTFL utilizan escalas de habilidades orales basadas en criterios de referencia, que son las más útiles para medir la actuación oral y escrita. Sin embargo, mientras la escala de ACTFL se usa de arriba a abajo, con referencia a la maestría, de forma holística y mostrando el dominio de lengua organizado, las del MCER se pueden utilizar de arriba abajo o de abajo arriba, de forma analítica u holística (según la escala utilizada), mostrando el dominio de lengua dependiendo de las escalas utilizadas (organizado o no) y con referencia a un continuum.

Estas diferencias conllevan distintas formas de interpretar los niveles y de ofrecer los resultados, como se verá en el capítulo 4 y en las conclusiones en el capítulo 5.

2.5.2 La escala de evaluación oral de ACTFL

De todo lo apuntado en el apartado anterior, se concluye que la escala utilizada en la OPI de ACTFL tiene estas características:

ACTFL tiene una sola escala para todos los casos, examinadores y usuarios (§2.6.2), aunque también tiene unas listas de preguntas de decisión binaria (§2.6.2- 5), para

contestar sí o no, para averiguar si el candidato cumple con ciertos objetivos de cada nivel, que se adjuntan para los calificadores en el manual de formación y ayudan en el proceso de calificación.

La escala es holística.

Define 10 niveles en 4 grandes bloques: inicial, intermedio, avanzado y superior; y su sistema de calificación tiene, entre sus características, tres importantes y diferentes de otras escalas:

A) No compensa elementos entre niveles.

B) La evaluación se hace “de arriba abajo”.

C) Se basa en puntos de maestría y no en un continuum de desarrollo de la habilidad.

Esto es porque se establece para medición de la habilidad sin tener en cuenta los procesos de aprendizaje y estableciendo los principales elementos que se deben cumplir en cada nivel, con independencia de los establecidos en el anterior, al que no se hace referencia sino para subsumirlos¹⁰⁵.

2.5.3 Las escalas del MCER

El MCER presenta 60 escalas¹⁰⁶ que deben ser seleccionadas para realizar los exámenes que se necesiten, combinadas según se estime adecuado para las competencias que se desee evaluar, redefinidas, resumidas o incrementadas para servir a los propósitos de cada situación e interpretadas indiferentemente de abajo arriba o de arriba abajo.

Cada una se divide en 6 niveles, aunque, como North apunta (2002: 96), la distancia entre el nivel A1 y el A2 y entre el nivel C1 y el C2, es decir, entre los niveles bajos y entre los niveles altos, es la mitad de la que representan el resto de niveles. Pues los niveles medios A2, B1 y B2 se escalonaron de forma que cada uno contenía subdescriptores de nivel (A2 y A2+, B1 y B1+, B2 y B2+) aunque finalmente no se trasladaron en forma de diferenciación de niveles a la escala final, ya que esta habría

¹⁰⁵ Se aclaran estos elementos en §2.6.2.

¹⁰⁶ Volvemos a decir que el sistema del MCER se explica detalladamente en el §2.7.

resultado ser una escala de 9 niveles que se estimó de difícil aceptación para las tradiciones de los sistemas educativos de los diferentes países europeos. Por ello, se dejó con seis niveles. Sus 9 niveles, no obstante, se han mantenido en algunas escalas en las que se diferencian A1, A2, A2+, B1, B1+, B2, B2+, C1 y C2 (North, 2002: 96)¹⁰⁷.

Resumiendo, diremos que el MCER define dos grandes áreas, la de las competencias y la de las habilidades o actividades, tres competencias (comprensión, expresión e interacción), y cinco criterios (alcance, corrección, fluidez, interacción y coherencia) y que, al contrario que la escala de ACTFL:

A) Puede compensar elementos entre niveles, así como estimar diferentes desarrollos de competencia y habilidad de forma horizontal, muy útiles pedagógicamente, analizando el aprendizaje y la progresión de diferentes características de la lengua a lo largo del tiempo, aunque todo el tiempo dentro del mismo nivel de dominio, reconociendo perfiles parciales y competencias desiguales (Consejo de Europa, 2002: 17)

B) Se puede interpretar de arriba abajo o al contrario.

C) Se puede usar tanto para dar resultados como una escala con puntos de maestría o de continuum.

En este trabajo pretendemos averiguar cómo la evaluación oral realizada mediante ellas puede homologarse con la hecha mediante las *guidelines* orales de ACTFL. En §3.2 procedemos a averiguar las posibilidades de equiparación entre los contenidos que desarrollan y describen, lo que sirve de base para establecer la pregunta de investigación de esta tesis, expuesta en el capítulo 1.

2.6 La entrevista oral de ACTFL

La entrevista llamada *Oral Proficiency Interview* (OPI) desarrollada por el *American Council on the Teaching of Foreign Languages* (ACTFL) en 1982 y 1986 y

¹⁰⁷ Estas diferencias no se han tenido en cuenta en nuestro sistema de calificaciones ni han aparecido en los resultados finales de las equiparaciones entre los sistemas del MCER y ACTFL. Véanse capítulos 4, de análisis de resultados, y 5, de conclusiones.

revisada en 1999 (Breiner-Sanders, 2000), es hoy día una de las más ampliamente usadas (Bordón, 2006: 210), comentadas y requeridas para usos académicos o sociales, fundamentalmente en Estados Unidos (búsqueda de trabajo, organización de clases, traslado a un país donde se habla otra lengua, diagnóstico o clasificación de estudiantes, investigaciones lingüísticas, etc.).

Incluso los críticos hablan de la OPI como el mejor sistema de evaluación oral que tenemos (Salaberry, 2000: 292) o proponen modelos similares con algunos cambios que intentan mejorarlo (Bachman / Savignon, 1983: 388-89, Bachman / Palmer, 1996: 298-307). La OPI goza hoy día de buena salud y no sólo se ha extendido entre los profesionales de Estados Unidos como uno de los sistemas de evaluación oral más fiables, sino que ha contribuido enormemente a entender la comunicación como el objetivo principal del aprendizaje de una lengua, haciendo que en muchos casos el término *proficiency* -generalmente sinónimo de otros como *competencia comunicativa* o *habilidad de uso de la lengua*- se emplee fundamentalmente para referirse a este tipo de entrevista (Bachman, 1990: 16 n. 2). “La noción de la competencia comunicativa propagada mediante los protocolos de ACTFL con la financiación de las agencias americanas, ha arraigado en gran parte de los profesionales de enseñanza de lenguas en las actividades de enseñanza, evaluación e investigación” (Liskin-Gasparro, 2003: 488).

En este capítulo vamos a explicar la OPI en detalle, ya que el trabajo de campo realizado intenta comparar los resultados de estudiantes evaluados mediante la OPI y mediante un nuevo test diseñado por nosotros de acuerdo con algunas escalas del MCER. Tras una introducción (§2.6.1- A) pasamos a explicar (§2.6.2- B) qué es la OPI, en qué medida posee un constructo teórico y práctico y unos sistemas de elicitación y calificación de los datos adecuados, qué niveles describe, cuáles son los procedimientos de realización, y qué otros aspectos importantes presenta.

Y al final (§2.6.3) se explican las principales críticas que han intentado poner de manifiesto -no siempre con éxito, conocimiento ni rigor crítico suficiente- sus limitaciones. Ello nos permitirá explicar mejor algunos conceptos de la OPI y ayudará a entender su alcance real y a evitar excesos de confianza o frustración, origen de no pocas críticas fundadas en malentendidos y desconocimiento.

2.6.1 Introducción

Comencemos por definir la OPI en palabras del manual de la propia organización americana que la diseñó¹⁰⁸:

La OPI de ACTFL es un procedimiento estándar para la calificación global de la habilidad funcional oral. Estandarizado, porque para garantizar la fiabilidad de las valoraciones de los ejemplos de lengua debe seguirse un procedimiento prefijado; y global, porque mide la producción lingüística de forma holística, delimitando muestras de fortaleza y debilidad (en el uso de la lengua), que fundamenten el nivel de consistencia de la habilidad funcional, así como las limitaciones de esa habilidad. No se miden puntos discretos de uso de la lengua o del conocimiento de la lengua. Hay cuatro tipos de criterios de calificación: las tareas globales o funciones que se realizan con la lengua; los contextos sociales y áreas de contenido en que se usa la lengua; las características de corrección que determinan cómo realiza el hablante determinadas tareas para cada contexto y área de contenido; y el tipo de texto oral utilizado, que puede variar desde palabras sueltas a discurso extendido (Swender, 1989:1-1).

Debe añadirse que en la valoración de las habilidades hecha mediante este test es irrelevante cuándo, cómo, dónde o de qué manera el candidato haya aprendido la lengua, ya que la OPI no es un test de aprovechamiento ni evalúa el grado de aprendizaje de la lengua con relación a cursos o contenido curricular, sino que se limita a valorar el rendimiento de la lengua “en términos de habilidad de uso efectivo y apropiado en situaciones de vida real” (Swender, 1989: 1-1).

Este test oral de ACTFL -como en este apartado veremos- no establece una lista de preguntas, sino que es un test abierto con el que se busca que el candidato comience hablando de las cosas que le interesan (por lo menos en la elicitación de discurso en los niveles bajo y medio). Pero esto no debe engañarnos ni llevarnos a pensar que se trata de una conversación natural, ya que el proceso está dirigido por el entrevistador para conseguir siempre una muestra de la mejor lengua posible del candidato, medida en términos de su capacidad de expresión, por un lado, y de los límites de sus posibilidades, por otro. Como el manual del evaluador dice “la OPI tiene que parecerse lo más posible a una conversación natural” (Swender, 1989: 1-2), lo que claramente implica que ni lo es -aunque debe parecerlo- ni lo puede ser en todo momento. La OPI es, en resumen, una entrevista cara a cara (o de forma telefónica) entre dos personas, un

¹⁰⁸ Puede verse también su página web www.actfl.org

candidato y un evaluador, este último entrenado y autorizado para elicitación los ejemplos necesarios de la habilidad funcional oral del candidato, calificarlos y evaluarlos.

La OPI se desarrolló a partir de otra entrevista anterior, la *Interagency Language Roundtable's (ILR) Foreign Service Institute Oral Proficiency Interview*, usada en los años 50 por el Departamento de Estado norteamericano y extendida a otros organismos durante los años 60 y 70¹⁰⁹. Esta poseía unos sistemas de elicitación cuidadosamente diseñados, la mejor garantía de que las notas eran fiables, significativas y útiles; y un procedimiento de elicitación con tareas, un conjunto de tipología de preguntas y temas y una escala de medición con la que calificar las muestras obtenidas mediante esos procedimientos (Bachman, 1990: 21). Como la entrevista era de uso gubernamental, ACTFL colaboró con el *Educational Testing Service (ETS)* a finales de los años 70 (junto a la ILR y el Cuerpo de Paz) para desarrollar una escala similar. Entonces se dio el primer paso para una evaluación independiente del gobierno. El *Educational Testing Service (ETS)* -cuyos miembros fueron entrenados por el FSI- asumió las actividades de evaluación de lenguas y formaron a voluntarios del Cuerpo de Paz, quien realizó la primera actividad de evaluación no gubernamental. A lo largo de los años 70, el sistema de evaluación oral se utilizó para la certificación de educación bilingüe (Liskin-Gasparro, 1987: 22).

La escala tuvo una redacción provisional en 1982 y ya ese mismo año se realizaron los primeros talleres de formación de evaluadores para profesores de español y francés. Debido al éxito obtenido y a su interés y demanda tanto institucional como académica, los talleres de formación proliferaron desde los 80 y 90 hasta nuestros días. El uso e influencia de este sistema de acercarse a la lengua desde una óptica de comunicación oral enfocando la enseñanza según la necesidad de usar la lengua y no solo de aprenderla, caló rápidamente y comenzaron a aparecer “libros de texto basados en las *guidelines* y currículos orientados en la *proficiencia* que están en uso en muchos institutos y escuelas y *colleges*” (Raffaldini, 1988: 197). En 1986 se editaron definitivamente las *guidelines* de ACTFL y nuevamente en 1999 se revisaron introduciendo algunas variantes necesarias tras años de análisis, estudios críticos y

¹⁰⁹ Spolsky menciona que la primera entrevista oral sobre interacción cara a cara fue en 1932 en la Universidad de Yale, para la evaluación de estudiantes de español. Pero que lo que se considera efectivamente el primer sistema de medida de la competencia lingüística es la evaluación realizada por el FSI entre 1952 y 1956, predecesora de la OPI de 1984 (Spolsky, 1995: 21).

recomendaciones de numerosos autores¹¹⁰, definiendo un nivel más -el avanzado medio- y reeditando las escalas.

Mediante esta nueva entrevista -en cualquiera de sus tres versiones, de 1982, 1986 o 1999- se evalúa la competencia comunicativa de los hablantes medida de forma holística “desde arriba hacia abajo” (Lowe, 1986: 396) empezando por tareas y funciones sencillas cuyo dominio hay que demostrar y ascendiendo hacia tareas de mayor dificultad hasta probar o descartar su uso.

Varios elementos conforman su constructo práctico:

- a. Los contextos de uso de la lengua y los temas asociados a ellos.
- b. El tipo de discurso empleado.
- c. Las funciones o tareas globales establecidas y definidas con un contenido de dificultad creciente.
- d. Los requerimientos de corrección referidos a la aceptabilidad, inteligibilidad, calidad y precisión del mensaje.

Define en total 10 niveles de actuación lingüística, establecidos en cuatro grandes macroniveles de habilidad (inicial, intermedio, avanzado y superior) cada uno de los cuales (excepto el superior) se subdividen en tres subniveles: bajo, medio y alto, lo que da la siguiente clasificación: inicial bajo, inicial medio, inicial alto, intermedio bajo, intermedio medio, intermedio alto, avanzado bajo, avanzado medio, avanzado alto y superior.

La escala de estos 10 niveles se diferencia de la original de la ILR en que aquella definía hasta 5 niveles, dos de los cuales quedan por encima de los que ACTFL considera útiles para medir la habilidad de los candidatos, por considerarlos no habituales entre los hablantes normales de una lengua (véase cuadro 2.6.1, comparativo entre ILR y ACTFL).

¹¹⁰ Para estas críticas, véase §2.6.3.

Cuadro 2.6.1. Relaciones entre las escalas de ACTFL y de la ILR

Escala de ACTFL	Escala de ILR
	5. Competencia nativa o bilingüe
	4+ Competencia destacada
	4. Competencia
	3+
Superior	3. Competencia para trabajo profesionalmente
Avanzado alto	2+
Avanzado	2. Competencia limitada para trabajar
Intermedio alto	1 +
Intermedio medio	1. Supervivencia
Intermedio bajo	1. Supervivencia
Inicial alto	0 +
Inicial medio	0. Sin competencia
Inicial bajo	0. Sin competencia

(Swender, 1989: 2-15).

Cada entrevista puede durar entre 10 y 30 minutos. Se trata de una conversación que debe parecerse lo más posible a una conversación natural; se graba y su contenido se precalifica mientras se realiza y se evalúa de forma definitiva tras una escucha posterior. En este proceso de calificar doblemente las entrevistas se puede producir tanto una confirmación como una modificación de la estimación inicial. Ambas cosas pueden ocurrir. Estudios realizados al respecto demuestran que los evaluadores mantienen la calificación en su segunda escucha, una vez realizada la entrevista, y que si cambian el resultado, suele ser para dar una calificación más baja (Thompson, 1995: 411); no obstante, esta modificación no puede llevar el resultado final a un macronivel diferente al apreciado durante el proceso de elicitación. Si tal cosa ocurriera la entrevista correría el riesgo de no ser válida¹¹¹.

La entrevista es diferente para cada persona y distinta incluso si se repitiera con el mismo candidato. El entrevistador realiza sus preguntas en relación con las respuestas obtenidas por el entrevistado respecto a sus intereses, formación, experiencias y opiniones mediante un sistema que se adapta de forma interactiva a sus respuestas y se ajusta al nivel de dificultad que cada candidato es capaz de desarrollar. Pero en su

¹¹¹ Lo veremos en el apartado donde se muestren los sistemas de elicitación y la estructura de la entrevista, §2.6.2- 3.

estructura y procedimiento es siempre igual, de forma que es un modelo estandarizado aunque se realice con preguntas y temas diferentes.

Los resultados se refieren a una escala basada en criterios de referencia y por lo tanto la actuación de los candidatos se compara con la definida en el constructo y no con la de ningún otro candidato. La entrevista establece un nivel mínimo en el que el candidato puede mantenerse y se pasa a establecer otro nivel en el que ya el candidato no pueda sostener la habilidad de uso de la lengua.

Para ser evaluador y realizar una OPI son necesarios una formación y un entrenamiento previos que da ACTFL a lo largo de numerosos talleres en Estados Unidos y que capacita a sus evaluadores para un periodo de solamente 4 años, transcurridos los cuales se debe proceder a un nuevo proceso de acreditación como evaluador. Así, se garantiza que la entrevista se realiza adecuadamente, de manera que su fiabilidad como prueba de evaluación oral de la lengua es grande¹¹².

2.6.2 Funcionamiento de la OPI

En lo que sigue vamos a tratar de explicar los siguientes aspectos de la entrevista:

1. Los principios de la escala de habilidades.
2. El constructo teórico de la lengua que se va a medir.
3. La estructura de la entrevista.
4. Los sistemas de elicitación de la lengua y el comportamiento del evaluador.
5. Las consideraciones necesarias para proceder a la calificación de las muestras obtenidas.

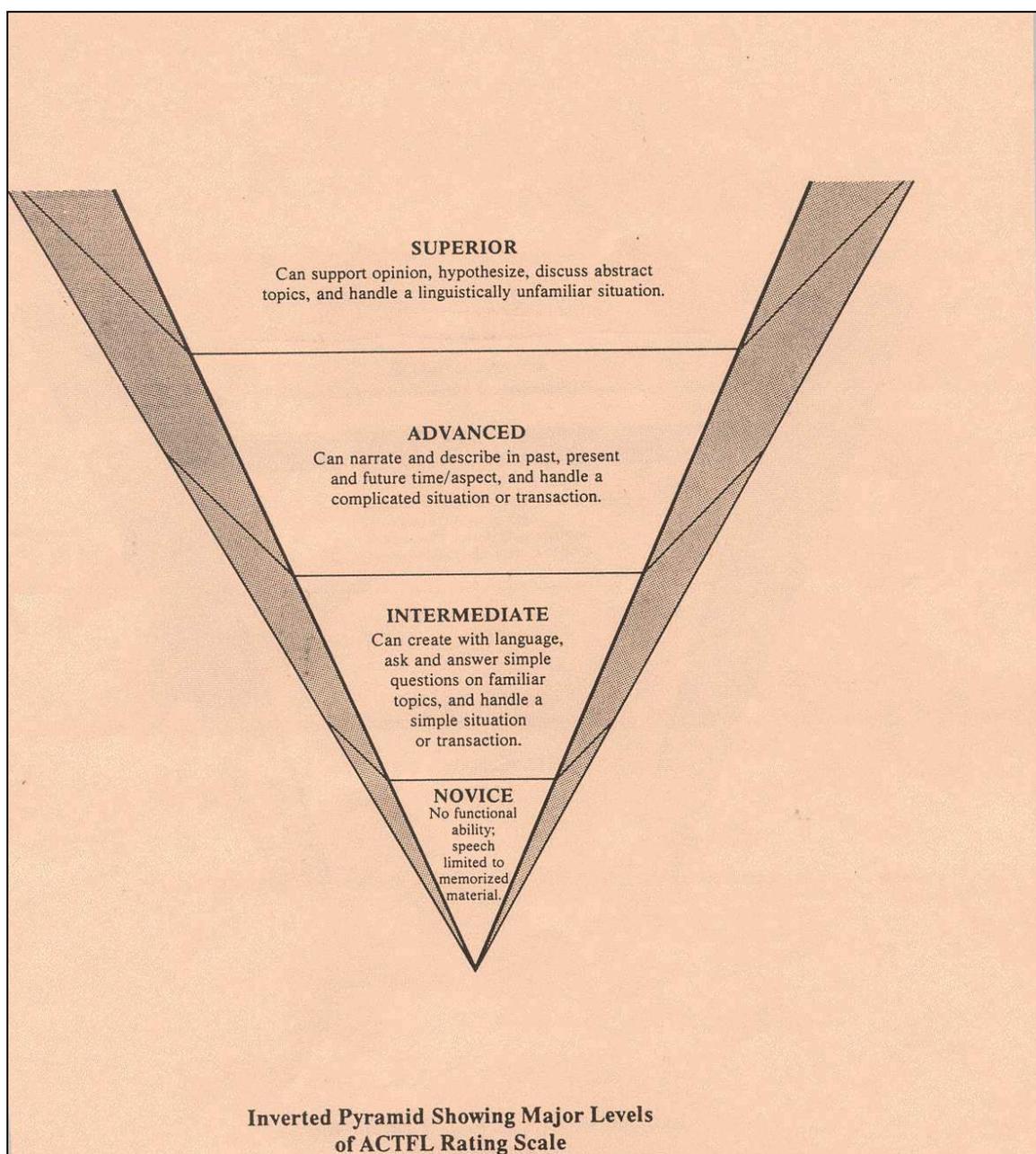
2.6.2- 1) Los principios de la escala de habilidades

La escala de habilidad de uso de la lengua definida por ACTFL pertenece a las escalas basadas en criterios y sirve para medir a los candidatos independientemente de

¹¹² Posteriormente, §2.6.3- D, estudiaremos estos aspectos de fiabilidad de las OPI que a veces han sido criticados.

cómo, dónde o cuándo hayan aprendido la lengua. Se basa en una jerarquía de perfiles de actuación que describe el comportamiento de los hablantes en cuatro macroniveles, desde los principiantes hasta quienes son capaces de intervenir efectivamente en la mayoría de conversaciones formales sobre temas prácticos, sociales, profesionales o abstractos. Tal jerarquía se fundamenta principalmente en una escala de tareas globales que se desarrollan a través de los cuatro grandes niveles (cuadro 2.6.2 a).

Cuadro 2.6.2 a. Macroniveles de la escala de ACTFL



(Swender, 1989: Apéndice, Tablas e ilustraciones. Ilustración 2 A).

Las siguientes consideraciones pueden ayudar a entender el sentido de esta escala:

a. Se asume que la evaluación de la lengua se realiza mejor desde el punto de vista holístico, es decir, de actuación general.

b. Se entiende que las habilidades lingüísticas aumentan con la complejidad de las tareas, de forma que se incrementan de forma exponencial en vez de consistir en una simple suma lineal de actividades. Por eso el dibujo del dominio de la lengua es una pirámide invertida en vez de un cuadro dividido en 4 niveles. En ella se aprecia la poca cantidad de lengua necesaria para desarrollar las tareas de los niveles bajos y cómo esta se incrementa considerablemente en cada subnivel.

c. Cada nivel contiene las características de los anteriores. Así, por ejemplo, un hablante de nivel avanzado bajo puede realizar todas las tareas del inicial y del intermedio.

d. Para acceder a un determinado nivel se deben dominar todas sus características. Por ejemplo, aunque un hablante pueda realizar ciertas funciones propias del avanzado - como narrar o describir- si no es capaz de realizarlas en sus contextos, tipo de texto y temática, no se califica como avanzado.

e. Los macroniveles permiten que diferentes candidatos hablen con estilos aparentemente muy diferentes que, no obstante, son clasificados dentro del mismo nivel, puesto que no se comparan las actuaciones de los hablantes entre sí sino con el cumplimiento de los criterios y características de la escala en cada nivel.

f. Junto a las tareas globales consideradas, se incluyen otros elementos para la evaluación¹¹³:

- Las circunstancias o contextos en que se desarrollan las actividades y las áreas de contenido usualmente asociadas a esos contextos.

¹¹³ Estos elementos se explican detenidamente en el siguiente punto, §2.6.2- 2, constructo teórico a medir.

- La corrección con que se realizan las tareas, tradicionalmente la gramática, el vocabulario y la pronunciación, pero también otros como la fluidez, la coherencia de los mensajes, la propiedad sociolingüística y aceptabilidad del discurso y estrategias de gestión y planificación discursivas.

- El tipo de texto utilizado en la realización de las tareas: desde palabras sueltas hasta sintagmas, frases sueltas o conectadas en párrafos y discurso extendido.

g. Además de definir cuatro grandes niveles, se definen tres umbrales dentro de estos -alto, medio y bajo- (cuadro 2.6.2 b).

Ya se ha dicho que si no se cumplen todos los criterios de un nivel no se puede estar calificado dentro de este. Así pues, si un candidato despliega una gran cantidad de las características de un nivel pero no las mantiene en todo momento, pertenece al umbral de ese nivel pero no a ese nivel que todavía queda por encima de sus posibilidades.

Por ejemplo, un candidato que demuestre un dominio significativo de ciertas características del avanzado pero no de forma suficiente, se mantiene en el nivel anterior (intermedio) dentro del subnivel intermedio alto (lo que viene a significar “casi avanzado”, cuadro 2.6.2 c). Ambos estilos de hablar, intermedio alto y avanzado bajo, podrían agruparse conjuntamente desde el punto de vista de la funcionalidad; pero desde el punto de vista del control y mantenimiento del resto de características del nivel son diferentes.

En estas fronteras grandes entre macroniveles (inicial alto, intermedio alto o avanzado alto), los hablantes pueden mostrar muchas diferencias entre sí, ya que podemos hallar tanto a candidatos que tengan un fuerte control de las funciones del nivel siguiente, pero grandes debilidades en el cumplimiento del resto de sus características como a otros candidatos que tengan una corrección grande, pero no puedan desenvolverse usando las funciones requeridas en el siguiente nivel.

Cuadro 2.6.2 b. Subniveles dentro de los macro niveles de la escala de ACTFL

SUPERIOR
AVANZADO ALTO
AVANZADO MEDIO
AVANZADO BAJO
INTERMEDIO ALTO
INTERMEDIO MEDIO
INTERMEDIO BAJO
INICIAL ALTO
INICIAL MEDIO
INICIAL BAJO

Cuadro 2.6.2 c. Frontera de diferenciación entre intermedio alto y avanzado

<p style="text-align: center;">Avanzado Cumple y mantiene en todo momento las características del avanzado.</p>
<p style="text-align: center;">Intermedio alto Muestra características del nivel avanzado pero no las mantiene en todo momento, aunque sí en más del 50% de las ocasiones requeridas. Realiza con total habilidad todas las actividades del intermedio y con un fuerte control de sus características.</p>
<p style="text-align: center;">Intermedio Cumple y mantiene las características del intermedio.</p>

En este mismo sentido, pero dentro de un mismo nivel, el desarrollo de sus características puede ofrecer diferencias de ritmo y desarrollo, a pesar de cumplirse con los criterios de habilidad que lo definen. Para quedar englobado dentro de un nivel determinado debe demostrarse habilidad suficiente al desarrollar sus tareas globales y el resto de características, pero una vez dentro, los progresos no se dan necesariamente de forma lineal: unos hablantes muestran mayores dudas, más nerviosismo, manejan la entonación con mayor esfuerzo, tienen más cantidad de inicios fallidos o reinicios de frases, más problemas para el uso de determinadas reglas gramaticales o características de la lengua, etc. Así, se dan pasos que van desde un conocimiento conceptual hasta un control parcial y a un control total que, dependiendo de su complejidad (concordancia en francés, aspecto en ruso, adjetivos pronominales en alemán, subjuntivo en español) o de otras características del hablante, pueden suponer un proceso muy largo.

Dentro de un mismo nivel se pueden diferenciar, por lo tanto dos nuevos subniveles, bajo y medio, ese caracterizado por una mínima habilidad y este por una habilidad más elaborada (cuadro 2.6.2 d). Las características de realización y de habilidad de uso de la

lengua dentro de estos dos subniveles que, no obstante, mantienen un uso de la lengua dentro del mismo nivel, se diferencian por un criterio de calidad y cantidad, es decir, por la cantidad de discurso que se produce, por un lado, y por la eficiencia y efectividad de ese discurso, por otro. Esta calidad también se establece por criterios de corrección. Dentro de estos subniveles puede haber diferencias de habilidad pero no de cambios en el uso de la habilidad funcional.

Cuadro 2.6.2 d. Subniveles dentro de un macronivel

Intermedio alto
Total control del nivel y además dominio -aunque no constante- de determinadas características del nivel siguiente en cantidad mayor a un 50% de las ocasiones requeridas.
Intermedio medio
Cumplimiento las características del nivel con mayor cantidad de discurso y más calidad de realización y corrección.
Intermedio bajo
Cumplimiento de todas las características del nivel aunque de forma mínima.

2.6.2- 2) El constructo teórico de la lengua que se va a medir

Como hemos visto en §2.2, un adecuado constructo teórico es uno de los principales elementos necesarios para establecer un buen test, constructo que debe estar claramente definido para poder proceder a elicitar y calificar la lengua en él definida. ACTFL (y la IRL) poseen, a juicio de Bachman, un adecuado constructo basado en lo que llama el *constructo teórico de la vida real*.

Los elementos a que hace referencia Bachman son los que en la escala de la OPI de ACTFL se llaman *criterios de evaluación*. Son cuatro:

1. Una secuencia de funciones o tareas globales de complicación mayor cuanto más alto es el nivel de uso de la lengua.
2. Unos contextos sociales o situaciones comunicativas diversos con unas áreas temáticas que les acompañan.
3. La corrección o precisión con que el hablante es capaz de realizar estas cosas.
4. El tipo de discurso o de texto que se produce.

Además de estos elementos, las nuevas escalas, reformadas en 1999, incluyen descripciones de otros elementos como fluidez y coherencia del discurso, estrategias de organización y planificación, estrategias de compensación, vocabulario, gramática, morfología o sintaxis, corrección, posibles errores dentro de cada nivel y pronunciación. Muchos de estos elementos han surgido de la profusa bibliografía crítica y de análisis a que se ha sometido la OPI durante sus años de funcionamiento¹¹⁴. Conjugando todos estos elementos, la OPI establece un test integrativo que mide una serie de habilidades desde una perspectiva global antes que atendiendo a la existencia o no de una característica lingüística concreta.

Todos estos componentes se analizan desde el punto de vista de su contribución a la habilidad de actuación en el uso de la lengua y en concreto, de las funciones específicas que conforman el constructo. Aunque puede ser tentador determinar una calificación basándose en fallos de vocabulario o gramática en realidad estos errores no son más que defectos llamativos; lo verdaderamente importante es detenerse en aquellas características que impidan la comunicación o provoquen un malentendido; eso es lo que hace que las tareas no se realicen eficientemente y lo que determina la calificación final.

Las tareas o funciones son el principal elemento de la medición de la habilidad comunicativa, pero están interrelacionadas con el resto de elementos, ya que no se producen aisladas sino en un contexto social y con unas temáticas concretas. Tales tareas, distribuidas en los cuatro niveles, son las siguientes:

¹¹⁴ Estos análisis críticos se describen en §2.6.3. De entre ellos destacamos, por haber sido tenidos en cuenta en esta reelaboración de la escala, algunos aspectos de la competencia estratégica, puntos de corrección gramatical presentados por Bachman (1988: 160) así como otros elementos sociolingüísticos y culturales (Lowe, 1998: 359).

	FUNCIONES O TAREAS GLOBALES
SUPERIOR	Puede discutir extensamente cualquier tema defendiendo su opinión y haciendo conjeturas e hipótesis.
AVANZADO	Puede llevar una conversación social o laboral y tomar la iniciativa, describir, narrar, explicar y opinar en la mayor parte de franjas de aspecto y tiempo.
INTERMEDIO	Puede llevar una conversación cara a cara, haciendo y respondiendo preguntas simples, y creando y combinando el material aprendido, para hacer sus mensajes, pedir información, preguntar horarios.
INICIAL	Puede comunicarse solo mínimamente, usando trozos aprendidos de frases o fórmulas, listas de palabras y enumeraciones.

Los contextos de uso se refieren a las circunstancias en que una persona usa la lengua; y conllevan unos contenidos determinados. Las *guidelines* o protocolos de ACTFL indican una jerarquía de estos temas y contextos basada en la cantidad y tipo de discurso que suponen. En niveles bajos, son situaciones en las que se utiliza material memorizado, porque es predecible el lugar y la naturaleza de los intercambios. Se basan sobre todo en frases en presente o sin marcas de tiempo reconocible (por ejemplo, la lista de elementos memorizados para pedir una comida en un restaurante). En los niveles altos, se trata de situaciones más complejas y se incrementan las necesidades de los interlocutores para usar flexiblemente la lengua. Esta complejidad surge de la imposibilidad de predecir los turnos de palabra y de la creciente complicación de las interacciones que se produzcan. Por ejemplo, siguiendo con el ejemplo del restaurante, podría ser la necesidad de protestar ante un servicio pobre o malo, o una discusión entre dos personas que mantienen puntos de vista diferentes sobre los efectos de la comida rápida.

El término *contenido* se refiere claramente a los temas de conversación, el elemento más variable de las entrevistas de ACTFL, ya que depende en gran medida de los intereses y formación de los candidatos. La cantidad de temas que se pueden abordar en una entrevista aumenta conforme aumenta el nivel de habilidad de los candidatos. En niveles bajos se limita a información autobiográfica y temas cercanos a la experiencia personal. En el superior, se espera que se pueda hablar, aunque no de forma especializada sobre variados temas, tanto concretos como abstractos. En todo momento

se evalúa la lengua usada y nunca los hechos expresados o las opiniones, aunque sean claramente erróneos.

Una de las primeras tareas del evaluador consiste en descubrir temas e intereses del candidato sobre los que basar buena parte de la entrevista con el objeto de conseguir que este hable. Posteriormente el entrevistador debe proceder a buscar mayor elaboración en los temas para probar un uso de la lengua apropiado a las características del nivel, y además, a lo largo de varias pruebas con temas diferentes, pues se busca comprobar si el nivel se mantiene en todo momento.

Las características de contexto y tema de cada macronivel son:

	CONTEXTO Y CONTENIDO
SUPERIOR	La mayoría de situaciones formales e informales. Una amplia gama de temas de interés general concretos y abstractos y algunos campos de su propia especialización y experiencia.
AVANZADO	La mayoría de contextos informales y formales de trabajo y relaciones sociales. Temas y hechos de interés general y otros no relacionados directamente con ellos: viajes, modos de vida, sistema de transporte, sistema educativo, etc.
INTERMEDIO	Escenarios informales habituales y limitados a intercambios básicos. Temas relacionados con su medio personal e inmediato: trabajo, familia, ciudad, amigos, etc.
INICIAL	Escenarios predecibles de su vida personal diaria. Aspectos concretos de la vida diaria: fechas y números, colores, meses, partes del cuerpo, ropa, comidas...

El tercer elemento, la corrección y precisión del discurso, se establece por varios parámetros entre los que se suelen citar la gramática, el vocabulario y la pronunciación. No obstante, estos no son los únicos. Parámetros sociolingüísticos, de fluidez o de pragmática son igualmente importantes. Un hablante que usa bien las reglas de la gramática puede cometer errores que causen incomunicación debido al fallo en otro tipo de reglas sociolingüísticas. Otro puede ser preciso y correcto a costa de perder fluidez, ser excesivamente breve, tener pausas demasiado largas y perder toda la naturalidad del

diálogo. O finalmente se puede perder el hilo de un discurso por falta de una palabra necesaria y carencias en compensarla mediante otras explicaciones más sencillas.

Estos son los elementos que deben considerarse:

- **Fluidez:** ritmo, cantidad de discurso y uso de elementos de cohesión.
- **Gramática:** normas de morfología y sintaxis.
- **Competencia pragmática:** la habilidad para usar mecanismos de gestión del discurso y de compensación ante situaciones de pérdida de control de la lengua.
- **Pronunciación:** habilidad para producir determinadas características del discurso, como una entonación adecuada, el acento de las palabras y el énfasis del sentido de las frases.
- **Competencia sociolingüística:** la habilidad para usar el lenguaje en registros y situaciones diferentes, apropiados a la cultura meta, junto al empleo de referencias culturales e idiomáticas.
- **Vocabulario:** amplitud y precisión del léxico empleado, así como cumplimiento de las normas de uso.

Las características de corrección para los cuatro macroniveles son:

	REQUISITOS DE CORRECCIÓN
SUPERIOR	Los errores nunca interfieren en la comunicación, o son propios de un nativo. No se preocupa de la forma de su discurso pues lo controla sin problemas.
AVANZADO	Es entendido sin dificultad por cualquier nativo.
INTERMEDIO	Es entendido, aunque con repeticiones e insistiendo para obtener información suficiente, por quien está acostumbrado a hablar con no nativos.
INICIAL	Puede ser difícil entenderle, incluso para hablantes acostumbrados a extranjeros. Hay que negociar el sentido de todo lo que dice.

No obstante, estas características se interpretan según cada nivel. En iniciales, la corrección se refiere principalmente a su inteligibilidad: el interlocutor debe llevar la

conversación y negociar o esforzarse por entender los mensajes debido a la poca cantidad de frases producidas, aunque quizás correctas por tenerlas memorizadas. Normalmente solo los hablantes acostumbrados a hablar con no nativos y que conozcan la lengua de sus interlocutores pueden lograr entenderles, debido a los fallos de pronunciación, vocabulario, gramática, normas sociolingüísticas y carencias estratégicas o pragmáticas.

Y el último gran elemento a considerar en cada macronivel, el tipo de discurso producido, se refiere a la cantidad y aspectos de organización del discurso. Comienza por palabras sencillas o sintagmas, pasa a frases independientes, llega al uso de párrafos y termina con una secuencia organizada de párrafos, haciendo con ellos un discurso amplio.

	TIPO DE DISCURSO
SUPERIOR	Amplio discurso.
AVANZADO	Párrafos.
INTERMEDIO	Frases sueltas.
INICIAL	Palabras sueltas y sintagmas.

Ahora que hemos visto estos elementos del constructo, podemos volver sobre la escala en que se sustentan y revisar sus niveles y subniveles. En concreto, dentro del nivel inicial, el cambio del subnivel bajo al medio se caracteriza por la cantidad de material memorizado, y la calidad juega un muy pequeño papel. Y el paso de medio a alto se caracteriza por la habilidad para usar creativamente la lengua, combinando y recomblando el material aprendido para transmitir contenidos personalizados y adaptados a la situación.

Dentro del intermedio, el cambio del subnivel bajo al medio se establece enteramente en los grados de cantidad y calidad, es decir, en la capacidad para mantener mayor número de temas de conversación y con mayor precisión. Y el paso al alto consiste en la emergencia de nuevas habilidades y el control de otras tareas globales propias del avanzado, con sus contextos de uso, contenido y tipo de discurso.

En el avanzado, las escalas de 1986 no establecían subniveles bajo y medio, pero estos se establecieron en la revisión de 1999, con los mismos criterios que para el resto de niveles. El paso al medio se sustenta en el control total de las marcas de tiempo y aspecto, en cuanto a la calidad del discurso, y en cuanto a la cantidad, al uso coherente y entrelazado de todas las funciones y tareas globales, es decir, a la capacidad para realizarlas conjuntamente y en párrafos con buena coherencia (narrar y explicar, analizando y describiendo, comparando y opinando, si es posible a un tiempo) en vez de ser simplemente capaz de utilizar estas funciones separadamente en párrafos pequeños. Y el paso al alto consiste en demostrar, además, habilidad para realizar discursos abstractos sobre temas propios del superior, demostrando ese nivel en más de un 50% de las ocasiones que lo requieren.

Es importante, ahora, fijarse en una característica de este tipo de entrevista, la capacidad de compensar flaquezas en determinados aspectos, con habilidades en otros, haciendo de la calificación un conjunto interrelacionado de todos sus elementos. Así, se puede mantener un nivel intermedio medio con muestras de gran control de la gramática, en detrimento de un vocabulario menos preciso o viceversa.

Por ello, un hablante con gran dominio de las reglas de gramática puede obtener la misma calificación de nivel que otro que no las controle igualmente, pero que en su caso posea un vocabulario más preciso o una habilidad estratégica y pragmática altas. Por lo tanto, la gramática no se conforma como el principal elemento de nivelación o calificación, ni puede ser el elemento principal de comparación externa (ya hemos dicho que en las OPI no se compara la actuación de candidatos entre sí, pero en la sociedad está claro que luego cada interlocutor va a juzgar a su manera a cada hablante).

Esta compensación se hace solamente dentro de cada macronivel. No entre niveles (entre un intermedio y un avanzado, por ejemplo) ya que en este caso cualquier fallo en un elemento hace que no se cumplan los requisitos del nivel inmediatamente superior y el candidato quede calificado como del inmediatamente inferior, por ejemplo intermedio alto. La gramática vuelve a no ser el elemento de calificación o de control más importante, ya que podemos encontrar hablantes con un buen control del aspecto y tiempos de la narración -propio de avanzado- pero que al incumplir otros elementos del nivel se queden en intermedio alto y otros que, sin tener ese buen control gramatical de las formas y aspecto, sin embargo, sí sean ya avanzados, por cumplir con todos los requisitos del avanzado, aunque mínimamente.

Los 10 niveles así establecidos, según las nuevas *guidelines* de 1999, son los que se ven en el siguiente cuadro (2.6.2 e.) que define, de forma general, todos los aspectos a considerar en el constructo de lengua manejado en la OPI, incluyendo no solamente los cuatro elementos ya señalados en las tablas anteriores, sino otros como la pronunciación, la inteligibilidad, los errores comunes de cada subnivel, las estrategias de planificación utilizadas en cada subnivel, y las estrategias compensatorias para cuando se intentan realizar tareas de nivel superior al que posee el candidato.

Cuadro 2.6.2 e. Descriptores de habilidad de la OPI de ACTFL (traducción propia; ver el original en inglés en Apéndice 2)

<p>Superior</p>	<p>Los hablantes del nivel superior son capaces de comunicarse con corrección y fluidez para participar plena y efectivamente en conversaciones sobre una gran variedad de temas en contextos formales e informales, tanto desde una perspectiva concreta como abstracta. Discuten sobre sus intereses y áreas de especialización, explican asuntos complejos en detalle y hacen narraciones extensas y coherentes, todo ello con facilidad, fluidez y corrección. Dan sus opiniones sobre un gran número de temas que les son importantes, como cuestiones sociales y políticas, y ofrecen argumentaciones bien estructuradas para defender sus opiniones. Pueden hacer y desarrollar hipótesis para explorar posibilidades alternativas. Cuando es necesario, utilizan un discurso extenso sin largas vacilaciones artificiales para defender sus puntos de vista, incluso cuando eso les lleva a elaboraciones abstractas. Tales discursos, aunque coherentes, pueden verse influidos por el sustrato de su propia lengua materna, antes que por los modelos de la lengua extranjera.</p> <p>Los hablantes del nivel superior dominan una gran variedad de estrategias interactivas y discursivas, como la toma de turnos de palabra, distinguen las ideas principales de las secundarias mediante el uso de técnicas sintácticas y léxicas, así como mediante otras características supraoracionales como el tono, acento y énfasis. Prácticamente no se aprecian errores sistemáticos en el uso de las estructuras básicas. Sin embargo, pueden darse errores esporádicos, especialmente con estructuras poco frecuentes o en algunas estructuras complejas más habituales aunque más propias del discurso formal o escrito. Y si tales errores aparecen no distraen a un interlocutor nativo ni interfieren en la comunicación.</p>
<p>Avanzado alto</p>	<p>Los hablantes del nivel avanzado alto ejecutan todas las actividades del avanzado con facilidad, seguridad y habilidad. Son capaces de usar la lengua con fundamento para explicar con detalle y de narrar en profundidad y con corrección en toda la gama de tiempos. Además, estos hablantes pueden manejar las actividades del nivel superior aunque no pueden mantenerse a ese nivel en todos los temas. No logran argumentar de forma estructurada para defender sus opiniones y pueden hacer hipótesis pero se hacen patentes</p>

	<p>algunos errores. Pueden discutir algunos temas de forma abstracta, sobre todo los de sus áreas de interés o campo profesional, pero se sienten más cómodos cuando discuten los temas de forma concreta.</p> <p>Los hablantes del nivel avanzado alto dejan constancia de sus recursos bien elaborados para compensar cualquier sentido defectuoso en algunas formas o limitaciones de vocabulario, mediante el uso seguro de estrategias comunicativas como parafrasear, hacer circunloquios y ejemplificaciones. Utilizan el vocabulario y la entonación precisos para dar el sentido y, a menudo, demuestran una gran fluidez y facilidad de palabra. Sin embargo, cuando tienen que realizar las tareas complicadas relacionadas con el nivel superior en diversos temas, es posible que su lengua se colapse o se muestre inadecuada o que eviten las tareas por completo, por ejemplo, echando mano de elementos más sencillos como la descripción o narración en vez de argumentar o hacer hipótesis.</p>
<p>Avanzado medio</p>	<p>Los hablantes del nivel avanzado medio pueden manejar con facilidad y seguridad una gran variedad de actividades comunicativas. Participan activamente en la mayoría de las conversaciones informales y algunas conversaciones formales sobre gran variedad de temas concretos relacionados con su trabajo, escuela, casa y actividades de tiempo libre, así como en acontecimientos públicos y de actualidad, de interés personal o importancia individual.</p> <p>Los hablantes del nivel avanzado medio dominan la habilidad de narrar y describir en la mayor gama de tiempos (pasado, presente y futuro) dando información completa, con un buen control del aspecto y adaptándose a las necesidades de la conversación con flexibilidad. La narración y la descripción tienden a combinarse y entretorse para relatar y sustentar los hechos en forma de párrafos.</p> <p>Los hablantes del nivel avanzado medio pueden manejar con éxito y relativa facilidad los cambios lingüísticos que se presenten ante una complicación o un vuelco inesperado del transcurso de los acontecimientos que surja en el contexto de una situación o actividad rutinaria con la que normalmente estén familiarizados. Para ello se suelen usar estrategias comunicativas como reexplicaciones o rodeos. El discurso de estos hablantes indica suficiente fluidez. Su vocabulario es considerablemente amplio, principalmente de carácter general, salvo en caso de algún área particular de especialización o de interés. Las estructuras de su lengua dominante tienden a desaparecer aunque el discurso puede aún reflejar la forma de la estructura oral de su propia lengua y no la de la lengua extranjera.</p> <p>Los hablantes del nivel avanzado medio participan en la conversación en una gran variedad de temas familiares tratados eficazmente con suficiente corrección, claridad, y precisión, como para transmitir las intenciones de sus mensajes sin confusión ni distorsiones. Son entendidos fácilmente por nativos que no estén acostumbrados a tratar con extranjeros. Cuando intentan realizar las acciones y manejar los temas del nivel superior, la calidad y cantidad de su discurso por lo general se deteriora significativamente. Estos hablantes son capaces de manifestar sus opiniones o alegar condiciones; sin embargo, carecen de la destreza para producir con consistencia una argumentación estructurada en un amplio discurso. Estos hablantes pueden usar algunas estrategias dilatorias, recurrir a la narración, descripción, explicación o reseña,</p>

	o sencillamente intentan evitar las exigencias de las actividades del nivel superior.
Avanzado bajo	<p>Los hablantes del nivel avanzado bajo pueden manejar una gran variedad de actividades comunicativas, aunque a veces con interrupciones. Participan activamente en la mayoría de conversaciones informales y un número limitado de conversaciones formales relacionadas con su escuela, casa, actividades de tiempo libre y en menor grado, otras relacionadas con acontecimientos del trabajo, actualidad, intereses personales y públicos u otros de importancia para ellos.</p> <p>Los hablantes del nivel avanzado bajo dominan la habilidad de narrar y describir en la mayor gama de tiempos (pasado, presente y futuro) con amplitud de los párrafos, aunque el control del aspecto puede a veces ser deficiente. Pueden manejar adecuadamente los cambios lingüísticos que se presenten ante una complicación o un vuelco inesperado del transcurso de los acontecimientos que surja en el contexto de una situación o actividad rutinaria con la que normalmente estén familiarizados, aunque a veces su discurso puede estar forzado o resultar pobre para el nivel. En esos casos se suelen usar estrategias comunicativas como reexplicaciones o rodeos. Al narrar o describir se combinan y encadenan las frases componiendo un discurso organizado y de la extensión del párrafo. Si se les insiste para que cuenten algo completamente, se pierden y tienden a empobrecer su discurso. La extensión de sus enunciados no suele sobrepasar el párrafo. La estructura de su lengua predominante se nota en el uso de falsos amigos, traducciones literales, y en que todavía su forma de hablar es la de su propia lengua y no la de la lengua extranjera.</p> <p>Mientras que la lengua de un avanzado bajo se caracteriza por su sucesión considerable aunque irregular, hay ciertas particularidades de artificialidad y vacilación, con apreciables autocorrecciones y rudeza gramatical. Su vocabulario es esencialmente genérico.</p> <p>Los hablantes del nivel avanzado bajo participan en la conversación con suficiente corrección, claridad, y precisión como para transmitir las intenciones de sus mensajes sin confusión ni distorsiones y pueden ser entendidos por nativos que no estén acostumbrados a tratar con extranjeros, aunque quizás se necesiten repeticiones y reexplicaciones. Cuando intentan realizar las acciones y manejar los temas del nivel superior, la calidad y cantidad de su discurso se deteriora significativamente.</p>
Intermedio alto	<p>Los hablantes del nivel intermedio alto pueden conversar con facilidad y seguridad cuando se trata de las actividades rutinarias y situaciones sociales propias del nivel intermedio. Pueden desenvolverse con éxito en muchas actividades sin complicación y situaciones sociales que requieren intercambio de información básica relacionada con el trabajo, escuela, entretenimiento, intereses particulares y áreas de su competencia, aunque las dudas y errores son evidentes.</p> <p>Los hablantes del nivel intermedio alto manejan las actividades del nivel avanzado pero no pueden mantener la actuación a ese nivel en toda la variedad de temas. Con cierta consistencia estos hablantes pueden narrar y describir en la mayoría de franjas temporales usando un discurso organizado y con la extensión del párrafo. Sin embargo, su actuación en las actividades del nivel</p>

	<p>avanzado muestra una o varias características de colapso, como la incapacidad de mantener la narración o descripción semántica o sintácticamente en la franja temporal apropiada, el desmoronamiento de las conexiones discursivas, el mal manejo de los elementos de cohesión, una disminución del alcance y propiedad del vocabulario, fracasar en el objetivo de uso de los circunloquios, o un significativo aumento de las dudas.</p> <p>Los hablantes del nivel intermedio alto se hacen entender sin problemas por hablantes que no están acostumbrados a tratar con extranjeros, aunque la lengua predominante todavía se nota (p. e. cambios de código, falsos amigos, traducciones literales, etc.), y puede haber lagunas en la comunicación.</p>
<p>Intermedio medio</p>	<p>Los hablantes del nivel intermedio medio pueden manejar con acierto una variedad de actividades comunicativas sencillas en situaciones sociales básicas. Su conversación se limita a algunos de los intercambios establecidos y predecibles necesarios para la supervivencia en la cultura de la lengua. Estos comprenden información personal relacionada consigo mismo, la familia, la casa, actividades diarias, preferencias e intereses personales, así como necesidades físicas y sociales como alimentación, compras, viajes y alojamiento.</p> <p>Los hablantes del nivel intermedio medio tienden a reaccionar al discurso, por ejemplo, contestando a preguntas directas y pidiendo información. Sin embargo son capaces de hacer una variedad de preguntas cuando necesitan obtener información simple para satisfacer sus necesidades básicas, como direcciones, precios y servicios. Cuando se les pide que desarrollen funciones o manejen temas de nivel avanzado, dan cierta información pero muestran dificultades para conectar las ideas, manipular el tiempo y aspecto y usar estrategias comunicativas como los circunloquios.</p> <p>Los hablantes del nivel intermedio medio pueden expresar contenidos personales siendo creativos con la lengua, en parte combinando y recomblando los elementos que ellos saben y el input de la conversación, para hacer enunciados de la duración de una frase y encadenando frases. Su discurso puede tener pausas, reformulaciones y autocorrecciones cuando se esfuerzan por buscar el vocabulario y formas apropiadas para expresarse sobre ellos mismos. Debido a las imprecisiones de vocabulario y/o pronunciación y/o gramática y/o sintaxis, pueden producirse malentendidos, pero estos hablantes son normalmente bien entendidos por interlocutores comprensivos, especialmente acostumbrados a tratar con extranjeros.</p>
<p>Intermedio bajo</p>	<p>Los hablantes del nivel intermedio bajo pueden manejar un número limitado de actividades en situaciones sociales simples siendo creativos con el lenguaje. Su conversación se limita a algunos de los intercambios establecidos y temas predecibles necesarios para la supervivencia en la cultura de la lengua. Temas relacionados con la información personal como, por ejemplo, él mismo y su familia, algunas actividades diarias y preferencias personales, y con algunas necesidades inmediatas, tales como pedir una comida y hacer compras sencillas. En el nivel intermedio bajo los hablantes, en buena medida, reaccionan y se esfuerzan por responder a preguntas directas y pedir información, pero también pueden preguntar adecuadamente unas cuantas cosas.</p>

	<p>Los hablantes intermedio bajo pueden expresar contenidos personales combinando y recomblando en enunciados cortos lo que ellos saben y lo que oyen a sus interlocutores. Sus intervenciones están llenas de dudas e incorrecciones en busca de la forma lingüística y el vocabulario mientras intentan componer sus mensajes. Su discurso contiene muchas pausas, reformulaciones ineficaces y autocorrecciones. Su pronunciación, vocabulario y sintaxis están poderosamente influidos por su lengua materna, pero a pesar de los continuos malentendidos que requieren repeticiones y aclaraciones, los hablantes de nivel intermedio bajo pueden entenderse con interlocutores comprensivos, especialmente con los que están acostumbrados a tratar con extranjeros.</p>
Inicial alto	<p>Los hablantes del nivel inicial alto pueden manejar una variedad de temas y actividades de nivel intermedio pero son incapaces de mantener el discurso en ese nivel. Pueden desenvolverse con éxito en un número de actividades comunicativas sencillas propias de situaciones sociales simples. Su conversación se limita a unos pocos temas predecibles necesarios para la supervivencia en la cultura de la lengua, tales como información personal básica, objetos básicos y un número reducido de actividades, necesidades inmediatas y preferencias. Los hablantes iniciales altos responden a preguntas simples y directas o piden información; cuando se les pide que hagan preguntas sólo son capaces de hacer unas pocas y memorizadas.</p> <p>Los hablantes iniciales altos pueden expresar contenidos personales basándose, sobre todo, en frases aprendidas o re combinaciones de frases y en lo que oyen de su interlocutor. Sus enunciados, que suelen consistir fundamentalmente en frases cortas y a veces incompletas en presente, pueden ser vacilantes o incorrectos. Sin embargo, como estos enunciados son una mera prolongación de material aprendido y un repertorio de locuciones, pueden parecer inesperadamente correctos y fluidos.</p> <p>Su pronunciación puede estar poderosamente influida por su lengua materna así como su vocabulario y sintaxis cuando intentan hacer enunciados personales. Pueden surgir muchos malentendidos pero los hablantes de inicial alto, mediante repeticiones y reformulaciones, pueden entenderse con interlocutores comprensivos y acostumbrados a extranjeros. Cuando se les pide que traten temas o realicen actividades de nivel intermedio, estos hablantes pueden responder a veces con frases inteligibles pero son incapaces de mantener el nivel basado en frases independientes.</p>
Inicial medio	<p>Los hablantes del nivel inicial medio se comunican mínimamente y con dificultad usando un número de palabras aisladas y locuciones memorizadas, ciñéndose al contexto concreto en que han aprendido la lengua. Cuando responden a preguntas concretas solamente pueden emitir dos o tres palabras cada vez u ocasionalmente alguna respuesta de repertorio. Hacen pausas constantemente, buscando palabras sencillas o repiten y reciclan las suyas y de las de su interlocutor. A causa de las dudas, falta de vocabulario, incorrección, o fallos en responder adecuadamente no logran hacerse entender fácilmente ni ante interlocutores acostumbrados a tratar con extranjeros. Cuando se les pide que manejen temas y realicen actividades de nivel intermedio suelen recurrir a repeticiones, palabras de su lengua materna o se callan.</p>
Inicial bajo	<p>Los hablantes del nivel inicial bajo carecen realmente de habilidad funcional y a causa de su pronunciación pueden ser ininteligibles. Si se les da suficiente</p>

	<p>tiempo e indicaciones que les sean familiares pueden intercambiar saludos, presentarse y nombrar unos cuantos objetos de su entorno familiar. Son incapaces de realizar actividades o manejar temas del nivel intermedio y no pueden participar en una auténtica conversación.</p>
--	---

(Breiner-Sanders, 2000: 14-17).

2.6.2- 3) La estructura de la entrevista

La entrevista se estructura de forma que se pueda elicitarse una muestra suficiente de las mejores habilidades del candidato. En ACTFL a esta muestra de lengua bien elicitada se le conoce con el nombre de *muestra calificable*, pues debe contener los elementos necesarios e indispensables para proceder a su calificación.

Los tiempos necesarios para obtener tales muestras dependen del nivel del candidato. Los más altos llevan más tiempo, ya que en cada nuevo nivel se incrementan los elementos evaluables. En los subniveles altos también se requiere más tiempo que en los bajos y medios, ya que se debe probar que existen las habilidades del nivel inmediatamente superior en cuanto a su tipo de discurso, tareas, contexto y contenido y características de corrección y precisión. Pero generalmente, un evaluador con experiencia puede evaluar en 10 minutos a un inicial, en unos 12 a un intermedio medio, hasta 20 si es intermedio alto, usa entre 20 y 25 si es avanzado, y 30 si es superior (Swender, 1989: 4-1).

La entrevista está estructurada en cuatro fases diferentes con objetivos lingüísticos, psicológicos y evaluativos claros, e incluye tres tipos de interacción: preguntas y repuestas, conversación y juegos de roles. No obstante, esta naturaleza conversacional propia de la entrevista se diferencia de una conversación real o no evaluativa en que el evaluador la guía y estructura cuidadosamente con el objeto de lograr del candidato una buena muestra de lengua calificable. En niveles bajos (inicial e intermedio) la conversación se limita a una serie de preguntas y respuestas. El *role play* les ayuda a realizar un tipo de conversación más natural dentro de una situación o contexto determinado. En niveles más altos, y en general durante toda la entrevista, el evaluador busca la existencia del lenguaje correspondiente a los criterios establecidos mediante el uso de una serie de tipos de preguntas sobre ciertos temas adecuados a la dificultad propia del nivel.

Lograr esa muestra calificable exige utilizar los criterios de evaluación del constructo teórico con el fin de encontrar un nivel de uso de la lengua cómodo (llamado *floor of performance*, o base de actuación), junto a sus características lingüísticas, de tipo de discurso y corrección apropiadas. Posteriormente pasa a encontrar el techo de posibilidades del hablante o limitaciones de su actuación (*ceilling of performance*) donde ya no se cumplan enteramente las características de un nivel. Con ello, la entrevista se estructura para encontrar modelos de fortaleza y de debilidad lingüística que determinen la base y las limitaciones del hablante.

La entrevista tiene siempre una estructura con cuatro fases (o cinco si contamos la salida) que guían la actuación a través de las tareas y temas descritos, desarrollándose hasta que se muestre que el hablante no logra la realización de alguna de ellas de forma adecuada al nivel descrito en las escalas. Y cada fase tiene unos propósitos claros, tanto lingüísticos, como evaluativos o afectivos¹¹⁵.

Las fases son:

1. Fase de calentamiento.
2. Fase de pruebas de nivel.
3. Fase de búsqueda de las limitaciones del hablante.
4. Fase de *role play*.
5. Fase de salida.

1. Fase de calentamiento

Es el comienzo de la entrevista: saludos e intercambio de información informal con el objeto de ayudar psicológicamente al candidato a sentirse tranquilo y cómodo y a que se habitúe a la manera de hablar y pronunciar del evaluador.

Lingüísticamente es una fase meramente formulista y con un lenguaje protocolario muchas veces, basado en material aprendido y de naturaleza relativamente automática.

Evaluativamente puede ser un momento en el que el evaluador tenga la oportunidad de observar ciertos temas que se puedan desarrollar posteriormente en la entrevista.

¹¹⁵ En el Apéndice 3 hay una entrevista transcrita con comentarios sobre cada fase como ejemplo de su estructura.

Además, la forma de responder o preguntar ya puede dar una primera idea respecto a un nivel de actuación del candidato, aunque este pueda variar a lo largo de la entrevista, o deba ser comprobado.

Cuando el candidato parece estar tranquilo y preparado para hablar, es momento de pasar a la siguiente fase.

2. Fase de pruebas de nivel

A veces un candidato ya responde a una primera pregunta sencilla y social (“¿Cómo estás?”) con una respuesta verdadera (“Pues no muy bien porque esta noche he tenido problemas para dormir y el ruido que hay en la calle me molesta”) en vez de la fórmula social “muy bien gracias ¿y usted/tú?”. Lo que en sí es una muestra de que está dispuesto a participar en una conversación más sustanciosa.

El objeto de esta fase desde el punto de vista evaluativo es probar un nivel de consistencia en el uso de la lengua dentro de un contexto particular y unos temas específicos, cumpliendo en todo momento con las características descritas para el nivel intermedio. Desde el punto de vista del candidato, es la oportunidad para mostrar su fortaleza y lo mejor de su habilidad lingüística.

Siempre se comienza una entrevista en el nivel intermedio, salvo que se vea ya en el calentamiento la imposibilidad de hacerlo así. El entrevistador debe llevar al candidato a tratar los temas del nivel y observar el cumplimiento de sus características:

- a) Ser creativo con el lenguaje.
- b) Ser capaz de responder y preguntar.
- c) Participar en una conversación sencilla.
- d) Participar con éxito en situaciones de intercambio social no complicadas.

Así, el entrevistador puede plantear una sencilla pregunta para hablar sobre su familia y, basándose en la respuesta, continuar pidiendo información más concreta como: “¿Tienes una hija?, ¿cómo es?” o “Dices que tienes una hermana mayor, ¿puedes hablarme de ella?” Estas preguntas abiertas permiten al candidato utilizar el lenguaje

más que dar respuestas cerradas como “sí” o “no” o utilizar lenguaje reciclando lo que dice el entrevistador¹¹⁶. Y así comienza a participar en una conversación.

Hay que tratar diversos temas con el fin de asegurarse de que ese modelo de uso de la lengua del candidato es realmente una habilidad y no solamente característico del tema que esté tratando. Así que habrá que tratar otros temas mediante más preguntas que continúen en esta fase de pruebas del nivel básico, como “háblame de tu clase de historia” o “cuéntame algo sobre lo que haces en tu tiempo libre” y volver a iniciar conversaciones al respecto.

Una vez comprobada la consistencia en las habilidades propias del nivel es el momento de comenzar la siguiente fase.

3. Fase de búsqueda de las limitaciones del hablante

El objeto de esta fase es averiguar las limitaciones o debilidades del hablante, para lo que se le lleva a probar su actuación en las situaciones y temas del siguiente nivel (el avanzado en este caso) generalmente subiendo el nivel de las preguntas al nivel inmediatamente superior sin cambiar de tema. Por ejemplo, si el candidato está hablando de los transportes que utiliza en su ciudad para ir al trabajo y no le gustan ni a él ni a la gente, podríamos pasar al nivel siguiente planteándole una pregunta como: “Vale, ¿puedes darme una explicación de qué pasa con los transportes en tu ciudad, de por qué los usan poco?”

Desde el punto de vista de la evaluación pueden pasar tres cosas en este momento:

1. El hablante no puede mantenerse en ninguna de las nuevas situaciones ni participar adecuadamente en ninguna de las nuevas conversaciones propuestas. Entonces termina la fase de exploración de limitaciones y se vuelve a la fase anterior.

2. El hablante logra participar adecuadamente en todas las conversaciones propuestas sobre temas nuevos propios del nivel y cumple con las características lingüísticas (tipo de discurso, corrección, temas y contextos y funciones o tareas) sin

¹¹⁶ De los tipos de preguntas utilizados en las OPI se habla en §2.6.2- 4.

problemas. Entonces esta fase de búsqueda de limitaciones funciona como una nueva fase de pruebas de nivel (fase 2) en la que se deben realizar las pruebas apropiadas para demostrar que esa habilidad y sus características se cumplen y mantienen en todo momento, proponiendo varios temas para las conversaciones.

3. El candidato parece cumplir, a veces, con los requisitos del nivel y otras veces no puede hacerlo con facilidad o muestra carencias. O sea, que aunque parece que puede tratar los temas del nivel y sus funciones o tareas no logra en todo momento hacerlo con las características adecuadas. Este suele ser un argumento para demostrar que está en el subnivel inmediatamente inferior, o intermedio alto. Es importante en estos casos tomarse el tiempo necesario para demostrar esto y establecer adecuadamente los fallos y las características de su discurso. Ya que si una persona responde a una pregunta con un nivel apropiado al siguiente nivel, aún no podremos decir que cumple con ese nivel inmediatamente superior. Para serlo deberá cumplir los protocolos y características definidas para ese nivel, sus funciones, su tipo de lengua, corrección y temas.

A continuación transcribo dos ejemplos de entrevistas (una con un candidato de nivel intermedio medio -Tiffany- y otra de avanzado medio -François-. Los ejemplos son nuestros) en las que se aprecia esta fase de búsqueda de las limitaciones. Al entrevistarles se puede observar cómo les hago las preguntas necesarias subiendo el nivel y me topo con imposibilidades de expresión que son una primera prueba de sus limitaciones. En el caso de Tiffany, hablante de nivel intermedio medio, se muestra la imposibilidad de expresarse a un nivel avanzado y narrar un incidente; en François, hablante de nivel avanzado medio, se aprecia la imposibilidad de expresarse a un nivel más alto, superior: no puede defender sus opiniones en discurso abstracto extendido.

Tiffany

Nivel final calificado: intermedio medio.

Va bien en las pruebas de intermedio y se colapsa cuando se le insiste en que narre algún incidente -prueba de limitaciones en avanzado- (señalado en negrita):

E.- ¿Puedes hablarme un poco de Chicago? Yo no conozco Chicago.

C.- Es al oeste de Chicago, casi 16 minutos al coche de Chicago, es un suburbio.

E.- *¿Y cómo es dónde tu vives?*

C.- *Es bonito, hay muchos árboles, las casas son muy cerca. No hay mucho espacio pero más que la ciudad. Y es aburrido porque no hay mucho para hacer. Hay más en la ciudad pero hay cines y tiendas, unos restaurantes, una piscina. Trabajo en una piscina.*

E.- *¿Sí? ¿Qué haces?*

C.- *Soy socorrista, también soy una entrenadora de los niños para nadar. Tengo un equipo de nadar.*

E.- *¿Puedes contarme como es un día normal en tu trabajo?*

C.- *Empiezo a las 9 y limpio los baños y en la piscina. Y entonces a las 10 la piscina abre y la gente vienen y yo veo la piscina para que nadie ¿muerte? Y si yo trabajo durante la tarde o durante la noche, necesito limpiar después de la piscina cierra.*

E.- *¿Y es un trabajo de mucha responsabilidad?*

C.- *No es difícil. Nunca hay personas que no pueden nadar.*

E.- *¿Y nunca has tenido ningún incidente en tu trabajo o con tu equipo? No hay muchos incidentes.*

E.- *¿No hay nada que me puedas contar?*

C.- *En verano yo y los otros entrenadores fui a un restaurante con más de 100 chicos y es loco.*

E.- *¿Sí, cuéntame qué pasó?*

C.- ***Unos de los chicos no tienen muchos años, como ellos tienen 8 años y 6 años y otros tienen más, como 18, pero ellos siempre cenar o comen en la restaurante y hacen contracciones, hacen un sucio en las mesas y es difícil. Si ceno no yo realizo todos.***

François

Nivel final calificado: avanzado medio.

Va bien en las pruebas de Avanzado pero se colapsa y limita la cantidad de información de forma muy destacable, cuando se le insiste en que defienda su opinión respecto a la mezcla culturas de la que habla -prueba de limitaciones en superior- (señalado en negrita):

E.- *Antes has dicho que la mezcla de culturas en tu país es diferente de la de EE.UU. ¿Qué diferencias hay entre la mezcla de culturas de Canadá y la de EE.UU?*

C.- *Pienso que en los EE.UU., aunque todavía siguen culturas definidas, pero pienso que en los EE.UU. lo que predomina es -ellos dicen melting pot, que es un tipo de mezcla bastante homogénea de muchas culturas- y es la cultura que es tal vez la cultura más fuerte de todo el mundo porque, bueno, se llevan tejanos en todos países, es tan fuerte que alguien que llega a EE.UU. con una cultura diferente no se puede como, no puede imponer su cultura, tiene que asimilarse a ese gigante. En Canadá no sé porque sería difícil porque somos mucho menos y menos poderosos pero las culturas se hacen frente sin, y se mezclan, la gente se mezcla pero cada uno puede conservar su cultura más que los EE.UU. No hay, no se impone una manera de ser tanto como en EE.UU. Por ejemplo creo que tendría que también ver con los pueblos fundadores de Canadá. En Estados Unidos había los ingleses pero en Canadá había también los franceses aunque perdieron la guerra, mucha gente se quedó en Canadá y muchos francoparlantes siguieron viviendo en Canadá.*

- E.- *¿Y qué ventajas tiene eso para Canadá? ¿Puedes hacer un análisis y explicarme las ventajas y los inconvenientes de ese modelo de mezcla cultural?*
- C.- *Tiene ventajas... pienso que puede ser más... pienso que tiene muchas ventajas como tiene también lados malos o incomodantes pero las ventajas serían es la riqueza que se destaca cuando por ejemplo imagino que un español que viniera a visitar Canadá le gustaría conocer una cultura nueva pero por eso por lo tanto le gustaría todavía más conocer dos. Y pienso que para un inmigrante es agradable de ver que se puede seguir viviendo como vivía en su país. Entonces eso está en también ese intercambio que tenemos, como vivimos juntos hay que intercambiar, no se puede vivir aislado. Cada uno, cada parte está se enriqueza con la cultura del otro.*
- E.- *Pero todo eso que me cuentas que es tan bonito es como idílico, ¿no crea problemas de guetos o contrastes y conflictos?*
- C.- *Eso es, es que la buena parte, la parte mala es difícil conseguir eso es que a menudo las dificultades hacen que la gente no se mezcle, la gente hay tensiones, entonces el contrario puede llegar a con las dificultades la gente no se mezcla, se aíslan y el aislamiento puede ser muy contraproducente.*

La fase de pruebas y de limitaciones busca patrones de comportamiento y no usos aislados de la lengua, por lo que se debe realizar buscándolos a lo largo de diversas muestras, temas y conversaciones sobre temas variados. Esto lo hace un evaluador experimentado de forma integrada y no como partes diferenciadas de la entrevista, de forma que el candidato apenas se da cuenta de que está pasando por diferentes fases o niveles de la prueba sino que siente mayor o menor dificultad para responder a las preguntas en relación a un mismo tema o una incomodidad lingüística, no emocional, cuando se le plantean ciertas preguntas. Si el evaluador comprueba una base sólida de actuación sube el nivel de las preguntas y si encuentra limitaciones vuelve al nivel de confianza progresivamente.

La distribución del tiempo entre pruebas y limitaciones es diferente. Para los subniveles bajo y medio se necesita más tiempo para demostrar la existencia de esta base y averiguar si dentro de ella se puede actuar a un nivel o a otro (basándose en los criterios de cantidad y calidad de discurso del mismo nivel) y muy poco tiempo para mostrar las limitaciones dentro del siguiente nivel. En cambio, para los subniveles altos se debe pasar mucho más tiempo en la fase de búsqueda de limitaciones, asegurándose que el lenguaje no cumple con el siguiente nivel pero que, en cambio, sí está la mayor parte del tiempo dentro de sus características; y muy poco tiempo para averiguar que la base de los anteriores subniveles (intermedio bajo y medio) es firme.

En esta fase de búsqueda de limitaciones se produce un fenómeno sicolingüístico que en ACTFL se conoce como de *ruptura de las habilidades* o *colapso* del uso de la

lengua (*breakdown*). El comportamiento del candidato puede caracterizarse por alguna de las siguientes pautas:

1. Anunciar que no puede responder a la pregunta (“eso es muy difícil”, “no puedo explicarlo”, etc.).
2. Utilizar su lengua materna u otras lenguas en las que se sienta cómodo al hablar.
3. Mostrar incomodidad mediante el comportamiento lingüístico: dudas, repetidos comienzos de frases, búsqueda de palabras, empezar una frase y no terminarla y cambiar a otra.
4. Dejar ver otros signos no lingüísticos de incomodidad como mover los hombros significativamente, reírse nerviosamente, fruncir el ceño, mover los ojos.
5. Continuar hablando pero disminuyendo la cantidad de corrección o precisión de sus mensajes, haciendo errores en cosas que anteriormente se habían logrado sin problemas, o -con hablantes de niveles altos- repitiendo la misma idea de forma diferente o con otras palabras, pero sin profundizar ni explicarla adecuadamente o perdiendo el control de la información.

A continuación se transcriben algunos ejemplos de todas estas situaciones de estrés que se producen en el candidato cuando no puede desarrollar ciertos temas por ser demasiado difíciles para él (los ejemplos son nuestros).

- Nerviosismo, risas, incomodidad, aumento de pausas y jadeos:

Como tenemos que memorizar algo con... y... ya los... la gente tiene muy... no cabeza muy duro pero piensa algo muy ah es muy no estrecho pero es muy pequeño pequeño, no pueden pen... pensar algo muy ah mmm... .. no pueden pensar algo diferente y ...

- Expresión de que es muy difícil, quejas por la pregunta:

No estoy segura pero eh los planes de urbano... no son bueno porque ah... es muy difícil ah gasolina es muy importante por ah compañías dinero.

El hombres los hombres grandes the big gays designe the cities or... no ah no puedo explicar.

-Uso de la lengua materna:

En Filadelfia las calles son muy grandes pero the drivers no piense sobre la gente en bicicletas.

No es ciudad ah designed para pedestriane (risa).

IBM y otras, tiene una sector de diversity y pero empresas pequeñas no tenga...

- Incremento de la búsqueda de palabras:

Es bueno pero soy joven, no tengo prisa por (risa) comitar or eh cas nos casamos

Para chicas y ah personas... ¿qué es la palabra?, de Europa, de África, de diferentes países, personas que trabajan en Estados Unidos, ¿qué es la palabra?

Hay muchas edificios viejos, árboles y... rij... riojos... los edificios son riojos y la gente es muy simpática.

Tiene ah ah tiene una lugares de ah católicos y ah los ah tienes muy grandes ah torres y ahh... el ela el lugar es en es en una... montaña.

- Aumento de los fallos, repetición de ideas, pérdida del control de las ideas, ruptura del inicio de las oraciones, imposibilidad para terminar las frases o ideas:

Pues ya tenemos problema en el político government y es porque nuestro sistema... como tenemos que memorizar algo con... y... ya los... la gente tiene muy... no cabeza muy duro pero piensa algo muy ah es muy no estrecho pero es muy pequeño pequeño, no pueden pen... pensar algo muy ah mmm... ... no pueden pensar algo diferente y ...

(Pregunta: ¿Sabes qué pasó?):

No sé qué pasó con el tren pero un autobús fui a, pero tomamos a Almería.

(Pregunta: ¿Y qué hicisteis?):

Por un poco tiempo estamos en el tren por la, otro tiempo estamos en la calle y entonces el autobús. Es un viaje muy largo. 5 horas.

Hay un río y entonces ah la tierra alto y muy alto es lejo y lo río y alto muy tierra. Y tiene una sistema de escaletores paran personas ah escalar ah alto arriba montaña.

- Escasez de información:

Los alemanes son muy, personas tienen estrés y son muy rápidos y ninguna es tranquilo algo así. Y eso es un diferencia muy grande entre Alemania y España. Porque me gusta mucho, un poco, la vida es muy corto y por eso es mejor cuando vives como tú o vosotros aquí en España.

Eso es, es que la buena parte, la parte mala es difícil conseguir eso. Es que a menudo las dificultades hacen que la gente no se mezcle, la gente hay tensiones, entonces el contrario puede llegar a con las dificultades la gente no se mezcla, se aíslan y el aislamiento puede ser muy contraproducente.

Cuando se han realizado las fases de prueba y limitaciones y se está seguro de un nivel de habilidad lingüística se procede a realizar un *role play*.

4. Fase de *role play*

Esta fase pretende introducir al hablante en situaciones y funciones lingüísticas que no se hayan podido probar durante las fases anteriores de una forma más conversacional.

En el nivel intermedio, permitiéndole que haga preguntas, inicie él mismo una conversación, la guíe y la finalice.

En el avanzado, haciéndole enfrentarse a una situación complicada que asimismo deberá manejar desde el principio hasta el final.

Y en el superior haciéndole que maneje una situación inusual.

Con un *role play* suele ser suficiente para comprobar estos hechos.

La selección del rol adecuado para comprobar los datos todavía no obtenidos es importante. Si un candidato está en los subniveles bajo o medio se le debe proponer un rol de ese mismo nivel. Si, en cambio, está en un subnivel alto, se le puede plantear un rol del nivel siguiente para comprobar sus limitaciones o, al contrario, darle uno de su mismo nivel para tener más pruebas de su habilidad. Una vez realizado el rol debe procederse a una fase de salida claramente diferenciada, ya que terminar en ese momento puede dar al candidato la impresión de que es solamente el final del rol y no de toda la entrevista y dejarlo desorientado.

ACTFL ofrece una serie de roles ya elaborados, en una baraja con 30 roles de diferentes colores para cada nivel, verde para intermedio, naranja para avanzado y amarillo para superior, con el fin de poder proceder a una selección sencilla y adecuada sobre la marcha. También da instrucciones y recomendaciones para elaborar otros personalmente (Swender, 1989: 6-6).

Estos son ejemplos de los *role play* propuestos, uno de cada nivel:

Intermedio:

Quieres alquilar un apartamento. Habla con el dueño del edificio y explícale lo que necesitas. Hazle cuatro o cinco preguntas para informarte de si tiene todo lo que necesitas.

Avanzado:

Te has quedado trabajando hasta tarde y eres el único en la oficina. Has salido a comprar una botella de agua y al volver no tiene nada que te identifique. Explícale la situación al guarda jurado y pídele que te abra la puerta.

Superior:

Te piden tu opinión sobre el hecho de que en muchos países esté aumentando la preocupación sobre las dietas y la necesidad de hacer ejercicio. Hay quien tiene objeciones y dice que simplemente es una obsesión fanática consecuencia de la opulencia de nuestra sociedad y su necesidad de llenar el tiempo. Debate el tema, expresando tu opinión sin dejarte convencer.

5. Fase de salida

El objeto de esta fase es, además, volver a llevar al candidato a las situaciones en las que se siente cómodo, es decir, a su óptimo nivel de actuación. Ya que es posible que en la fase anterior haya tenido que realizar una actividad en la que se haya sentido incómodo. En todas las OPI la entrevista debe terminar con el candidato hablando en el nivel en el que se puede manejar adecuadamente, con preguntas de su nivel a las que pueda responder tranquilo.

Estas fases se cumplen en toda entrevista. Aquí se han explicado comenzando por el intermedio y haciendo pruebas en intermedio y búsqueda de limitaciones en avanzado. Si el avanzado se mantiene se considera una fase de pruebas y se hará la búsqueda de limitaciones en superior y si nuevamente parece sustentarse, el candidato deberá probar el cumplimiento de las características de ese nivel para quedar calificado como avanzado alto o superior. Si, en cambio, el candidato no sustenta el intermedio se procede a hacer las pruebas de inicial y a establecer si es inicial bajo, medio o alto.

Todo el proceso está guiado, además, por unas normas de conducta del entrevistador que ayudan a enfrentarse a situaciones problemáticas y por una tipología de preguntas que permite lograr que el candidato dé buenas muestras de sus habilidades. De esto trata el siguiente punto.

2.6.2- 4) Los sistemas de elicitación de la lengua y el comportamiento del evaluador

A pesar de ser una entrevista grabada y un examen de lengua, el entrevistador debe hacerla de forma que se parezca a una conversación natural, cómoda, sin estrés y adaptada a las necesidades, temas y gustos del candidato y no a los temas que el "profesor" desee tratar y que quizás el candidato no sepa manejar.

Hay que evitar el nerviosismo típico de estos casos o el que el saber o no saber respecto a un tema, o la suerte, sean factores que influyan en el resultado de la entrevista o en la evaluación final. Para ello, se debe evitar hacer de "profesor". Hay que entrevistar al candidato con normalidad, evitando actuaciones típicas tales como corregir, enseñar, terminar frases, ayudar con léxico, cortar la conversación, expresar sus opiniones, atosigar a preguntas, variar los temas abruptamente o hacer juicios de la lengua durante la entrevista.

En cambio, sí se deberá hacer preguntas con un lenguaje normal, preocuparse por las respuestas y hacer comentarios breves, interesarse y pedir clarificaciones sobre lo que está diciendo el candidato (siempre que el nivel lo permita), esperar y ser paciente y aceptar el silencio como parte de la conversación (pues mediante él el candidato puede estar pensando la respuesta o preparando sus ideas). Conseguir, en fin, que sea una conversación y que el candidato olvide que hay un profesor que le juzga o controla. No obstante, no olvidemos que esta conversación tiene el objetivo de demostrar ciertas habilidades. Así que habrá que intentar que estas surjan a través de los diversos.

Mientras se pregunta y se evalúan las respuestas, pero sin desviarse de la conversación natural, hay que mantener el control de la entrevista sin permitir que el candidato haga lo que quiera o hable de lo que mejor sabe, pues si eso ocurre no lograremos una muestra con todos los elementos necesarios para la evaluación. Hay que motivar al tímido, tranquilizar al nervioso, centrar al evasivo y parar al controlador. Para todo lo cual hay un protocolo de actuación, una estructura y pasos a seguir, así como unos modelos de preguntas y formas de tratar los temas, que ayudan a lograr esa muestra de la mejor lengua posible del candidato, por encima de su personalidad e inteligencia, de su timidez o de la suerte. Vamos a exponerlo.

Elicitar una conversación natural determina una serie de condiciones sobre los tipos de interacción que se consideran apropiadas para una OPI (Swender, 1989: 5-1). Una de

las principales es, sin duda, considerar como lo más importante la transmisión del sentido y, consecuentemente, hacer que el entrevistador se comporte de forma que el candidato vea que el sentido de lo que dice se está teniendo en cuenta.

El entrevistador debe mostrar interés por el discurso del candidato mediante signos verbales y no verbales, como el uso de expresiones faciales, gestos, movimiento de los ojos, expresiones de sorpresa, acuerdo o desacuerdo y pequeños comentarios sobre lo que dice, a pesar de las dificultades que el candidato tenga para elaborarlo. Ya que de lo contrario, una actitud antinatural puede lastrar la motivación del candidato rebajando su nivel de compromiso. Interesarse por el mensaje significa interesarse por más elementos que la mera evaluación. A veces, este interés se muestra mediante referencias a algo que se ha dicho anteriormente, preguntando por algo para lograr una clarificación, reuniendo determinados temas en una cadena o preguntando más sobre determinados aspectos de un tema.

No obstante, no hay que confundir este reconocimiento del mensaje del candidato con el juicio respecto a su nivel de conceptualización, ni valorar su contenido como una característica de nivel. Las OPI no son una evaluación basada en el contenido, no se evalúan puntos concretos de información de ningún tipo.

En ACTFL, de acuerdo con numerosas concepciones pragmáticas y sociolingüísticas de la comunicación, se asume que el sentido en el lenguaje natural no se crea desde el principio sino que se establece mediante una negociación entre los hablantes que cooperan para lograr el sentido de sus enunciados e intervenciones. Durante las entrevistas, el evaluador debe situarse ante este comportamiento cooperativo esforzándose por crear y transmitir el sentido: mostrar brevemente su postura respecto a un tema (“A mi también me gusta mucho eso”), gestionando la negociación cuando la comunicación no resulte clara, por ejemplo, indicando que no ha comprendido del todo y pidiendo una aclaración o repetición (“¿Qué quieres decir con eso?”) o verificando que lo que ha entendido es lo que se quiere decir (“¿Quieres decir que eso ya lo sabías?”). Usando estas técnicas se puede incrementar la voluntad del candidato para que sea más abierto y participativo.

Muchas de estas técnicas son distintas según el nivel. Por ejemplo, no es apropiado pedir clarificaciones y explicaciones a candidatos en niveles inicial o intermedio, ya que eso equivale a complicar más las preguntas; en avanzados y superiores, sin embargo, esta técnica lleva a probar la habilidad de explicarse, describir o defender una opinión y lleva el discurso a niveles de conceptualización y expresión más elevados.

Es muy difícil evitar que una entrevista de evaluación pierda su formato en el que el evaluador posee una cuota de poder mucho más grande que el candidato; no obstante, hay que intentar manejar las intervenciones y los turnos de forma que, en la medida de lo posible, la jerarquía decaiga o no sea un impedimento para la conversación. Si, una vez iniciado un tema, se permite que el candidato lo explique sin interrumpirlo; y, si cuando el candidato da muestras de estar dispuesto a ser participativo se le permite que lo sea, se logrará que la conversación pierda gran parte de su artificialidad y ganará en efectividad.

El silencio es otro elemento importante que hace que la conversación sea natural. No significa necesariamente que el candidato tenga déficit de expresión. Por parte del evaluador, usar ciertos silencios estratégicamente, en vez de aprovechar cada pausa del candidato para introducir una nueva pregunta, ayuda al candidato a incrementar su grado de participación y a continuar hablando libremente. El silencio del candidato puede ser muy productivo: le permite pensar en profundidad el alcance de la pregunta que se le haya formulado; ayuda a planear una estrategia para realizar una tarea completa bien organizada; y sirve para reflexionar sobre algo ya dicho que debe ser aclarado o sobre la forma de introducir un nuevo tema relacionado con lo que se está tratando. Es, en definitiva, muestra de que el hablante está dispuesto a expresar ciertas cosas de la mejor forma posible, implicando aspectos cognitivos, gramaticales, pragmáticos, léxicos, etc. Claro que también puede ser señal de limitaciones; pero en tal caso se acompaña de otros elementos que permiten identificarlo: gestos, expresión de la cara, dudas, autocorrecciones.

Se ha señalado más arriba, en este mismo punto, que la OPI no es un test basado en el contenido. Efectivamente. Y eso hace que cada OPI tenga un contenido diferente, ya que cada hablante y cada evaluador entablarán su propia conversación. Lo que no es diferente en cada OPI es su sistema de evaluación, ya que en cada una se realizan las mismas pruebas, ya explicadas, sobre la existencia de la habilidad de uso de la lengua, sea a través de los temas que sea. En este sentido, lo que sí hay establecido son unos tipos de tipos de preguntas -no una lista de preguntas- y una forma de trabajar para desarrollarla a lo largo de los diferentes temas. Los evaluadores con experiencia entienden las OPI como una serie de temas desarrollados a lo largo de los diferentes niveles. Para ello se utiliza la técnica de realizar preguntas en una espiral de dificultad creciente, manteniendo el mismo tema hasta que se haya obtenido material suficiente.

Un ejemplo muy claro, puede ser el de tratar el tema de las comidas rápidas, por ejemplo, comenzando por el nivel inicial, pidiendo al candidato una lista de sus comidas preferidas; en el intermedio se puede preguntar al candidato sobre los restaurantes locales que sirven mejor esos tipos de comida; en avanzado, pasar a que se describan y comparen con cierto detalle las diferencias entre esas comidas en diferentes restaurantes, y en el superior llegar a discutir la conveniencia de alimentarse a base de esas comidas o de hacerse vegetariano y hacer hipótesis sobre las transformaciones que se producirían en la vida o en la sociedad si la gente cambiara de hábitos alimenticios.

Como se ve, la información aparecida en las respuestas de un nivel se lleva, mediante otras preguntas más elaboradas y difíciles, a otro nivel. Los cambios de tema deben realizarse con más cuidado y siempre empezando dentro de un nivel en el que el candidato muestre confianza en sus características de expresión; posteriormente se sube el nivel de las preguntas, continuando con el mismo tema, para probar las limitaciones de discurso. Y -ya se ha dicho, pero es momento de realzarlo significativamente- es necesario probar las habilidades del candidato a través de varios temas diferentes, para evitar que sea la familiaridad con alguno de ellos la causa de una aparente muestra de fortaleza en el uso de la lengua.

Para realizar todo esto conscientemente y guiar una buena conversación que logre obtener todos los elementos necesarios, ACTFL dispone de 9 tipos de preguntas, con cuyas técnicas, ventajas y desventajas conviene familiarizarse. Las preguntas directas siempre son una buena ayuda y se realizan a lo largo de toda la entrevista. Pero no hay que basarse solamente en ellas ya que la conversación se convertiría en un interrogatorio que no serviría para calificar la lengua. En realidad, hay que saber seleccionar en cada momento el tipo de pregunta adecuado a la situación y a la habilidad mostrada por el candidato.

Estas son esas 9 maneras de preguntar, explicadas con sus características, ventajas y desventajas:

1. Preguntas cerradas: son buenas para comenzar a dar confianza en la fase de calentamiento. De ellas se obtiene como respuesta un “sí/no”, o una mera enumeración de términos.

“¿Vives con una familia española?”

En los niveles iniciales reducen la ansiedad. Y en otros niveles ayudan a explorar nuevos temas.

Tienen en su contra que, usadas en gran cantidad, convierten la entrevista en un interrogatorio e impiden que el candidato tenga oportunidades para hablar y explicarse. Está bien comenzar por preguntas cerradas, pero una vez centrado el tema conviene cambiar rápidamente a otras abiertas

2. Preguntas alternativas: son las que ofrecen dos posibilidades de respuesta que el candidato debe elegir.

“¿Vives con una familia o con otros estudiantes?”

En niveles bajos ayudan al candidato a dar una respuesta fácil y a sentirse eficiente, ya que permiten que repita una de las dos cosas que el evaluador le ha ofrecido. También ayudan a comprobar si el candidato tiene cierto nivel de comprensión. En los niveles altos permiten delimitar un nuevo tema o llamar la atención del candidato sobre determinados aspectos de lo que se esté tratando para que continúe explicándolos más sustancialmente. Las disyuntivas planteadas suelen ser más complejas.

En su contra, pueden producir respuestas breves, sin información útil para la evaluación de las habilidades, por lo que se debe pasar a otras más abiertas, relacionadas con el tema.

3. Preguntas sobre hechos o informaciones concretas: son aquellas formadas por quién, dónde, cuándo, etc. (Es mejor cambiarlas por otras más abiertas, tipo cómo o por qué).

Ayudan a empezar a hablar a candidatos tímidos de niveles medios; centran las respuestas en los tiempos y hechos deseados por el entrevistador y permiten al candidato comenzar a hablar. No obstante, en niveles bajos no son muy provocadoras de largas respuestas y obligan al evaluador a continuar preguntando para obtener más información.

En contra, y si se repiten mucho, pueden llevar la entrevista al formato de interrogatorio y no producen discurso libre del candidato. Es aconsejable -como con los dos anteriores tipos de preguntas- pasar rápidamente a otro tipo más abierto.

4. Preguntas de entonación: una pregunta de este tipo es la que convierte un enunciado en pregunta mediante una adecuada entonación interrogativa.

“¿No fue divertido vivir con ellos?”

“Esa experiencia, ¿fue muy diferente de lo que esperabas?”

Este tipo de preguntas se usa en las conversaciones naturales indicando sorpresa, desacuerdo o acuerdo. Aunque no se trata de una pregunta directa, sin embargo produce al candidato la necesidad de aportar más información sobre la aserción central y es un modo muy productivo de elicitar discurso. Evita que la entrevista parezca un interrogatorio, se hace más conversacional y exige del candidato elaborar respuestas más explicativas sobre el tema que estaba respondiendo.

En contra, para que sean efectivas, hay que elegir el momento apropiadamente, expresando sorpresa o incredulidad para que la acción resulte creíble y apropiada pragmáticamente. Y siempre tiene el peligro de llevar a una simple respuesta cerrada tipo “sí/no”.

5. Enunciados terminados en “¿no?”, “¿no es verdad?”, o “creo yo”: añadiendo a un enunciado producido por el candidato un final que fuerce más conversación o le obligue a tomar una posición respecto a lo dicho.

“Pero hizo frío, ¿no?”

“Se podría esperar eso, ¿no?”

Sirven para conseguir un posicionamiento del hablante y, a menudo, llevan a explicar una opinión o defenderla frente a una potencial posición adversa del entrevistador. Por ello, son útiles en niveles altos.

En contra, en niveles bajos no son efectivas, pues los candidatos no tienen soltura para explicar en detalle las situaciones o manifestar su opinión; o llevan a una simple respuesta cerrada: “sí/no”.

6. Preguntas o demandas educadas: son preguntas directas y demandas claras que se convierten en pregunta debido a la entonación adecuada, el uso de marcas modales o de cortesía.

“Por favor, descríbemelo”

“¿Puedes decirme cómo es tu ciudad?”

“¿Podrías describirme el lugar?”

“¿Me puedes explicar esas ideas con un poco más de detalle?”

Evitan la apariencia de interrogatorio y dan un tono conversacional, llevando a respuestas amplias. Bien usadas en diferentes formas verbales, modos y tonos, pueden servir en niveles altos.

En contra, realizadas con demasiada educación, pueden llevar a aumentar la relación de poder y jerarquía de toda entrevista; y, en niveles bajos, a reforzar la idea de una jerarquía no deseada.

7. Preguntas con una introducción: comienzan con una breve introducción sobre algo que se utiliza para marcar o centrar la situación y sobre lo que luego se formula la pregunta.

“Has dicho antes que... ¿me podrías decir algo más sobre ello?”

“Últimamente he leído que... ¿qué opinas de ellos?”

“Antes has mencionado que tu hermano... ¿puedes decirme algo más sobre eso?”

Son buenas para -refiriéndose a algo previo- explorarlo en más profundidad. En niveles bajos ayudan a hacer transiciones temáticas con tranquilidad. En niveles altos pueden servir para elevar el nivel de la lengua, establecer una controversia o aumentar la seriedad del tema. O permiten introducir un tema, centrándolo en algo especial para exigir un análisis más detallado.

En contra, si se hacen sin mucho orden o son demasiado largas, pueden ser malinterpretadas o confundir a los candidatos.

8. Preguntas hipotéticas: son las que tratan de hechos irreales o circunstancias que se proyectan especulativamente, tratando las causas y consecuencias de determinados hechos bajo ciertas condiciones. El hablante marca sus respuestas con señales de que lo que dice son creencias personales y no hechos sólidos. No se trata de hacer hipótesis simplemente mediante la morfología, sino de centrar el tema en una complicación que requiera un análisis especulativo. Normalmente, estas preguntas se traen desde otras anteriores. Por ejemplo, tras explicar un asunto, se pide una opinión sobre él, que se defiende tal opinión y, finalmente, que se hagan hipótesis al respecto.

“Si no hubieras venido a estudiar aquí, ¿cómo habría cambiado tu vida en tu futuro laboral?”

“Imaginemos que se prohíbe eso en tu país, ¿cómo cambiarían las costumbres de la gente y qué consecuencias tendría para la economía?”

Estas preguntas llevan la conversación del aquí y ahora a temas más abstractos necesarios para medir a los candidatos del nivel superior o avanzado alto.

Recalamos que la mera utilización de la morfología adecuada para hacer hipótesis no muestra una habilidad funcional, ya que esta se basa en el cumplimiento de tareas

globales, analizando los conceptos tratados de una forma especulativa y no solo en su corrección morfosintáctica.

9. Preguntas reformuladas: sin ser un tipo de pregunta en sí, la necesidad de reformular preguntas también es parte del sistema de elicitación del discurso.

Pregunta: “¿Puedes describirme tus actividades diarias?”

Replanteamiento de la pregunta: “¿Qué haces cada día?”

Son necesarias para asegurarse de la comprensión en cualquier nivel, aunque son más frecuentes en niveles bajos. No obstante, no debe hacerse salvo que sea necesario, porque el candidato no haya entendido bien una pregunta.

Los aspectos negativos de esta pregunta son que si se hacen repetidamente se pierde la visión de la necesaria amplitud del discurso, ya que promueven su simplificación. Igualmente, si se hacen muchas variaciones de una misma pregunta, los candidatos de bajo nivel pueden perderse intentando averiguar los matices diferenciales entre una y otra y no saber qué se les está preguntando.

No saber usar correctamente estos tipos de preguntas, de acuerdo con el momento y nivel adecuado, puede llevar a provocar en el candidato estrés, lastrar sus posibilidades y poner en peligro el resultado de la entrevista.

A continuación mostramos tres ejemplos de un uso erróneo de preguntas y la solución que se debería dar en cada caso¹¹⁷. En ellos se aprecia la creación de estrés debido a una inadecuada progresión o selección de las preguntas, lo que lleva a la ruptura del lenguaje (las malas preguntas se dan subrayadas).

Ejemplos de un mal comportamiento elicitor del evaluador, con preguntas que crean estrés

1.

Pregunta: *¿Puedes hablarme de tu ciudad?*

Respuesta: *Mi ciudad es pequeña, no es tan interesante como Chicago porque en Chicago hay más cosas y en mi ciudad es muy aburrida.*

¹¹⁷ Estos ejemplos son de elaboración propia.

Pregunta: Vale, tú ciudad no es como Chicago, pero ¿puedes contarme algo más?

Error:

No saber replantear adecuadamente la nueva pregunta. El estudiante no sabe qué decir porque no tiene claro qué espera el entrevistador mediante esa pregunta, o porque, al intentar responder, pretende hacer una comparación entre su ciudad y otra más interesante y cercana a la suya, Chicago, en este caso. Lo cual es mucho más difícil de lo que el entrevistador quería.

Solución:

La solución está en intentar que el estudiante se centre en el tema de su ciudad y evite hacer una comparación con otra. Por lo tanto, habrá que hacerle ver que la pregunta no es tan compleja y que es suficiente con que responda a cómo es la ciudad, sin intentar discursos mucho más interesantes pero difíciles de estructurar.

Para rebajar el nivel de ansiedad del candidato se debe centrar la pregunta claramente. Podemos utilizar una pregunta cerrada. Algo como: “Vale, tú no vives en Chicago. ¿Cómo se llama donde tú vives?” Y después, cuando se centre en su ciudad: “Cuéntame un poco, ¿cómo es?”

O también así: “¿Te gusta tu ciudad?”

Respuesta: “Sí/no”

Pregunta: “¿Por qué?”

2.

Pregunta: “¿Puedes hablarme de tus planes?”

Respuesta: “Eh, bueno, nunca he pensado en eso, la vida es muy..., yo soy estudiante...”.

Pregunta: “¿De verdad no tienes planes? ¡Cuéntame un poco más!”

Error:

No entender que el estudiante ha interpretado como demasiado abstracta la pregunta.

Solución:

No hay que permitirle que se agobie intentando pensar en su vida, lo cual le puede parecer demasiado genérico, filosófico o serio. En su lugar hay que centrar la pregunta bajando el nivel a algo concreto y cercano a su vida real, con una pregunta cerrada.

Algo así:

“Vale, la vida es muy larga y hay muchas posibilidades, naturalmente. Pero, ¿tienes planes para las próximas vacaciones o los próximos meses?”

Respuesta: “Sí”

Y luego: “¿Puedes contármelos?”

3.

Pregunta: “¿Y la gente, cómo es diferente en Detroit o en otro estado?”

Respuesta: “En el norte de Detroit, la gente, lo mismo. Pero en el sur hay, no lo mismo, son vida más menos comparativa”.

Pregunta: “¿Puedes explicarme por qué?”

Respuesta: “En el norte hay mucho comercio y industria. En el sur no. Hay playas y más campo”.

Error:

Si a una persona de nivel intermedio, o que está empezando a responder a las preguntas del entrevistador, se le piden clarificaciones o explicaciones de algo que no se entiende bien pero que es interesante, probablemente se le esté pidiendo más de lo que pueda dar en este momento, o se le esté dando a entender que no se ha expresado lo suficientemente bien.

Si intenta reexplicarse y reorganizar su mensaje para hacerse entender mejor, probablemente -si no tiene más nivel- no lo logrará y será peor su discurso. La intriga e interés natural de un hablante no es buena guía de preguntas para un entrevistador

Solución:

Hacer las preguntas que correspondan en cada momento y adecuadas al nivel. Como en el ejemplo 1, centrando la pregunta en su ciudad, y luego pedir otras cosas al respecto.

El manual de ACTFL cita, además, otros elementos negativos que se deben evitar en el comportamiento del evaluador:

- Hablar demasiado despacio o de forma antinatural. Salvo en el caso de candidatos de nivel muy bajo.

- Preocuparse excesivamente por la forma del discurso del candidato y corregirle o hacer juicios de valor evaluativo.

- Corregir respuestas. Esta actitud que es de buena ayuda en el aula, no es bienvenida en la OPI, porque no es momento de enseñar y hacerlo puede interrumpir la comunicación. Solamente cuando la comunicación no es efectiva se pueden pedir aclaraciones, y esto solamente si lo permite el nivel de la fase de la entrevista y del candidato, algo muy diferente a corregirle o terminar sus enunciados.

- Repetir las respuestas resumidamente o haciéndose eco de sus ideas y terminar las frases o el vocabulario. Esto es claramente contraproducente para el candidato, pues no ayuda a que se esfuerce en la construcción de su discurso.

- Meter prisa en responder. En la OPI debe quedar claro que no disponer de suficiente tiempo de respuesta dificulta uno de sus objetivos, el de lograr muestras del mejor discurso posible de los candidatos, tanto cognitiva como lingüísticamente. La prisa impide que se obtengan buenas evidencias de las características positivas del discurso o muestras de sus limitaciones. El silencio, por tanto, es aceptable.

- Hacer preguntas sueltas y sin relación entre ellas o preguntas falsas sobre las que ya se sepa la respuesta. Esto es contraproducente respecto al nivel de compromiso que se requiere del candidato en sus intervenciones.

- Abusar de preguntas cerradas, pues impiden que el candidato profundice en el tratamiento de los temas.

- Expresar juicios sobre el discurso del candidato. Aunque puede usarse en contextos educativos, durante una entrevista para evaluación no es momento de dar pistas sobre su tipo de lenguaje. Esto puede llevar a los candidatos a asumir concepciones falsas sobre lo que se espera de ellos, lastrando las posibilidades de obtener la mejor muestra de sus posibilidades comunicativas.

Para terminar, diremos que el evaluador, a pesar de todas las técnicas que le ayudan a hacer la entrevista más natural, no debe perder el control de la entrevista, ya que es responsable de obtener todas las muestras necesarias para poder calificarla posteriormente.

2.6.2- 5) Las consideraciones necesarias para proceder a la calificación de las muestras obtenidas

Ya se ha dicho que la OPI es una evaluación con referencia a un criterio. Para evaluarlas deben considerarse de forma holística las características del lenguaje de los candidatos, de acuerdo con las características de las escalas dentro de cada nivel. La comparación de hablantes entre sí no es útil, ya que la evaluación asigna rangos de características y no evalúa elementos discretos. Por tanto, un hablante de un nivel puede caracterizarse por un discurso diferente al de otro hablante considerado dentro del mismo nivel, que cumple igualmente con las funciones, tareas y demás características del nivel, pero produce ejemplos de lengua diferentes en corrección, fluidez, cantidad de temas que trata, etc.

Para proceder a la calificación de las muestras de lengua se debe proceder a identificar modelos de uso de la lengua y de sus limitaciones, no elementos sueltos de fortaleza o debilidad, y sin dejarse llevar por otros elementos que sin duda se utilizan en el aula para caracterizar la lengua de los estudiantes. La calificación es global y los fallos de corrección no pueden servir por sí solos para calificar a un candidato como de nivel bajo. La escala describe detalladamente las características de cada nivel de forma global y es a ella y a todas sus características a la que hay que dirigirse para dar las calificaciones.

Se han hecho análisis del sistema de calificaciones para averiguar cuáles son, en efecto, los elementos que más pesan en los entrevistadores y evaluadores para realizar su calificación y, a pesar de las críticas que se hicieron a la OPI, insinuando que fundamentalmente se basaba en la corrección formal, las conclusiones muestran que lo preponderante¹¹⁸ es principalmente una suma de factores comunicativos.

Los niveles bajos recuerdan un tipo de entrevista de demostración de material aprendido en una clase de lengua, debido a las enormes limitaciones de comunicación que tienen todavía estos candidatos. No obstante, pueden darse frases con un nivel de corrección muy alto, por usar material aprendido de forma memorística. Ello no debe llevar a creer que el discurso del candidato posee esas características de corrección alta.

¹¹⁸ En §2.6.3 se tratan las críticas hechas a la OPI.

Los análisis hechos de los sistemas de evaluación con la OPI (Thompson, 1995) muestran que hay una gran consistencia entre evaluadores y que cuando hay discrepancias, suelen ser porque los segundos calificadores tienden a dar calificaciones más bajas que los primeros; que la fiabilidad es alta, con independencia de las lenguas que se estén evaluando o los niveles de los candidatos. No obstante parece que hay dos niveles más difíciles de evaluar, el avanzado alto y el intermedio alto. Asimismo las principales discrepancias surgen entre los subniveles altos y sus inmediatos siguientes entre intermedio alto y avanzado y entre avanzado alto y superior (Thompson, 1995: 411- 13).

Para proceder a calificar una grabación, es necesario que la muestra sea estándar y contenga todos los elementos necesarios estipulados. De lo contrario, debe volver a repetirse la entrevista. Una grabación puede no ser válida si presenta los siguientes fallos:

1. No se han obtenido suficientes muestras de la base de actuación oral del candidato ni de sus limitaciones de expresión.
2. No contiene las cuatro fases necesarias en toda OPI.
3. No se incluyen pruebas en suficiente número de temas o áreas de contenido.
4. La grabación es demasiado breve, sin dar oportunidades al candidato.

A continuación se acompaña un cuadro con los elementos básicos a repasar para asignar a cada candidato a un nivel; repaso de elementos que debe realizar cada evaluador, una vez que tiene una opinión, para poder dar la calificación definitiva.

Superior (S)	Participa en conversaciones formales e informales y profesionales, defiende sus opiniones y elabora hipótesis, mantiene un discurso extenso sin errores, se enfrenta a temas concretos y abstractos, maneja situaciones no habituales y habla con un alto nivel de corrección incluso en las estructuras menos frecuentes y sin molestar a un hablante nativo ni provocar malentendidos.
Avanzado alto (AH)	Realiza todos los requisitos del avanzado sin problemas y además se maneja como un hablante de nivel superior por lo menos un 50% del tiempo que se requiere.
Avanzado medio (AM)	Realiza todas las actividades del nivel avanzado manifestando riqueza y calidad de discurso y cantidad de información, así

	como una corrección alta en el manejo de tiempos y aspectos en la narración y detalle en las explicaciones.
Avanzado bajo (AL)	Es autosuficiente en las necesidades de las situaciones de trabajo y escuela, narra y describe en la mayor parte de tiempos y aspectos, mantiene su discurso en párrafos constantes, habla de una amplia variedad de temas de su trabajo y actualidad y es capaz de iniciar, mantener y concluir una conversación en la que participa como hablante activo respondiendo y guiando las intervenciones así como en conversaciones en las que se presenten complicaciones, sin necesidad de que sus interlocutores estén acostumbrados a tratar con hablantes extranjeros.
Intermedio alto (IH)	Realiza todos los requisitos del intermedio sin problemas y además se maneja como un hablante de nivel avanzado por lo menos un 50% del tiempo que se requiere.
Intermedio medio (IM)	Realiza todas las actividades del nivel intermedio manifestando riqueza y calidad de discurso y cantidad de información y vocabulario así como una corrección y variedad alta en el manejo de tiempos.
Intermedio bajo (IL)	Es creativo en el uso del lenguaje, combinando y recombina material aprendido, para producir todo tipo de información personal, habla y hace preguntas sobre una gran variedad de temas personales y familiares, utilizando frases sencillas, iniciando, manteniendo y terminando tareas sociales sencillas o interacciones que requieran intercambio de información. Se hace entender generalmente por hablantes acostumbrados a hacerlo con extranjeros y repitiendo la información cuando se necesita.
Inicial alto (NH)	Lleva a cabo las actividades del nivel intermedio aunque no consistentemente usando el lenguaje de forma creativa para hablar de temas personales más del 50 % del tiempo.
Inicial medio (NM)	Habla fundamentalmente usando fórmulas breves, frases rutinarias, listas y enumeraciones de elementos sobre temas altamente predecibles de su vida diaria. Se le entiende con dificultad y solo si se está acostumbrado a tratar con hablantes no nativos.
Inicial bajo (NL)	Contesta a preguntas sobre temas predecibles de su vida diaria con monosílabos, palabras sueltas memorizadas o repeticiones. Conoce y maneja sintagmas breves y listas de palabras.

(Swender, 1989: Apéndice: Rating Guide).

Para terminar este apartado, ofrecemos una traslación del modelo de test de la OPI de ACTFL según la forma de analizar tests y tareas de Bachman (1990), expuesta por Yofe (1997)¹¹⁹.

FACETS OF THE TEST RUBRIC	Time Allocation	15 to 35 minutes, depending on the level
FACETS OF THE INPUT FORMAT	Channel of presentation Mode of presentation Form of presentation Vehicle of presentation Language of presentation Identification of problem Degree of speededness	Aural Receptive Language (aural) Live L2 N/A No speed
NATURE OF LANGUAGE	Length <u>Propositional content</u> Vocabulary Degree of contextualization Distribution of new information Type of information Topic Genre <u>Organizational characteristics</u> Grammar Cohesion <u>Rhetorical organization</u> <u>Pragmatic organization</u> Illocutionary force Sociolinguistic characteristics	Varied, usually relatively short. Familiar Context embedded N/A Concrete <-> abstract Nominated by the testee Conversation Formal to semi-formal Yes Yes Functional Usually standard var. of the TL, semi-formal register
FACETS OF THE EXPECTED RESPONSE FORMAT	Channel Mode Type of response Form of response Language of response	Aural Productive Unrestricted Language - oral L2
NATURE OF LANGUAGE	Length <u>Propositional content</u> Vocabulary Degree of contextualization Distribution of new information Type of information Topic Genre <u>Organizational characteristics</u> Grammar	Varied, mostly relatively short sentences Familiar/general Mostly embedded N/A Concrete & abstract, depending on interviewee Varied Conversation Yes (from formal to informal)

¹¹⁹ Ello permite hacer una comparación rápida entre el test de la OPI y el nuevo test que hemos diseñado mediante el MCER para realizar esta tesis, que aparece en el capítulo 3.

	Cohesion <u>Rhetorical organization</u> <u>Pragmatic organization</u> Illocutionary force Sociolinguistic characteristics	registers) Yes Yes Functional Standard variety of the TL
RESTRICTIONS ON RESPONSE	Channel Format Organizational characteristics Propositional and illocutionary Time or length of response	Restricted Unrestricted Restricted Fairly unrestricted characteristics Fairly unrestricted
RELATIONSHIP BETWEEN INPUT AND RESPONSE		Reciprocal

(Yoffe, 1997: 9).

2.6.3 Críticas al modelo de entrevista oral de ACTFL

Desde el momento de su publicación surgieron voces críticas y escépticas o pesimistas que cuestionaron varios, si no todos, los elementos, criterios y conceptos del sistema de ACTFL, junto a adeptos que lo defendían o ensalzaban entusiastas por el impacto que les suponía en sus clases de lengua (Liskin-Gasparro, 2003: 484). Como consecuencia de estas críticas, se realizaron numerosas investigaciones que avalaban una u otra perspectiva y aconsejaban cambios según los resultados e intereses de las investigaciones y de los investigadores (Chalhoub-Deville / Fulcher, 2003). Todo ello lo han tenido en cuenta en ACTFL¹²⁰ en las sucesivas revisiones y análisis que se han hecho de las *guidelines* y los sistemas de evaluación (Lowe, 1998; Liskin-Gasparro, 2003; Swender, 2003, Surface / Dierdroff, 2003).

Algunas de las críticas se han tenido en cuenta, otras han obligado a la organización a dar nuevas explicaciones que aclararan los conceptos. Finalmente, los mismos investigadores han aceptado esas explicaciones o los resultados de investigaciones y han terminado cambiando de opinión o suavizando sus posturas. Un ejemplo de todo ello es

¹²⁰ En ACTFL, así como en los otros dos organismos encargados de administrar los diferentes tests de los que surgió la OPI, basados en las mismas orientaciones (el ETS y la ILR, que junto a ACTFL llamaremos en adelante AEI. Lowe (1998: 359), explica el acrónimo: la A se refiere a ACTFL, la E al *Educational Testing Service* y la I a la *Interagency Language Roundtable* del gobierno americano, respectivamente; que son las organizaciones (que en orden inverso), junto al FSI (Foreign Service Institute), han contribuido históricamente al desarrollo, elaboración y definición de la *foreign language proficiency*.

la consideración que en su momento hiciera Bachman respecto a la escala de AEI, aduciendo que se basaba en una norma (Bachman / Savignon, 1986: 385, Bachman, 1990: 308-350), aunque ACTFL dijera que era una escala basada en criterios de referencia, como el propio Bachman ha aceptado finalmente (Bachman, 2004: 301) (véase más adelante §2.6.3- E).

Las principales críticas que se le hicieron fueron respecto a sus fallos en la validez como sistema de evaluación (validez de autenticidad, de contenido, de predicción, de constructo), en su escala y sistema de calificaciones, en la definición de la competencia comunicativa, en su fiabilidad, en el tipo de lenguaje que elicitaba y su representatividad, en el sistema de elicitarlo y si este es una conversación auténtica o no, o, en fin, en las posibilidades de uso como herramienta de programación y asesoramiento del rendimiento académico (Kramsch, 1987: 355). Y continúan haciéndose tras más de treinta años de experiencia probada, a pesar del aumento de su influencia y uso en todos los ámbitos.

Esto es lógico, importante y necesario, porque, como acabamos de decir, es imprescindible estar en constante análisis y revisión, acomodándose a las necesidades de los usuarios, sin que tales críticas supongan una desacreditación del test o sus procedimientos, aunque sí hayan servido tanto para iniciar numerosos programas de investigación respecto a la naturaleza de la OPI y de la evaluación oral en general, como para centrar adecuadamente el valor y alcance real de este test, aclarando que la competencia que definen las *guidelines* no es tan absoluta y extensa como en principio se pensaba (Liskin-Gasparro, 2003: 488).

De todos estos trabajos críticos e investigaciones vamos a ocuparnos en adelante¹²¹.

2.6.3- A) Críticas sobre si la OPI es una conversación natural

Una de las primeras críticas que se realizó a la OPI fue respecto a si la entrevista es en realidad una conversación natural o no y si el formato y estructura de preguntas utilizado permite elicitar suficiente cantidad de elementos interactivos sobre el lenguaje de los candidatos. En ella se basaban, además, para hacer otra serie de críticas

¹²¹ Ello nos va a servir para analizar más profundamente el sistema de evaluar de ACTFL, además de para ampliar la perspectiva sobre la problemática respecto a la evaluación oral en general.

relacionadas y restarle validez de apariencia y de constructo y, con ello, de todo lo demás. La justificación de estas críticas se debía a que en las propias directrices de ACTFL se decía que la entrevista debía asemejarse lo más posible a una conversación. En §2.4 se ha explicado el problema de la conversación en las entrevistas de evaluación oral en general, LPI¹²². Ahora solo vamos a resumir algunos de los trabajos que se han centrado en la crítica o defensa de las entrevistas de ACTFL en particular.

La disputa sobre si es conversación o no, con los consiguientes análisis a que se sometió la OPI, ha quedado saldada a su favor: primero, porque hay suficientes trabajos que dan a la OPI una mezcla de elementos de conversación y de entrevista (Ross, 1992; Ross / Berwick, 1992); segundo, porque la OPI no pretende ser una conversación sino tener características de esta, ya que siéndolo del todo se impide la validez y consistencia de sus calificaciones (Brown, 2003: 2); y, en tercer lugar, porque los últimos estudios sobre el proceso de interacción de una entrevista oral¹²³ muestran que no sólo en la OPI, sino en cualquier otro modelo de entrevista oral se producen muchos elementos conversacionales, pero que son de difícil control y que se debe hacer un gran esfuerzo en la formación de los evaluadores, en el diseño de las escalas y en la estructura de las entrevistas¹²⁴ para impedir que su influencia reste validez al proceso¹²⁵.

Los primeros comentarios respecto a la OPI (Clark / Clifford, 1988; Raffaldini, 1988) comentaban que la escala y tipo de test era más útil que las escalas basadas en puntos discretos, pero que había diferencias con lo que es una conversación. Las diferencias las señalaban en la diferencia de jerarquías entre los participantes, en el lugar donde se producía el diálogo, en el estatus dado entre participantes, en las identidades o roles de los interlocutores, etc. (Clark / Clifford, 1988: 141-142) o en que la entrevista refleja solamente en parte las características de una conversación y no representa lo que es una conversación natural (Raffaldini, 1988: 201).

Para responder a estas críticas habría que plantearse si ACTFL deseaba que la OPI fuera una conversación o si sus intenciones habían sido malinterpretadas, así como si las características de una conversación son deseables para una buena entrevista oral para examen de lenguas. En realidad, ACTFL no establece que la OPI sea una conversación, sino algo parecido a una conversación, porque en ella los temas y la libertad de su

¹²² Allí se ve cómo estas críticas han ido perdiendo o cambiando su sentido.

¹²³ Ya comentados en §2.4.

¹²⁴ Estos aspectos se han visto en §2.4.3- 4.

¹²⁵ Esfuerzo que ACTFL hace desde el principio llevando en esto mucha ventaja a otras asociaciones y modelos de entrevistas existentes. Por ejemplo del MCER que carece de un sistema formal institucionalizado al respecto.

desarrollo son impredecibles. Simplemente dice que es aconsejable que la OPI tenga ciertas características de una conversación pero dentro de un marco estructurado en forma de entrevista¹²⁶. Brown (2003) critica el deseo de que las entrevistas sean una conversación ya que la variación natural entre hablantes nativos y la imprevisión de la interacción pueden parecer buenas y cercanas a lo que es una conversación natural y aumentar la validez aparente del test, pero al mismo tiempo pueden llevar a variar el constructo del test (Brown, 2003: 20) haciendo que pierda fiabilidad.

También Moder / Halleck (1998), en su defensa, dicen que la OPI no es el espejo de una conversación porque no es ese su cometido, y que es un tipo de entrevista altamente estructurada con el fin de obtener los datos que necesita (Moder / Halleck, 1998: 118). Defienden que el hecho de que la OPI no mida la competencia lingüística mediante una conversación informal no significa que no la mida. Incluso, su sistema puede ser más efectivo que el de una conversación informal: “el marco de la entrevista es un contexto mucho más apropiado para evaluar la lengua de lo que es una conversación” (Moder / Halleck, 1998: 122).

Se desee, o no, y se necesite, o no, que sea una conversación natural, varios han sido los trabajos dedicados a analizar específicamente si la OPI lo es. Ross (1992) y Ross / Berwick (1992) han analizado algunos aspectos de la interacción y otros elementos conversacionales tales como la propuesta, rechazo o reformulación de nuevos temas o las características de modificación del lenguaje y la acomodación al discurso de los no nativos (peticiones de clarificación, despliegue de preguntas y simplificaciones) que se dan entre nativos y no nativos. Concluyen que la OPI comparte elementos de conversación natural con otros propios de las entrevistas (Ross 1992: 173) y usa los mismos sistemas de acomodación que en las conversaciones entre nativos y no nativos (Ross / Berwick, 1992: 170)¹²⁷.

2.6.3- B) Críticas sobre si la OPI es un test de competencia general o gramatical

Otras críticas se basaron en un mal entendimiento de lo que las *guidelines* decían, en una lectura parcial o errónea de sus presupuestos o en análisis de resultados obtenidos

¹²⁶ Esto ya se ha señalado en el apartado referente a la evaluación de ACTFL, §2.6.

¹²⁷ Remitimos a §2.4 donde se analizan los elementos interactivos y de conversación de las entrevistas en general, LPI y OPI, para más datos al respecto y para ver cómo este problema ha quedado superado.

con entrevistas no bien realizadas ni por personas formadas para ello. Así, se pretendió que, a pesar del esfuerzo puesto en enfatizar el uso de la lengua real en el aula y en los exámenes, la OPI medía principalmente gramática y corrección formal. Y los críticos incurrieron en contradicciones, pues mientras unos achacaban que el sistema era de puntos discretos, otros lo invalidaban por no estar basado en esos mismos puntos discretos.

Savignon juzgaba a la OPI como un sistema demasiado gramatical y centrado en la corrección, viendo en ella “un renovado énfasis en puntos discretos de la gramática al nivel de la oración o la corrección” (Savignon, 1985: 130) y se llegaba a preguntar:

¿Por qué, entonces, en una campaña para estimular el uso espontáneo de la lengua en las clases de lengua americanas, una campaña que se dice que se ha inspirado en la tradición funcionalista, se pone tanto énfasis en puntos discretos de gramática y corrección? La respuesta está para mí en el origen de la escala de calificación del FSI en el que las “guidelines” de ACTFL están basadas (Savignon, 1985: 131).

Y, junto a Bachman, pretendía que el “énfasis (en la gramática y corrección) está obviamente incluido en el proceso de la entrevista de manera que asegura que el examinador busque o intente elicitarse estructuras gramaticales específicas” (Bachman / Savignon, 1986: 387).

Las respuestas y trabajos que desmentían estos supuestos no se hicieron esperar. Primero haciendo ver cómo el nuevo modelo de competencia comunicativa calaba entre los profesionales, dando aire nuevo a la enseñanza, frente al uso demasiado extendido de programaciones basadas en puntos discretos de gramática:

La mayoría (de profesores asistentes a los talleres de formación) estaba de acuerdo en que les había clarificado sus ideas respecto a la enseñanza y al aprendizaje. Les aportaba unos objetivos claros, facilitaba el diseño de materiales seleccionados con relación a las escalas y mostraba cómo sus estudiantes progresaban de un nivel a otro (Liskin-Gasparro, 1987: 212).

Magnan (1988) estudia en qué medida la OPI se centra en la gramática, partiendo del concepto mismo de gramática y analizando siete categorías gramaticales (conjugación verbal, tiempos, determinantes, adjetivos, preposiciones, pronombres y relativos) en 40 entrevistas de OPI realizadas en francés a las que se analizaron los errores en esas siete categorías. Su conclusión fue que no hay datos de que con la OPI haya excesivo énfasis en la gramática, en la forma de corregir y establecer niveles y, por lo tanto, esta no puede considerarse un test de gramática. Desmiente los comentarios respecto a que el aprendizaje de la gramática implícito en la OPI sea lineal, Argumenta

que la corrección mostrada en el uso de tiempos verbales a lo largo de los diferentes niveles no tiene mucha incidencia en las calificaciones finales obtenidas, y que tampoco otros elementos, como las funciones o la corrección formal, presentan esa linealidad gramatical a lo largo de los niveles, sino que todos están interrelacionados y esa interrelación ha sido malentendida.

Sí halló una relación entre los errores cometidos y las calificaciones obtenidas, pero sin que ello supusiera que la corrección lingüística formal fuera el principal elemento del sistema de calificaciones sino que, muy lógicamente, conforme aumenta la competencia en general, decrecen los errores. Pero, incluso en este aumento de la competencia no encontró linealidad a lo largo de los niveles; al contrario, lo que encontró fue que en el nivel inicial el discurso es más correcto que en otros niveles, debido al uso memorístico del lenguaje; y que luego comienza una curva de errores que aumentan hasta el nivel intermedio alto y, a partir del avanzado, desciende la curva de errores porque aumenta el nivel del discurso considerablemente.

También comenta que puede que la causa de los malentendidos sea el concepto de gramática manejado, achacándolo al “tradicional y ortopédico concepto de uso correcto de la morfología y de la corrección formal frente a otro concepto más moderno que se fija en la propiedad del discurso y otros elementos comunicativos, pragmáticos o sociolingüísticos”. Y termina diciendo que es básico “entender la noción más compleja de qué es la gramática, no solo para la evaluación sino para cualquier propuesta de revisión curricular que intente mejorar la competencia de nuestros estudiantes” (Magnan, 1988: 267).

Para Magnan el desarrollo de la corrección lingüística depende de diferentes factores tales como la orientación de las metodologías utilizadas en el aula, el nivel de los estudiantes, el riesgo que asuman al desarrollar las tareas o el concepto del error que tenga el profesor. Y en la OPI, una persona con una calificación alta no tiene necesariamente más corrección en cada elemento de la lengua que otra en un nivel menor.

Por dar un ejemplo extremo, en su estudio halló que se pueden tener los mismos porcentajes de errores en el uso de adjetivos y preposiciones en el nivel avanzado medio que en el intermedio medio (Magnan, 1988: 271); a pesar de esos errores comunes, el avanzado logra usar la lengua en situaciones más complejas y con sentidos más elaborados. Esto es porque las bandas de errores sobrepasan los niveles de las escalas

definidas en los estándares de ACTFL, lo que lleva, definitivamente, a no poder considerarse a su test, la OPI, como un test de gramática.

A este respecto Lowe (1998: 372) aclara que los protocolos de ACTFL no se basan en la corrección gramatical sino en un concepto general de la corrección lingüística, porque sirven tanto para quien tenga una base académica como para quienes aprendan en contextos de calle donde se da más importancia a otros factores. Y que, por ejemplo, un candidato calificado como intermedio alto puede tener un buen control morfológico de los verbos en pasado mientras que otro candidato puede no tenerlo, pero que ambos caben en el mismo nivel.

También se han analizado los criterios de que se sirven los correctores y evaluadores de los tests de ACTFL para ver si estos son puntos gramaticales o no. Halleck (1992, 1995) dice que estos son principalmente comunicativos y que “la mayor parte de las razones para calificar (a un candidato en un nivel u otro) se sacan directamente de las *guidelines* con sus propias palabras, como “no se aprecia que aparezca discurso organizado”, “es capaz de preguntar y responder preguntas”, o se parafrasean”. Además, la mayoría de los evaluadores dio razones relacionadas con estrategias comunicativas (169 veces), frente a una pequeña parte de razones basadas en puntos gramaticales (solo 11 veces). Los evaluadores estaban más interesados en dar razones que tuvieran que ver con la lengua que en resaltar errores gramaticales, (Halleck, 1992: 230).

En otro trabajo posterior llega a la misma conclusión y dice que a la hora de hacer los juicios mediante la OPI, los factores comunicativos son más importantes que la gramática; y que la corrección formal no es un factor determinante a la hora de determinar el nivel (Halleck, 1995: 230).

2.6.3- C) Críticas sobre su valor como asesoramiento didáctico

También se puso en duda el uso de la OPI en ámbitos académicos o como herramienta de enseñanza (Kramsch, 1997: 355), que ayudara a los profesores de lengua a organizar y preparar sus clases, que sirviera para realizar programaciones; o que fuera útil como elemento de asesoramiento y evaluación educativa. Savignon comentaba “la esquizoide visión del uso de la lengua promovida por los estándares de ACTFL” (1985: 130) y le negaba cualquier posibilidad para desarrollar sílabus o diseños curriculares.

Por el contrario, los comentarios de los participantes en los cursos de formación se mostraban entusiastas al respecto:

Tras los talleres de formación sobre las directrices de ACTFL los participantes se declaraban muy satisfechos respecto a las posibilidades de su uso para planificar cursos de todos los niveles. Señalaban que su concepción de la lengua como una suma de elementos funcionales, contextuales, temáticos y de gramática a lo largo de sus escalas graduadas en niveles que cubrían la totalidad de la producción de la lengua les era muy útil para llevar a sus estudiantes a lo largo de una serie de pasos a través de funciones, temas y situaciones que requerían diferentes usos de la lengua con diferente nivel de maestría (Liskin-Gasparro, 1987: 212).

En realidad, su utilidad era evidente allí donde se pusiera en práctica y el uso de evaluaciones orales enfocadas a la competencia se convirtió rápidamente en un elemento indispensable de las programaciones, a veces adaptando los contenidos de los cursos, otras adaptándose las entrevistas para encajar en los cursos y siendo la guía para numerosos programas (Liskin-Gasparro, 1987: 213). Hasta en aspectos interculturales como el modo de enseñar contrastes culturales para que ayuden a los estudiantes a ver las diferencias entre las lenguas y culturas (Rings, 2006: 43).

Meredith (1990) realizó una investigación para comprobar si el test servía como test de fin de curso o si existía una relación entre las calificaciones de una OPI y los años previos de estudio de la lengua y concluyó que, aunque es impreciso como instrumento de medida del aprendizaje de un curso (Meredith, 1990: 295), efectivamente los niveles representados en la pirámide invertida de ACTFL representan niveles reales de incremento de la habilidad. Además, se vio que el uso de los estándares de ACTFL en el aula incrementa la motivación de los estudiantes para desarrollar sus capacidades de conversación (Meredith, 1990: 293).

La crítica de que los estándares de ACTFL no eran válidos para uso académico, programación de cursos y evaluación del rendimiento ha quedado totalmente descartada con el paso del tiempo y el uso cada vez más extendido de estos estándares de ACTFL y la OPI en todos los ámbitos educativos e instituciones académicas de EE.UU., tanto en niveles K-6 (enseñanza infantil), K-12 (enseñanza secundaria), como universitarios. Nos parece que esta crítica de su inadecuación como instrumento pedagógico se debía a las diferencias entre los objetivos académicos tradicionales y los objetivos que ACTFL propugnaba, que se han ido abriendo paso día a día y han triunfado finalmente.

En 1988 ACTFL entró a formar parte del NCATE (*National Council for Accreditation of Teacher Education*). Este organismo presta servicio como organismo acreditado en unos 623 programas de formación de profesores en 50 estados, además del distrito de Columbia y Puerto Rico (McAlpine / Dhonau, 2007: 248). Y en 2002, ante la necesidad de acordar un currículo de enseñanza que sirviera para todos los estados a nivel nacional y que pusiera fin a la larga cadena de fracasos en la enseñanza de lenguas, la cual tras 12 años de aprendizaje solo era capaz de situar a los adolescentes en el nivel intermedio bajo, el NCATE pidió a ACTFL que propusiera estándares de lenguas extranjeras para incluir entre los criterios de acreditación de los programas de formación de profesores (Shulz, 2002: 290). En 2002 ACTFL presentó sus protocolos de preparación y formación de profesores de enseñanza secundaria (K-12) que, tras ser discutidos, recibieron su versión final en octubre de 2002 como *Program Standards for the Preparation of Foreign Language Teachers* (McAlpine / Dhonau, 2007: 248).

Consecuencia de esta nueva presencia de ACTFL en NCATE y su participación activa en el asesoramiento y formación del profesorado de lenguas extranjeras, comenzaron a producirse importantes cambios, tanto para los candidatos a profesor, como en las instituciones que los preparan; y en los programas de profesores se ha impuesto como requisito la obtención del nivel avanzado bajo de la OPI para obtener el certificado nacional de profesor de lenguas (Pearson *et ál.*, 2006: 508).

2.6.3- D) Críticas sobre la validez y fiabilidad de la OPI

De todas las críticas, las principales y más elaboradas fueron respecto a la falta de consistencia en los sistemas de evaluación y las consiguientes falta de fiabilidad y falta de validez en numerosas áreas del proceso. Estas críticas se centraron en problemas relacionados con la escala; la definición implícita en ella respecto a las habilidades; la competencia comunicativa y los sistemas de calificar que de ella se derivan; el comportamiento del evaluador; y el sistema de administración del test.

Vamos a verlas organizadas desde dos puntos de vista, uno cualitativo y otro cuantitativo.

Desde el punto de vista cualitativo los críticos principales han usado siempre los mismos argumentos, centrados en la falta de un buen concepto de competencia lingüística (Savignon, 1985; Bachman / Savignon; 1986, Kramsch, 1987; Bachman, 1988; Salaberry, 2000). Comenzaban cuestionando el concepto de competencia lingüística, porque esta se fundamenta mediante la validez aparente, y esto les llevaba a criticar falta de autenticidad¹²⁸. Luego, se desautorizaba su validez predictiva, ya que con estas faltas señaladas no se puede demostrar que el uso de la lengua del test exista también en la vida real y, consecuentemente, se llegaba a la falta de validez de constructo y a la falta de una escala basada en criterios.

Savignon hizo referencia a la necesidad de una definición de la competencia comunicativa (1985: 129), aspecto en el que se centraron las críticas posteriores de estos autores (Bachman / Savignon, 1986; Bachman, 1988; Bachman, 1990; Salaberry, 2000). Pero la crítica principal partía de Bachman (1988), quien decía que el hecho de que en ACTFL el test varíe de candidato en candidato puede influir en las notas, porque ello introduce un error de medida aleatorio y amplía la falta de validez; algo que la propia organización, ACTFL, intentaba reducir mediante los minuciosos procedimientos de formación de evaluadores que estandarizan su procedimiento. Pero, para Bachman, desgraciadamente, esos mismos procedimientos generaban un modelo de test que causaba el mismo efecto en todas las entrevistas y notas, similar para todos los candidatos, lo que constituía un sesgo que, así se cerraba el círculo, le restaba validez (Bachman, 1988:153; 1990: 223)¹²⁹.

La única manera de eliminar esos factores como fuente de error, continúa Bachman, es incluirlos en la definición de las habilidades a evaluar, con lo que el test se reduce a una medición limitada de elementos: únicamente aquellos incluidos en la definición de la habilidad. “No podemos decir entonces que es una medida de la competencia lingüística sino de determinadas habilidades. Y estaremos obligados a revelarles a nuestros candidatos tales limitaciones” (Bachman, 1988: 154).

También Bachman / Savignon (1986: 385) critican lo que llaman el problema de la circularidad, a lo que también alude Bachman (1988: 155). Esto es, que la escala define los niveles de acuerdo a lo que en ella se dice; y el test solo mide lo que dice su escala y

¹²⁸ En §2.3 se han explicado estos conceptos y sus relaciones.

¹²⁹ La explicación de validez y fiabilidad y sus interrelaciones, tanto según Bachman como según otros autores y organizaciones, se ha hecho en §2.3 Baste ahora resumir que la validez se refiere a la capacidad de definir qué mide el test y la fiabilidad a la propiedad de este de dar sistemáticamente los mismos resultados y no variarlos cada vez.

no algo que exista en la vida real. Para estos autores, todo esto lleva a una confusión entre la habilidad y el método de evaluarla (Bachman / Savignon, 1986: 385; Bachman, 1988: 155) restando validez a las calificaciones, porque el método se convierte en un factor de error (Bachman / Savignon, 1986: 384).

Por todas estas razones, estiman que es una forma limitada de medir las actividades y competencia y que es necesario definir de forma diferente la habilidad comunicativa y los criterios de calificación utilizados en la evaluación. Y aconsejan incluir en las escalas y en la definición de la habilidad otros elementos más abstractos relacionados con diferentes habilidades (Bachman, 1988: 159), puntos discretos (Bachman / Savignon, 1986: 385), así como elementos sociolingüísticos y discursivos (Bachman / Savignon, 1986: 388).

Bachman / Savignon opinan que el sistema de evaluar de ACTFL carece de una teoría sobre la competencia comunicativa (Bachman / Savignon, 1986: 388); y Bachman (1988: 155) opina que el concepto de habilidad unitaria de Oller está superado y que el sistema global de calificación de AEI no se basa en ninguna teoría lingüística. También critican que las especificaciones de ACTFL se basan en contextos y áreas temáticas concretas, lo cual es un problema porque para que la evaluación mida de forma clara un rango de habilidades en diferentes lenguas y contextos, la escala debería definir la competencia lingüística independientemente del lenguaje, contexto y habilidad. La OPI, al mezclar las habilidades con los contextos, no puede hacer una generalización de sus resultados a otras situaciones diferentes a las evaluadas en sus contextos y se convierte en un sistema limitado “solamente a los contextos y bajo las condiciones y situaciones incluidas en los procedimientos” (Bachman / Savignon, 1986: 384). El tipo de lenguaje que con ello se logra elicitar se define solamente en términos de habilidad práctica pero no de competencia.

Raffaldini critica, con similares argumentos, el concepto de competencia implícito en la OPI, diciendo que falla en medir importantes elementos de la habilidad comunicativa, sobre todo elementos sociolingüísticos e interactivos (por ejemplo, cómo influir en el discurso del otro) y que es un sistema parcial de medir la habilidad comunicativa. Le parece un sistema limitado y centrado en solo dos cosas, intercambio de información y opiniones (Raffaldini, 1988: 201).

En definitiva se criticaba su concepto de competencia junto al hecho de que este surgiera sin bases teóricas aparentes ni validaciones empíricas (Valdman, 1988: 121).

Los principales elementos críticos que acabamos de señalar venían paralelos al nuevo modelo de definición de la competencia comunicativa propuesto por Bachman (1986, 1988, 1990), diferente al implícito en las *proficiency guidelines* de ACTFL. Bachman, y el resto de críticos que le siguen, basándose en su definición de competencia comunicativa (que llaman CLP, *Communicative Language Proficiency*), ya explicada en §2.2, invalidan gran parte del constructo de la OPI, así como otros aspectos como su autenticidad, su capacidad de predicción o el uso de su escala de evaluación; y con ello pretenden invalidar todo el proceso de calificación.

Algunos de sus consejos, críticas y comentarios, no obstante, se han introducido en la revisión de la escala de ACTFL de 1999. Fundamentalmente los referentes a la introducción de más elementos sociolingüísticos en las escalas, así como a la descripción más detallada de otros aspectos tanto gramaticales como pragmáticos y estratégicos, lo cual se ha hecho con éxito (Lowe, 1998: 376).

En su defensa, Lowe aclara que las descripciones de la escala se diseñaron para la ILR tras la segunda guerra mundial y que, desde su primera implementación a mediados de los 50 hasta la actualidad, han sido revisadas, estudiadas y especificadas cuidadosamente en varias ocasiones (Lowe, 1998: 359); que, a pesar de lo que pueda parecer a los legos, no se hicieron al azar sino bien pensadas, de acuerdo con un número de principios que ya se explicaron en Lowe (1986) que probablemente se han malinterpretado, principios que trató de volver a explicar en 1998.

Esos criterios seguidos para definir los niveles fueron:

a) Seguir pautas de parquedad, describiendo lo fundamental y más característico y no lo periférico, y dando esquemas claros que compusieran un marco del que extraer lo principal:

En vez de usar definiciones muy específicas se opta por dar esquemas con lo esencial de cada habilidad pero que permitan evaluar la suficiencia de cada definición. El asunto reside en saber definir la esencia de lo central de manera que ello informe de lo periférico (Lowe, 1998: 362).

b) Reflejar perspectivas congruentes que contribuyan a la descripción de los niveles. Pero basándose en hallazgos o descubrimientos, no en constructos:

Estos ejemplos deben recogerse de la experiencia y no de constructos teóricos que deberían funcionar pero que no se ha probado que funcionen; algo que contribuya a un nivel, no algo hipotético que se construye pensando que puede contribuir al nivel (Lowe, 1998: 365).

c) Resaltar hechos recientes, no teorías lingüísticas: cosas que sean comprensibles por usuarios que no pertenezcan solamente a ámbitos académicos. Cita, por ejemplo, que los conceptos de coherencia y cohesión no estaban inicialmente en la escala de ILR pero se consideraban, sin duda, en la evaluación de los niveles altos. Y que luego se introdujeron en la escala de ACTFL. Pero que no se pueden introducir otros elementos de gramática y corrección porque puede que sean útiles en escenarios académicos pero no en otro; y pueden penalizar a candidatos que procedan de ámbitos no académicos. Por ejemplo, un intermedio alto puede tener en escenarios académicos buen control de elementos formales, pero un candidato formado en la calle puede tener menos control formal y, sin embargo, más léxico. Definir un nivel usando esa terminología académica penalizaría a estos otros candidatos (Lowe, 1998: 372).

Lowe remarca que estos elementos, así como los socioculturales y algunos culturales, ya se consideraban en las definiciones, aunque no de forma explícita.

Y por ello, se decidió no incluir con más detalle los elementos culturales:

El reflejo de la cultura, ya sea la cultura con minúscula como con mayúscula, no aparece en los niveles bajos pues el candidato, en estos niveles, carece de espacio en su "caja negra" para estas cosas. Sin embargo, hacia el avanzado alto o superior, incluso aparece la cultura con mayúsculas automáticamente. Pero, desafortunadamente, mientras la cultura con minúsculas es un proceso orientado y, por consiguiente, abarcable, la cultura con mayúsculas está orientada a los hechos y se corresponde más con tests de rendimiento. Por supuesto que se podrían escribir en las definiciones, pero su contenido es abierto (Lowe, 1998: 377).

Respecto a la crítica de que el sistema de definir la competencia no sirve y todo lo que de ello se deriva, que es un test contextual, limitado y sin capacidad de predicción fuera del ámbito del test, la circularidad, etc., Liskin- Gasparro (1987: 37) explica que los descriptores de la escala de ACTFL se diseñaron basándose en la experiencia y la observación del progreso de estudiantes de segundas lenguas a lo largo de las funciones que podían entender o expresar, en las áreas temáticas con las que podían tratar y en la exactitud con que podían entender o comunicar un mensaje. Y que por ello son descriptivas y no normativas.

Ross (1998) cree que parte de las críticas surgían de que se prefería pensar en la competencia como en el grado en que el discurso está bien construido, formal o morfosintácticamente, antes que en la manera que emplean dos personas que interactúan conjuntamente negociando su contexto y significado para dar sentido (Ross, 1998: 334).

Lowe aclara que la OPI evalúa la lengua en general, no la lengua para fines específicos, negando que el lenguaje que elicitada sea solo la habilidad práctica para actuar en contextos determinados. Para él (1998: 374) hablar de competencia excluye hablar de actuación, porque esto último supondría evaluar actividades concretas, como por ejemplo, leer un telegrama en alemán. Esto puede ser una actividad que requiera más nivel del que el candidato posea en realidad, pero que pueda hacerlo no lo adscribe a ese nivel de dominio de lengua.

Hablar de actuaciones prácticas supone evaluar actividades concretas que quizás el hablante pueda realizar, por familiaridad, a más nivel del que posee en realidad. ACTFL, en cambio, se preocupa de evaluar la capacidad general evitando que sean las actividades habituales que el hablante conoce bien las que marquen la calificación. Igualmente, hablar de actividades supone hacer tests con temas; mientras que hablar de competencia supone evaluar mediante temas libres, porque no se evalúa en qué medida se sabe tratar ciertos temas sino el dominio general de la lengua empleada en ellos. Por lo tanto, evaluada de forma libre, la habilidad es transferible a otros temas diferentes a los utilizados en el test.

Muy clarificador, Lowe (1988: 12) sugiere que los dos conceptos de uso de la lengua, el de AEI y el de Bachman, suponen dos tipos de concepción de competencia comunicativa opuestos. El de ACTFL es holístico y de arriba abajo, mientras que el otro tiene una visión de la habilidad de abajo a arriba. Bachman acepta esta afirmación y dice lo mismo, desde su punto de vista: “El test de ILR y la OPI presenta un tipo diferente de concepto de habilidad lingüística que se deriva del constructo de sus escalas y que es diferente al de habilidad comunicativa de la lengua, ya que este es un proceso de abajo a arriba en la habilidad de la lengua, y atomístico; mientras que el de ACTFL es de arriba a abajo y holístico” (Bachman, 1990: 6-13).

Otro aspecto de estas críticas se refiere a la validez de la escala para calificar y a si permiten hacer juicios adecuados sobre el nivel; es decir, por un lado, a la ambigüedad o imposibilidad de entenderla y usarla apropiadamente, y, por otro, a si la escala es en sí apropiada.

Las críticas en este caso se centran en que es necesario especificar el grado de control de la lengua en términos que sean suficientemente claros y específicos como para distinguir los niveles, de forma que los criterios puedan ser interpretados por los calificadores de forma consistente (Bachman / Savignon, 1986: 388). Salaberry (2000: 304) critica, igualmente, que no están claros los componentes mediante los que se toman las decisiones de calificación y arguye que falta un conjunto detallado de criterios para evaluar la actividad oral global (Salaberry, 2000: 297).

Estas ambigüedades, a juicio de los críticos, impiden una calificación consistente, lo que le resta al test validez, por la falta en la definición del constructo, y le resta, a su vez, fiabilidad, por la falta de consistencia de los evaluadores al hacer las calificaciones. Por ello la rechazan y proponen un método distinto de calificar y unas escalas diferentes que establezcan criterios y contenidos o habilidades separados para cada área (Bachman / Savignon, 1986: 388).

Es Lowe (1998: 364-365), una vez más, el que responde a estos argumentos. Respecto a que no están claros los criterios de calificación explica que el sistema de AEI se basa en ir analizando la competencia de arriba abajo y sin hacer compensaciones entre elementos. Lo importante está en remarcar umbrales de dominio, no puntos medios. Aunque sabe que esto es difícil, ya que el mundo real se compone de puntos medios. Pero si aceptamos definir niveles como puntos medios, los puntos de cambio entre niveles diferentes se ven como algo fluido. Esto no es así en las escalas de AEI porque queremos que se pueda asignar cada actuación claramente a un nivel o a otro. Para ello, los bordes no deben parecer un continuum, deben estar claramente delimitados.

Por ello, la escala no compensa deficiencias o fallos en unos elementos con una mayor destreza en otros. Se debe hacer todo bien para estar en un nivel determinado: hacer todo lo del avanzado para estar en él, o quedarse en *casi avanzado*, es decir, en intermedio alto. Y esto, aunque es difícil porque no es sencillo trabajar con rangos, y se tiende a trabajar con puntos medios, debe quedar claro para los evaluadores. Aunque Lowe entiende que es duro y que los mayores fallos de fiabilidad se dan en las fronteras entre grandes niveles (acepta que el hecho de llamarlos *bajo*, *medio* y *alto* es algo que no ayuda y mejor habría que cambiar esas nominaciones por otras como *básico*, *completo* y *plus*).

Respecto a la necesidad de incorporar otros puntos discretos en las definiciones, ACTFL, en su reforma de la escala del año 1999, ha remarcado algunos elementos,

resaltando en sus definiciones de nivel y su articulación, alineando los enunciados de las funciones en una parte de la definición, los del contenido/contexto en otra, y la corrección en otra diferente. No obstante, Lowe se muestra escéptico respecto a si eso ayudará a calificar mejor, si este sistema será más sencillo para los calificadores o si mantendrá la óptica general (Lowe, 1998: 364). De cualquier manera, concluye, eso no varía la forma de evaluar del sistema AEI, que no es por puntos discretos sino por rangos (Lowe, 1998: 366).

Para terminar este punto, vamos a centrarnos en otras investigaciones que abordan las críticas de la validez y fiabilidad de las escalas y del método de evaluar de ACTFL desde un punto de vista cuantitativo.

Muchas de las críticas que se habían realizado a la evaluación y escala de ACTFL necesitaban nuevas investigaciones para ser mantenidas o rechazadas. Algunas de estas investigaciones las ha promovido la propia organización (Surface / Dierdorff, 2003; Swender, 2003). Todas ellas han ido disolviendo con claridad las dudas existentes, expuestas en las páginas anteriores, sobre la capacidad de la OPI como medida de la competencia oral.

Ha habido varios análisis que han estudiado la validez de la escala para dar calificaciones, la ambigüedad de sus interpretaciones o la dificultad de su uso por parte de los evaluadores, con la consiguiente, o no, falta de consistencia y fiabilidad.

Barnwell (1989), en una investigación sobre si los nativos son capaces de entender la escala de ACTFL y servirse de ella para dar las calificaciones, halló una alta correlación entre las calificaciones de los evaluadores oficiales y los evaluadores no nativos y concluyó que la escala sirve, ya que todos los nativos fueron capaces de ver las diferencias de nivel entre los candidatos en el mismo orden que los evaluadores de ACTFL.

Meredith (1990) analizó la escala para ver si efectivamente servía para apreciar y calificar diferencias de habilidad entre estudiantes en ámbito académico y concluyó que aunque era un instrumento insuficiente para calificar el desarrollo de un curso, sí servía para establecer diferencias reales entre las habilidades de los estudiantes (Meredith, 1990: 293).

Halleck (1995) también analizó los juicios holísticos hechos con la OPI, en los que la gramática tiene poco que ver, y otros juicios de madurez gramatical y sintáctica

independientes y discretos, encontrando que había una relación entre ambos por lo que “la madurez sintáctica tiene un grado de realidad psicológica con respecto a los niveles de competencia” definidos con la OPI (Halleck, 1995: 224).

En otro estudio posterior (Halleck, 1996) investigó la fiabilidad del uso de la escala con evaluadores no oficiales entrenados para ello y concluyó que, aunque era mejor utilizar evaluadores oficiales que daban mayor fiabilidad a las correcciones, estudiantes entrenados eran capaces de calificar la OPI suficientemente (Halleck, 1996: 232). Si se puede entrenar a gente no formada directamente por ACTFL para hacer una buena evaluación según sus escalas, es lógico y justo deducir que estas y su sistema de calificación son inteligibles.

Young (1995), en un estudio sobre las características de conversación entre los diferentes modelos de entrevistas orales (la OPI y otros modelos de LPI), explica que la OPI tiene más validez que otras LPI, porque tiene un formato abierto de hacer preguntas sobre temas que surgen durante la interacción, acomodados para cada candidato, lo que hace a cada entrevista diferente y aumenta su validez sobre el uso de la lengua (Young, 1995: 29). No obstante, este hecho le lleva a pensar que ello provoca menos fiabilidad, al ser más difícil la comparación entre los resultados de diferentes entrevistas.

Este término de la fiabilidad de la OPI está ya saldado a su favor tras los análisis que la propia organización y otros investigadores han llevado a cabo, que han concluido en dar un alto grado de fiabilidad y seriedad en los protocolos y administración del test (Thompson, 1995; Swender, 2003; Surface / Dierdorff, 2003).

Thompson (1995) realizó una amplia investigación con 882 entrevistas en varias lenguas planteándose cuestiones de fiabilidad: la correlación entre diferentes evaluadores, en qué medida se varía del primer al segundo evaluador de una OPI, si es más difícil hacer la evaluación en unos niveles que en otros o si varía la fiabilidad de unos niveles a otros. Los resultados son favorables al sistema de formación de evaluadores de ACTFL y a sus sistemas de calificación, ya que encontró unos altos grados de fiabilidad en todos los casos y una alta similitud entre las evaluaciones en las 5 lenguas usadas.

Concluye que la fiabilidad entre evaluadores entre la primera y segunda calificación en las cinco lenguas era grande, tanto con el coeficiente de Pearson¹³⁰ como con otras fórmulas utilizadas y que su grado de consistencia es muy alto (Thompson, 1995: 414).

También entre estas conclusiones destaca que la fiabilidad es mayor en el nivel superior que en los demás, seguido del intermedio medio, avanzado, intermedio alto y avanzado alto, que posee el menor índice de fiabilidad, por lo que es el nivel más difícil de calificar (Thompson, 1995: 413). Las mayores discrepancias se encontraron en las *fronteras mayores entre niveles* (entre intermedio alto o avanzado, y avanzado alto y superior, donde se comparten muchas características de la habilidad de ambos niveles), (Thompson, 1995: 411).

Surface / Dierdorff (2003), ante la responsabilidad de informar y actualizar los datos sobre la fiabilidad de los procedimientos de ACTFL periódicamente, analizaron casi seis mil entrevistas (5881 exactamente), hechas y evaluadas por evaluadores certificados de ACTFL en 19 lenguas y sometieron los resultados a varios procedimientos estadísticos de análisis de la consistencia, fiabilidad, probabilidad y acuerdo entre evaluadores y otros factores de las calificaciones (correlación de Pearson, correlación de Spearman, Tau de Kendall, y las medidas Kappa, Raw y Gamma¹³¹).

Tras este exhaustivo análisis concluyen que los educadores y empleadores que usan la OPI de ACTFL pueden confiar en los resultados y uso de las notas con toda seguridad (Surface / Dierdorff, 2003: 519).

Los resultados fueron los siguientes:

- La consistencia entre evaluadores en todas las lenguas alcanza grados muy altos; el 80 por ciento de las veces hay total acuerdo entre evaluadores y el 18 por ciento siguiente muestra discrepancias entre solo un nivel (intermedio bajo, intermedio medio, por ejemplo) (Surface / Dierdorff, 2003: 511).

- La consistencia es similar ya sea con las lenguas más comúnmente evaluadas como con otras consideradas más raras (Surface / Dierdorff, 2003: 512). Igualmente no varía respecto de unas lenguas a otras en general (Surface / Dierdorff, 2003: 514).

¹³⁰ De este coeficiente de averiguar correlaciones para fiabilidad y otros tipos de validación de test hemos hablado en §2.3

¹³¹ En su artículo explica brevemente las características de cada uno de estos tipos de análisis. Para una explicación sencilla de estos elementos estadísticos se debe consultar bibliografía como Bachman (2004), capítulo 3.

- La fiabilidad no varía de unos niveles a otros, a lo largo de los cuatro grandes niveles (inicial, intermedio, avanzado o superior) (Surface / Dierdorff, 2003: 512), aunque se muestra mayor discrepancia entre niveles avanzado alto y superior (Surface / Dierdorff, 2003: 515), como ha sido puesto de manifiesto en trabajos anteriores.

Swender (2003) también analiza la fiabilidad respecto de dos modelos de OPI, la realizada en directo y la realizada por teléfono. Concluye que ambas tienen el mismo grado de fiabilidad. Para obtener resultados objetivos y evitar el factor de que los candidatos pudieran obtener mejores resultados en la segunda entrevista realizada, al haber aprendido durante la primera, se hicieron dos grupos: la mitad de candidatos tenía un modelo como primera entrevista y la otra mitad el otro. Los resultados dieron el mismo nivel en cada entrevista a cada candidato (Swender, 2003: 522), entre ellos se resalta que en español (en la actualidad la mayoría de entrevistas se realizan en esta lengua) hubo un acuerdo total (Swender, 2003: 523).

2.6.3- E) Críticas sobre si la escala se basa en criterios o en una norma

Otra crítica era respecto a la escala. El argumento era que esta no era una escala basada en criterios de referencia sino que estaba basada en una norma¹³², pues toma como modelo de la lengua la de un nativo educado (Savignon, 1985: 131; Bachman / Savignon, 1986: 385; Bachman, 1990).

Bachman (1990: 308-350) comparaba ambos tipos de escalas (las referidas a un criterio y las referidas a una norma) desde su propia óptica. Según él, para evaluar la habilidad de uso de la lengua hay dos caminos: uno, llamado de la *vida real* (RL en sus siglas inglesas); y otro, el llamado enfoque de *habilidad de interacción* (IA en sus siglas inglesas).

El primero era el utilizado por ILR y ACTFL y, según la opinión de Bachman, no define un habilidad específica sino que establece un dominio de uso de la lengua en la vida real, identificándolo con lo que se considera el uso de la lengua de los hablantes competentes nativos.

¹³² Para las escalas basadas en una norma o en criterios, véase §2.5.1

Para diseñar una escala de este tipo se comienza por identificar el dominio de lengua de un grupo de hablantes nativos y luego se establecen los diversos niveles referidos a ellos. Las características más relevantes para describir la lengua de esta manera son las que aparecen en cualquier instancia de uso de la lengua; es decir, el contexto, el contenido, la situación, la gramática, el vocabulario o la pronunciación. Y citaba como ejemplo de todo ello la definición del nivel avanzado dada por ACTFL, que se establece mediante definiciones de contexto, contenido y otros elementos de la lengua (Bachman, 1990: 326):

Contexto:

“Puede realizar tareas complicadas en diferentes situaciones sociales...”

Contenido:

“Puede comunicar hechos y hablar sobre temas de interés público y personal...”

Bachman englobaba a este tipo de escalas generales de habilidad lingüística entre las escalas basadas en la norma de un grupo de hablantes, ya que es este grupo, en última instancia, el verdadero usuario de la lengua que se define con las características de la escala. Son, por lo tanto, escalas basadas en una norma, aunque sus autores pretendan que son escalas basadas en criterios de referencia.

Para él las únicas escalas consideradas como verdaderas escalas de referencia eran las que se definían de acuerdo con puntos discretos de uso de la lengua, pues cualquier calificación que se haga con estos puntos discretos se basa en la adecuación a los descriptores recogidos en la escala, independientes de los grupos reales de hablantes. Y por el contrario, cualquier calificación hecha con otras escalas, aunque se diga que están basadas en criterios, si estos están basados en la vida real, tienen un grupo al que acercarse de forma normativa y con el que compararse, y dejan de ser escalas de criterios para pasar a ser escalas basadas en una norma.

Esta necesidad de que las escalas sean de puntos discretos no ha sido, sin embargo, aceptada por otros autores. Para Carroll (1983) una escala referida a criterios debe dejar fuera los tradicionales elementos en que se basa la lengua (gramática, vocabulario y pronunciación), ya que estas categorías no hacen referencia a criterios de la actuación comunicativa sino a la contribución y subordinación a las habilidades que la hacen posible, lo cual no es equivalente a las habilidades comunicativas (Carroll, 1983: 10).

La escala referida a un criterio se debe referir a un conjunto de criterios de comportamiento comunicativo tales como “Puede iniciar, desarrollar y elaborar un

tema” o “puede mantener el diálogo aunque de forma pasiva”. Y solamente si se necesitan aclarar los niveles de la actuación de tales comportamientos podemos usar características (discretas) como “Puede interpretar con corrección usos de vocabulario y gramática y el sentido de variaciones en los modelos de acento y tono”.

De cualquier forma fue Lowe (1998: 374-75) quien contestó a la mayor parte de críticas respecto a la escala. Acepta que las llamadas a los nativos como elementos de referencia sobre la inteligibilidad de los discursos de los candidatos han sido un problema para las tres organizaciones (AEI). Pero recalca que su inclusión en las escalas no las convierte en una norma, pues el concepto de hablante nativo es solo un ideal procedente de la escala de la ILR y que no está en los conceptos de la OPI, ya que estos se cambiaron al diseñar esta nueva escala porque llevaban a malas interpretaciones, demasiado exigentes, por parte de ciertos evaluadores.

Para ser una escala basada en una norma le falta, además, y esto es concluyente, una de las características más importantes de los tests basados en una norma: que estos organizan a los candidatos de mejor a peor y los distribuyen a lo largo de una línea continua. Esto, dice, no lo hacen los tests AEI pues dan a cada candidato el nivel que le corresponde aglomerándolos según sus habilidades (Lowe, 1998: 378-79)

Un test orientado a la norma asigna una puntuación a cada candidato y hace un ranking organizándolos de peor a mejor. No en el sistema AEI. En un grupo de 30 estudiantes, continúa, 5 pueden ser inicial bajo, 7 inicial medio, 15 inicial alto, 3 intermedio bajo. Esto no es esparcirlos según una norma. ¿Cuál es el ranking de los 15 de nivel inicial alto? Las calificaciones de AEI no dicen a los profesores, supervisores, etc. quién está mejor, peor, en la media o cerca de la media dentro de un rango asignado.

El sistema AEI, concluye, no cumple uno de los mayores requerimientos de los tests con referencia a una norma. Da las calificaciones referidas a criterios, basándose en aspectos como “¿funciona el candidato dentro del nivel asignado?” (sin compararlo con otros candidatos) (Lowe, 1998: 379). Los tests de AEI, como tests referidos a un criterio, se definieron para medir si un candidato cumple un estándar determinado de maestría, controla ciertos elementos de habilidad, y cruza un umbral determinado de habilidad funcional. En resumen, si cumple el criterio.

Finalmente Bachman (2004) retira esta crítica e incluye a la escala entre las escalas con referencia a un criterio cuando dice que entre los tests basados en criterios hay dos

tipos, los tests que presentan el dominio sin organización y los que tienen el dominio organizado. Estos últimos informan de las calificaciones en términos de niveles de habilidad, que definen lo que los candidatos son capaces de hacer en diferentes niveles, como las escalas de ACTFL o del IELTS (*International English Language Testing System* de la Universidad de Cambridge) (Bachman, 2004: 301).

2.6.3- F) Conclusiones sobre las críticas al sistema de evaluar de AEI y a la OPI

Como hemos visto, las dudas sobre el modo en que la OPI definía la habilidad lingüística llevaron a realizar numerosas investigaciones sobre el contenido de los tests, el método y la validez de las interpretaciones, los resultados extraídos y el uso que se hace de ellos. Todas ellas han ayudado a entender y expandir su uso y han clarificado conceptos sobre la evaluación oral en general.

Sus defensores han mostrado en todo momento cómo la influencia del modelo de competencia comunicativa de AEI ganaba adeptos entre los profesores, instituciones y usuarios y al mismo tiempo han ido desarrollando una serie de trabajos para contestar, comprobar o desmentir las críticas, que han concluido que tanto sus escalas como su sistema de evaluación son válidos y fiables. A pesar de las críticas, el impacto y la expansión de los protocolos de la OPI ha sido enorme y el uso de las entrevistas orales se ha extendido no solo a ámbitos académicos sino entre empleadores, empresas comerciales y negocios de importación (Liskin-Gasparro, 1987: 215).

Este éxito se debe a que el concepto de competencia oral implícito en las *guidelines* y propagado por ACTFL y otras agencias federales americanas se ha incrustado en muchas de las actividades de enseñanza, evaluación, investigación y programación de un segmento importante de la profesión (Liskin-Gasparro, 2003: 488) y sigue bien vivo, institucionalizado y utilizado hasta el punto de ser evidente en muchos textos y en los *Standards for Foreign Language Learning in the 21st Century* (ACTFL, 1999) y en el trabajo ya citado dentro del NCATE.

Este minucioso escrutinio al que ha estado sometida la OPI ha demostrado su fortaleza. ACTFL, en una actitud responsable, ha aceptado todas las críticas y las ha incorporado a una agenda de investigaciones que ha terminado apoyando lo que ACTFL decía respecto a las interpretaciones y usos de la OPI (Chalhoub-Deville, 2003: 504).

2.7 El Marco común europeo de referencia y la evaluación

2.7.1 Introducción

El *Marco común europeo de referencia para las lenguas: aprendizaje, enseñanza, evaluación* (MCER) (Consejo de Europa, 2002) ha sido, sin duda, el hecho reciente más significativo en la escena de la educación lingüística en Europa (Alderson, 2005: 257). Publicado originalmente en 2001, año europeo de las lenguas, el MCER supone la culminación de un proyecto iniciado por el Consejo de Europa en los años 70 para delinear los objetivos que podrían ser apropiados para los estudiantes de lenguas en el contexto europeo.

Tras la amplia implementación del nivel umbral en currículos, libros de texto e incluso pruebas y exámenes, se sintió la necesidad de definir objetivos lingüísticos de aprendizaje a niveles más bajos (los llamados acceso y plataforma) así como a niveles más altos: avanzado, eficaz y maestría... Estos seis niveles se han recogido ahora juntos publicados en el MCER y por problemas de traducción a alguna de las lenguas europeas se han llamado A1 (para el nivel más bajo), A2, B1, B2, C1 y C2 (para el nivel más alto). La publicación el MCER cumple la necesidad, no solo de describir los niveles de lengua como se representan en tests, sino también de definir objetivos de aprendizaje de lenguas, y como punto de referencia para la elaboración de materiales de aprendizaje, enseñanza y evaluación en general (Alderson, 2005: 257-58).

El MCER pretende ser un documento de consulta obligada que incluye una descripción detallada de lo que implica aprender una lengua y poder utilizarla en situaciones de comunicación y define niveles de dominio que permiten identificar las distintas etapas en el aprendizaje de una lengua (Figueras *et ál.*, 2005: 5). Su aspecto más familiar son las numerosas escalas, debido a su valor obvio en evaluación y medición de los logros del aprendizaje (Alderson / Banerjee, 2002: 81). Pero los objetivos del MCER son mucho más amplios. En realidad, no es sólo un conjunto de escalas, sino un compendio de lo que se sabe del aprendizaje de la lengua, del uso de la lengua y de la competencia lingüística. Así como una guía para construir sillabus, desarrollar criterios y especificaciones de tests, diseñar materiales y libros de texto y formación de profesores¹³³ (Alderson / Banerjee, 2001: 220):

¹³³ Veremos enseguida, al analizar su aplicación práctica y las críticas que ha recibido el documento (§2.7.4) cómo estas afirmaciones entran dentro de lo que un documento político plantea como intereses y desideratum que no siempre son realistas, como algunos de los propios escritores del MCER o sus exegetas iniciales han terminado por reconocer aceptando sus numerosas e importantes limitaciones.

El Marco común europeo de referencia proporciona una base común para la elaboración de programas de lenguas, orientaciones curriculares, exámenes, manuales, etc. en toda Europa. Describe de forma integradora lo que tienen que aprender a hacer los estudiantes de lenguas con el fin de utilizar una lengua para comunicarse, así como los conocimientos y destrezas que tienen que desarrollar para poder actuar de manera eficaz. La descripción también comprende el contexto cultural donde se sitúa la lengua. El Marco de referencia define, asimismo, niveles de dominio de la lengua que permiten comprobar el progreso de los alumnos en cada fase del aprendizaje y a lo largo de su vida (Consejo de Europa: 2002: 1).

Según Alderson / Banerjee, “uno de los aspectos de su influencia y de su utilidad es el planteamiento del modelo de la lengua, uso y aprendizaje, a menudo basado explícitamente en el modelo de Bachman” (2002: 81) o “en las ideas del mundo de la enseñanza y lingüística aplicada angloamericana. La influencia de Wilkins (1976), Canale / Swain (1980) y Bachman (1990) es particularmente evidente”¹³⁴ (Alderson, 2007: 660). Estos modelos se pueden apreciar en varios capítulos, apartados y conceptos del MCER: por ejemplo, en la descripción de la competencia comunicativa, en la competencia lingüística y en la sociolingüística y pragmática.

Desde su publicación, su uso se ha extendido a todas estas actividades, incluido como referencia para la construcción de tests, creando una amplia familia de documentos y componentes¹³⁵; y “va a continuar incrementando su influencia debido a la necesidad de reconocimiento internacional de certificación en Europa para garantizar la movilidad educativa y laboral” (Alderson / Banerjee, 2001: 219). A través de sus numerosas escalas “desarrolla una dimensión importante de la enseñanza, la dimensión horizontal, que proporciona un mayor grado de detalle y de descripción para cada uno de los niveles comunes de referencia (...) que va a facilitar a los profesionales de todas las lenguas extranjeras la descripción de lo que esperan que los alumnos sean capaces de realizar” (Verdía, 2002: 5). Todo lo cual facilita el conocimiento de los aspectos más fuertes y débiles del aprendizaje de lenguas de cada estudiante. Con esto en la mano, cada uno puede presentarse en otro centro europeo, e incluso podría hacerlo ante un agente laboral y mostrar sus conocimientos. También Parrondo (2004: 980) hace

¹³⁴ El modelo de Bachman se ha explicado en §2.2.

¹³⁵ Entre sus desarrollos más interesantes se encuentran el PEL o Portfolio europeo de lenguas; el proyecto Dialang de autoevaluación por internet relacionado con el MCER sin fines de certificación; o ALTE (Asociación europea de evaluadores de lenguas). Para más datos sobre estos “hijos” del Marco, como los denomina Teresa Bordón, véanse los anexos C (Escala de Dialang) y D, del MCER, así como otros artículos y documentos (Little / Perclová, 2003; Cassany, 2002; Cassany, 2004; Little, 2005; Alderson / Huhta, 2005; el monográfico de la revista Mosaico, núm. 9, de diciembre de 2002, editada por la Consejería de Educación en Bélgica, Países Bajos y Luxemburgo; o la web www.mec.es/programas-europeos).

hincapié en este valor pedagógico, por su enorme versatilidad para adaptarse a una gran diversidad de situaciones de enseñanza-aprendizaje, y por parecer motivador para los estudiantes¹³⁶.

Sin embargo, y a pesar de que la evaluación es uno de sus principales objetivos y de que pretende ayudar a una “adecuada conceptualización y activación del constructo de lenguas a evaluar, y evitar las interferencias derivadas de los métodos de evaluación” (Consejo de Europa, 2002: 182), es desde el campo de la evaluación general de donde han venido las principales críticas, como se verá en §2.7.4. Pues hacer tests sobre la habilidad lingüística, objetivos, fiables, útiles y para poblaciones extensas, exige una metodología de diseño, pilotaje y análisis exigente y en el MCER no hay información suficiente para realizar este trabajo.

Por eso los resultados del MCER van por buen camino en cuanto a progresos curriculares y al desarrollo del PEL, que se está poniendo en práctica en numerosas escuelas de Europa y España (Cassany, 2004; Cassany, 2007; Instituto Cervantes, 2006), pero no en el campo de la evaluación general ni de equiparación de niveles. Las críticas en este sentido obligaron al Consejo de Europa a elaborar un nuevo documento, una especie de manual para relacionar exámenes con el MCER¹³⁷ (Consejo de Europa, 2002; 2005; 2007), aunque este es informativo y burocrático, más que una verdadera guía para diseñar y validar exámenes.

¹³⁶ Efectivamente, se ha visto útil para aspectos relacionados con el aula y la enseñanza institucional, así como para evaluación educativa y para reflexionar sobre la autoevaluación, elementos que han sido bien desarrollados en documentos como el Portfolio europeo de lenguas. Esto es así porque el MCER y su desarrollo a través del PEL ayuda a hacer un seguimiento del aprendizaje y los logros en diferentes campos, no sólo vertical sino horizontalmente; a que el profesorado realice otras actuaciones y evaluaciones de acuerdo con los criterios y niveles del MCER; a que el mismo estudiante se haga su autoevaluación (Little, 2002: 64). En los capítulos 4 y 5, explicando y comentando el uso práctico de las escalas en la evaluación, comentaremos la importancia de este desarrollo horizontal de la descripción de la lengua. No obstante, su uso en todos los campos, e incluso el uso del PEL, ha sido criticado. El tema no está acabado y fluctúa entre quienes le ven enormes posibilidades de desarrollo y quienes ven grandes dificultades en su aplicación, por su ambigüedad, por la excesiva dificultad de aplicación en los modelos educativos más habituales en Europa o por su dificultad de adecuación a los diferentes contextos escolares. Véase Westhoff, 2007: 678; Little, 2007: 652; y, en general, el número 91 (4), de *The Modern Language Journal* (2007: 641- 685).

¹³⁷ Figueras explica el nacimiento de este manual así: “(el) rápido apareamiento entre tests de nivel y niveles de salida con el MCER, hecho por agencias de evaluación y ministerios, sin evidencias para hacer tales enlaces, provoca numerosas exigencias de transparencia en los profesionales de la evaluación (...) En la conferencia de Barcelona de 2003 se pusieron de manifiesto los peligros de las prisas e infundadas afirmaciones hechas respecto a la relación entre el MCER y los exámenes. En respuesta a ello se realizó la primera versión del manual piloto para relacionar exámenes con el Marco en 2003” (Figueras, 2007: 673). Ya hubo una primera versión en la web del Consejo de Europa en 2002. El original en inglés, en su última versión de 2007, se titula *Relating Language Examinations to the CEF, Manual Preliminary Pilot Version*. De este manual, así como de las críticas al MCER, tratamos en §2.7.4 y §2.7.5.

El MCER, en su enfoque general y en su necesidad de desarrollo y adaptación a las necesidades de cada situación educativa o evaluativa, promueve un enfoque de aprendizaje de lenguas modernas orientado a la acción (Consejo de Europa, 2004: 13), como lo explica en el capítulo 2.1 titulado “Un enfoque orientado a la acción” (MCER; 2002: 9). A través de sus otros capítulos, se repasan y señalan muchos aspectos y conceptos de la enseñanza, aprendizaje y evaluación de lenguas desde una perspectiva que, intentando aportar los conceptos necesarios para ayudar al logro de los objetivos, pretende “promover ante todo la reflexión y discusión; sin pretender ser un intento de introducir un sistema uniforme paneuropeo” (Figueras *et ál.*, 2005: 261). Este enfoque de reflexión y discusión se hace explícito en diferentes partes del documento, por ejemplo, el apartado 2.3, cuando declara que una de sus funciones principales es “animar y capacitar a todos los que participan en la enseñanza de lenguas y en los procesos de aprendizaje para que informen a otros, con la mayor transparencia posible, no sólo de sus objetivos, sino también sobre los métodos que utilizan y los resultados conseguidos” (Consejo de Europa: 2002: XIII).

Por ello, todo su contenido se presenta como recomendaciones para que el usuario pueda utilizarlas de acuerdo con sus objetivos y las necesidades de sus productos, respetando su origen cultural y las tradiciones del contexto educativo en que se encuentre. A este respecto se dice claramente:

Los ejemplos son sugerencias. Puede que el usuario desee mantener algunos, rechazar otros y añadir los propios. No hay que tener reparo en ello, pues el usuario decide sus objetivos y su producto. Hay que recordar que lo que a un usuario pueda parecerle innecesario tiene su sitio en el Marco de referencia porque otra persona de distinto origen cultural puede considerarlo esencial... Un usuario puede considerar necesario añadir categorías y exponentes que podrían resultarle útiles a otro usuario. Por este motivo, el esquema taxonómico presentado en los capítulos 4 y 5 del Marco no se contempla como un sistema cerrado, sino que está abierto a futuros desarrollos en función de la experiencia.

Este principio también se aplica a la descripción de niveles de dominio de la lengua... los seis niveles utilizados... se presentan como recomendaciones y no son, de ninguna manera, obligatorios (Consejo de Europa: 2002: XIII).

La publicación del MCER se tomó con optimismo e interés, pues se sintió y agradeció la publicación de un documento que iba a servir de notable ayuda a la profesión. El MCER se publicó primero en 2001 en inglés (Cambridge University Press), francés (Didier), alemán (Langenscheidt) y portugués (ASA). Y en 2002 en checo (Univerzita Palackého y Olomouci), húngaro (PMIK), italiano (La Nuova Italia,

Oxford) y español (Instituto Cervantes y ANAYA). Posteriormente, ha sido traducido a muchas otras lenguas, entre las que están el vasco, catalán y gallego¹³⁸. Pero, “una vez disipada la euforia inicial surgieron las reservas sobre su “infalibilidad”, incluso sobre la necesidad de su existencia, y empiezan las críticas a la complejidad del texto, a la ausencia de concreción y a algunas inconsistencias desde distintos niveles educativos y desde distintos campos” (Figueras, 2005: 6). Críticas y reservas sobre las que trataremos más adelante (§2.7.4).

Nada de eso ha sido problema para que se haya utilizado, con más o menos acierto, en todo tipo de acciones educativas. También se han intentado analizar las correspondencias entre los niveles del MCER y exámenes existentes en otros países. Entre ellos, los trabajos del *Centre international d'études pédagogiques* francés para relacionar exámenes con el MCER¹³⁹, el proyecto del CITO (organismo del Estado holandés para evaluación)¹⁴⁰ de construcción de un banco de ítemes europeo, o el interés por homologar el MCER con el sistema de ACTFL en Estados Unidos¹⁴¹; así como trabajos de diseño curricular (Llorián / Rodrigo, 2005; Instituto Cervantes, 2006), de análisis de la competencia léxica (Baralo, 2005) o de autoevaluación por medio del Portfolio (Cassany, 2007).

Después de todo, y en definitiva, el MCER no soluciona problemas, sino que plantea retos, algunos quizás imposibles técnicamente de realizar, como la equiparación de exámenes diferentes entre sí. Pero, en realidad, si de sus 9 capítulos quitamos aquellos que se dedican a explicar el compendio de lo que hay sobre la enseñanza y aprendizaje de lenguas, y si, de lo realmente llamativo, sus escalas, nos quedamos con el objetivo de

¹³⁸ Además del inglés y francés, lenguas oficiales del Consejo de Europa, el documento se ha traducido a las siguientes lenguas: albanés, alemán, armenio, búlgaro, catalán, checo, croata, esloveno, español, esperanto, finés, francés, friulián, gallego, georgiano, griego, holandés, húngaro, italiano, japonés, moldavo, polaco, portugués, ruso, serbio, ucranio y vasco. Esta lista, y sus editoriales o páginas web donde se puede conseguir cada documento, está en la página web del Consejo de Europa: www.coe.int/T/DG4/Portfolio/?L=E&M=/documents_intro/common_framework.html; una vez en la página dirigirse a: CEFR and related documents (consultada el 21 de enero de 2008).

¹³⁹ Christine Tagliante, “Liens entre les certifications de français langue étrangère du MENESR et le CECR”, comunicación ante el *Policy Forum The Common European Framework of Reference for Languages (CEFR) and the development of language policies: challenges and responsibilities*, organizado por el Consejo de Europa en Estrasburgo durante los días 6 - 8 de febrero de 2007.

¹⁴⁰ *Mapping the Dutch Foreign Language State Examinations onto the Common European Framework of reference, Report of a Cito research project commissioned by the Dutch Ministry of Education, Culture and Science*, (Noijons / Kuijper, 2006).

¹⁴¹ Interés reseñado también en el Intergovernmental Policy Forum de Estrasburgo (2007), *The Common European Framework of Reference for Languages (CEFR) and the development of language policies: challenges and responsibilities*, organizado por el Consejo de Europa en Estrasburgo durante los días 6 - 8 de febrero de 2007, con la comunicación “Language Learning Policies in the United States”, de Jacque Bott Van Houten, del *National Council of State Supervisors for Languages*.

que sirvan para equiparar y homologar actuaciones educativas entre sí, cosa que el MCER no dice cómo hacer, solamente nos queda el deseo de impulsar la calidad y posibilidad de comparar certificaciones y currículos, para lo que todavía es necesaria mucha investigación y muchos trabajos futuros. Un plan político educativo ideado para el futuro, al que todavía hay que dedicar muchos trabajos e investigaciones que ayuden a su desarrollo¹⁴².

En conclusión, en vez del texto que esperábamos de ayuda a la profesión y a la evaluación, es un documento político para reflexión y movilización de voluntades, al que le queda mucho camino por recorrer.

En este apartado, y tras esta introducción (§2.7.1), vamos a explicar los principales contenidos e innovaciones del MCER (§2.7.2); a continuación, nos centraremos en el papel de la evaluación dentro del MCER (§2.7.3); expondremos las críticas que todo ello ha producido (§2.7.4) intentando abarcar tanto sus logros como sus principales limitaciones, igual que hemos hecho con el sistema de evaluar de ACTFL, la OPI; y terminaremos (§2.7.5) con el desarrollo del manual, publicado por el Consejo de Europa, especialmente dedicado a ayudar a desarrollar y utilizar tests relacionados con el MCER así como con una conclusión final sobre el MCER (§2.7.6).

2.7.2 Contenidos del MCER

El MCER se divide en 9 apartados más 4 anexos que aportan datos sobre numerosos conceptos del campo de enseñanza y aprendizaje de lenguas y sobre el desarrollo y diseño de las escalas y descriptores del dominio de la lengua.

Muestra dos bloques de contenido muy diferenciados: el bloque descriptivo, capítulos 1, 2, 6, 7, 8, y 9, que recoge de manera exhaustiva los elementos que contribuyen al aprendizaje de una lengua, y el estatus actual de la profesión, desde una óptica de la enseñanza comunicativa; y los capítulos 3, 4 y 5, donde presenta sus 6 niveles comunes de referencia, junto a los descriptores ilustrativos de los niveles de habilidad y dominio, así como las actividades comunicativas y estrategias de comunicación. Para ello se han diseñado 60 escalas escalonadas en seis niveles, sobre

¹⁴² Volvemos a remitir al número de 2007 de *The Modern Language Journal*, 91 (4): 641- 685, donde se exponen sus principales limitaciones, o a §2.7.4 donde nos hacemos eco de alguna de ellas.

diferentes aspectos de la competencia y habilidades lingüísticas. Todo ello acompañado de explicaciones sobre la competencia y las competencias generales del individuo, las actividades de la lengua, el aprendizaje de lenguas y sus procesos, el contexto del aprendizaje de las lenguas, sus condiciones y restricciones, opciones metodológicas, el uso de estrategias de expresión, comprensión, interacción y mediación, los procesos comunicativos, el papel de las tareas en la enseñanza, opciones de diseño curricular y evaluación¹⁴³.

Antes de centrarnos en los aspectos relacionados con la evaluación mediante el MCER, en su capítulo 9, vamos a señalar los principales contenidos de cada capítulo centrándonos fundamentalmente en aquellos aspectos que puedan ser interesantes para el diseño de tests y la evaluación.

El **capítulo 1** define los fines, objetivos, funciones y criterios del MCER, entre ellos ser integrador, transparente y coherente, según los principios generales del Consejo de Europa respecto a la lengua y el fomento del plurilingüismo. Se titula *Marco común europeo de referencia en su contexto político y educativo*.

El **capítulo 2**, titulado *Enfoque adoptado*, explica el enfoque adoptado, centrado en la acción, pues considera a los usuarios y alumnos como agentes sociales, miembros de una sociedad que tiene tareas (no solo relacionadas con la lengua) que llevar a cabo en una serie de circunstancias. Este capítulo es un buen compendio de los conceptos y parámetros que describen el uso de la lengua, la habilidad del usuario, el aprendizaje y la enseñanza, preparativo de los siguientes donde se introducen ya los niveles de competencia y actividades comunicativas. El capítulo sirve también para aclarar aspectos del contenido y alcance del MCER, como, por ejemplo, que cuando se describen especificaciones de objetivos de aprendizaje:

No se describen los procesos que capacitan a los alumnos para actuar como es debido, ni los que les permiten desarrollar las competencias que posibilitan las acciones. No

¹⁴³ En palabras de Alderson, que resumen irónicamente el documento: “una enciclopedia del aprendizaje y uso de la lengua (alguna vez descrita como todo lo que John Trim sabe respecto a la lengua)” (Alderson, 2007: 661). A este respecto véase la opinión del propio Trim en entrevista con Saville (2005), a partir de la página 283 y especialmente en la 285, donde se da pie para decir lo que Alderson dice.

describen las formas en que los profesores facilitan los procesos de adquisición y aprendizaje de lenguas (Consejo de Europa: 2002: 18).

El **capítulo 3**, titulado *Niveles comunes de referencia*, establece una serie ascendente “de seis niveles comunes de referencia para describir el dominio que el alumno tiene de la lengua” (Consejo de Europa: 2002: 16). El uso de seis niveles se justifica en “la práctica normal de varias entidades públicas que se dedican a realizar exámenes ... en que se ha comprobado que son útiles, transparentes y adecuados por parte de grupos de profesores que son hablantes nativos y no nativos” (Consejo de Europa: 2002: XIII), así como en la experiencia e investigación de los grupos de trabajo, cuyos resultados se validaron con diversos métodos de escalonamiento y desarrollo de escalas, de los que se habla en North / Schneider, (1998).

Los 6 niveles se definen en su capítulo 3.6 (Consejo de Europa: 2002: 36-39), de forma genérica, y luego se ofrecen en diversos cuadros descriptivos centrados en diferentes utilidades: para el usuario o estudiante, el cuadro de *Niveles comunes de referencia: escala global*, (Consejo de Europa, 2002: 26), para autoevaluación, el cuadro de *Niveles comunes de referencia: autoevaluación*, (Consejo de Europa, 2002: 30-31), o para el examinador, el cuadro de *Niveles comunes de referencia: aspectos cualitativos del uso de la lengua hablada* (Consejo de Europa, 2002: 30-31), que además sirve para acotar cada nivel dentro de 5 elementos: alcance, corrección, fluidez, interacción y coherencia. Todo ello hace del MCER un sistema muy descriptivo¹⁴⁴.

Además, en los siguientes capítulos, se dan otras escalas centradas en las habilidades (capítulo 4) y en la competencia (capítulo 5).

Es interesante la aclaración que se da en el capítulo 3.5 y figuras 2, 3, 4 y 5 (Consejo de Europa: 2002: 29, 34-35) respecto al número de 6 niveles, pues en él se habla de las ventajas de la flexibilidad del enfoque ramificado de determinación de los niveles; esto es, del hecho de que cada usuario puede establecer más niveles o subniveles en relación con sus necesidades y el uso que se vaya a hacer de la información resultante. Aquí se mantienen con coherencia los principios ya citados

¹⁴⁴ Este capítulo ofrece otras interesantes consideraciones que tienen directa relación con la evaluación, como por ejemplo que las escalas se pueden leer de arriba abajo o de abajo a arriba, lo que tiene consecuencias, como hemos visto en §2.6.3- D).

respecto a que el objetivo es promover debate y ofrecer elementos para que este sea transparente¹⁴⁵. Así se posibilitan subdivisiones de los niveles hasta llegar a una escala de 13 niveles (aunque podrían ser incluso más):

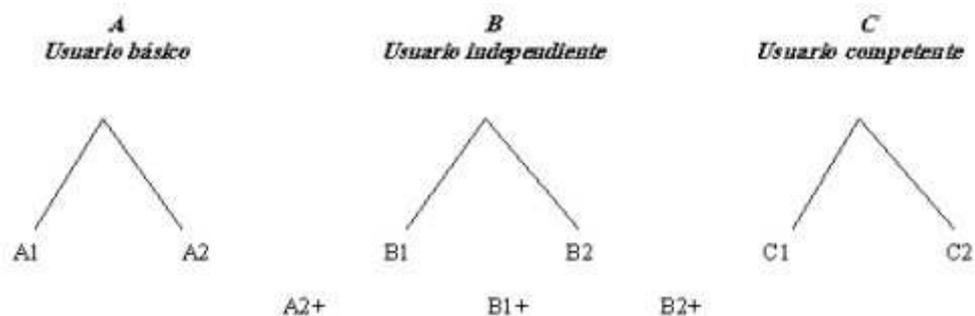


Figura 2

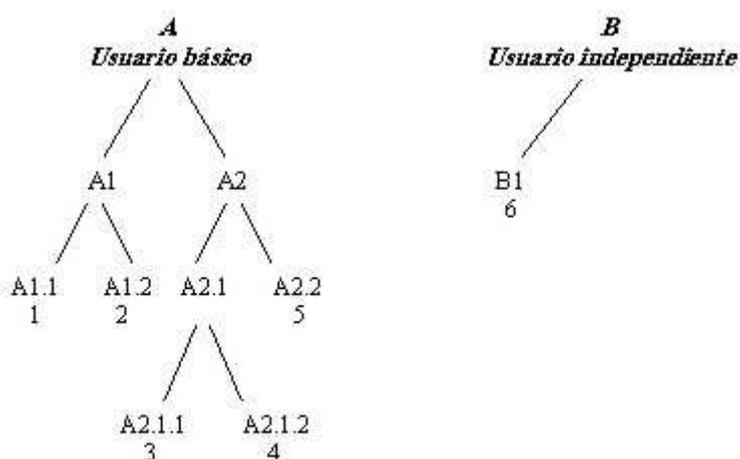


Figura 3

¹⁴⁵ Lo que no sabemos es si esas buenas intenciones tienen fundamento práctico; ya que ampliar el número de niveles y descriptores implica aumentar la complejidad de la evaluación, la formación de evaluadores y los criterios de calificación. Y si con los 6 niveles existentes hay problemas de interpretación, no vemos cómo van a poder hacerse más subdivisiones que sean necesarias o eficaces. De esto se trata también en el apartado de las críticas al MCER (§2.7.4), en este capítulo, así como en los análisis sobre sus escalas (§3.1) y los resultados prácticos de evaluar mediante ellas (§4.5.1).

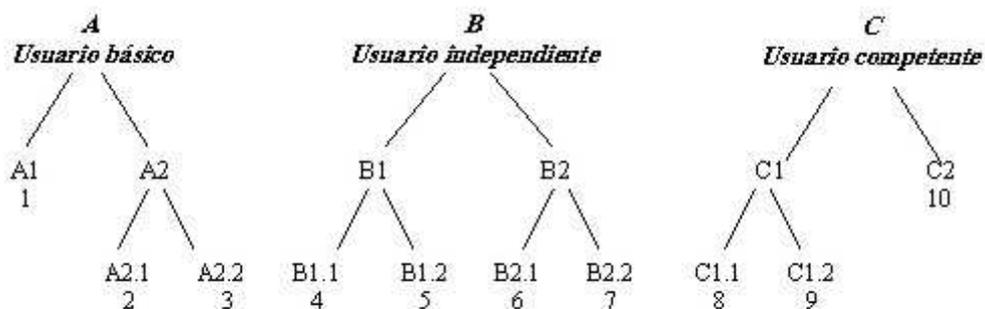


Figura 4

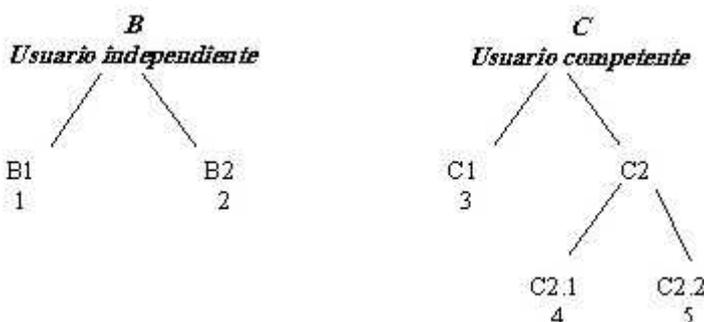


Figura 5

(Consejo de Europa: 2002: 34-35)

En el mismo MCER se dan escalas en las que ya se han definido algunos de estos subniveles, por ejemplo la escala de *Expresión oral en general*, analizada en §3.1.4, que subespecifica los niveles B2.1 y B2.2.

A pesar de estas posibilidades de autoelaborar escalas, el MCER justifica sus acciones y decisiones diciendo que “los profesionales de todo tipo agradecen especialmente que se les ofrezca un conjunto de niveles comunes de referencia como instrumento de clasificación; profesionales a los que, como en muchos otros campos, les resulta ventajoso trabajar con niveles de exigencia estables y aceptados en cuanto a medida y formato” (Consejo de Europa, 2002: XIII).

También se explica que, además de estas subdivisiones que ramifican los 6 niveles hasta donde se crea necesario, en algunas de las escalas que se ofrecen en los capítulos 4 y 5, hay niveles en los que se hacen subdivisiones internas aportando otras características de la habilidad que están por encima del dominio del nivel, pero sin llegar a ser del nivel (o subnivel) inmediatamente superior. Esto se ha hecho, por ejemplo, en la escala de *Monólogo sostenido: argumentación* (Consejo de Europa,

2002: 63), donde en el nivel B1 se dan dos descriptores, uno de menor nivel (“Ofrece breves razonamientos y explicaciones de opiniones, planes y acciones”) y otro de más nivel, pero todavía B1, (“Desarrolla argumentos lo bastante bien como para que se puedan comprender sin dificultad la mayor parte del tiempo”).

Por el contrario, hay otras escalas que carecen de descriptor para algunos niveles, lo que se aclara y explica con razones como que: a) en algunos casos los descriptores “se descartaron en el control de calidad”, b) en otros casos “posiblemente los descriptores podrían ser escritos pero no se han escrito”, c) “aunque el área existe, su formulación parece muy difícil”, o d) “el área no existe o no es adecuada para ese nivel y no se puede realizar una distinción” (Consejo de Europa, 2002: 40)¹⁴⁶.

Esto ocurre en las siguientes escalas (tomadas de los capítulos 4 y 5), (véanse cuadros 2.7.2 -A y 2.7.2- B).

Cuadro 2.7.2- A. En las siguientes escalas y niveles *no hay descriptor posible*

No hay descriptor posible de nivel C2 de las escalas de:	<i>Monólogo sostenido: argumentación</i> <i>Declaraciones públicas</i>
No hay descriptor posible de nivel C1 de las escalas de:	<i>Monólogo sostenido: argumentación</i>
No hay descriptor posible de nivel A2 de las escalas de:	<i>Monólogo sostenido: argumentación</i> <i>Informes y redacciones</i> <i>Control y Corrección</i> <i>Escuchar conferencias y presentaciones</i> <i>Notas, mensajes y formularios</i> <i>Tomar notas (en conferencias, seminarios, etc.)</i>
No hay descriptor posible de nivel A1 de las escalas de:	<i>Monólogo sostenido: argumentación</i> <i>Declaraciones públicas</i> <i>Informes y redacciones</i> <i>Planificación</i> <i>Compensación</i>

¹⁴⁶ No entendemos cómo un documento que pretende servir de ayuda a la discusión transparente entre diferentes colectivos de diferentes países y lenguas deja sin definir niveles con estos argumentos. Entre ellos nos parece alarmante el de que *parece difícil de describir*, dado que el MCER pretende, en su flexibilidad, permitir que los usuarios tengan la posibilidad de establecer subniveles de acuerdo con sus necesidades educativas o evaluativas. Si los expertos que lo han escrito no han sido capaces de hacer ese trabajo, ¿cómo se puede pensar que otras personas van a poder hacerlo? El caso es al contrario en otras situaciones, porque hay escalas donde no se han definido algunos niveles cuando fácilmente se podría haber hecho. Véase un ejemplo de esto en el análisis hecho de la escala de *Monólogo sostenido: argumentación* (§3.1.5) escala a la que fácilmente se le podrían haber introducido descriptores para los niveles C1 y C2 equivalentes a los que hay para esos niveles en otras escalas.

	<i>Control y Corrección</i> <i>Comprender conversaciones entre hablantes nativos</i> <i>Escuchar conferencias y presentaciones</i> <i>Escuchar retransmisiones y material grabado</i> <i>Ver televisión y cine</i> <i>Identificación de las claves e inferencia (oral y escrita)</i> <i>Conversación informal</i> <i>Conversación formal y reuniones de trabajo</i> <i>Notas, mensajes y formularios</i> <i>Tomar la palabra (turnos de palabra)</i> <i>Cooperar</i> <i>Pedir aclaraciones</i> <i>Tomar notas (en conferencias, seminarios, etc.)</i> <i>Dominio de vocabulario</i> <i>Flexibilidad</i> <i>Turnos de palabra</i> <i>Desarrollo de descripciones y narraciones</i> <i>Precisión</i>
--	---

En las escalas que figuran en el siguiente cuadro solo se definen niveles hasta el nivel que se señala en esta tabla, no por encima, ya que se equiparan.

Cuadro 2.7.2- B. Escalas que presentan un nivel máximo sobre el que no hay descrita mayor habilidad

C1 es el límite de dominio o habilidad para las escalas de:	<i>Comprender conversaciones entre hablantes nativos</i> <i>Escuchar avisos e instrucciones</i> <i>Escuchar retransmisiones y material grabado</i> <i>Leer correspondencia</i> <i>Leer en busca de información y argumentos</i> <i>Leer instrucciones</i> <i>Ver televisión y cine</i> <i>Identificación de las claves e inferencia (oral y escrita)</i> <i>Conversación informal</i> <i>Interacción escrita en general</i> <i>Escribir cartas</i> <i>Tomar la palabra (turnos de palabra)</i> <i>Cooperar</i> <i>Dominio de la pronunciación</i>
---	--

	<i>Desarrollo de descripciones y narraciones</i>
B2 es el límite de dominio o habilidad para las escalas de:	<i>Planificación Compensación Leer para orientarse Colaborar para alcanzar un objetivo Interactuar para obtener bienes y servicios Intercambiar información Pedir aclaraciones Turnos de palabra</i>
B1 es el límite de dominio o habilidad para las escalas de:	<i>Notas, mensajes y formularios</i>

El **capítulo 4**, titulado *El uso de la lengua y el usuario o alumno*, se dedica principalmente a describir las escalas de actividades comunicativas, pero también a explicar los ámbitos, contextos y situaciones en que se usa la lengua, con sus condiciones y restricciones, temas y tareas.

En él se dan 40 escalas que no solo describen lo que se puede hacer, sino en qué medida. Figueras explica el conjunto de estas escalas y sus descriptores:

En la descripción de los niveles, se distinguen seis grandes categorías en un conjunto de tres por dos: recepción, producción e interacción en sus modos escrito y hablado. Cada uno de ellos se vuelve a subdividir ocasionando una colección de 56 conjuntos de descriptores¹⁴⁷ que ponen en práctica una precisa definición de actividades como, por ejemplo, Escuchar conferencias y presentaciones, Leer cartas, Tomar la palabra. Además de esto, se describen categorías para recepción, producción y estrategias de interacción y se dan, también, categorías para los aspectos de la competencia del estudiante (esto es, fluidez hablada, precisión gramatical, adecuación sociolingüística)... No se consideran otras áreas muy especializadas, como por ejemplo una descripción en detalle de las actividades de mediación (Figueras et ál., 2005: 264-65).

Pero en los apéndices C y D se dan otras escalas que sirven para ver actividades concretas que se pueden realizar dentro de cada nivel, mediante la formalización de los descriptores según un sistema activo “Puedo hacer”. Estas escalas son las del proyecto de

¹⁴⁷ Son 55 escalas, contando también las que se ofrecen en el capítulo 5, centradas en el dominio de la lengua. Si a ello le sumamos otras escalas generales como la de *Autoevaluación*, aparecidas en el capítulo 3, hay un total de 60 escalas en el MCER, sin contar con las 9 de DIALANG y ALTE de los anexos, que también se pueden utilizar. No extraña que, como ya hemos dicho, Alderson destaque que lo más conocido del MCER sean sus escalas.

evaluación de DIALANG y las de la Asociación Europea de Examinadores de Lenguas (*Association of Language Testers in Europe, ALTE*) que están ancladas al MCER.

Estas escalas del capítulo 4 pueden utilizarse para definir las especificaciones para el diseño de tareas de evaluación, para informar de los resultados de una evaluación, para autoevaluación por parte del alumno o evaluación por parte del profesor (Consejo de Europa, 2002: 180).

Son las siguientes, agrupadas por destrezas, según el MCER:

- Actividades de expresión oral (5 escalas): *Expresión oral en general, Monólogo sostenido: descripción de experiencias; Monólogo sostenido: argumentación (por ejemplo en un debate); Declaraciones públicas; y Hablar en público.*

- Actividades de expresión escrita (3 escalas): *Expresión escrita en general; Escritura creativa; Informes y redacciones.*

- Estrategias de expresión (3 escalas): *Planificación; Compensación; Control y corrección.*

- Actividades de comprensión auditiva (5 escalas): *Comprensión auditiva en general; Comprensión de conversaciones entre hablantes nativos; Escuchar conferencias y presentaciones; Escuchar avisos e instrucciones; Escuchar retransmisiones y material grabado.*

- Actividades de comprensión lectora (5 escalas): *Comprensión de lectura en general; Leer correspondencia; Leer para orientarse; Leer en busca de información y argumentos; Leer instrucciones.*

- Actividades de comprensión audiovisual (1 escala): *Ver televisión y cine.*

- Estrategias de comprensión (1 escala): *Identificación de las claves e inferencia (oral y escrita).*

- Actividades de interacción oral (9 escalas): *Interacción oral en general; Comprender a un interlocutor nativo; Conversación; Conversación informal (con amigos); Conversación formal y reuniones de trabajo; Colaborar para alcanzar un objetivo (por ejemplo, reparar un coche, discutir sobre un documento, organizar un acontecimiento); Interactuar para obtener bienes y servicios, Intercambiar información; Entrevistar y ser entrevistado.*

- Actividades de interacción escrita (3 escalas): *Interacción escrita en general; Escribir cartas; Notas mensajes y formularios.*

- Estrategias de interacción (3 escalas): *Tomar la palabra* (turnos de palabra); *Cooperar*; *Pedir aclaraciones*.

- Tipos de textos (2 escalas): *Tomar notas* (en conferencias, seminarios, etc.); *Procesar textos*.

El **capítulo 5**, titulado *Las competencias del usuario o del alumno*, se centra en escalas relacionadas con el dominio de la lengua, es decir, las competencias lingüísticas. En total se presentan 13 (*Competencia lingüística general, Riqueza de vocabulario, Dominio de vocabulario, Corrección gramatical, Dominio de la pronunciación, Dominio de la ortografía, Adecuación sociolingüística, Flexibilidad, Toma de turnos de palabra Desarrollo de descripciones y narraciones, Coherencia y cohesión, Fluidez oral y Precisión*). Junto a ellas se describen los conceptos de las diferentes competencias que se desgranar a lo largo de las escalas, así como las capacidades de aprender y las destrezas de los individuos en relación con la lengua. Pero se advierte de que los sistemas lingüísticos son enormemente complejos y los sistemas utilizados para describir tradicionalmente la competencia han resultado inadecuados, y los nuevos modelos propuestos tampoco lo han logrado de una forma efectiva.

Las escalas que se dan en este capítulo se pueden utilizar para autoevaluación del alumno y evaluación por parte del profesor, o para ofrecer puntos de referencia en el diseño de criterios de evaluación. Se pueden usar las escalas combinándolas para hacer un párrafo holístico para cada nivel, como una lista de control, como una parrilla de categorías seleccionadas, etc. (Consejo de Europa, 2002: 181).

Son las siguientes, agrupadas por competencias, según el MCER:

- Competencias lingüísticas (1 escala general): *Competencia lingüística general*.
- Competencia léxica (2 escalas): *Riqueza de vocabulario, Dominio de vocabulario*.
- Competencia gramatical (1 escala): *Corrección gramatical*.
- Competencia fonológica (1 escala): *Dominio de la pronunciación*.
- Competencia ortoépica (saber articular un discurso partiendo de la forma escrita), (1 escala): *Dominio de la ortografía*.
- Competencia sociolingüística (1 escala): *Adecuación sociolingüística*.
- Competencia discursiva (4 escalas): *Flexibilidad, Turnos de palabra, Desarrollo de descripciones y narraciones, Coherencia y cohesión*.

- Competencia funcional (2 escalas): *Fluidez oral, Precisión.*

El **capítulo 6**, titulado *El aprendizaje y la enseñanza de la lengua*:

Estudia los procesos de aprendizaje y enseñanza de la lengua, referidos a la relación existente entre la adquisición y el aprendizaje, y a la naturaleza y el desarrollo de la competencia plurilingüe, así como a las opciones metodológicas de carácter general o específico relativas a las categorías establecidas en los capítulos 3 y 4 (Consejo de Europa, 2002: XV).

El **capítulo 7**, titulado *Las tareas y su papel en la enseñanza de la lengua*, estudia este tema atendiendo a su diseño, su dificultad, diversos factores que influyen en su realización por los usuarios (afectivos o lingüísticos), las condiciones y restricciones que conllevan o las estrategias que provocan.

El **capítulo 8**, titulado *La diversificación lingüística y el currículo*, es un breve capítulo dedicado a señalar “las consecuencias de la diversificación lingüística para el diseño curricular y analiza asuntos como el plurilingüismo y pluriculturalismo; objetivos de aprendizaje diferenciados; principios del diseño curricular; escenarios curriculares; el aprendizaje de idiomas a lo largo de toda la vida, modularidad y competencias parciales” (Consejo de Europa, 2002: XVI).

El **capítulo 9**, titulado *La evaluación*, trata cuestiones de evaluación y en él se explica “la utilidad del MCER para la evaluación del dominio de la lengua y del aprovechamiento, así como los criterios de evaluación y los distintos enfoques del proceso de evaluación” (Consejo de Europa, 2002: XIV).

En él, aunque se explican aspectos relativos a la evaluación de lenguas, no se aporta ningún test para evaluar con el MCER ni se explica cómo se debe evaluar con este documento. Por ello la pregunta que a todo lector y usuario del MCER le viene a la cabeza es la misma que Alderson ha cifrado en su famosa pregunta “¿Cómo sé yo que mi nivel B1 es tu nivel B1?”. Como el MCER no da respuesta a este y muchos otros aspectos de la evaluación, tras su publicación se sacó un documento relacionado con esto, a través de la página de internet del Consejo de Europa (www.coe.int/), *Manual para vincular exámenes de lengua al MCER* (Consejo de Europa, 2002 b). Nosotros vamos a dedicar los siguientes apartados (§2.7.3, §2.7.4 y §2.7.5) a analizar este asunto.

El **anejo A** trata algunos temas generales y teóricos que serán de utilidad para los usuarios que deseen elaborar escalas adaptadas a una población concreta de alumnos.

El **anejo B** ofrece información relativa al proyecto suizo, que desarrolló las escalas de descriptores utilizadas.

Finalmente **los anejos C y D** presentan escalas desarrolladas por DIALANG y ALTE.

De todo ello destacamos como más novedoso, además de las escalas ya señaladas, el hecho de delimitar 4 tipos de actividades lingüísticas: producción, expresión, interacción y mediación, en vez de las 4 destrezas tradicionales; el hecho, no exento de críticas, de que el documento esté fundamentado en una concepción de la enseñanza y de la descripción del dominio lingüístico comunicativas y en ningún momento se hagan referencias a contenidos gramaticales necesarios para alcanzar tal dominio; o que se den indicadores comparativos con la intención de ayudar a la discusión y de que se puedan establecer equiparaciones entre lenguas y niveles, ya que ahora, como afirma North, (2007: 659), “la gente está dispuesta a discutir estos asuntos mucho más fructíferamente”.

2.7.3 La evaluación dentro del MCER: el capítulo 9

Vamos a centrarnos en tres aspectos de este capítulo¹⁴⁸:

- 1) La explicación de su posible uso para especificar contenidos y criterios.
- 2) La posibilidad de utilizarlo para facilitar la comparación entre pruebas.
- 3) La explicación de cómo se puede hacer un test viable con el contenido del documento.

¹⁴⁸ Nos limitamos aquí a explicar lo que el MCER dice al respecto. Las críticas y las dificultades o imposibilidades para realizarlo, que han sido señaladas con profesionalidad y acierto, las dejamos para §2.7.4. Véase, por ejemplo Fulcher (2004 b).

1.

Respecto al primero, su uso para especificar contenidos y criterios, explica que se pueden consultar los capítulos 4 (*El uso de la lengua y el usuario o alumno*), 5 (*Las competencias del usuario o del alumno*) y 7 (*Las competencias del usuario o del alumno*).

En el apartado 4.4 se especifican las actividades lingüísticas comunicativas para proceder a la especificación de los contenidos de una prueba y de las tareas usadas para la evaluación; en el apartado 4.1 se habla del contexto de uso de la lengua y de sus restricciones, ámbitos y condiciones; en el apartado 4.6 se explican los textos que se pueden usar.

En el apartado 5.2 se explican *las competencias lingüísticas comunicativas* y sirve para perfeccionar la elaboración de ítems de la prueba o de las fases de una prueba oral, con el fin de conseguir evidencias de las competencias lingüísticas, sociolingüísticas y pragmáticas relevantes.

Y el capítulo 7 analiza las tareas y su papel en la enseñanza de lenguas y estudia la dificultad y otros conceptos necesarios a tener en cuenta a la hora del diseño de las tareas de las pruebas.

También explica cómo funcionan las escalas, con el fin de usarlas como contribución a la valoración de objetivos de evaluación y formulación de criterios de calificación. Así aclara posibles diferencias de su uso: por ejemplo, que las escalas de actividades comunicativas (capítulo 4) pueden servir para la evaluación que realiza el profesor o para autoevaluación respecto a tareas de la vida real “porque estos dos tipos de evaluación se realizan sobre la base de una representación muy pormenorizada de la capacidad lingüística de los alumnos desarrollada durante el curso” (Consejo de Europa, 2002: 180). O que las escalas de dominio (capítulo 5) “pueden servir para ofrecer puntos de partida para el desarrollo de criterios de evaluación. Transformando las impresiones personales -no sistemáticas- en juicios ponderados, dichos descriptores pueden contribuir al desarrollo de un marco compartido de referencia entre un grupo de examinadores”¹⁴⁹ (Consejo de Europa, 2002: 181).

En cuanto a la especificación de exámenes, da un ejemplo de uno de expresión oral, y aconseja utilizar una evaluación auténtica que tenga muestras de una serie de

¹⁴⁹ El aparente lenguaje abstracto y la falta de concreción real de cómo utilizar el MCER es en realidad debido a que este documento no ayuda más y se limita a dar este tipo de consejos. Veremos esta crítica más adelante, con Weir, en §2.7.4.

discursos relevantes, una conversación, un debate informal y una tarea de cooperación centrada en el objeto.

Por ejemplo:

Una conversación simulada que funciona como introducción; un debate informal sobre temas por los que el candidato demuestra su interés; una fase de transacción en forma de actividad, cara a cara o simulada, de búsqueda de información por teléfono; una fase de expresión, basada en un informe escrito en el que el examinado ofrece una descripción de su área de especialidad y de sus planes; y una cooperación centrada en el objetivo, una tarea de consenso entre los examinandos (Consejo de Europa, 2002: 179)

2.

Respecto al segundo aspecto, el cual es central para esta tesis, la posibilidad de utilizar el MCER para facilitar la comparación entre pruebas, apunta que es difícil comparar sistemas de evaluación y relacionar evaluaciones, ya que suelen evaluar elementos radicalmente distintos, incluso cuando abarcan los mismos ámbitos. Primero porque el constructo suele estar poco definido, es decir presenta *escasa conceptualización y activación del constructo*, y segundo, porque hay interferencias derivadas de los métodos de evaluación.

Pero estas dificultades se pueden reducir ya que el MCER ofrece unos principios para solucionar el problema de la conceptualización (lo acabamos de ver en el punto 1) mediante un esquema descriptivo con el que intenta categorizar de forma práctica el uso de la lengua; y porque mediante las escalas de descriptores se configura un cuadro conceptual que se puede utilizar tanto para relacionar sistemas nacionales e institucionales entre sí mediante el MCER como para planear los objetivos de exámenes concretos utilizando las categorías y los niveles de las escalas (Consejo de Europa, 2002: 182).

Un ejemplo de la necesidad y utilidad del MCER para la equiparación entre exámenes es el del proyecto de equiparación entre los exámenes del English-Speaking Union (ESU) y el MCER. La ESU procedió situando un objetivo concreto en un nivel determinado, por ejemplo, un examen que pretende cubrir los niveles 4 y 5 de una escala. Para proceder a equiparaciones es necesario que los exámenes que se quiera comparar hayan sido situados en una escala común de dominio de la lengua.

Entonces sería posible durante un tiempo establecer la relación entre las calificaciones de un examen de la serie con las calificaciones de los otros

*exámenes... De esta manera es posible establecer relaciones entre calificaciones de exámenes y niveles de dominio de la lengua, porque los exámenes tienen, por definición, un determinado nivel de exigencia y un grupo de examinadores preparados capaces de interpretar su nivel. Es necesario conseguir que los niveles comunes sean explícitos y transparentes, proporcionar ejemplos que hagan operativos tales niveles y después someterlos a una gradación*¹⁵⁰ (Consejo de Europa, 2002: 44-45).

3.

Respecto a cómo se puede hacer un test viable con el contenido del documento, apunta que el conjunto de escalas debe ser usado como una herramienta y que “no se pretende que cualquiera utilice, en un enfoque práctico de evaluación todas las escalas en todos los niveles; pues a los examinadores les resulta difícil abordar un gran número de categorías, y además la serie total de niveles presentada puede que no sea adecuada al entorno concreto”. Se aconseja reducir el número de categorías a un número viable y se añade que la experiencia demuestra que más de cinco provocan una sobrecarga cognitiva (Consejo de Europa, 2002: 193).

Así, con respecto a una evaluación oral, si se considera que las estrategias de interacción son un aspecto cualitativo de la comunicación, las escalas ilustrativas contienen 14 escalas, por lo que hay que hacer una selección. “Deben ser combinadas, reformuladas y reducidas a un conjunto más pequeño de criterios de evaluación que sea adecuado para las necesidades de los alumnos, y para los requisitos de la evaluación y para el estilo de la cultura pedagógica específica” (Consejo de Europa, 2002: 194).

Esto es lo que hace Cassany (2007) al proponer un modelo de portfolio de autoevaluación de la conversación mediante una tabla orientativa. Es un buen ejemplo de autoconfección de un sistema de evaluación mediante el MCER, seleccionando los elementos que interesen en cada situación. En este caso, de las escalas que da el MCER, se seleccionan principalmente, de entre las escalas de *Competencia lingüística*, las de *Pronunciación*, *Entonación*, *Corrección*, *Fluidez*, *Turnos de palabra* y *Estrategias conversacionales*. Y de ellas se selecciona el contenido mínimo requerido para hacer un análisis personal, con preguntas directas sacadas del contenido de las escalas en sus diferentes niveles.

¹⁵⁰ Está claro que la equiparación es uno de los objetivos del MCER. Pero el MCER no explica cómo hacerlo, salvo que uno ya sea un examinador preparado, capaz de interpretar el nivel y relacionarlo con las escalas del MCER, así como diestro en el manejo de las técnicas necesarias de diseño y validación de tests, de las que tampoco se dice nada.

Por ejemplo:

Estrategias conversacionales: *¿puedes hacer preguntas?, ¿responderlas?, ¿cambiar de tema?, ¿hacer repetir una intervención no comprendida?, ¿inferir por el contexto algo que no entiendes?, ¿sabes hacer un circunloquio para aclarar un concepto difícil?, ¿dar rodeos en una respuesta que no te interesa dar?, etc.* (Cassany, 2007: 17).

El MCER también da varios ejemplos, cuatro en concreto, de este tipo de autodiseño de la evaluación.

Ejemplo 1: *Cambridge Certificate in Advanced English (CAE)*, Parte 5: Criterios para la evaluación (1991)

Criterios para la prueba	Escalas ilustrativas	Otras categorías
Fluidez	<i>Fluidez</i>	
Corrección y alcance	<i>Alcance general Riqueza de vocabulario Corrección gramatical Control de vocabulario</i>	
Pronunciación	<i>Control fonológico</i>	
Aprovechamiento de la tarea	<i>Coherencia Adecuación sociolingüística</i>	Éxito en la tarea Necesidad del apoyo del interlocutor
Comunicación interactiva	<i>Estrategias de turnos de palabra Estrategias de colaboración Desarrollo temático</i>	Alcance y facilidad del mantenimiento de la intervención

(Consejo de Europa, 2002: 194).

Ejemplo 2: *Internacional Certificate Conference (ICC)*: certificado de inglés para los negocios, Prueba 2: Conversación de negocios (1987)

Criterios de la prueba	Escalas ilustrativas	Criterios de la prueba
Escala 1 (sin denominación)	<i>Adecuación sociolingüística Corrección gramatical Control de vocabulario</i>	Escala 1 (sin denominación)
Escala 2 (Uso de las características del discurso para iniciar y mantener la fluidez de la conversación)	<i>Estrategias de turnos de palabra Estrategias de colaboración Adecuación sociolingüística</i>	Escala 2 (Uso de las características del discurso para iniciar y mantener la fluidez de la conversación)

(Consejo de Europa, 2002: 195)

Ejemplo 3: *Eurocentres*: Evaluación de la interacción en grupo reducido (RADAI) (1987)

Criterios de la prueba	Escalas ilustrativas	Otras categorías
Alcance	<i>Alcance general</i> <i>Riqueza de vocabulario</i>	
Corrección	<i>Corrección gramatical</i> <i>Control de vocabulario</i> <i>Adecuación sociolingüística</i>	
Pronunciación	<i>Fluidez</i> <i>Control fonológico</i>	
Interacción	<i>Estrategias de turnos de palabra</i> <i>Estrategias de colaboración</i>	

(Consejo de Europa, 2002: 195).

Ejemplo 4: Consejo Nacional Suizo e Investigación: Evaluación de actuaciones en vídeo

Criterios de la prueba	Escalas ilustrativas	Otras categorías
Alcance	<i>Alcance general</i> <i>Riqueza de vocabulario</i>	
Corrección	<i>Corrección gramatical</i> <i>Control de vocabulario</i>	
Fluidez	<i>Fluidez</i>	
Interacción	<i>Interacción global</i> <i>Turnos de palabra</i> <i>Colaboración</i>	
Coherencia	<i>Coherencia</i>	

(Consejo de Europa, 2002: 196).

Otro sistema muy útil de evaluar mediante el MCER, sin duda, lo vemos en el uso y en la filosofía del *Portfolio* y del *Portafolio electrónico*, que supone una sincera y seria reflexión sobre el aprendizaje y sus logros a lo largo de un tiempo determinado y ayuda a ser consciente de numerosos elementos que forman parte del proceso y a utilizarlos de forma activa en la retroalimentación, o estudio y aprendizaje de lenguas. Ya que la evaluación acostumbra a dirigir la educación, el portfolio y el PEL pueden ser, sin duda,

útiles para la clase de E/LE porque aportan congruencia, autenticidad y conexión con la realidad (Cassany, 2007: 19).

Otro ejemplo puede ser la propuesta de Davison / Fulcher (2007) de diseñar tests dirigidos a obtener un efecto en sus usuarios. O nuestra propuesta (Martínez Baztán, 2007) para una evaluación de la expresión escrita de inmigrantes escolarizados. Al utilizar el MCER hay que plantearse varios pasos de definición de objetivos, elaboración de tareas, validación de estas, uso y pilotaje, análisis cuantitativo y estadístico de resultados, y formación de evaluadores para la calificación adecuada y objetiva.

La primera fase la describe así:

Para un test (de expresión escrita) del nivel A2 habrá que acudir a los cuadros del MCER y buscar entre ellos los deseados para el test (Martínez Baztán, 2007: 8).

La segunda pretende diseñar las tareas y pilotarlas objetivamente:

Evaluar con referencia a estas escalas supone comenzar por escribir ítems o preguntas y exámenes que provoquen un discurso que sea prueba de la existencia de una habilidad determinada que se desea evaluar, de acuerdo con la escala.

Luego habrá que pasar a la fase de validación del test, a comprobar si este es objetivo y fiable.

Por ejemplo, para medir la destreza de escritura del nivel A2 (que se definen en el MCER como: “Escribe una serie de frases y oraciones sencillas enlazadas con conectores sencillos tales como y, pero y porque”) unas tareas podrían ser pedir al candidato que describa cómo es su país, la gente de su país o su ciudad.

Tras administrar estas preguntas se procede a valorar en qué medida contienen los elementos que se buscaban y a juzgarlos con un sistema claro y compartido por todos los correctores (Martínez Baztán, 2007: 8-9).

Y la tercera, de formación de evaluadores y establecimiento de criterios objetivos de calificación con relación al MCER, así:

En la corrección hay que usar otras escalas relacionadas con la expresión escrita. Habrá que llegar a un acuerdo sobre la manera en que vamos a analizar el discurso (las respuestas de los estudiantes) de acuerdo al cumplimiento de la tarea propuesta, según los criterios y escalas del MCER y sin dejarnos llevar por otros criterios no escritos o no considerados ni en el MCER ni en el diseño y objetivos del test (Martínez Baztán, 2007: 9).

Pero el MCER no aporta datos que ayuden a diseñar estos tests. Por ello, a pesar del interés, o debido a él, pronto se resaltaron “las dificultades que presenta el documento a

la hora de concretar diseños curriculares o especificaciones de pruebas de evaluación” (Figueras, 2005: 7). Mencionando, en particular “las listas aparentemente cerradas que en realidad no lo son... Además, estas listas no presentan (ni se prestan) a una fácil jerarquía de dificultad que permita graduar actividades y tareas de aprendizaje o de evaluación... Pero las críticas más duras surgen de la comunidad de evaluadores, cuando se constata que precisamente a causa de la gran aceptación del MCER por la sociedad, las calificaciones, los certificados y los exámenes europeos están abocados a relacionarse con los niveles del MCER” (Figueras, 2005: 7).

De estas limitaciones del MCER, de las críticas directas de los evaluadores y de las respuestas del Consejo de Europa, nos ocupamos en los dos siguientes apartados.

2.7.4 Críticas y limitaciones del MCER

- Os digo que yo soy ciego o vos vais desnudo

(Cuento XXXII, *Libro de los enxiemplos del Conde Lucanor et de Patronio*,
Don Juan Manuel, 1337).

Sin duda con la publicación del MCER se formó un revuelo en la profesión de enseñanza de lenguas. Su carácter enciclopédico sobre lo que supone la enseñanza, el aprendizaje y la evaluación, afectó a todos los ámbitos; en consecuencia se comenzó a leer y utilizar apoyándose en sus propias afirmaciones de flexibilidad y posibilidades de adaptación a cada contexto educativo y cultural. Pero, en sí, el documento no es sencillo de leer e interpretar, menos de poner en práctica, y mucho menos de usar para evaluación. Por ello, lo que más destacó fueron sus escalas en 6 niveles y la posibilidad de subdividirlas según las necesidades de cada uno.

Pero el fondo comunicativo definido en sus dimensiones de la competencia comunicativa y la novedosa descripción de la lengua según 4 categorías de actividades lingüísticas no han calado todavía en el uso en las aulas o libros de texto, según destaca Little (2007: 646). Como acepta North, poca gente ha cambiado la visión de las destrezas acomodándola a la que el MCER propone (North, 2007: 659). Y, como dice

Alderson, no ha tenido un impacto real en un radical rediseño de tests o libros de texto (2007: 661)¹⁵¹.

Byrnes apunta, siguiendo a Westhoff¹⁵², que esto se deba quizás a que el MCER, pese a la pretensión comunicativa de las prácticas de enseñanza europeas, sea un tipo de revolución que va a la contra de la corriente normal de textos y clases de lengua (Byrnes, 2007: 645). Aunque también apunta a que en todo proceso dinámico de implantación o aceptación de nuevos principios de educación lingüística es difícil la adaptación, debido a los diversos contextos educativos existentes (Byrnes, 2007: 642).

Sobre esta difícil puesta en marcha también pueden haber influido la complejidad y limitaciones del propio documento y el hecho destacado por Fulcher (2004 b: 259) de que no es tal marco sino un modelo falto de herramientas que permitan hacerlo operativo.

Sea como sea, la verdad es que las dificultades se han ido manifestando al querer trabajar con él en diferentes ámbitos educativos y finalmente se han puesto de relieve en el número 91 de *The Modern Language Journal*, de 2007¹⁵³, dedicado a analizar el MCER: en él se hace una profunda revisión del actual estado del MCER y sus usos y abusos, concluyendo con la necesidad de dedicar tiempo y esfuerzo a numerosas implementaciones y análisis empíricos del documento que desarrollen y mejoren los campos en los que este está poco fundamentado.

El propio North acepta sus limitaciones y dificultades de uso en contextos escolares porque, dice que “el MCER solamente define comportamientos para cada nivel” (North, 2007: 643) cuando la escuela necesita que se especifiquen las habilidades y competencias que emergen y (estas deben) estar relacionadas con el tipo de educación que promueven. En este sentido afirma “en el marco de la escuela se necesita situar el desarrollo del aprendizaje de las competencias lingüísticas dentro del conjunto de desarrollo cognitivo y social del niño. La mayoría de los descriptores ilustrativos del MCER pueden ser inadecuados para esto” (North, 2007: 659).

¹⁵¹ Este investigador va más lejos y señala que se están produciendo recelos y dudas a la hora de aplicarlo, como consecuencia de la falta de investigación empírica (Alderson, 2007: 660). Y llega a pedir, previa denuncia de sus malos usos, más responsabilidad en su práctica, diciendo que hay que lograr que, se haga lo que se haga, se haga con decoro y responsabilidad profesional (Alderson, 2007: 662).

¹⁵² Westhoff afirma que las escalas del MCER y el PEL son incompatibles con los modos de enseñar al uso, ya que los libros y currículos de enseñanza de lenguas extranjeras tienen en su mayoría un canon gramatical basado en enseñar puntos discretos uno tras otro una vez que el anterior se considera aprendido, mientras que el MCER no especifica lo que se debe aprender para poder hacer lo que en sus escalas se describe (Westhoff, 2007: 676-678).

¹⁵³ *The Modern Language Journal*, 91 (4): 641-685. Véanse los artículos de Byrnes, Little, North, Alderson, Krumm, Bonnet, Figueras y Westhoff.

Las primeras críticas al MCER vinieron de la parte de la evaluación y solamente años después se han comenzado a señalar con fuerza estas otras deficiencias de concepto y falta de fundamentos teóricos del documento y sus limitaciones. Las principales limitaciones que ahora se señalan son:

- La inadecuación para su uso en contextos escolares, con jóvenes, e incluso para adultos inmigrantes, sin antes proceder a una readaptación de sus especificaciones y contenidos (North, 2007; Krumm, 2007: 668; Little, 2007: 642).

- La falta de fundamentos teóricos basados en diversas teorías lingüísticas, de adquisición de lenguas o sociolingüística y la carencia de especificaciones basadas en el desarrollo del aprendizaje (North, 2007: 657; Hulstijn, 2007; Alderson, 2007).

- La necesidad de realizar categorizaciones de la lengua en función de diferentes enfoques de enseñanza o incluso especificar los elementos gramaticales que sean necesarios para adquirir la competencia descrita en las escalas; junto a la necesidad de extensiones y revisiones del MCER mediante más investigaciones y análisis empíricos que permitan desarrollar aspectos concretos del lenguaje, del contenido, las actividades, los exponentes funcionales, lingüísticos y socioculturales (Alderson, 2007: 660-662; Westhoff, 2007: 661; Figueras, 2007: 674).

- Las ambigüedades y fallos en el diseño de los propios descriptores y escalas. El propio North acepta que a algunos les falta validación empírica, en los niveles C1 y C2, y sobre todo a muchos de los descriptores de actividades comunicativas (North, 2007: 657); cosa que se une a otras ya señaladas como la falta de claridad y definición de muchos de sus términos, junto a la opacidad, solapamientos y ambigüedad señaladas por Alderson “cuando las escalas se examinan de cerca” (Alderson, 2007: 661); además de la dificultad que supone el definir los niveles C1 y C2, cercanos a la lengua nativa, de una forma exclusivamente lingüística, que deseche factores cognitivos y educativos (Alderson, 2007: 662).

Todo esto no ha hecho sino dificultar la aplicación del MCER, cuando no ha llevado a malas prácticas por parte de diferentes sectores que han intentado aparentar su puesta al día haciendo afirmaciones infundadas sobre la adecuación de sus prácticas y productos al documento. Es notoria la rapidez con que editoriales de libros de texto o examinadores y administradores, se han apresurado a decir que sus trabajos y productos

eran acordes con el MCER, sin análisis empíricos que justifiquen ni validen sus hechos (Figueras, 2007: 673; Alderson, 2007: 643)¹⁵⁴. Alderson lo denuncia explícitamente:

A pesar de sus limitaciones e imperfecciones los proveedores de exámenes, las editoriales de libros de texto y los diseñadores de currículos hacen afirmaciones sobre las relaciones entre sus productos y el MCER. Ello consuela a los usuarios. Pero el problema es que no hay evidencias empíricas de esas afirmaciones (...) Y faltan evidencias sobre lo que se necesita para aprender una lengua y el nivel requerido en determinados contextos (afirmaciones que se han hecho) sin asegurarse de si esos niveles pueden ser efectivamente logrados o justificados (Alderson, 2007: 661)

Hemos comenzado señalando las limitaciones generales a todo el MCER. Ahora vamos a centrarnos en las críticas vertidas sobre la imposibilidad efectiva y fiable de evaluar con él, con lo que esto conlleva de imposibilidad para equiparar tests entre sí (Fulcher, 2004 b: 261); o sobre el hecho de que, a pesar de que pretende ayudar a la evaluación, certificación y equiparación entre niveles y exámenes, su uso es imposible para quienes no estén instruidos en evaluación, pues es necesaria una información más detallada con la que llevar a cabo diseño de tests (Fulcher, 2004 b: 261).

Vamos a centrarnos en las principales críticas que se relacionan con la imposibilidad de evaluación mediante el MCER.

Weir ha sido quien más profusamente ha reseñado y criticado sus limitaciones¹⁵⁵. La principal, pero no la única, que no sirve como modelo para el diseño de tests (Weir, 2005: 296). Los fallos principales que proceden del documento e inciden en las dificultades de realizar una evaluación son, según él (Weir, 2005: 282):

¹⁵⁴ Esto, no obstante, no ha llegado a desarrollos curriculares importantes. Salvo el del Instituto Cervantes, de 2006, ningún otro país ha realizado su adaptación de contenidos al MCER. Esto es, sin duda, debido a su enorme dificultad y a que el MCER carece de directrices concretas que doten de contenido al dominio descrito en sus escalas, pues es un documento deliberadamente general que no se centra en aspectos concretos del lenguaje (Alderson, 2007: 660). A esto hay que sumar una deficiencia intrínseca, el hecho de que, al ser solamente una categorización de dominio dada por profesores de lengua y no una explicación del desarrollo del aprendizaje ni de la adquisición de la lengua (aunque pueda parecerlo y llevar a errores de concepción de uso y aprendizaje de la lengua), su traslado a materiales concretos sea difícil e intuitivo. Como dice Figueras, “desarrollar un currículo que determine el contenido de las especificaciones para los descriptores y niveles, partiendo de las necesidades del aprendizaje, es un trabajo pendiente que solo en la Universidad de Cambridge se han planteado y que tardará años en realizarse” (Figueras, 2007: 674).

¹⁵⁵ Críticas que suscribimos. En §3.1 se analizan algunas escalas del MCER y se muestran sus limitaciones. En el capítulo 4, dedicado a la calificación de pruebas orales realizadas mediante el MCER, se comentan las limitaciones y problemas que han surgido a la hora de evaluar y calificar las entrevistas utilizadas en este trabajo.

- falta de validez de contexto: las escalas postulan un rango de variables, respecto a los contextos y condiciones de actuación, incompleto y desigual;
- falta de una teoría válida del aprendizaje: hay poca información sobre la naturaleza del procesamiento cognitivo en los diferentes niveles de habilidad;
- falta de justificación de las calificaciones: las actividades casi nunca ofrecen más información sobre el tipo de actuación que se espera que produzcan;
- la forma de algunos de sus descriptores no es suficientemente consistente o transparente.

Algunos de estos fallos ya habían sido anunciados antes y los hemos reseñado de forma general. Little *et ál.*, llamaban la atención sobre la necesidad de una descripción mas detallada del dominio lingüístico en determinadas áreas del aprendizaje de la lengua¹⁵⁶ (Little *et ál.*, 2002: 64) y aconsejaban hacer una revisión de las escalas para uso didáctico, incrementándolas con materiales más claros para los estudiantes, como actividades y tareas apropiadas, con las que luego se sintieran familiares a la hora de autoevaluarse, evitando malentendidos respecto al logro de sus objetivos (Little *et ál.*, 2002: 64). Aunque Little acepta que el MCER se puede utilizar para analizar las necesidades de los estudiantes y especificar los objetivos de su repertorio comunicativo de forma que conlleve implicaciones pedagógicas (Little, 2007: 649).

Es, creemos, una cuestión de saber hacer y saber trabajar con ello. Aunque este trabajo sea, en principio, posible para cualquier profesional de la enseñanza, se torna una tarea de grandes dificultades si se carece de previos y serios conocimientos sobre evaluación. “No se puede asegurar que los test que dicen ser B1 realmente lo son, ni que un B1 de inglés esté midiendo lo mismo que un B1 de alemán. No está claro cómo tales equivalencias pueden establecerse sin dudar” (Little, 2007: 64). Problema que también señala Bonnet, quien aplaude el establecimiento de un banco de ítemes europeo que facilite la labor de calibrar exámenes¹⁵⁷, ya que debido a las propias escalas del MCER

Hay dificultades inherentes relacionadas con la calidad del asesoramiento. Se dice habitualmente que tal y tal certificado o diploma se basa en el MCER o demuestra una

¹⁵⁶ Ya hemos señalado (§2.7.2) las escalas que no dan descriptores para algunos de sus niveles; además, varias no han sido validadas de manera empírica o calibradas (*Declaraciones públicas, Conversación formal y reuniones de trabajo*, y todas las de escritura: *Producción escrita en general, Escritura creativa, e Informes y redacciones*) y algunas se han construido combinando elementos de otras sin la calibración que necesitarían (*Hablar en público*, y todas las de escritura) (Little *et ál.*, 2003: 81-90).

¹⁵⁷ Véase la nota 12, en el capítulo 1.

competencia según tal nivel del MCER. Sin embargo, tales equivalencias están muy lejos de ser sencillas de realizar y demostrar. (...) Es muy difícil saber si un test calibrado en Finlandia para B1 de griego es equivalente a otro calibrado en Portugal para B1 de polaco (Bonnet, 2007: 671).

Las críticas llegan también hasta otras escalas que están supuestamente calibradas y completas y sí definen los seis niveles.

Huhta comenta su elevado nivel de abstracción (2002: 133), lo que hace difíciles las interpretaciones de nivel. Algo que también encuentra Jones (2002), para quien la gente tiende a interpretar las afirmaciones “Puedo hacer” de forma diferente; por ejemplo las declaraciones sobre el nivel más alto, C2, no se distinguen bien de las del nivel inmediatamente inferior, C1. Lo que indica la necesidad de mayor precisión en la especificación de los tests de la que se da en el MCER (Jones, 2002: 176).

También Alderson ve dificultades en las definiciones de los niveles más altos, C1 y C2, pues “definir descriptores de acuerdo a la lengua nativa es dudoso dada la dificultad de describir los niveles C1 y C2 solamente mediante elementos lingüísticos y no cognitivos o educativos” (Alderson, 2007: 662).

A este respecto, Bordón (2007) señala que, en las autoevaluaciones hechas mediante el PEL, alumnos procedentes de diferentes contextos de aprendizaje tienen percepciones notablemente diferentes y exigencias diferentes respecto a sus logros y limitaciones al utilizar la escala de *Autoevaluación* (Bordón, 2007: 94). Algo que también Little había señalado (Little, 2002: 64).

Igualmente, Bordón señala un fallo incomprensible en la descripción de la escala de *Autoevaluación*¹⁵⁸, pues el MCER describe autoevaluación dividida en sólo tres categorías (comprender, hablar y escribir), subdivididas en cinco: comprensión auditiva y comprensión de lectura, hablar y conversar, y escribir. Esto, argumenta, no se corresponde con la concepción del uso de la lengua que el MCER propone, ni con la más lógica división de destrezas y habilidades en comprensión, expresión, interacción y mediación, que en sus vertientes oral y escrita darían 8 habilidades (comprensión auditiva y comprensión de lectura; expresión oral y expresión escrita; interacción oral e interacción escrita; interpretación (mediación oral) y traducción (mediación escrita) (Bordón, 2007: 84-85).

¹⁵⁸ Escala de *Niveles comunes de referencia: cuadro de autoevaluación* (Consejo de Europa, 2002: 30-31).

De acuerdo con el propio MCER, hablar y escribir no indica la misma categoría que comprensión y expresión. Se nos ocurre que se podría argumentar que escribir equivale a expresión escrita, pero entonces, hablar, ¿por qué se desglosa en hablar (expresión oral) y conversar (interacción oral), si escribir es solo expresión escrita y no interacción escrita? (Bordón, 2007: 86).

Esta crítica, aunque ella no se hace en este sentido, incide en algo ya dicho por otros autores y que es una de las debilidades del MCER para servir como referente en la construcción de tests, pues deja claro que, con algunas de sus escalas, se hace difícil definir un constructo de lengua para evaluar.

En un trabajo del Ministerio holandés para desarrollar un banco de ítems de comprensión (Alderson *et ál.*, 2004), se hacen análisis sobre la capacidad del MCER para guiar el trabajo de diseño y elaboración de especificaciones de tests para los diferentes niveles, se intentan identificar sus lagunas y, en sus conclusiones, se incluye una necesaria revisión crítica del MCER.

Entre los defectos encontrados por el grupo de expertos que trabajaron en el proyecto se encuentran:

- La experiencia de proyectos previos, incluyendo el Proyecto Dialang, sugiere que el MCER en su actual forma no da suficiente guía teórica y práctica de la que extraer las especificaciones para los niveles (Alderson *et ál.*, 2004: 1).

- Las definiciones de algunas tareas son problemáticas para desarrollar ítems y para los constructores de tests (Alderson *et ál.*, 2004:3).

- Hay deficiencias en la concepción de la comprensión y en la teoría de cómo esta funciona (Alderson *et ál.*, 2004: 9).

- En el texto se introducen muchos conceptos que luego no se reflejan en las escalas, como el de competencia, competencia general o competencia comunicativa (Alderson *et ál.*, 2004: 9).

- Aunque hay un capítulo entero dedicado a las tareas (el capítulo 7) en ningún momento se especifica cómo se pueden desarrollar distinguiendo su nivel o cómo distinguir niveles entre las diferentes tareas (Alderson *et ál.*, 2004: 10).

- Hay inconsistencias diversas y definiciones poco desarrolladas, como los términos -entre otros- “simple”, “frecuente”, “común”, “familiar”, “predecible” (Alderson *et ál.*, 2004: 12).

Y se concluye con 12 recomendaciones entre las que figura la necesidad de una guía más clara sobre los textos y tareas apropiadas para cada nivel¹⁵⁹, en términos de vocabulario y gramática necesaria, así como del desarrollo de la competencia en aspectos sociolingüísticos y pragmáticos (Alderson *et ál.*, 2004:19).

Ante las rápidas críticas por parte de evaluadores que entendían que en el MCER no había suficientes elementos para diseñar y evaluar, y el hecho de que se estaban realizando afirmaciones sin la debida cautela, sobre la adecuación y relación de tests con el MCER, el Consejo de Europa publicó un documento llamado *Relating Language Examinations to the CEF, Preliminary Pilot Version* (Consejo de Europa: 2002)¹⁶⁰.

La polémica, tras la publicación del documento llegó a la prensa en un artículo en *Guardian Weekly TEFL Supplement* del 18 de marzo de 2004, en el que Fulcher (2004 a) advertía del peligro de tomar el MCER como un sistema válido de evaluación¹⁶¹. Por lo que el mismo North (uno de los autores del MCER) contestó en el *Guardian Weekly* del 15 de abril (Alderson, 2005: 258).

Fulcher (2004 a) decía algunas cosas que son evidentes para los profesionales de la evaluación, pero que si no se hacen patentes pueden pasar inadvertidas para otros profesionales de la enseñanza de lenguas y para sus principales usuarios. En primer lugar, relataba el sistema de cómo se había diseñado la escala de seis niveles y concluía que los editores reconocen que el MCER,

No tiene una base teórica, e incluso lo que se recoge en las escalas no es la competencia del hablante sino la percepción que los profesores tienen de su

¹⁵⁹ Alderson *et ál.* (2004: 284), citando un artículo de Weir, inédito, (2005), “Wastage and Threshold. Or does the emperor have any clothes?” dice que “los documentos del Consejo de Europa respecto a los niveles Plataforma, Umbral y Avanzado no son de ayuda para diferenciar entre niveles: los autores dicen que es imposible especificar diferencias (al distinguir entre los niveles de los estudiantes de Plataforma y Umbral) (...) (Como consecuencia) las diferencias entre Plataforma y Umbral parecen mínimas”. Nosotros, en una investigación realizada con 50 profesores de lenguas (§3.1.7) hemos encontrado que las diferencias entre los que sería en A2 y B1 (plataforma y umbral, respectivamente), todavía parecen difíciles a algunos de los evaluadores. Y entre otros niveles, como el B2 y C1, parecen indiferenciadas o difíciles de interpretar. Pero las diferencias que hemos hallado más inconsistentes y difíciles son entre los niveles C1 y C2, según los propios comentarios de los evaluadores utilizados en nuestra investigación (§4.2) y según los resultados finales obtenidos (§4.5).

¹⁶⁰ Posteriormente han ido apareciendo nuevos documentos y versiones de este mismo documento junto a guías para ofrecer seminarios para calibrar exámenes de acuerdo con este documento, editadas por el Consejo de Europa en colaboración con la Universidad de Cambridge y Eurocentres (North / Lepage, 2005; North / Lepage, 2005 b; Jones, 2005). Además, el Consejo de Europa ha comenzado a publicar ejemplos de diferentes modelos de evaluación con sus calificaciones y comentarios justificando esas calificaciones. Véase la nota 12, en el capítulo 1. Como hemos señalado en ella, estos ejemplos no incluyen muestras en español.

¹⁶¹ Respuesta que contiene similares argumentos a los expuestos en Fulcher (2004 b).

competencia. En otras palabras, el MCER no es más que un conjunto de descriptores que reflejan eso sobre lo que un grupo de profesores de toda Europa se han puesto de acuerdo en que representa ser “más” competente o “menos” competente (Fulcher, 2004 a).

Pasaba luego a apuntar el peligro de que el MCER,

Se estuviera convirtiendo en un sistema a utilizar como norma. Así (señalaba) para los profesores es un problema porque les hace pensar que la escala representa una jerarquía de adquisición de la lengua, (cuando no lo es, como el mismo North admitía en su respuesta -North, 2004); para los usuarios, el peligro es que cualquier test que ya no dé sus resultados de acuerdo al MCER va a ser visto como inválido o sin reconocimiento;... y para los diseñadores de tests el peligro es que deseen relacionar las notas de sus tests con el sentido que esas notas tengan según el MCER, sólo para lograr ser reconocidos en Europa (Fulcher, 2004 a).

Pero terminaba diciendo: “esto ya ha comenzado”¹⁶².

Y advertía de que, además, hacer exámenes relacionados con el MCER no es una tarea sencilla, que la relación es sobre todo de forma intuitiva, y que para enseñar a hacerlo se había editado en la web un manual específico para paliar esto.

Todos estos fallos comentados los resume Weir (2005) en la imposibilidad de definir un constructo para evaluación, de diferenciar entre niveles¹⁶³ y, consecuentemente, de diseñar tests fiables para establecer un sistema de calificaciones y, en definitiva, para equiparar exámenes entre sí.

El MCER, en su forma actual, no es suficientemente comprensible, coherente o transparente para un uso profano en evaluación de lenguas. Las escalas no ofrecen suficiente información sobre cómo la variación en términos de parámetros contextuales puede afectar a la actuación subiendo o bajando el nivel o cumpliendo los objetivos de las declaraciones “puedo hacer”. Y falta un sistema de validación basado en una teoría de la evaluación. (Estos fallos en explicar) esos parámetros de validación basados en una teoría de evaluación y contexto- esto es, definiendo de forma clara el constructo que se desea evaluar- vician los intentos de usar el MCER como base para el desarrollo comparable de tests dentro, y a través, de diferentes lenguas y niveles. Lo que provoca falsas asunciones de equivalencia al situar en el mismo punto de la escala tests

¹⁶² Figueras aclara los términos de esta polémica: “La publicación de este manual (para relacionar exámenes con el MCER) no calma los ánimos, y el debate llega a la prensa. La respuesta por parte de North no se hace esperar: como uno de los autores del documento (North) insiste en que los objetivos del MCER son precisamente los de facilitar la referencia a unos niveles comunes y dar pautas para entender la docencia, el aprendizaje y la evaluación de lenguas, con la voluntad de sugerir y describir, no de prescribir” (Figueras, 2005: 7-8).

¹⁶³ Hemos encontrado algunas de estas críticas muy acertadas. Ya hemos señalado que en §3.1.7 presentamos una investigación sobre las escalas, que llegan a la misma conclusión respecto a su dificultad para diferenciar niveles. Y, en cierta medida, se llega a esa misma conclusión en la dificultad para utilizar las escalas al realizar las calificaciones y en los resultados obtenidos. Véase §4.2.

construidos para diferentes propósitos y audiencias, con diferentes constructos y de diferentes manera (Weir, 2005: 281-282).

Lo cual es, evidentemente, un espejismo.

Ante todas estas críticas, es fácil ver lo difícil que es diseñar un examen según el MCER. Los diseñadores de tests tienen que desarrollar ellos mismos el contenido de los tests ya que el MCER, aparte de una lista genérica de taxonomías, no ayuda a determinar qué puede o no ser apropiado para cada nivel, pues sólo se dan de forma explícita las actividades comunicativas y las estrategias¹⁶⁴ (Weir, 2005: 288). “El diseñador de tests que se encuentra con que tiene que decidir qué contenidos y actividades incluir en sus diferentes niveles no encuentra suficiente ayuda en el MCER para seleccionar materiales, textos o actividades relacionadas con el nivel de conocimiento requerido en un nivel particular” (Weir, 2005: 293) ni sobre el tipo de sintaxis que puede ayudarle a definir el nivel.

Y es igualmente complicado organizar un sistema válido de calificaciones (Weir, 2005: 296). En definitiva “unas comparaciones basadas solamente en las escalas podrían ser engañosas. El MCER tal y como se presenta actualmente no nos permite decir que unos tests son comparables, ni nos permite diseñar tests comparables” (Weir, 2005: 298).

Y el problema se agrava cuando las reticencias sobre lo que puede estar haciéndose, a fecha de 2007, son grandes, si leemos las palabras de advertencia de gente como Alderson (2007) o Little (2007), quienes muestran cautela respecto a los propios resultados que agencias como DIALANG o ALTE estén dando respecto a sus tests.

Little dice que aunque el MCER no ha tenido mucha influencia en exámenes oficiales de entrada a cursos en los sistemas de educación europeos y aunque DIALANG ha probado que se pueden realizar tests según las bases de los descriptores y escalas del MCER¹⁶⁵, las agencias de exámenes pueden hacer usos indebidos del MCER (Little, 2007: 649).

Y Alderson (2007: 662) va más lejos y se permite criticar a ALTE diciendo que “es claro su interés en aceptar las afirmaciones de sus miembros de que sus tests están bien

¹⁶⁴ Estas se explican en el MCER en el apartado 4.4.

¹⁶⁵ Algo que nosotros también hemos podido comprobar mediante nuestro test y de lo que tenemos resultados estadísticos fiables; véase el capítulo 4 y el Apéndice 10.

validados, y que ello le resta credibilidad”. Por ello le “parece que solo EALTA¹⁶⁶ es suficientemente independiente y en sus directrices para una buena práctica¹⁶⁷ se refiere de forma explícita a la necesidad de evidencias empíricas en las que apoyar las afirmaciones que anclan las prácticas al MCER”.

2.7.5 El documento para relacionar exámenes con los niveles del MCER

De todas las críticas señaladas, nos quedamos con una que consideramos grave: el hecho de que el MCER no sea una guía adecuada para evaluar ni para equiparar exámenes, que es uno de sus objetivos. Grave porque esta carencia no se aprecia claramente, dando a entender que con él, mediante la reflexión y discusión que propone y promueve, se puede hacer esta difícilísima labor.

Este hecho se reconoció finalmente en el documento *Relating Language Examinations to the CEF, Preliminary Pilot Versión of the Manual* (Consejo de Europa, 2002), donde se advertía de su uso para relacionar tests con las escalas del MCER (tanto los existentes con anterioridad como otros creados de forma independiente utilizando el MCER), pero no para probar que esos tests, sean los que sean, son tests válidos y fiables¹⁶⁸. Ya que antes de proceder a cualquier relación o prueba de un test con el MCER, es responsabilidad del examinador demostrar la validez de su examen mostrando que este evalúa el constructo deseado. Y, además, se aclara que este hecho es de enorme dificultad (Consejo de Europa, 2003: 2) pues requiere unos conocimientos sobre evaluación que no están al alcance de todos¹⁶⁹.

El capítulo 6 del manual explica algunos conceptos del diseño de tests: validación externa e interna (pág. 100), teoría clásica de tests e TRI (*Item Response Theory*) (pág.

¹⁶⁶ EALTA es una organización independiente formada en 2004 que tiene entre sus objetivos ayudar a la formación de profesionales de la evaluación.

¹⁶⁷ *Directrices de EALTA para una buena práctica en el desarrollo de pruebas y la evaluación de lenguas* (EALTA, 2004). En: www.ealta.eu.org/documents/archive/guidelines/Spanish.pdf.

¹⁶⁸ En realidad, el manual es parte de un proyecto piloto para adecuar tests al MCER, “del que se dio la oportuna información en *Language Testing*, 7, 22, 2005, 262-279 o en *Language Assessment Quarterly: An International Journal*, Vol. 3, 1, 2006, 3-30” (Consejo de Europa, 2007: 6) y que ha continuado desde 2003 con nuevas versiones o resúmenes del manual (Consejo de Europa, 2005) y al que se le han ido aportando trabajos y ejemplos que se recogen en la red o en CD, DVD y CD-ROM, sobre tests y muestras de lengua de inglés, alemán, francés, portugués, italiano y español (2007: 4), (de esta lengua sólo con ítems de lectura y comprensión oral) que se encuentran en www.coe.int/portfolio.

¹⁶⁹ Que un test no lo puede diseñar cualquiera, ni usando el MCER, se reconoce en la siguiente cita: “la validación empírica requiere considerable conocimiento sobre sicometría y construcción de tests. Si el proveedor de tests carece de tal experiencia se recomienda que se forme o solicite los servicios de cualificados sicometristas” (Consejo de Europa, 2003: 4).

103), análisis cualitativo (pág. 105), métodos de recogida de datos sobre el funcionamiento del test y análisis factorial (pág. 106), correlaciones (pág. 109), etc. Con esto queda clara la necesidad de profesionales formados para desarrollar tests y validarlos de acuerdo con todos los criterios científicos.

Ya hemos visto que el MCER no enseñaba a diseñar tests. Pero, si este manual tampoco enseña a diseñar tests de acuerdo con el MCER, ¿cuál es su sentido? Dos fundamentalmente: 1) aclarar que el diseño correcto y validado de un test, de acuerdo con los estándares normales de la teoría de la evaluación, es una tarea previa a cualquier otra tarea, y que, por lo tanto, todo test que pretenda estar diseñado mediante los contenidos u objetivos del MCER no puede escapar a esta obligación previa; y 2) ofrecer unos mecanismos y procedimientos burocráticos para que cualquier test, por ejemplo, uno que acabáramos de diseñar siguiendo los objetivos de determinadas escalas del MCER y hubiéramos validado siguiendo los criterios de la profesión de evaluador, pueda demostrarse que es un test que se corresponde con el MCER, no porque lo diga su diseñador sino porque cumpla con unos pasos que el Consejo de Europa determina para aceptar tal test.

Para esta segunda actividad, la única de la que se ocupa este manual, se determinan una serie de pasos y actividades y unas fichas para cumplimentar con los resultados de esos pasos, que tienen como objeto demostrar que se han seguido adecuadamente¹⁷⁰. Las fichas-formulario son 28, además de un formulario base para solicitar la idoneidad final del test en relación con el MCER, así como otras para hacer llegar opiniones en forma de feedback a la organización, sobre el procedimiento.

No obstante, el manual también reconoce y acepta que los pasos, procedimientos y formularios son sólo uno de los caminos que se puede seguir para demostrar la adecuación del test al MCER, ya que en la teoría de la evaluación existen otros procedimientos que se pueden seguir para demostrar que un test evalúa un constructo determinado o es paralelo a otro test, sea el MCER o cualquier otro (Consejo de Europa, 2003: 24).

Los objetivos del documento son principalmente tres:

¹⁷⁰ Estos pasos tienen mucho parecido con los establecidos en el documento de ALTE, presentado en la conferencia de EALTA, 2005, por los miembros del grupo de trabajo del código de prácticas de ALTE, Piet van Avermaet, Henk Kuijper, Nick Saville, titulado *Setting and Monitoring Professional Standards: a QMS approach* (www.cambridgeesol.org/rs_notes/rs_nts22.pdf, 21 de diciembre de 2007).

- Describir los procedimientos de análisis, administración y alcance de un test.
- Explicar los resultados y asociarlos a las escalas del capítulo 3 del MCER.
- Facilitar las evidencias que prueben que los procedimientos para hacer esa relación entre el test y el MCER se han seguido debidamente (Consejo de Europa, 2003: 2).

Pero desde el principio se dice que no establece normas sobre ningún procedimiento para construir tests. Sino que “se deja bien claro que es de suma importancia informar sobre lo que se ha hecho para relacionar un test o examen con el MCER” (Figueras, 2005: 268).

Este objetivo de informar adecuadamente, se consigue mediante cuatro pasos a los que se dedica cada uno de los capítulos centrales del documento¹⁷¹.

1- Familiarización con el MCER, mediante una serie de actividades que prueban que se ha reflexionado sobre su contenido y se comprende (capítulo 3); por ejemplo, mediante una discusión sobre el concepto y contenido de los niveles del MCER como un todo, utilizando la tabla 1 del capítulo 3 del MCER¹⁷².

2- Especificación del contenido y tareas del test que se quiere relacionar con el MCER, mostrando que esto se refiere verdaderamente a lo que contiene el MCER (capítulo 4); por ejemplo, describiendo el contenido de las tareas diseñadas en el test e indicando la relación con las escalas del MCER. Así, en el formulario A9 se debe señalar qué tipo de actividades va a medirse con el nuevo test (Consejo de Europa, 2004: 42), o el formulario A3 pide información respecto a los procedimientos de calificación que se han determinado para juzgar el resultado de las tareas (Consejo de Europa, 2004: 38).

3- Estandarización (capítulo 5): procedimientos para asegurarse de que todos los participantes en el proceso, en la elaboración del test o en su administración conocen el MCER y lo entienden y utilizan de la misma manera, así como para asegurarse de que no va a haber variaciones en la formas de actuar. Este capítulo se ocupa de lo que en

¹⁷¹ Sus capítulos son 1. Introducción. 2. El MCER. 3. Familiarización. 4. Especificación. 5. Estandarización. 6. Validación empírica. 7. Protocolo para dar parte. En el Apéndice 4 se dan los formularios oficiales del Consejo de Europa para demostrar que un test se adecua a lo que el MCER pretende. Los hemos rellenado para mostrar que nuestro nuevo test del MCER cumple sus condiciones. En el capítulo 3 de esta tesis hemos utilizado un sistema de validación independiente.

¹⁷² Cuadro 1 de *Niveles comunes de referencia: escala global*.

teoría de evaluación se llama fiabilidad ya que advierte respecto a la invalidez del test y de su no relación con el MCER si se carece de esta (Consejo de Europa, 2004: 66).

4- Validación empírica: recopilación de los datos y análisis que prueban que el test es un test válido y que además se relaciona con el MCER (capítulo 6).

Todo lo cual supone la toma de conciencia sobre la necesidad de un curso de formación de evaluadores, necesario pero todavía insuficiente para lograr diseñar tests adecuadamente.

2.7.6 Conclusiones sobre el MCER

El MCER, como buen documento de política lingüística, tiene el objeto de marcar un camino a recorrer. Eso es lo que está pasando con sus derivaciones en los otros documentos que se han señalado, el PEL, el documento para referir tests al MCER, diseños curriculares, etc. Pero tiene el peligro sobre el que ya advirtió Fulcher, en el suplemento educativo de *Guardian Weekly* del 18 de marzo de 2004, de que en vez de que se tome como referencia para debate y reflexión se tome como una norma o sea malentendido.

Advierte de que desde que existe se está malentendiendo el proceso de validación de tests. Este proceso debe consistir en demostrar que un test evalúa lo que se pretende. Pero, en su lugar, se está derivando hacia la mera demostración de que un test está diseñado de acuerdo con el MCER, lo cual es insuficiente y no prueba que el test sea válido. El mismo manual para relacionar exámenes con el MCER advierte de que eso es erróneo, ya que dice claramente que, antes de relacionarse con el MCER, debe demostrarse que un test está bien diseñado, es válido y fiable (Fulcher, 2004 a).

Este documento político tiene su punto más alto en la descripción de las escalas de seis niveles, que no obstante han sido también criticadas. Pero, sean o no aceptables, el hecho de que estén ahí no es suficiente para llegar a ellas, pedagógicamente, sino que representan el final de un trayecto; y tampoco son efectivas desde el punto de vista de la evaluación ya que carecen de contenido que permita desarrollar tests. Además, más importante, para diseñar y poder usar un test es necesario dar todos los pasos señalados en el capítulo 2 de esta tesis y que el MCER no facilita.

Lo mismo podríamos decir de otras prácticas que se están llevando a cabo basadas en el MCER, como la publicación de materiales didácticos y de diseños curriculares. Estos últimos no se pueden tomar por la norma, pues esto es contrario al espíritu y concepto del mismo MCER de servir para prácticas diversas y estar abierto a adaptaciones a diferentes situaciones. El primero, el diseño de materiales, también debería estar sometido, aunque así no se haya hecho, a ciertas garantías de objetividad como las que se exige a los tests, para demostrar que un texto es de B1 o B2 porque sus actividades, contenidos y tareas se relacionan con los niveles del MCER y no solo porque lo decidan¹⁷³ sus editores, cosas que están, desgraciadamente sucediendo, sin datos objetivos cuantificables. Pero esta petición, ya expresada antes por Alderson, no ha surtido efecto¹⁷⁴.

2.8 Equiparaciones propuestas entre los niveles del MCER y las *guidelines* de ACTFL

Otros autores también han pronosticado, sin trabajos empíricos que los avalen, equiparaciones especulativas entre los niveles del MCER y otras formas de calificar y presentar los niveles de habilidad en el uso de la lengua, o están haciendo intentos de equiparación entre ambas escalas mediante otras investigaciones similares a la nuestra, en otras lenguas¹⁷⁵.

Bordón (2006: 86) expone las correspondencias entre los niveles del MCER y los clásicos niveles de cursos de ELE.

Buitrago (2006, artículo no publicado) especula sobre las relaciones, en un cuadro en el que incluye también los niveles del DELE, ALTE y el Consejo de Europa.

¹⁷³ El MCER necesita debate, reflexión y desarrollos; no directrices. Como dijo North en respuesta a Fulcher, también en *Guardian Weekly*, el 15 de abril de 2004.

¹⁷⁴ Alderson (2007: 661-662) se queja de que el portfolio tiene un comité europeo de validación de los portfolios que se le presentan, pero no ha querido hacer lo mismo con los tests ni materiales de enseñanza. Es una queja amarga porque dice que “Desgraciadamente, y a pesar de la influencia que los exámenes tienen en el currículo de lenguas -y en las vidas-, el Consejo de Europa ha rechazado establecer un mecanismo semejante para validar o inspeccionar las afirmaciones hechas por proveedores de exámenes o diseñadores de libros de texto. Y esta renuncia es, a mi modo de ver, un serio abandono de sus responsabilidades profesionales”.

¹⁷⁵ Hemos sabido que el Profesor Arthur Mosher, del departamento de Lenguas de la Universidad de Dayton (Estados Unidos) está realizando una investigación con el alemán. Nos hemos intentado poner en contacto con él para obtener los datos de su investigación, pero no hemos obtenido respuesta.

Y Tschirner (2007) también da su opinión especulativa sobre una posible equiparación entre los niveles de ACTFL y los del MCER, con la lengua alemana¹⁷⁶.

Vamos a dar aquí los cuadros de estas especulaciones, que volveremos a comentar en §4.5, una vez hayamos terminado la investigación y podamos contrastar los resultados con los de estos autores.

Bordón (2006: 86) dice que los niveles del MCER se corresponden con los de los cursos de ELE, así:

C2	superior
C1	avanzado
B2	avanzado: un intermedio alto próximo a un avanzado
B1	intermedio medio
A2	nivel inicial o de principiante
A1	nivel inicial o de principiante

Buitrago (2006) hace un cuadro con cinco comparaciones entre los niveles del MCER, de los DELE, de las escalas de ALTE, de los niveles de Wilkins para el Consejo de Europa y la escala de ACTFL¹⁷⁷:

ACTFL	NL	NM	NH	IL	IM	IH	AL	AH	S
Consejo de Europa	Nivel ruptura	Nivel supervivencia.	Nivel umbral.	Nivel ventaja.	Nivel comp. operativa efectiva.	Nivel dominio.	Bilingüismo.		
ALTE	Acceso.	Plataforma.	Umbral.	Usuario independiente.	Usuario competente.	Usuario óptimo.			
DELE		Nivel inicial.	Nivel Intermedio.			Nivel superior.			
MCER	A1 Acceso.	A2 Plataforma.	B1 Umbral.	B2 Avanzado.	C1 Dominio operativo eficaz.	C2 Maestría.			

¹⁷⁶ Agradezco los comentarios respecto a este artículo al Dr. De Graff a quien debo su traducción del alemán.

¹⁷⁷ Extrañamente, en esta escala, los niveles de ACTFL son solamente 9, cuando desde 1999 se ha introducido el avanzado medio y la escala contiene 10 niveles.

Este cuadro muestra cómo las escalas del MCER son las mismas que en ALTE, de donde efectivamente proceden, ya que su coincidencia es total. A este respecto cabe decir que los descriptores de dominio del MCER son deudores de esta organización y de cuantos han trabajado colaborando con ella y con el Consejo de Europa para establecer niveles de dominio de la lengua. Las escalas proceden, por lo tanto, de otras dedicadas a evaluación, aunque ahora se presentan para todo tipo de usos pedagógicos.

Comparando los niveles del MCER y ACTFL, quedarían las siguientes equivalencias¹⁷⁸:

MCER	A1 Acceso	A2 Plataforma		B1 Umbral		B2 Avanzado		C1 Dominio operativo eficaz.		C2 Maestría
ACTFL	NL	NM	NH	IL	IM	IH	AL	AH	S	

Por último Tschirner (2005), que es miembro de ACTFL y evaluador de OPI, pronostica una posible equiparación entre los niveles del MCER y ACTFL, que tendría los siguientes resultados:

MCER	ACTFL
C2	Superior
C1	Advanced High
B2	Advanced Mid
B1	Intermediate High
A2	Intermediate Mid
A1	Novice High

De todos estos intentos especulativos de comparación nos llama la atención el hecho de que se equipare siempre el máximo nivel en una escala con el máximo de otra, dando por supuesto que se evalúan los mismos constructos de lengua hasta los mismos niveles de dominio, como si no se pudieran establecer nuevas delimitaciones del dominio por encima de lo ya delimitado o como si todos los sistemas establecieran el nivel máximo de la misma manera.

¹⁷⁸ Estas no están tan lejos de lo que nuestra investigación finalmente ha encontrado, aunque hay ciertas diferencias. Véase §4.5.5.

En ACTFL vemos, (§2.6), que el nivel superior establece una serie de habilidades que se pueden superar por encima y así lo hace la escala de ILR. Igualmente, la escala de ACTFL parece más alta que otras, pues exige un dominio de lenguaje abstracto defendiendo opiniones y haciendo hipótesis, es decir, habilidades por encima de las habituales a cualquier hablante. Bien podría ser que cualquiera de las escalas que se intentan equiparar llegara a unos niveles más altos o más bajos que las otras, según las diferentes definiciones de sus constructos.

Pero esto no parece considerarse nunca, lo que denota una carencia en la concepción de la evaluación oral de la lengua, dando por supuesto que el nivel más alto no necesita definiciones, ya que es el más alto posible, y olvidando que la descripción de la lengua es una tarea difícil y complicada, también en los niveles altos, sean de bilingüe, de maestría, de dominio eficaz, etc. y que con cada test se evalúa solamente aquello que se ha definido.

CAPÍTULO 3

TRABAJO DE CAMPO. ANÁLISIS DE LAS ESCALAS DEL MCER Y ACTFL Y DISEÑO, VALIDACIÓN Y USO DE UN TEST ORAL PARA EVALUAR SEGÚN EL MCER

En este capítulo se explica el trabajo de campo realizado para llevar a efecto la comparación entre los dos sistemas de evaluación oral: el de la OPI de ACTFL y el establecido mediante las escalas de habilidad y dominio del MCER.

En §3.1 se realiza un análisis crítico de algunas escalas del MCER, con el fin de dilucidar el concepto de lengua que subyace a ellas, y se presenta un experimento realizado con 50 profesores y expertos con el fin de comprobar nuestras interpretaciones respecto a su ambigüedad.

En §3.2 se hace un análisis comparativo entre algunas escalas del MCER y la de las *guidelines* de ACTFL y con sus resultados hacemos la primera propuesta sobre la equiparación entre ambos modelos de evaluar y escalonar la lengua; propuesta en la que nos hemos basado para hacer las hipótesis de investigación de esta tesis.

En 3.3, y a partir de él, se describe el trabajo de diseño del nuevo test realizado a partir del MCER: el test, su forma de administración, protocolo de realización de las entrevistas y los candidatos que participaron en la investigación.

En §3.4 se describen las tareas de este nuevo test diseñado expresamente para la investigación, basado en el MCER. Se describen las tareas de que se compone, el

lenguaje que pretenden elicitarse y evaluar, las escalas de medición de habilidades que se utilizaron para diseñarlas y las que se usó para calificar a los candidatos.

En §3.5 se explican las precauciones tomadas para obtener suficientes garantías para que fuera posible hacer las comparaciones entre los resultados obtenidos con él y con la OPI.

El §3.6 se dedica a explicar la validación del test mediante las pruebas piloto hechas con cada tarea, con profesores expertos y con candidatos similares a los que luego serían sus usuarios, para comprobar su utilidad; y para demostrar que es efectivo, válido y objetivo.

Estos pasos los hicimos siguiendo los métodos normales que se utilizan en evaluación¹⁷⁹.

Por último, §3.7 presenta el test del MCER con una ficha de sus características.

En los capítulos siguientes se analizan los resultados de utilizarlo para evaluar, junto a los de la OPI, y se elaboran las conclusiones y recomendaciones a las que se ha llegado con esta investigación.

3.1 Análisis de las escalas del MCER

3.1.1 Análisis crítico de los niveles establecidos por el MCER

El MCER “establece una serie ascendente de niveles comunes de referencia para describir el dominio que el alumno tiene de la lengua” (Consejo de Europa, 2002: 16). Estos niveles están descritos siguiendo unos criterios y una teoría sobre la competencia lingüística con el fin de que:

b. Las diferencias puedan ser apreciadas razonable y coherentemente.

¹⁷⁹ Este análisis se hizo siguiendo a Alderson (1995) y Bachman / Palmer (1996) cuyos conceptos de evaluación y sistemas de asegurarse de que esta es válida, fiable y objetiva ya han sido expuestos en el capítulo 2.

- c. *Las descripciones sean breves, claras, transparentes y formuladas en términos positivos.*
- d. *Los descriptores tengan una integridad independiente y no necesiten a otros descriptores para su interpretación.*
- e. *Todas las escalas sean coherentes desde el punto de vista del contenido, es decir, que entre sí tengan unos valores muy similares (Consejo de Europa, 2002: 197-199).*

Esto se pretende, según confiesa el principal creador de las escalas del MCER, con unas escalas que han sido realizadas para no repetir los fallos de las escalas existentes hasta el momento, entre los cuales está que:

- *La forma de definir los descriptores es relativa. Sólo se entiende el sentido de un descriptor del nivel 4 cuando se lee en relación con los del nivel 3 y 5.*
- *Son circulares: las tareas del nivel 4 se consideran del nivel 4 porque los estudiantes del nivel 4 las pueden realizar. Y los estudiantes son de nivel 4 porque pueden realizar las tareas del nivel (North, 2002: 88).*

Asimismo la confección de las escalas se realizó con unas ambiciones muy altas:

Identificar categorías que puedan, por un lado basarse en una teoría y, al mismo tiempo, sean categorías con las que se puedan relacionar los profesores y estudiantes; que, además, sean categorías escalonables; y que puedan ser interpretadas de forma común en varios contextos de forma que pueda decirse que poseen en común los mismos valores (North, 2002: 88)

Con el fin de proceder a comparar el MCER con la escala de ACTFL, hemos tenido que comenzar por analizar las escalas del MCER para ver si estas entre sí definen siempre los mismos niveles de referencia y presentan las equivalencias que se suponen, tal y como acabamos de decir. Para comprobar y comentar estos extremos hemos hecho una cata en varias escalas y hemos visto que el resultado se aleja de lo afirmado tanto por North como por el MCER en varios aspectos que nos parecen delicados y que pueden comprometer el sistema de evaluación¹⁸⁰.

Lo primero que nos llama la atención es que creemos que los seis niveles descritos no están claros ni muestran una progresión suficientemente diferenciada. Aunque se presentan en tres grupos, denominados usuario básico (A1 y A2), usuario independiente (B1 y B2) y usuario competente (C1 y C2) (Consejo de Europa, 2002: 25), estos no están claramente delimitados ni agrupados. Más bien se aprecia una continuidad entre

¹⁸⁰ Nos parece muy bien el intento de North y su equipo de Eurocentres, así como el sistema utilizado que explica en North (2002). No obstante creemos, y en este apartado se demuestra mediante nuestra propia investigación con 50 profesores de lengua, que los resultados no son ni los esperados, ni los que se anuncian en el MCER.

A1, A2 y B1, un cambio real al pasar a B2, y otra vez una continuidad entre B2, C1 y C2. Lo que nos lleva a dudar de la diferenciación básica en tres grupos y de la existencia de unos criterios comunes a cada uno. No vemos que las diferencias marcadas entre los niveles sean suficientes como para mostrar seis: podríamos establecer tres fácilmente pero, a partir de ahí, sería muy difícil observar diferencias objetivas para diferenciar más, suficientes diferencias como para que todas las personas que tengan que utilizar el MCER puedan estar de acuerdo.

Así el MCER establecería en realidad solamente tres niveles:

- A1 y A2.
- B1, todavía con algunos problemas para diferenciarlo de A2.
- B2, C1 y C2¹⁸¹.

Lo segundo que queremos destacar es que entre la descripción de las diferentes escalas no encontramos el cumplimiento de las características señaladas en los puntos b., c., d. y e. del MCER. Las diferencias no se ven fácilmente, sus descriptores no son claros y transparentes, sino que necesitan a otros para su interpretación y no nos parece que las escalas sean coherentes o tengan unos valores similares. Esto lleva a la necesidad de recibir un curso de formación que ayude a interpretar unívocamente los niveles¹⁸², pero ni tal curso está planeado en el MCER ni es muy posible hacerlo, ni siquiera mediante el manual piloto para relacionar exámenes con el MCER (Consejo de Europa, 2003), ya que el MCER es una guía de la que se pueden extraer numerosos tests y formas de evaluar y ello imposibilita el diseño de un curso general de evaluación.

Para comprobar estos términos mostramos los análisis realizados con varias de las escalas de habilidad del MCER y sus descriptores: la de *Autoevaluación: Expresión*

¹⁸¹ Claro que estos tres niveles son fácilmente subdivisibles hasta 5 si se utilizan exclusivamente criterios de corrección formal, como apuntamos en el capítulo 4, al hacer la corrección de entrevistas reales: así se puede diferenciar entre un C1, con errores, y un C2, sin errores; igualmente entre un A1, cuyas carencias de vocabulario y estructura serían enormes, y un A2 que ya lograría solucionar algunos de los problemas del A1. Pero no creemos que estos criterios sean suficientes ante una verdadera evaluación oral, donde la habilidad de usar la lengua debe definirse con elementos y criterios funcionales y comunicativos.

¹⁸² Este curso de formación es muy difícil de realizar, ya que con el MCER se pueden elaborar muchos diferentes tests para objetivos y candidatos diversos. No existe un modelo de test del MCER, por lo que no se puede dar un curso de formación que pueda homogeneizar las formas de evaluar y de usar los criterios y escalas para ese test inexistente.

oral, la de *Expresión oral en general*, la de *Monólogo sostenido: argumentación* y la de *Niveles cualitativos de referencia*¹⁸³.

Como, tras estos análisis, nos parecieron muchas las ambigüedades y fallos encontrados en las descripciones de las escalas, nos vimos obligados a realizar también un experimento objetivo al respecto, el cual nos llevó a reafirmarnos en nuestras conclusiones. Esto se ve en el capítulo 3, (§3.1.7). Allí se expone este experimento llevado a cabo con 50 profesores de español como lengua extranjera, en tres grupos que se explican en §3.1.7- A, §3.1.7- B y §3.1.7- C. Sus resultados conjuntos se ofrecen en §3.1.8.

3.1.2 Análisis general de los descriptores y de las escalas del MCER

A continuación, presentamos un análisis sobre algunos de los descriptores de nivel usados en las escalas del MCER, para ver si cumplen las características señaladas de claridad, independencia y coherencia, que deben permitir a los usuarios establecer objetivamente las calificaciones cuando utilicen el MCER. Como acabamos de decir, creemos que tales características no se cumplen adecuadamente. Vamos a verlo.

El descriptor de nivel B1 en la escala de *Expresión oral en general* dice que el estudiante “Puede llevar a cabo, con razonable fluidez, una descripción sencilla de una variedad de temas que sean de su interés, presentándolos como una secuencia lineal de elementos” (Consejo de Europa, 2002: 62). Esto se corresponde, en grado de habilidad y dominio, con el descriptor de B1 de la escala de *Desarrollo de descripciones y narraciones*, que dice de las habilidades del estudiante: “Realiza con razonable fluidez narraciones o descripciones sencillas mediante una secuencia lineal de elementos” (Consejo de Europa, 2002: 122).

¹⁸³ Además de estas escalas, en el capítulo 4, dedicado a las calificaciones y evaluación de las entrevistas y a los análisis de los resultados, emplearemos otras como las de *Coherencia y cohesión*, *Flexibilidad*, *Corrección gramatical*, *Control y corrección*, *Compensación*, *Dominio de vocabulario*, *Riqueza de vocabulario*, *Expresión oral en general*, *Autoevaluación* y *Competencia lingüística general*.

	<i>Expresión oral en general</i>		<i>Desarrollo de descripciones y narraciones</i>
B1	Puede llevar a cabo, con razonable fluidez, una descripción sencilla de una variedad de temas que sean de su interés, presentándolos como una secuencia lineal de elementos	=	Realiza con razonable fluidez narraciones o descripciones sencillas mediante una secuencia lineal de elementos

Si analizamos estos dos descriptores, se ve claramente que casan entre sí sin problemas. Ambos dicen del hablante que “con razonable fluidez” “puede llevar a cabo” o “realiza” “una descripción sencilla” o “descripciones sencillas”, “presentándolos como una secuencia lineal de elementos” o “mediante una secuencia lineal de elementos”. En lo único que difieren es en que en el descriptor de la escala de *Expresión oral en general* se habla además de “una variedad de temas que sean de su interés” y en el de la escala de *Desarrollo de descripciones y narraciones* no se habla de qué temas, pero se amplía la habilidad también a narraciones sencillas, ya que la escala trata de ello. En el siguiente cuadro se aprecia esta diferencia y las demás similitudes:

	<i>Expresión oral en general</i>		<i>Desarrollo de descripciones y narraciones</i>
B1	“con razonable fluidez” “puede llevar a cabo” “una descripción sencilla” “secuencia lineal de elementos” “una variedad de temas de su interés”	= = = = =/ =	“con razonable fluidez” “realiza” “descripciones sencillas” “secuencia lineal de elementos” “narraciones sencillas”

En este caso, vemos que, efectivamente, ambas escalas utilizan descriptores sencillos y equivalentes para sus niveles. Pero esta certera definición entre habilidades y competencias no es tan sencilla de encontrar en otros casos. En otras escalas hay problemas y ambigüedades de interpretación que obligan a tomar decisiones a menudo muy difíciles o a intentar llegar a acuerdos entre evaluadores¹⁸⁴.

¹⁸⁴ Y no solamente en la competencia oral, sino en otras competencias, como por ejemplo con las escalas de competencia escrita, según Vázquez Macías, (2006), (memoria dirigida por este autor, Martínez Baztán, no publicada, del I Máster en enseñanza de español como lengua extranjera de la Universidad de Granada).

Así, este descriptor citado de la escala de *Expresión oral en general*, de nivel B1, “Puede llevar a cabo, con razonable fluidez, una descripción sencilla de una variedad de temas que sean de su interés, presentándolos como una secuencia lineal de elementos” (Consejo de Europa, 2002: 62), no tiene tan fácil correspondencia con el B1 de la escala de *Monólogo sostenido: argumentación*, que dice “Ofrece breves razonamientos y explicaciones de opiniones, planes y acciones” (Consejo de Europa, 2002: 63). Ambos se enfrentan en el siguiente cuadro:

	<i>Expresión oral en general</i>		<i>Monólogo sostenido: argumentación</i>
B1	Puede llevar a cabo, con razonable fluidez, una descripción sencilla de una variedad de temas que sean de su interés, presentándolos como una secuencia lineal de elementos (B1)	=/=	Ofrece breves razonamientos y explicaciones de opiniones, planes y acciones (B1.1)

En primer lugar, porque no hay nada que los asemeje (léanse de nuevo: “Puede llevar a cabo, con *razonable fluidez*, una *descripción* sencilla de una *variedad de temas* que sean *de su interés*, presentándolos como una *secuencia lineal* de elementos” y “Ofrece breves razonamientos y explicaciones de opiniones, planes y acciones”). Y además, porque en uno se habla del grado de fluidez (razonable), de la variedad de temas, alcance y grado de complejidad de estos (de su interés), de una función (descripción) que se define por su complejidad (sencilla), y de la organización secuencial del discurso (secuencia lineal de elementos), mientras el otro descriptor se limita a hablar de temas (opiniones, planes y acciones), tareas que se puede hacer respecto a ellos (razonamientos y explicaciones) y grado de complejidad de esas tareas (breves). Es decir, no son comparables en contenido ni alcance, como en el siguiente cuadro se muestra:

	<i>Expresión oral en general</i>		<i>Monólogo sostenido: argumentación</i>
B1	<ul style="list-style-type: none"> - grado de fluidez >> razonable. - temas y alcance >> su interés. - función >> descripción - complejidad >> sencilla - discurso >> secuencia lineal. 		<ul style="list-style-type: none"> - temas >> opiniones, planes, acciones. - funciones >> razonamientos y explicaciones. - complejidad de las tareas >> breves.

Si no nos hubieran advertido de que ambos descriptores marcan el mismo nivel B1, difícilmente podríamos saberlo. Es decir, si decimos de un candidato del que tenemos su discurso grabado en una cinta, que es capaz de “ofrecer breves razonamientos y explicaciones de opiniones, planes y acciones” (B1.1) y utilizamos su mismo discurso para hacer una apreciación general de su habilidad, utilizando la escala de *Expresión oral en general*, no será sencillo llegar a la conclusión de que “lleva a cabo, con razonable fluidez, una descripción sencilla de una variedad de temas que sean de su interés, presentándolos como una secuencia lineal de elementos” (B1).

Aunque una solución puede ser interpretar el descriptor de la escala aceptando o deduciendo que “ofrecer breves razonamientos” es en cierta medida equivalente a presentar “una variedad de temas”, “con razonable fluidez” en una “secuencia lineal de elementos”, como se muestra en el cuadro siguiente:

<i>Expresión oral en general, B1</i>		<i>Monólogo sostenido: argumentación, B1.1</i>
- “una variedad de temas” - “con razonable fluidez” - “presentándolos como una secuencia lineal de elementos”	=	- ofrecer breves razonamientos...

Ya que razonar lleva implícito, como mínimo, cierta secuenciación. Así, habríamos equiparado ambos descriptores, aunque un tanto forzadamente. El resultado dependerá de la interpretación que cada usuario evaluador haga de la escala¹⁸⁵.

Además, esta interpretación que acabamos de proponer nos lleva a un nuevo problema que vamos a intentar explicar. El nivel B1 de la escala de *Monólogo sostenido: argumentación* está dividido en dos subniveles (B1.1 y B1.2), mientras que la otra escala (de *Expresión oral en general*) con la que estamos comparándola no hace subdivisiones. Acabamos de equiparar el B1 de esta escala con el B1.1 de la otra, pero si ahora atendemos al B1.2 de esta (*Monólogo sostenido: argumentación*), que dice “Desarrolla argumentos lo bastante bien como para que se puedan comprender sin dificultad la mayor parte del tiempo”, entonces ya este no se puede equiparar con el B1

¹⁸⁵ Esto es así porque el marco no establece las diferencias basándose en mayores o menores habilidades funcionales, sino por cantidad de discurso o tipo de discurso, lo que es difícil de averiguar en la práctica. ACTFL, sin embargo, marca las diferencias basándose en elementos que hacen que un discurso sea claramente de más nivel: elementos de discurso muy delimitados, ya explicados en §2.6.2.

de la de *Expresión oral en general*, que parece de menor nivel aunque ambas sean B1. El B1.2 de *Monólogo sostenido: argumentación* parece más cercano al B2 de *Expresión oral en general*, como se muestra en el siguiente cuadro:

	<i>Monólogo sostenido: argumentación</i>		<i>Expresión oral en general</i>
B1.2	Desarrolla argumentos lo bastante bien como para que se puedan comprender sin dificultad la mayor parte del tiempo.	B2 ==	Realiza descripciones y presentaciones claras y detalladas sobre una serie de asuntos relacionados con su especialidad, ampliando y defendiendo sus ideas con aspectos complementarios y ejemplos relevantes.
B1.2	Desarrolla argumentos lo bastante bien como para que se puedan comprender sin dificultad la mayor parte del tiempo.	B1 =/=	Puede llevar a cabo, con razonable fluidez, una descripción sencilla de una variedad de temas que sean de su interés, presentándolos como una secuencia lineal de elementos.
B1.1	Ofrece breves razonamientos y explicaciones de opiniones, planes y acciones.	B1 ==	Puede llevar a cabo, con razonable fluidez, una descripción sencilla de una variedad de temas que sean de su interés, presentándolos como una secuencia lineal de elementos.

En general, es bastante difícil interpretar y diferenciar los niveles que expresan algunos de estos descriptores. Por ejemplo, a la luz de lo dicho hasta ahora, es difícil ver si el descriptor B2 de la escala de *Monólogo sostenido: argumentación* equivale al B2 de la escala de *Expresión oral en general* o representa más habilidad. Es lo que intentamos mostrar en el cuadro siguiente:

	<i>Monólogo sostenido: argumentación</i>		<i>Expresión oral en general</i>
B2	Desarrolla argumentos con claridad, ampliando con cierta extensión y defendiendo sus puntos de vista con ideas complementarias y ejemplos adecuados. Construye cadenas argumentales razonadas. Explica puntos de vista sobre un tema, proponiendo las ventajas y desventajas de varias opciones.	B2 =?=	Realiza descripciones y presentaciones claras y detalladas sobre una serie de asuntos relacionados con su especialidad, ampliando y defendiendo sus ideas con aspectos complementarios y ejemplos relevantes.

B1.2	Desarrolla argumentos lo bastante bien como para que se puedan comprender sin dificultad la mayor parte del tiempo.	B2 = =	Realiza descripciones y presentaciones claras y detalladas sobre una serie de asuntos relacionados con su especialidad, ampliando y defendiendo sus ideas con aspectos complementarios y ejemplos relevantes.
------	---	-----------	---

De cualquier forma, estas interpretaciones propuestas debemos demostrarlas. El problema que surge ahora, nos parece, es grave. Es el hecho de que el MCER basa muchas de las diferencias de nivel en criterios de contenido o detalle respecto al nivel inmediatamente inferior, lo que convierte a las definiciones de nivel en una escalera basada en matices y en dar en cada nivel una especificación más concreta de los mismos temas que en el nivel anterior, antes que en demostrar una habilidad diferenciada o en definir efectivamente la calidad del discurso. Y eso hace que, contrariamente a lo que pretende el MCER, los descriptores no sean independientes unos de otros, ya que el sentido de uno depende del alcance de lo definido en el inmediatamente inferior.

Esta definición de niveles basada en detalles, cantidad de información y mayor especificación de los temas sobre los que hablar, es muy subjetiva, no nos parece suficientemente aceptable para delimitar diferencias y probablemente sea poco operativa al tratar realmente con ejemplos de discurso de candidatos cuando haya que valorarlos¹⁸⁶.

Para comprobarlo comenzamos el análisis de las escalas de *Autoevaluación, Expresión oral en general, Monólogo sostenido: argumentación y Niveles cualitativos de referencia*. Y terminamos con un experimento realizado con 50 profesores.

3.1.3 Análisis crítico de la escala de *Autoevaluación: expresión oral*

A similares conclusiones se llega tras los análisis de la escala de *Autoevaluación* (Tabla 1).

¹⁸⁶ Veremos en el capítulo 4 si esto es así o no.

Tabla 1. Autoevaluación: expresión oral

A1	A2	B1	B2	C1	C2
Utilizo expresiones y frases sencillas para describir el lugar donde vivo y las personas a las que conozco.	Utilizo <u>una serie de</u> expresiones y frases para describir <i>con términos sencillos</i> a mi familia y otras personas, <u>mis condiciones de vida, mi origen educativo, mi trabajo actual o el último que tuve.</u>	Sé enlazar frases de forma sencilla con el fin de describir experiencias y hechos, mis sueños, esperanzas y ambiciones. Puedo explicar y justificar brevemente mis opiniones y proyectos. Sé narrar una historia o relato, la trama de un libro o película, y puedo describir mis reacciones.	Presento descripciones claras y detalladas de una amplia serie de temas relacionados con mi especialidad. Sé explicar un punto de vista sobre un tema exponiendo las ventajas y los inconvenientes de varias opciones.	Presento descripciones claras y detalladas <u>sobre temas complejos que incluyen otros temas.</u> <i>desarrollando ideas concretas</i> y <u>terminando o con una conclusión apropiada.</u>	Presento descripciones <u>o argumentos de forma clara y fluida, y con un estilo que es adecuado al contexto, y con una estructura lógica y eficaz que ayuda al oyente a fijarse en las ideas importantes y a recordarlas.</u>

(Consejo de Europa, 2002: 30-31)

Comparamos paso a paso los descriptores de sus seis niveles para ver su alcance y significado.

A1 y A2.

A1 y A2 solamente se diferencian en la cantidad de información que debe aportar el hablante y en la especificación de los temas a tratar. En la tabla, aparece subrayada la nueva información diferente entre A2 y A1. Como se ve, son prácticamente iguales y no aparece ninguna indicación de mayor dominio o habilidad. Únicamente se aumenta el contenido: en vez de “Utilizo expresiones y frases”, se dice “Utilizo *una serie de* expresiones y frases”; y se citan nuevos temas de conversación: “mis condiciones de vida, mi origen educativo, mi trabajo actual o el último que tuve”.

B1.

B1 se establece como un nivel de transición entre el nivel A (que podríamos llamar de principiante) y el B2 y siguientes, C1 y C2. Efectivamente, nos parece que presenta características que lo diferencian claramente de ellos. Se profundizan los temas

personales del nivel A (“experiencias y hechos, mis sueños, esperanzas y ambiciones, reacciones personales”), se aumentan las funciones (“describir, explicar, justificar opiniones, narrar”) y se mejora el sistema de organización textual, comenzando a realizar párrafos (“enlazar frases de forma sencilla”).

En este nivel encontramos una pequeña contradicción, que en algunas otras escalas solucionan convirtiéndolo en dos subniveles (B1.1 y B1.2). Observamos que entre lo que se dice al principio del descriptor y lo que se implica al final, hay un notable aumento de nivel. Creemos que se necesita menos nivel para hacer lo que dice al principio (“Sé enlazar frases *de forma sencilla* con el fin de describir experiencias y hechos, mis sueños, esperanzas y ambiciones. Puedo explicar y justificar brevemente mis opiniones y proyectos”) que para hacer lo que se dice al final (“Sé narrar una historia o relato, la trama de un libro o película, y puedo describir mis reacciones”), ya que aquí no aparece matizado con “de forma sencilla” sino como algo completo. Quizás hubiera sido mejor diferenciar dos subniveles, B1.1 y B1.2, como se hace en otras escalas.

B2.

B2 introduce un nuevo campo de dominio de la lengua: temas de especialidad, ya no directamente personales o privados. En B1 se trataba de experiencias (“mis sueños, esperanzas y ambiciones”) mientras que en B2 se habla de “una amplia serie de temas relacionados con mi especialidad”. Y para ser capaz de usar estos temas específicos y explicarlos adecuadamente se amplían las funciones, que ahora no son solamente “explicar y justificar brevemente mis opiniones y proyectos” (B1) sino “explicar un punto de vista sobre un tema, exponiendo las ventajas y los inconvenientes de varias opciones”, con lo que la amplitud del tema crece hasta un análisis de pros y contras de diversas opciones. Así queda muy clara la diferencia entre B1 y B2, de una forma que hace que B1 se mantenga dentro de los campos temáticos y funcionales del nivel A, y B2 inicie una diferencia que lo mantiene más cercano al C. Por esto creemos que el nivel de usuario independiente no queda bien aclarado en el MCER y la diferencia entre B1 y B2 es mayor que la existente entre A2 y B1 o la que veremos entre B2 y C.

C1.

C1 no presenta una gran diferencia respecto al anterior B2, ya que (véase lo subrayado en la tabla núm. 1) la información nueva de C1 consiste en ampliar los temas,

pasando de temas “de mi especialidad” (B2) a “temas complejos que incluyen otros temas” (C1). Dos diferencias más, aunque menores, son el saber terminar las descripciones “con una conclusión apropiada” (elemento este que no estaba excluido directamente del B2, ya que puede estar implícito en su definición “sé explicar un punto de vista sobre un tema”, B2) y la aclaración de que se debe poder “desarrollar ideas concretas”, lo que introduce el matiz de concreción cognitiva y, lógicamente, léxica, en el tratamiento de los temas. Pero, aunque esto supone un dominio de vocabulario alto y especializado, no se sale del dominio de lo concreto, no llega a lo abstracto.

C2.

C2 se diferencia de C1 en dos cosas: a. estilo y b. estructura en el tratamiento de los temas. Es decir, en la forma de expresión más rica y variada y en que aparece una nueva función, la de argumentar. Respecto a lo primero se dice: “de forma clara y fluida”, “con un estilo que es adecuado al contexto”, y “con una estructura lógica y eficaz”. Y se introduce al oyente como juez del discurso para aclarar que este debe entenderlo todo sin problemas: la expresión y organización de las ideas, la selección de estas, la forma de exponer los asuntos debe ser tal que “ayuda al oyente a fijarse en las ideas importantes y a recordarlas”. Es decir, hay que tener la capacidad de hacer notar qué ideas están por encima de las otras. Respecto a lo segundo, se cita ahora explícitamente la función de argumentar, aunque se suponía en B1 y sobre todo en B2 y C1.

Tras las catas a esta escala del MCER mantenemos la importante conclusión, señalada al comienzo de este apartado, de que creemos que diferencian en realidad solamente tres grandes niveles y no los seis que se pretende:

- a. A1 y A2
- b. B1
- c. B2, C1 y C2.

Y que las diferencias existentes entre los niveles están basadas en gran medida en la cantidad, con la ampliación de temas y estilo, y en menor medida de calidad: se amplían los temas, se amplían las funciones y se amplía la organización textual. Y esto nos parece, por un lado, una definición pobre del uso general de la lengua, con pocos elementos, además de ser de difícil uso por parte de los aprendices o candidatos para su

autoevaluación, y por otro, muy ambigua y llena de posibles interpretaciones, a pesar de su aparente sencillez.

3.1.4 Análisis crítico de la escala de *Expresión oral en general*

La misma conclusión se mantiene en los análisis de la escala de *Expresión oral en general* (Tabla 2).

Tabla 2. *Expresión oral en general*

C2	Produce discursos claros, fluidos y bien estructurados cuya estructura lógica resulta eficaz, ayuda al oyente a fijarse en elementos significativos y a recordarlos.
C1	Realiza descripciones y presentaciones claras y detalladas sobre temas complejos, integrando otros temas, desarrollando ideas concretas y terminando con una conclusión adecuada.
B2.2	Realiza descripciones y presentaciones claras y <u>sistemáticamente desarrolladas, resaltando adecuada-mente los aspectos significativos y los detalles relevantes que sirvan de apoyo.</u>
B2.1	Realiza descripciones y <u>presentaciones</u> claras y detalladas sobre una serie de asuntos relacionados con su especialidad, <u>ampliando y defendiendo sus ideas con aspectos complementarios y ejemplos relevantes.</u>
B1	Puede llevar a cabo, con razonable fluidez, una descripción sencilla de una variedad de temas que sean de su interés, presentándolos como una secuencia lineal de elementos.
A2	Sabe hacer una descripción o presentación sencilla de personas, condiciones de vida o trabajo, actividades diarias, cosas que le gustan o no le gustan, en una breve lista de frases y oraciones sencillas.
A1	Puede expresarse con frases sencillas y aisladas relativas a personas y lugares.

(Consejo de Europa, 2002: 62)

Comparamos y analizamos la progresión de sus descriptores a lo largo de los seis niveles para observar si estos son claros y marcan adecuadamente las diferencias:

A1.

La habilidad de candidatos del nivel A1 se limita a la capacidad para hacer frases sueltas sobre temas cercanos y sencillos, personas y lugares. Nos parece que esta definición de nivel es pobre y quiere suponer mucho menos nivel del que esas palabras definen realmente. La interpretación del alcance de este nivel se deduce más por ser el más bajo de la escala que por las palabras usadas para definirlo, ya que con lo que define (“Puede expresarse con frases sencillas y aisladas relativas a personas y lugares”) se puede elaborar un discurso de más nivel, pues no se aplican restricciones, ni se explica su alcance u otras características. Nos parece que se expresan las características formales del tipo de discurso (un modo de enlazar frases sueltas sin coherencia textual) pero que no se define la habilidad.

A2.

A2 continúa el mismo tipo de estructura oracional y sencillez (“en una breve lista de frases y oraciones sencillas”), y amplía los temas, aunque siguiendo con lo personal, cercano y conocido (“personas, condiciones de vida o trabajo, actividades diarias, cosas que le gustan o no le gustan”). La diferencia se marca ahora citando dos funciones nuevas (“descripción o presentación sencilla”), unos temas más concretos y definiendo la coherencia textual. No obstante, estas cosas no quedaban descartadas con la pobre definición dada para A1, que adquiere su valor, como habíamos anunciado, por comparación con lo descrito en A2, antes que por lo que para él mismo se defina.

B1.

B1 mantiene la función de describir, su amplitud de sencillez y los temas: “una variedad de temas cercanos o de interés”. Pero presenta una organización más compleja: enlazar ideas en una secuencia (“una secuencia lineal de elementos”). Según esta definición, las ideas se encadenarían para organizar una información más rica, pero este sistema de organización no necesitaría todavía de conectores oracionales ni textuales, sino la linealidad de los acontecimientos y su comprensión lógica.

B2.

B2 vuelve a parecernos el nivel en el que se marcan más diferencias, separándolo conceptualmente de B1. Los temas pasan a ser “de su especialidad” y a estar desarrollados “ampliando y defendiendo sus ideas con aspectos complementarios y

ejemplos relevantes”. Se sigue en la misma línea apuntada respecto a la escala anteriormente comentada, de *Autoevaluación*. Lo que cabe destacar ahora es la disimilitud entre la escala de *Autoevaluación* y esta, ya que siendo la misma definición en sus términos fundamentales, donde en aquella dice “sé explicar un punto de vista sobre un tema exponiendo las ventajas y los inconvenientes de varias opciones” (B2 de *Autoevaluación*) ahora se dice “ampliando y defendiendo sus ideas con aspectos complementarios y ejemplos relevantes” (B2 de *Expresión oral general*). De donde deducimos que ambas cosas se equiparan sin forzar los significados sin que suponga cambiarles el sentido, aunque está descrito con mayor precisión en esta última escala.

Pero esta escala define un nuevo subnivel, el B2.2, al que añade lo siguiente (subrayado en la tabla): “sistemáticamente desarrolladas, resaltando adecuadamente los aspectos significativos y los detalles relevantes que sirvan de apoyo”. Y esto supone realmente mayor nivel respecto a lo definido en la escala de *Autoevaluación*, lo que implica falta de coherencia entre escalas y va a dificultar la evaluación cuando haya que referirse a varias de ellas para valorar el discurso de un mismo candidato.

Véanse en subrayado los elementos que no se pueden equiparar:

(*Autoevaluación* B2): “Presento descripciones claras y detalladas de una amplia serie de temas relacionados con mi especialidad. Sé explicar un punto de vista sobre un tema exponiendo las ventajas y los inconvenientes de varias opciones”.

(*Expresión oral en general* B2.1): “Realiza descripciones y presentaciones claras y detalladas sobre una serie de asuntos relacionados con su especialidad, ampliando y defendiendo sus ideas con aspectos complementarios y ejemplos relevantes”.

(*Expresión oral en general* B2.2) “Realiza descripciones y presentaciones claras y sistemáticamente desarrolladas, resaltando adecuadamente los aspectos significativos y los detalles relevantes que sirvan de apoyo”.

Nos parece que este último descriptor B2.2 supone mayor nivel que el anterior B2 de la escala de *Autoevaluación*. El asunto se hace más incoherente porque se asemeja mucho más al descriptor C1 de *Autoevaluación*: “Realiza descripciones y presentaciones claras y detalladas sobre temas complejos, integrando otros temas, desarrollando ideas concretas y terminando con una conclusión adecuada”, con lo que

este B2.2 está más cerca de aquel C1 que de su correspondiente B2, como mostramos en el cuadro siguiente:

	<i>Expresión oral en general</i>		<i>Autoevaluación</i>
B2.2	<u>Realiza descripciones y presentaciones claras y sistemáticamente desarrolladas, resaltando adecuadamente los aspectos significativos y los detalles relevantes que sirvan de apoyo.</u>	C1	Realiza descripciones y presentaciones claras y detalladas sobre temas complejos, <u>integrando otros temas, desarrollando ideas concretas y terminando con una conclusión adecuada</u>
B2.1	Realiza descripciones y presentaciones claras y detalladas sobre una serie de asuntos relacionados con su especialidad, ampliando y defendiendo sus ideas con aspectos complementarios y ejemplos relevantes.	B2	Presento descripciones claras y detalladas de una amplia serie de temas relacionados con mi especialidad. Sé explicar un punto de vista sobre un tema exponiendo las ventajas y los inconvenientes de varias opciones.

“Integrar temas, desarrollarlos adecuadamente, resaltar los aspectos significativos y los aspectos relevantes” (B2.2 de *Expresión oral en general*) está muy cerca (en todo caso el matiz va a ser muy difícil de diferenciar ante ejemplos reales del discurso de un extranjero -o incluso de un nativo-) de “resaltar adecuadamente los aspectos significativos y los detalles relevantes que sirvan de apoyo (a una descripción o presentación”, C1 de *Autoevaluación*). Si alguien quiere hacer esto último podrá estar haciendo lo anterior, salvo que la diferencia para el descriptor de C1 se centre exclusivamente en el hecho de la integración de temas en una presentación. Mayor complejidad que si no hay tal integración de temas o si se limita a ser la presentación de un solo tema.

Basar la diferencia de niveles en este hecho nos parece confiar en un dato cuantitativo que quizás sea difícil de elicitar en una prueba oral y de difícil concreción. Podremos encontrar discursos de nivel C1 que tengan las características de ese nivel C1 pero que hagan un análisis de un solo tema, y otros discursos que, intentando analizar varios temas (o integrarlos), se queden en B2 por sus características estilísticas, léxicas u organizativas.

C1.

El descriptor de C1 es una copia del C1 de *Autoevaluación*, cambiando exclusivamente algunas palabras por sinónimos (las que damos subrayadas): “*Presento* descripciones claras y detalladas sobre temas complejos *que incluyen* otros temas, desarrollando ideas concretas y terminando con una conclusión *apropiada*”.

Esto nos lleva a criticar lo que creemos que es una mala descripción y una carencia importante en las escalas, ya que impedirá definir adecuadamente si una persona es de nivel B2.2 o C1. Ya habíamos visto esto mismo en el análisis de la escala de *Autoevaluación*, cuando decíamos que en el descriptor de C1 se definen elementos que se suponen en B2, aunque no se expresen claramente en B2. Es decir, la diferencia entre B2 y C1 no está ahora -ni tampoco lo estaba en la anterior escala- suficientemente clara.

C2.

Para el descriptor del nivel C2 tenemos el mismo comentario que hemos hecho respecto a la escala de *Autoevaluación*, salvo que aquí se habla de ‘discursos’ y queremos entender que el término no se usa con el sentido general de ‘enunciado oral’, sino con el sentido mucho más concreto de ‘discurso social o político’, es decir, una argumentación extensa basada en secuencias de párrafos organizados.

Véanse los descriptores de C2 de *Autoevaluación* y *Expresión oral en general*:

(*Expresión oral en general*): “Produce discursos claros, fluidos y bien estructurados cuya estructura lógica resulta eficaz, ayuda al oyente a fijarse en elementos significativos y a recordarlos”.

(*Autoevaluación*): “Presento descripciones o argumentos de forma clara y fluida, y con un estilo que es adecuado al contexto, y con una estructura lógica y eficaz que ayuda al oyente a fijarse en las ideas importantes y a recordarlas”.

	<i>Expresión oral en general, C2</i>	<i>Autoevaluación, C2</i>
C2	“Produce discursos claros, fluidos y bien estructurados cuya estructura lógica resulta eficaz, ayuda al oyente a fijarse en elementos significativos y a recordarlos”	“Presento descripciones o argumentos de forma clara y fluida, y con un estilo que es adecuado al contexto, y con una estructura lógica y eficaz que ayuda al oyente a fijarse en las ideas importantes y a recordarlas”

En definitiva, la capacidad de un nivel alto de expresión o C2 sería la de hacer discursos. No obstante, llegar a esta conclusión e interpretar adecuadamente el sentido de “discursos” no es algo fácil ni evidente, salvo que se quiera ver en cada descriptor una expresión de más nivel que en el anterior, lo diga con las palabras que lo diga. Porque esta palabra “discursos” aparece en varios descriptores de otras escalas, por ejemplo en la de *Niveles cualitativos de referencia*, en la columna dedicada al análisis de la coherencia, con un sentido diferente y más abierto, utilizado para definir el nivel B2 y C1. Por lo que llegar a la conclusión de que ahora, en esta escala de *Expresión oral en general*, al decir “discurso” se refiere a un discurso pleno y por ello a una habilidad de nivel C2, es algo que no tiene realmente una razón clara, salvo por el hecho de tener que definir mayores habilidades que en el nivel anterior, C1.

Como venimos diciendo de los análisis de las anteriores escalas, esto representa un importante fallo en el sistema de definir los niveles, deudor de unos conceptos de nivel preestablecidos y de unas interpretaciones, ni claras ni aseguradas, y dependientes de lo que se crea que significan los niveles anteriores.

3.1.5 Análisis crítico de la escala de *Monólogo sostenido: argumentación*

La misma conclusión se mantiene en los análisis de la escala de *Monólogo sostenido: argumentación* (Tabla 3).

Tabla 3. *Monólogo sostenido: argumentación*

A1 y A2	B1.1	B1.2	B2.1	B2.2	C1 y C2
No hay descriptor posible	Ofrece breves razonamientos y explicaciones de opiniones planes y acciones.	Desarrolla argumentos lo bastante bien como para que se puedan comprender sin dificultad la mayor parte del tiempo.	Desarrolla argumentos con claridad, ampliando con cierta extensión y defendiendo sus puntos de vista con ideas complementarias y ejemplos adecuados. Construye cadenas argu-	Desarrolla argumentos sistemáticamente dando un énfasis apropiado a los aspectos importantes y apoyándose en detalles adecuados.	No hay descriptor posible

			mentales razonadas. Explica puntos de vista sobre un tema, proponiendo las ventajas y desventajas de varias opciones.		
--	--	--	---	--	--

(Consejo de Europa, 2002: 63).

Comparamos y analizamos los descriptores de esta tabla.

La escala de *Monólogo sostenido: argumentación* no ofrece descriptores de nivel para 4 de los 6 niveles: A1, A2, C1 y C2. Pero ofrece dos subniveles extras para B1 y B2: B1.1, B1.2, B2.1 y B2.2

B1.

B1.1 no es claro en su definición, ya que se limita a decir que el hablante debe poder “ofrecer breves razonamientos y explicaciones de opiniones planes y acciones”. El tema no se especifica más y las funciones del lenguaje no se aclaran en su profundidad, grado de alcance, organización textual, etc. más allá de la cantidad: “breves”. Si las interpretáramos en todo su sentido deberíamos pensar que quien haga eso tendría la capacidad de argumentar, explicar y opinar utilizando todos los recursos de la lengua, lo que llevaría a un tipo de discurso y organización textual que despliegue datos y los comente, analice o contraste, aunque limitados cuantitativamente (“breves razonamientos”). ¿Es que el hecho de ser “breves” invalida el que posean las características textuales de un razonamiento, que por definición supone un conjunto de frases con hilo argumental, inicio, planteamiento de problemas, explicación, análisis y conclusión?

Una vez más sale a la luz el fallo apuntado en los análisis realizados con las escalas anteriores: la ambigüedad es tal que tenemos que interpretar las variaciones de nivel por conocimientos o juicios de nivel preestablecidos. Así, en vez de describir adecuadamente las características del discurso de B1, explicando el modo de saber hacer esos razonamientos y diferenciándolos de otros modos que se puedan desarrollar en otros niveles, se supone que el hablante debe saber hacer razonamientos, aunque con pocos recursos estilísticos, textuales y organizativos, porque es solamente de nivel B1 y

porque se dice que en A1 y A2 no se posee capacidad para hacerlo. Es decir, no se define el nivel sino que se presupone que el usuario de las escalas lo debe deducir de una preconcepción de la división de las habilidades del lenguaje en 6 niveles.

B1.2 no es mucho más claro. Se limita ahora a introducir una nueva función, “desarrollar argumentos”, y a cifrar el juicio sobre su óptima realización mediante el grado de comprensión que respecto a estos argumentos tenga el oyente. No nos parece oportuno dejarlo en manos de una comprensión subjetiva como la del oyente y sin aclarar en qué debería basarse ese juicio. Se intuye que el descriptor se está refiriendo a cierta coherencia textual y organizativa, ya que se habla del “desarrollo” de los argumentos y porque en otras escalas, por ejemplo la de *Autoevaluación*, se ha hablado del tipo de discurso que se supone debe hacer bien un candidato de B1 (“Sé enlazar frases de forma sencilla... Puedo explicar y justificar brevemente mis opiniones”). Pero, al mismo tiempo y sin referirse explícitamente a ningún tipo de organización textual, se dice que el desarrollo de los argumentos debe ser “lo bastante bien como para que se puedan comprender”, de donde se puede interpretar que el grado de profundidad y coherencia textual debe ser grande, porque argumentar exige todo un dispositivo de técnicas que requieren bastante nivel.

Nuevamente esta ambigüedad de nivel se dirime por un prejuicio: sabemos que no se requieren habilidades excesivamente complejas, ya que este nivel es un nivel todavía bajo, el tercero de seis. Otro camino para interpretarlo como algo no demasiado elevado es cotejarlo con otras escalas, pero con ello se complica y dificulta el trabajo de llegar a acuerdos entre los profesionales que quieran usarlas.

B2.

B2.2 se diferencia de B2.1 simplemente en el alcance de los detalles: ahora se exige más “claridad, ampliando con cierta extensión, sistemáticamente, construyendo cadenas argumentales”.

El conjunto del nivel B2 es similar al ya definido en las escalas de *Expresión oral general* y *Autoevaluación*, solamente que ahora se hace referencia a un nuevo elemento, las argumentaciones y las cadenas argumentales razonadas, que introduce un nuevo factor de distorsión en la interpretación del nivel. Porque este elemento se deduce de las escalas de *Autoevaluación* y *Expresión oral en general* en unos niveles mayores, el C1

de Expresión oral general y el C2 de *Autoevaluación*, donde podríamos equiparar sus descriptores (“Presentaciones claras y detalladas sobre temas complejos, integrando otros temas, desarrollando ideas”, C1 de Expresión oral general) y (“Presento descripciones o argumentos de forma clara”, C2 de *Autoevaluación*) a las cadenas de argumentos de las que se habla ahora. Esto es porque esta nueva escala equipara los niveles B2, C1 y C2, por lo que no hace diferencias entre ellos.

A continuación -para ver claras estas equiparaciones entre los tres niveles- damos la descripción, poniendo entre paréntesis los sinónimos con que se describe eso mismo en las otras escalas y el nivel al que pertenecen (véase también el cuadro que sigue, para esta misma explicación):

B2, “Desarrolla argumentos con claridad, (“argumentos de forma clara” -C2 de *Autoevaluación*-) ampliando con cierta extensión (“ampliando y defendiendo sus ideas con aspectos complementarios” -B2 de *Expresión oral en general*-) y “defendiendo sus puntos de vista” (“defendiendo sus ideas” -B2 de *Expresión oral en general*-) con “ideas complementarias y ejemplos adecuados”, (“con aspectos complementarios y ejemplos relevantes” -B2 de *Expresión oral en general*-).

B2, “Construye cadenas argumentales razonadas” (“sobre temas complejos que incluyen otros temas”, -C2 de *Autoevaluación*-).

B2, “Explica puntos de vista sobre un tema”, (“Sé explicar un punto de vista sobre un tema” -B2 de *Autoevaluación*-) “proponiendo las ventajas y desventajas de varias opciones” (“exponiendo las ventajas y los inconvenientes de varias opciones” -B2 de *Autoevaluación*-).

B2, “Desarrolla argumentos sistemáticamente” (“claras y sistemáticamente desarrolladas, complementarios” -B2 de *Expresión oral en general*-) “dando un énfasis apropiado a los aspectos importantes” (“resaltando adecuadamente los aspectos significativos complementarios” -B2 de *Expresión oral en general*-) “y apoyándose en detalles adecuados” (“detalles relevantes complementarios” -B2 de *Expresión oral en general*-).

	<i>Expresión oral general</i>	<i>Autoevaluación</i>	<i>Monólogo sostenido: argumentación.</i>
C2		Presento descripciones o argumentos de forma clara y fluida, y con un estilo que es adecuado al contexto, y con una estructura lógica y eficaz que ayuda al oyente a fijarse en las ideas importantes y a recordarlas.	
C1	Realiza descripciones y presentaciones claras y detalladas sobre temas complejos, integrando otros temas, desarrollando ideas concretas y terminando con una conclusión adecuada.	Presento descripciones claras y detalladas sobre temas complejos que incluyen otros temas, desarrollando ideas concretas y terminando con una conclusión apropiada.	
B2	Realiza descripciones y presentaciones claras y sistemáticamente desarrolladas, resaltando adecuadamente los aspectos significativos y los detalles relevantes que sirvan de apoyo. (B2.2) Realiza descripciones y presentaciones claras y detalladas sobre una serie de asuntos relacionados con su especialidad, ampliando y defendiendo sus ideas con aspectos complementarios y ejemplos relevantes. (B2.1)	Presento descripciones claras y detalladas de una amplia serie de temas relacionados con mi especialidad. Sé explicar un punto de vista sobre un tema exponiendo las ventajas y los inconvenientes de varias opciones.	Desarrolla argumentos sistemáticamente dando un énfasis apropiado a los aspectos importantes y apoyándose en detalles adecuados. (B2.2) Desarrolla argumentos con claridad, ampliando con cierta extensión y defendiendo sus puntos de vista con ideas complementarias y ejemplos adecuados. Construye cadenas argumentales razonadas. Explica puntos de vista sobre un tema, proponiendo las ventajas y desventajas de varias opciones. (B2.1)

Se ha querido dar la idea de que esas funciones solo se pueden realizar a dos niveles (B1 y B2) y no existiría una posibilidad de diferenciación de mayor nivel. Esto, en principio, nos parece poco efectivo y es en sí un elemento más de falta de concreción y ambigüedad en las escalas, peligroso por dos razones que hemos venido ya apuntando y aquí se vuelven claramente delicadas: en primer lugar, porque al dar la definición se han mezclado elementos de B2 de otras escalas -como hemos visto- con otros de C2, lo que

supone pensar que una persona que realiza un discurso con los elementos o características de B2 posee al mismo tiempo el C1 o C2. Y en segundo lugar, porque lleva a la confusión a la hora de diferenciar niveles y hacer las interpretaciones y evaluaciones usando al mismo tiempo otras escalas: no se sabe qué es efectivamente B2, C1 y C2.

Si al realizar una evaluación sobre el nivel del discurso de un candidato se emplean otras escalas, ya que en muchas ocasiones hay que echar mano de otras para ver el alcance definitivo del discurso del candidato (*Dominio de vocabulario* a la vez que *Cohesión textual*, *Monólogo sostenido* a la vez que *Expresión oral general*, etc.), y en unas se ofrecen descriptores que también están en otras pero en un nivel superior o inferior, las interpretaciones resultantes van a llevar a error indefectiblemente.

Habría sido fácil tomar elementos de los descriptores de nivel C1 y C2 de otras escalas para componer descriptores diferenciados también en esta escala de *Argumentación*: el trabajo resulta conveniente no solo por la coherencia interna de toda la formulación de escalas de habilidad y dominio, sino por las razones ya apuntadas sobre la interpretación y uso a la hora de proceder a hacer juicios y calificaciones en una evaluación concreta a un candidato.

Concluyendo, la diferenciación que venimos señalando de solamente tres niveles, se reduce ahora de forma explícita, ya que los mismos redactores del MCER equiparan, B2, C1 y C2. Con ello queda solamente la diferencia entre ser menos que B1, ser B1 o ser más que B1.

Estos son:

- B1.
- y el macronivel B2-C1-C2.

3.1.6. Análisis crítico de la escala de *Niveles cualitativos de referencia*

Esta escala (tabla 4) presenta bastantes cosas dignas de comentar: en primer lugar, como su nombre indica, es una escala de referencia general cualitativa, por lo que debería ser la que ayudara a sacar de dudas a las otras escalas o la que se debiera usar para matizarlas, con lo que nos parece más importante que las otras. Y, en segundo

lugar, presenta algunas contradicciones con otras escalas del MCER ya analizadas, respecto a las que sube o baja el nivel¹⁸⁷.

Tabla 4. Niveles cualitativos de referencia

	ALCANCE	CORRECCIÓN	FLUIDEZ	INTERACCIÓN	COHERENCIA
C2	26. Muestra una gran flexibilidad al reformular ideas diferenciando formas lingüísticas para transmitir con precisión matices de sentido, enfatizar, diferenciar y eliminar la ambigüedad. También tiene un buen dominio de expresiones idiomáticas y coloquiales. (AM)	27. Mantiene un consistente dominio gramatical de lengua complejo, aunque su atención esté pendiente de otros aspectos (por ejemplo, de planificación o del seguimiento de las reacciones de otros). (AM)	28. Se expresa espontánea y detalladamente con fluidez natural y coloquial, evitando o <i>sorteando la dificultad con tanta discreción que los demás apenas se dan cuenta.</i> (AM)	29. Participa en una conversación con facilidad y destreza, captando y utilizando claves no verbales y de entonación sin esfuerzo aparente en la conversación esperando su turno, dando referencias, haciendo alusiones, etc. de forma natural. (AM)	30. Crea un discurso coherente y cohesionado, haciendo uso completo y apropiado de estructuras organizativas variadas y de una amplia serie de conectores, y de otros mecanismos de cohesión. (AM)
C1	21. Tiene un buen dominio de una amplia serie de aspectos lingüísticos que le permiten elegir una formulación para expresarse con claridad y con un estilo apropiado sobre diversos temas generales, académicos, profesionales o	22. Mantiene con consistencia un alto grado de corrección gramatical; <i>los errores son escasos, difíciles de localizar y, por lo general, los corrige cuando aparecen.</i> (AM)	23. Se expresa con fluidez y espontaneidad sin apenas esfuerzo. Solo un tema conceptualmente difícil puede obstaculizar la fluidez natural de su expresión. (AM)	24. Elige las frases adecuadas de entre una serie disponible de funciones del discurso para introducir sus comentarios con el fin de tomar o mantener la palabra y relacionar hábilmente sus propias intervenciones con las de los demás interlocutores.	25. Produce un discurso claro, fluido y bien estructurado con el que demuestra un uso controlado de funciones organizativas, conectores y mecanismos de cohesión. (AM)

¹⁸⁷ Quizás sea el hecho de que antes ante las ambigüedades que encontrábamos habíamos interpretado un nivel que ahora se rebaja o matiza. De cualquier manera, el que ahora se matizan ciertos elementos no invalidan las ambigüedades, incoherencias y fallos analizados en las anteriores.

	de ocio sin tener que restringir lo que quiere decir. (AM)			(AM)	
B2	16. Tiene un nivel de lengua lo bastante amplio como para ofrecer descripciones claras y expresar puntos de vista sobre temas generales sin evidenciar la búsqueda de palabras, y sabe utilizar oraciones complejas para conseguirlo. (Av)	17. Demuestra un control gramatical relativamente alto. No comete errores que provoquen la incompreensión y corrige casi todas sus incorrecciones (Av)	18. Es capaz de producir fragmentos de discurso con un ritmo bastante uniforme; aunque puede dudar mientras busca estructuras o expresiones. Se observan pocas pausas largas. (Av)	19. Inicia el discurso, toma su turno de palabra en el momento adecuado y finaliza una conversación cuando tiene que hacerlo, aunque puede que no siempre lo haga con elegancia. Colabora en debates sobre temas cotidianos, confirmando su comprensión, invitando a los demás a participar, etc. (IH/Av)	20. Utiliza un número limitado de mecanismos de cohesión para convertir sus frases en un discurso claro y coherente, aunque puede mostrar cierto “nerviosismo” si la intervención es larga. (AL/AM)
B1	11. Tiene un repertorio lingüístico lo bastante amplio como para desenvolverse y un vocabulario adecuado para expresarse, aunque dubitativamente y con circunloquios, sobre temas tales como su familia, sus aficiones e intereses, su trabajo, sus viajes y acontecimientos actuales. (IL/IM)	12. Utiliza con razonable corrección un repertorio de fórmulas y estructuras de uso habitual y asociadas a situaciones predecibles. (IL/IM)	13. Puede continuar hablando de forma comprensible, aunque sean evidentes sus pausas para realizar una planificación gramatical y léxica y una corrección, sobre todo en largos períodos de expresión libre. (IL/IM)	14. Es capaz de iniciar, mantener y terminar conversaciones sencillas cara a cara sobre temas cotidianos de interés personal. Puede repetir parte de lo que alguien ha dicho para confirmar la comprensión mutua. (IL/IM)	15. Es capaz de enlazar una serie de elementos breves, diferenciados y sencillos para formar una secuencia lineal de ideas relacionadas. (IL/IM)
A2	6. Utiliza estructuras compuestas por	7. Utiliza algunas estructuras	8. Se hace entender con	9. Sabe contestar preguntas y responder a	10. Es capaz de enlazar grupos de palabras con

	oraciones básicas con expresiones, grupos de unas pocas palabras y fórmulas memorizadas, con el fin de comunicar una información limitada en situaciones sencillas y cotidianas. (NM/NH)	sencillas correctamente, pero todavía comete sistemáticamente, errores básicos. (NH/IL)	expresiones muy breves, aunque resultan muy evidentes las pausas, las dudas iniciales y la reformulación. (NH/IL)	afirmaciones sencillas. Sabe indicar cuando comprende una conversación, pero apenas comprende lo suficiente como para mantener una conversación por decisión propia. (NM/NH)	conectores sencillos tales como “y”, “pero” y “porque”. (IL/IM)
A1	1. Dispone de un repertorio básico de palabras y frases sencillas relativas a sus datos personales y a situaciones concretas. (NM)	2. Muestra un control limitado de unas pocas estructuras gramaticales sencillas y de modelos de oraciones dentro de un repertorio memorizado. (NM)	3. Sólo maneja expresiones muy breves, aisladas y preparadas de antemano, utilizando muchas pausas para buscar expresiones, articular palabras menos habituales y corregir la comunicación. (NM)	4. Plantea y contesta preguntas relativas a datos personales. Participa en una conversación de forma sencilla, pero la comunicación se basa totalmente en la repetición, reformulación y corrección de frases. (NH/IL)	5. Es capaz de enlazar palabras o grupos de palabras con conectores muy básicos y lineales como “y” y “entonces”. (NH/IL)

(Consejo de Europa, 2002: 32-33).

Para una mejor comparación entre esta tabla (tabla 4) y otras del MCER hemos numerado cada cuadro para hacer referencia a sus descriptores y contenido en las siguientes explicaciones.

A1.

Los niveles descritos aquí se diferencian de los descritos en las otras escalas en forma bastante evidente. Por ejemplo, mientras en la escala de interacción oral de *Autoevaluación* (Consejo de Europa, 2002: 30-31) se dice en A1 “Planteo y contesto a preguntas sencillas sobre temas de necesidad inmediata o asuntos muy habituales”, ahora se describe A1 con mayores limitaciones, que hacen imposible esa habilidad: no

vemos cómo se puede ser capaz de plantear y contestar a preguntas sencillas sobre temas de necesidad inmediata o asuntos muy habituales si solo se dispone de un repertorio básico de palabras y frases (1), o solo se manejan expresiones muy breves, aisladas y preparadas de antemano, utilizando muchas pausas para buscar expresiones (3).

Creemos que plantear y contestar preguntas, aunque sea sobre temas habituales, requiere mayor habilidad y conocimientos. Entre una y otra cosa media un terreno grande, el que dista de ser capaz de usar material aprendido de forma creativa, aunque se necesite tiempo para comunicarse y encontrar expresiones, aunque se cometan errores de todo tipo, y el de ser capaz de usar solamente un breve repertorio de frases memorizadas o comunicarse solo en contextos para los que antes se ha preparado lo que se pensaba decir (3).

A2.

Igualmente en A2 se presenta la misma contradicción (o aclaración) de niveles con otras escalas comentadas: ahora se marcan unas limitaciones mayores: “expresiones, grupos de unas pocas palabras y fórmulas memorizadas” (6), “apenas comprende lo suficiente como para mantener una conversación por decisión propia” (9), mientras en la escala de *Autoevaluación* se pueden hacer cosas más complicadas como “Soy capaz de realizar intercambios sociales..., comunicarme en tareas sencillas y habituales que requieren un intercambio simple y directo de información sobre actividades y asuntos cotidianos”.

En relación con las definiciones de A1 esta nos parece de, incluso, menor nivel, ya que “participa en una conversación de forma sencilla, pero la comunicación se basa totalmente en la repetición, reformulación y corrección de frases” (A1) no parece de más nivel que simplemente “apenas comprende lo suficiente para mantener una conversación” (A2).

B1.

Este nivel no presenta rasgos diferentes de los hasta ahora analizados y coincide, tanto en los diferentes campos descritos en esta escala como con otras escalas comentadas, en marcar un nivel claro basado en la calidad y cantidad, adscrito a temas cotidianos donde el hablante es el centro, “su familia, sus aficiones e intereses, su

trabajo, sus viajes” (11) y se limita a expresar ideas en presente, “acontecimientos actuales (11), enlazando elementos en una secuencia lineal” (15).

B2.

Al contrario que en la escala de *Autoevaluación*, donde se hablaba de los temas de conversación diciendo que estos eran “cotidianos”, ahora en este mismo nivel se dice que se puede hablar “sobre temas generales” (16). En aquel caso decíamos que eso rebajaba el nivel con respecto a otras escalas. Ahora esto se vuelve a equiparar con las demás escalas, marcando decididamente una característica diferencial de nivel, excepto en la columna dedicada a la interacción, donde vuelve a aparecer la palabra “cotidianos” respecto a los temas de los que se puede hablar: “colabora en debates sobre temas cotidianos” (19). Esta es la única vez que aparece en las escalas la palabra “cotidianos” en este nivel B2¹⁸⁸, introduciendo nuevos elementos de ambigüedad en los descriptores.

Otras dos cosas llaman la atención en la definición de la interacción: el uso de términos poco claros y sumamente necesitados de interpretación como “no siempre con suficiente elegancia” (19), y el hecho de terminar la descripción con un “etc.” rarísimo (no se usa esto en otras escalas) que no sabemos a qué se puede referir ni consideramos necesario.

El descriptor que da para la columna de la coherencia nos parece ambiguo por el uso de la palabra “discursos”, que no sabemos con qué alcance se usa, si con el de discurso extenso de los niveles C1 y C2 o con el más técnico de enunciados complejos. Creemos que pasar de frases a discursos, pues aquí se dice expresamente que se convierten las frases en discurso, es demasiado abrupto y se podrían haber definido estadios intermedios basados en párrafos, párrafos organizados y encadenados, etc. No obstante, creemos que utiliza la palabra “discursos” aproximándose al sentido que tiene en C1 y C2, o sea, con el fin de dar una marca de habilidad que aumente el nivel.

C1.

El descriptor (21), sin ser más explícito, amplía el dominio de los temas y capacidades de expresión (general y profesional). En general, se trata de un dominio para expresarse sin problemas sobre todo tipo de temas sin restricciones, aunque

¹⁸⁸ Lo que no deja de sorprendernos por falta de coherencia y homogeneidad, además de por lo que supone de tener que interpretar este término como algo de pretendido mayor alcance del que tenía hasta el momento, en niveles menores.

también se apunta en el descriptor del campo de la fluidez (23) que “un tema conceptualmente difícil puede obstaculizar la fluidez”. Lo más ambiguo e interpretable de estos descriptores es el alcance de los temas y del término ‘discurso’.

Realmente nos preguntamos si existen diferencias palpables entre B2 y C1 a la hora de enfrentarse a discursos de hablantes. Lo veremos en el momento de realizar las calificaciones y evaluaciones. Por el momento, solo unos pocos datos de escasa apreciación práctica fundamentan las diferencias. Nos parece destacado el hecho de que en el campo de la corrección en B2 aparezcan errores aunque los corrija, mientras que en el nivel C1 los errores sean escasos y difíciles de localizar; asimismo que en el campo de la fluidez, en B2 haya dudas evidentes a la hora de buscar expresiones, mientras que en C1 solo haya problemas ante un tema conceptualmente difícil; y que en el campo de la coherencia, en B2 se pierda el control si la intervención es larga. Pero ¿qué sucede si no es larga y no se pierde ese control, o si las dudas y autocorrecciones son tan sutiles que apenas se notan? En tal caso B2 y C1 van a ser muy difíciles de diferenciar.

C2.

Creemos que este nivel contiene aspectos que se definen de forma muy ambigua. Por ejemplo, cuando se dice que tiene “un consistente dominio gramatical de lengua complejo, *aunque* su atención esté pendiente de otros aspectos” (27) pensamos que puede tomarse como carencia y mejor habría sido definirlo en positivo, diciendo algo así como que mantiene un dominio gramatical alto *que le permite* atender a otros aspectos del discurso.

Lo más destacable del análisis de los descriptores de este nivel es que en poco se diferencia de los de C1 y que llegar a diferenciar un discurso de C1 y C2 puede ser altamente subjetivo.

En la tabla hemos subrayado los elementos más claramente diferenciales. Viéndolos, se aprecia claramente que la diferencia entre ambos niveles, salvo lo apuntado para el C1 respecto a limitaciones ante temas de dificultad, es muy difícil de llevar a la práctica:

- En el campo del alcance, la diferencia entre los descriptores (21) y (26) se limita a dominar expresiones idiomáticas y coloquiales y al matiz de que en 21 se dice “un buen dominio de una amplia serie”, que expresa en positivo mucho dominio, y, en negativo, que puede existir un dominio todavía más completo y sin limitaciones. Entre los descriptores de *Corrección gramatical* (22) y (27), la diferencia está en la existencia de

errores o no, lo que no impide que alguien no los tenga y sea todavía C1, ya que en C1 estos se esconden y no se notan o se autocorrigien.

- En el campo de la fluidez, descriptores (23) y (28), la diferencia parece clara entre tener problemas ante temas difíciles o no tenerlos. No obstante, en C2 todavía se apunta que puede haber ciertos problemas que se solucionan sin que se noten, lo que lo acercaría más al C1, pues ese es el problema de la corrección gramatical.

- En el campo de la interacción, aunque se definen ambos niveles con descriptores diferentes, no vemos la diferencia que estos suponen a la hora de evaluar un discurso.

- Y en el campo de la coherencia, descriptores (25) y (30), lo subrayado nos parece igualmente de difícil apreciación en la práctica, cuando haya que calificar discursos concretos de alto nivel, C1 o C2.

3.1.7 Experimento sobre la interpretabilidad de los niveles del MCER

Para probar todos estos extremos, fallos y ambigüedades que señalamos tras los análisis de las escalas y sus descriptores, decidimos realizar una prueba con 50 profesores de español como lengua extranjera. Les dimos, desordenada, la escala de *Expresión oral en general*, y les pedimos que la ordenaran por niveles¹⁸⁹. Los resultados demuestran empíricamente todo lo que venimos diciendo con nuestros análisis hasta el momento, la inexistencia de seis niveles reales en el MCER y la excesiva ambigüedad de sus descriptores de nivel. Pero nos aportan más datos: el primero, que la mayoría de expertos consultados consideran al descriptor del nivel C1 como definidor del C2, es decir, como definidor de mayores habilidades discursivas orales. Y en segundo lugar, que se aprecia una dificultad para diferenciar entre A2 y B1¹⁹⁰, lo que prueba nuestra suposición de que la diferencia entre niveles A, B y C no es efectiva ni operativa.

Agrupamos a los profesores de tres formas: un grupo (grupo núm. 1) compuesto por 10 profesores de un mismo centro de trabajo que opinaron individualmente; otro (grupo núm. 2) compuesto por 10 profesores de diferentes centros de enseñanza de toda España, entre los que había unos cuantos que tenían cargos directivos (jefes de estudio o directores), al que se pidió que hiciera el trabajo en parejas; y otro (grupo núm. 3) de 30

¹⁸⁹ Ver en el Apéndice 5 la escala desordenada que se les dio y las preguntas planteadas.

¹⁹⁰ Esta dificultad ya la anunciaba Fulcher (2004: 256) al analizar las dificultades para ver diferencias entre los niveles *Waystage* y *Threshold*, en los que se ha basado el MCER, dando lugar al A2 y B1 respectivamente.

profesores de toda España a quienes se pidió que realizaran el mismo trabajo en grupos de tres, comentando y consensuando de forma razonada sus opiniones.

El objetivo de esta variedad de profesores y el hecho de pedirles que trabajaran en diferentes grupos era doble: primero, evitar que la costumbre de impartir unos cursos o niveles en un mismo centro de trabajo o con una metodología concreta viciara los resultados e interpretaciones; y, por otra parte, averiguar si las interpretaciones logran acercarse más al MCER cuando se trabaja en grupos en los que hay que argumentar las opiniones y consensuar los resultados.

A todos los grupos se les pidió que ordenaran los 7 descriptores de la escala de *Expresión oral* y que establecieran el nivel al que se referían; además, como había siete descriptores y no seis, porque un nivel aparece subdividido en dos (en concreto el B2.1 y B2.2) debían decidir qué 2 descriptores pertenecían al mismo nivel.

El primer grupo estaba formado por 10 profesores del CLM de la Universidad de Granada, que imparten habitualmente cursos intensivos de lengua española para extranjeros pero que no trabajaban con el MCER, estaban poco acostumbrados a sus descriptores y escalas de habilidad y dominio y no conocían, en concreto, la escala que se les pidió que ordenaran.

El grupo segundo eran otros 10 profesores y directivos de centros de enseñanza de diversas partes de España (Barcelona, Madrid, Cantabria, País Vasco, Galicia, Sevilla...) que acudieron a recibir un taller del Máster de Enseñanza de Español para extranjeros de la UIMP (Universidad Internacional Menéndez Pelayo), que aunque conocían el MCER no estaban muy familiarizados con su uso.

Y el tercero eran 30 profesores de diversos centros de España y el extranjero (algunos del Instituto Cervantes) de diferentes lenguas (profesores de inglés y de español) igualmente de diversas partes de España, y bastante más familiarizados con el MCER¹⁹¹.

A continuación se resumen los resultados más importantes.

¹⁹¹ Entre ellos se contaba, por ejemplo, con algunos de los que habían hecho la traducción al gallego.

3.1.7- A) Pruebas realizadas con expertos. Grupo primero

En este primer experimento vamos a ver que efectivamente los descriptores son muy ambiguos y están sometidos a todo tipo de interpretaciones sobre su verdadero nivel; y que los seis niveles que el MCER pretende medir se reducen a los tres que nosotros hemos postulado.

Veremos que la claridad hallada en la identificación de los tres primeros descriptores de nivel (A1, A2 y B1) desaparece completamente a partir de B1, pasándose a una indiferenciación entre los siguientes tres niveles. Ni siquiera sumando los dos descriptores que se dan para el nivel B2, que está subdividido en B2.1 y B2.2, se acierta a escalonar adecuadamente y solamente el 50% de los expertos indica los descriptores de B2 dentro del nivel B2. Además, el descriptor considerado de más nivel es el de C1 (y no el C2).

Vamos a ver los datos porcentuales de las respuestas dadas por estos 10 profesores (véase el cuadro 3.1.7- A):

No hubo ninguna duda respecto a cuál era el descriptor definidor del nivel A1: las 10 personas decidieron acertadamente.

Respecto al nivel A2, 8 de los 10 acertaron. El fallo de los otros dos indica que no tenían clara la diferencia de las habilidades entre los niveles A2 y B1, por lo tanto entre usuario básico y usuario independiente.

El nivel B1 lo aciertan 7. Los tres que fallan prefieren pensar que este descriptor se mantiene todavía en el nivel A2, mostrando la misma dificultad para diferenciar A2 de B1, que ya hemos señalado anteriormente.

En general, se aciertan estos descriptores para A y B1, o, cuando menos, se ordenan adecuadamente de menor a mayor habilidad. Lo hacen el 70, 80 y 100% de profesores.

Respecto a los descriptores para los niveles B2, C1 y C2, los errores son mayores y se aprecia un alto grado de indiferenciación, con confusiones entre las habilidades de los tres, pues hay quien identifica el descriptor de B2 como definidor de B1 (cuadro 3.1.7

aa.), como definidor de C1 o incluso como definidor de C2; quien C1 lo da como definidor de C2; o quien C2 lo identifica como definidor de B2 o de simple B1.

De toda esta aparente confusión queda clara una cosa: el descriptor considerado de más nivel es el de las habilidades descritas bajo C1, ya que el 80% lo identificaron como definidor de nivel C2.

B2.1: indicaron correctamente el B2.1 solamente 2 personas de las 10 (como B2 y B2.1); las otras 8 se repartieron así: 5 lo dieron por C1 (como C1, C1.1 y C1.2), 2 por C2 (como C2 y C2.1) y 1 por B1.

B2.2: solamente 3 acertaron (como B2 y B2.1); las otras 7 se repartieron así: 4 lo dieron por C1 (como C1 y C1.2), 2 por B2.1 y 1 por B1.

C1 y C2: parece que se prefiere C1 como definidor de mayores habilidades ya que solamente 2 personas aciertan; y los otros 8 profesores piensan que corresponde a la definición del nivel C2. Los descriptores de C2, en cambio, muestran mayor variedad de interpretaciones: solo una persona acierta al afirmar que se refieren al nivel C2, otra los interpreta como definidores del nivel B1, otra como de C1, mientras la mayoría (7 de 10) lo interpretan como definidor del nivel B2 (como B2, B2.1 y B2.2).

Cuadro 3.1.7 a. Resultados del primer grupo de expertos

Nivel	Corrección	Resultados señalados (en negrita: correcto) (subrayado: lo preferido)	Comentarios
A1	100%	<u>A1</u> 90% A1.1 10%	Todos aciertan. - Una persona subdividió el nivel A1 en A1.1 y A1.2
A2	80%	A1.2 10% <u>A2.1</u> 10% <u>A2</u> 70% B1 10%	- Esta persona no lo diferenció del nivel A y creyó que el siguiente descriptor, del B1, correspondía al nivel A2. Lleva una adecuada gradación de niveles pero falla en diferenciar A1 de A2 y A2 del B. - Fallan en diferenciar nivel A de nivel B y cree que el siguiente descriptor, del B1, correspondía al nivel A2
B1	70%	A2 20% A2.2 10% <u>B1</u> 70%	- Fallan en diferenciar nivel A de nivel B.
B2.1	20%	B1 10% <u>B2</u> 10% <u>B2.1</u> 10% C1.1 10% <u>C1</u> 30% C1.2 10% C2 10% C2.1 10%	- Ambos descriptores parecen no participar del mismo tipo de habilidades, ya que quienes fallan optan por pensar que se refieren a habilidades de más nivel (o de menos en el caso de marcar B1 -un 10%-) pero no que se refieren al mismo nivel en grados diferentes. Lo que se deduciría si se decidiera que el B2.1 es B2.2 o viceversa, que solamente ocurre en el 20% de casos.
B2.2	30%	B1 10% <u>B2</u> 20% <u>B2.1</u> 20% <u>B2.2</u> 10% C1 30% C1.2 10%	- Del B2.2 se ha subrayado el B2 en todas sus variantes como preferido antes que el C1 (que es marcado por el 30%) porque entre todos los que marcan B2 suman el 50%.
C1	20%	<u>C2</u> 80% <u>C1</u> 20%	- Este es el descriptor preferido para definir el más alto nivel de habilidad de la lengua. Y el más claramente entendido dentro de los de nivel alto, ya que las confusiones son homogéneas y nadie lo identifica con habilidades de menor nivel.
C2	10%	B1 10% <u>B2</u> 30% <u>B2.1</u> 10% <u>B2.2</u> 30% C1 10% <u>C2</u> 10%	- Del C2 se ha subrayado el B2 en todas sus variantes como preferido, porque entre todos los que marcan B2 suman el 70%.

Cuadro 3.1.7 aa. Resumen de resultados del primer grupo de expertos

Personas que aciertan (de 10)	10	10	7	2	1	2	1
NIVEL	A1	A2	B1	B2.1	B2.2	C1	C2
Trasvases			< 3 A2	< 1 B1	< 1 B1 < 4 B2.1		<<<< 1 B1 <<< 4 B2 << 3 B2.2 < 1 C1
a otros				4 C1 >>	3 C1 >>		
niveles				1 C1.2 >>>	1 C1.2 >		
				2 C2 >>>>		8 C2 >	

Hasta ahora -como habíamos concluido tras el análisis de la primera escala del MCER- continuamos apreciando una división de niveles de solamente tres grados, con grandes ambigüedades en su interpretación, siendo C1, como hemos dicho, el representante de unas mayores habilidades:

- A1 y A2
- B1
- B2, C1 y C2.

3.1.7- B) Pruebas realizadas con expertos. Grupo segundo

En este segundo grupo, compuesto por 10 profesores agrupados en 5 parejas, también vamos a ver que los descriptores son muy ambiguos y están sometidos a todo tipo de interpretaciones; y, de nuevo, que los seis niveles que el MCER pretende medir se reducen a los tres que nosotros hemos postulado.

Repetimos -como en el análisis anterior- que la claridad en la identificación de los tres primeros descriptores de nivel (A1, A2 y B1) desaparece a partir de B1, con una indiferenciación entre los otros tres niveles, B2, C1 y C2. Y que C1 sigue pareciendo el definidor de mayores habilidades.

Vamos a ver los datos porcentuales sobre las respuestas dadas por estos 10 profesores (véanse cuadros 3.1.7 b y 3.1.7 bb).

Al igual que con el anterior grupo, no hubo ninguna duda respecto a cuál era el descriptor definidor del nivel A1 (100% de aciertos).

Respecto al nivel A2, también todos acertaron, pero un grupo decidió hacer en este nivel la subdivisión A2.1 y A2.2 y pensó que el descriptor de B1 era definidor del A2.2, con lo que se sigue viendo esa indiferenciación entre A2 y B1 en determinadas ocasiones.

El nivel B1 lo indican correctamente 4 grupos. El otro prefiere pensar que este descriptor se mantiene todavía en el nivel A2.2; lo que indica que este grupo no ve clara la diferencia de habilidades entre A y B.

Y respecto a los descriptores para los niveles B2, C1 y C2, los índices de acierto vuelven a ser muy bajos y se aprecia un alto grado de indiferenciación entre los tres niveles, con confusiones entre las habilidades de los tres: sigue habiendo quien identifica el descriptor de B2 como definidor del nivel B1 o del nivel C1 (aunque en este grupo nadie lo llevó hasta pensar que era el definidor del nivel C2); quien da el descriptor de C1 como definidor del nivel C2; y quien piensa que el descriptor de C2 define el nivel C1 o incluso un nivel más bajo, el nivel B2.

De esta aparente indeterminación de habilidades en los descriptores volvemos a ver que el descriptor considerado de más nivel es el de las habilidades descritas bajo C1, ya que el 60% lo identificaron como descriptor de C2.

En este grupo de profesores se aprecia una mayor homogeneidad en la interpretación de las habilidades de B2, pero el 40% da el descriptor de B2.1 como definidor de las habilidades de B2.2, es decir, se interpreta como descripción de mayor habilidad el descriptor de B2.1 que el de B2.2.

B2.1: ningún profesor indicó el descriptor de B2.1 como definidor de ese nivel (dos grupos lo marcaron como definidor de B2.2, 2 grupos como definidor de C1 y un grupo como definidor de B1).

B2.2: aunque todos lo identificaron como definidor del nivel B2, nadie lo marcó correctamente como definidor del nivel B2.2: 4 grupos lo marcaron como definidor de nivel B2.1 y uno como simple B2, sin subdividirlo.

Respecto a los niveles C1 y C2, se vuelve a preferir al descriptor de C1 como definidor de mayores habilidades de toda la escala: solamente 2 grupos lo marcan adecuadamente como el definidor de C1; en cambio, el resto piensa que define al nivel C2.

Y el descriptor del C2 vuelve a mostrar una mayor divergencia de interpretaciones: 2 grupos lo interpretan como definidor del nivel C2, otro lo da como definidor del nivel C1, y dos más como definidores del nivel B2.2.

Cuadro 3.1.7 b. Resultados del segundo grupo de expertos

Nivel	Corrección	Resultados señalados (en negrita: correcto) (subrayado: lo preferido)	Comentarios
A1	100%	<u>A1</u> 100%	Todos lo aciertan.
A2	100%	A2.1 20% <u>A2</u> 80%	- Un grupo -dos personas- subdividió el nivel A2 en A2.1 y A2.2 Este grupo creyó que el siguiente descriptor, del B1, correspondía al nivel A2.2. Lleva una adecuada gradación de niveles pero falla en diferenciar A2 del B1.
B1	80%	A2.2 20% <u>B1</u> 80%	- Un grupo falla en diferenciar nivel A2 de nivel B1.
B2.1	0%	B1 20% B2.2 40% <u>C1</u> 40%	- Un grupo lo marca como B1 pero en la gradación absoluta ocupan el 4 lugar en dificultad (si este grupo no hubiera decidido que había A2.1 y A.2.2, habría seleccionado este descriptor como B2, correctamente).
B2.2	0%	B2 20% <u>B2.1</u> 80%	- Los grupos seleccionan adecuadamente los descriptores de B2 pero fallan al graduar sus habilidades por cambiarlos y dar el B2.1 como B2.2 y viceversa. Por ello en el porcentaje sale 0% de aciertos. En un 80% se considera que el B2.2 es el descriptor de habilidad de menos nivel y se da por B2.1.
(B2)	(40%)		- Si consideráramos el número de grupos que dan ambos descriptores por descriptores de B2 a secas el número de aciertos ascendería al 40%
C1	40%	<u>C2</u> 60% C1 40%	- Este es el descriptor preferido para definir el más alto nivel de habilidad de la lengua, ya que se da por C2 en vez de por C1. Y el más claramente interpretado como descriptor de habilidades de nivel alto, ya que nadie lo marca como descriptor de habilidades de menor nivel.
C2	40%	B2.2 40% C1 20% <u>C2</u> 40%	

Cuadro 3.1.7 bb. Resumen de resultados del segundo grupo de expertos

Personas que aciertan (de 5)	5	5	4	0	0	2	2
NIVEL	A1	A2	B1	B2.1	B2.2	C1	C2
Trasvases			< 1 A2.2	< 1 B1	< 4 B2.1 << 1 B2		< 1 C1 <<< 2 B2.2
a otros niveles				2 B2.2 > 2 C1 >>		3 C2 >	

3.1.7- C) Pruebas realizadas con expertos. Grupo tercero

Este tercer grupo, compuesto por 30 profesores agrupados en 10 tríos, reafirma lo ya comprobado en los casos anteriores. Y C1 les sigue pareciendo el definidor de mayores habilidades. Vamos a ver los datos porcentuales sobre las respuestas dadas por los 30 profesores (véanse cuadros 3.1.7 c y 3.1.7 cc):

Al igual que con el anterior grupo, no hubo ninguna duda respecto a cuál era el descriptor definidor del nivel A1.

Respecto al nivel A2, también todos acertaron, pero nuevamente un grupo decidió hacer en este nivel la subdivisión A2.1 y A2.2, mostrando las mismas dificultades para diferenciar el nivel de usuario básico, A, del independiente, B1, ya señaladas.

El nivel B1 lo indican correctamente 9 grupos. El otro grupo prefiere pensar que este descriptor define el nivel A2.2; lo que indica que este grupo no ve clara la diferencia de habilidades entre A2 y B1.

Y respecto a los descriptores para los niveles B2, C1 y C2, los índices de aciertos vuelven a ser bajos y se aprecia el mismo grado alto de indiferenciación entre los tres niveles, con confusiones entre las habilidades de los tres: sigue habiendo quien

identifica el descriptor de B2 como definidor del nivel B1 o del C1 (aunque sin llegar a marcarlo como definidor de C2), quien interpreta el descriptor de C1 como definidor del nivel C2; y quien ve en el descriptor de C2 la definición de los niveles C1 o B2.

De esta indeterminación de habilidades en los descriptores volvemos a ver que el descriptor considerado como definidor de más nivel es el de las habilidades descritas bajo el descriptor C1, ya que el 70% lo identificaron como definidor del nivel C2.

B2.1: el 20% indicaron al descriptor de B2.1 como propio de su nivel; un grupo lo marcó como definidor de B1, dos grupos lo marcaron como definidor del nivel B2.2, 2 grupos como de nivel C1, un grupo como de nivel C1.2 y otro como de nivel C2.

B2.2: solamente el 20% lo marcó como correctamente B2.2: 2 grupos (20%) lo marcaron como definidor de B2 simplemente, el 50% (5 grupos) como definidor de B2.1 y un grupo (10%, tres personas) como definidor de C2.

Respecto a los niveles C1 y C2, se vuelve a preferir al C1 como descriptor definidor de mayores habilidades: solamente 2 grupos lo marcan como C1; frente a 7 grupos (70%) que piensan que corresponde al C2 y un grupo lo da como definidor de C1.2.

El C2 muestra de nuevo mayor variedad de interpretaciones: solo un grupo lo da efectivamente como C2, frente a una mayoría de 4 grupos (40%) que lo interpretan como definidor del nivel C1; otro grupo como definidor del nivel B2; tres grupos como definidor del nivel B2.2; y uno como definidor del nivel C1.1.

Cuadro 3.1.7 c. Resultados del tercer grupo de expertos

Nivel	corrección	Resultados señalados (en negrita: correcto) (subrayado: lo preferido)	Comentarios
A1	100%	<u>A1</u> 100%	Todos aciertan.
A2	100%	A2.1 10% <u>A2</u> 90%	- Un grupo -tres personas- subdividió el nivel A2 en A2.1 y A2.2 Este grupo creyó que el siguiente descriptor, del B1, correspondía al nivel A2.2. Lleva una adecuada gradación de niveles pero falla en diferenciar A2 del B1.
B1	90%	A2.2 10% <u>B1</u> 90%	- Un grupo falla en diferenciar nivel A2 de nivel B1.
B2.1	20%	B1 10% B2.1 20% B2.2 20% <u>C1</u> 20% <u>C1.1</u> 10% C2 20%	- Un grupo lo marca como B1 pero en la gradación absoluta ocupan el 4º lugar en dificultad (si este grupo no hubiera decidido que había A2. 1 y A.2.2, habría seleccionado este descriptor como propio de B2, correctamente).
B2.2	20%	B2 20% <u>B2.1</u> 50% B2.2 20% C1.1 10%	
C1	20%	C1 20% C1.2 10% <u>C2</u> 70%	- Este es el descriptor preferido para definir el más alto nivel de habilidad de la lengua ya que se da por C2 en vez de por C1. Y el más claramente interpretado como descriptor de habilidades de nivel alto, ya que nadie lo marca como descriptor de habilidades de menor nivel.
C2	10%	B2 10% B2.2 30% <u>C1.1</u> 10% <u>C1</u> 40% C2 10%	

Cuadro 3.1.7 cc. Resumen de resultados del tercer grupo de expertos

Personas que aciertan (de 10)	10	18	9	2	2	2	1
NIVEL	A1	A2	B1	B2.1	B2.2	C1	C2
Trasvases			<1 A2.2	<1 B1	< 2 B2 < 5 B2.1		<<< 1 B2 << 3 B2.2 < 1 C1.1 < 4 C1
a				2 B2.2 >			
otros				1 C1.1 >	1 C1.1 >	1 C1.2 >	
niveles						7 C2 >>	

3.1.8 Conclusiones del experimento sobre la ambigüedad en el MCER

Acabamos de ver que hay bastantes coincidencias entre todos los profesores, sean de donde sean, trabajen en el mismo centro o no y trabajen independientemente o en grupos.

Tras estos análisis se comprueban todos los postulados que veníamos comentando:

- El alto grado de ambigüedad de los descriptores y la necesidad de formación para entender a qué se refiere cada descriptor y eliminar las interpretaciones y dudas: solamente el A1, A2 y B1 (este último con algunas vacilaciones entre A2 y B1) se identifican claramente.

- Que el MCER diferencia tres en vez de los seis niveles que pretende.

- Que es insuficiente la diferencia entre los niveles B2, C1 y C2.

- Que la definición del C2 basada en la capacidad de hacer discursos no es clara porque la palabra *discursos* es insuficiente para identificar un nivel alto y muchas personas lo identifican con niveles más bajos, B2, C1 o incluso B1.

- Que se determina al descriptor de C1 como el que define las habilidades de mayor nivel, y al de B2.2 como de menor nivel que el de B2.1.

Del total de profesores que participaron, solamente el 10% identifica al descriptor de C2 como propio de C2, y solamente un 30% acierta a identificar los de nivel B2 y C1 como correspondientes a esos niveles de habilidad (descriptores de posiciones 4, 5 y 6).

Por esto mantenemos, ahora no como resultado de nuestras interpretaciones subjetivas, sino con los datos de estos 50 profesores, que las escalas de esta habilidad del MCER no diferencian 6 niveles de habilidad, sino solamente tres: A1 y A2 por un lado, B1 por otro y el resto (B2, C1 y C2) por otro.

El cuadro siguiente (3.1.8) muestra los datos expresados porcentualmente, según las respuestas dadas por los expertos. Los porcentajes de acierto en definir la sencillez o complejidad de las habilidades descritas se dan eliminando la dificultad de tener que determinar el subnivel B2.1 y B2.2 adecuadamente dentro del nivel B2 (que se compone de dos descriptores de habilidad) y considerando solamente el orden en que marcaron los descriptores en una escala de más sencillo a más difícil, del 1 al 7.

El cuadro considera solamente si se acertó a graduarlos y ordenarlos de más sencillo a más difícil, sin tener que delimitar que dos descriptores correspondían al mismo nivel. Así, se da por acertada la gradación de B1 al situarlo en tercer lugar (efectivamente es el tercero ya que A1 es primero y A2 es segundo) aunque se diga que ese descriptor corresponda al nivel A2.2 (pensando que el primero es para A1 y el segundo para A2.1, como ocurrió en varias ocasiones).

La primera columna indica los descriptores ordenados desde el de habilidad menor (el 1 es el de A1) al de habilidad mayor (el 7, que corresponde al descriptor de habilidad de C2).

Las columnas dos, tres y cuatro, son los porcentajes de profesores de los grupos uno, dos y tres que marcaron adecuadamente el orden de cada descriptor; y damos entre paréntesis el número de descriptor que se marcó por la mayoría para esa posición, el que se prefiere como definidor de ese nivel. Por ejemplo, en el grupo 1 aparece el 30% como porcentaje de personas que marcaron el descriptor 4 como acertado, o sea en cuarto lugar.

Este porcentaje es muy bajo y hubo más personas que dieron al descriptor 5 la cuarta posición, interpretándolo de menor habilidad que el cuarto. Por ello ponemos entre paréntesis el 5, acompañado del porcentaje de personas que lo marcó así. Igualmente el descriptor 5 solamente lo colocaron en la quinta posición el 10% de personas de ese grupo 1, y hubo más personas (50%) que lo colocaron en la posición 4

(interpretándolo como descriptor de una menor habilidad, o lo marcaron en la posición 6 (el 40% de personas), interpretándolo como descriptor de más habilidad).

Asimismo la columna de la derecha da los resultados del conjunto de los 50 profesores que realizaron esta prueba globalmente.

Cuadro 3.1.8. Resultados conjuntos de los tres grupos de expertos

Escala de dificultad de los descriptores ordenados del 1 al 7	Grupo 1	Grupo 2	Grupo 3	Total de los 50 profesores
1	100%	100%	100%	50 100%
2	90%	100%	100%	49 98%
3	90%	100%	100%	49 98%
4	30% (5 – 40%)	20% (5 – 40%) (6 – 40%)	30% (6 - 30%)	15 30%
5	10% (4 – 50%) (6 – 40%)	20% (4 – 80%)	40% (4 - 60%)	15 30%
6	20% (7 – 80%)	40% (7 – 60%)	30% (7 - 70%)	16 32%
7	10% (5 – 70%)	10% (5 – 40%) (6 – 40%)	10% (5 - 40%) (6 - 40%)	5 10%

3.2 Comparación entre los niveles de las *guidelines* de ACTFL y los establecidos por las escalas del MCER

Tras esta cata en varias escalas de descriptores de habilidad y dominio del MCER y con las opiniones de los expertos que reafirmaron nuestras suposiciones respecto a la ambigüedad con que están descritas las escalas del MCER, vamos a proceder a la

comparación de algunas escalas del MCER con las de la OPI de ACTFL, para definir las equivalencias planteadas en la investigación.

En este apartado se hace una comparación entre la escala de la OPI de ACTFL y tres escalas del MCER: la *Escala global* (§3.2.1), la de *Niveles cualitativos de referencia* (§3.2.2), y la de *Autoevaluación: interacción oral* (§3.2.3).

Posteriormente, (§3.2.4) muestra los contenidos generales del MCER según los criterios de ACTFL, con el fin de tener una más clara relación entre ambos sistemas y escalas.

Las conclusiones de la comparación entre ambos sistemas se ofrecen en §3.2.5, donde se presenta un cuadro de posibles correspondencias entre las escalas del MCER y la de ACTFL¹⁹². Con ellas vamos a justificar las equivalencias que hemos planteado en nuestra investigación¹⁹³.

3.2.1 Comparación con la escala de *Niveles comunes de referencia: escala global*

Comparamos ambas escalas: la de la OPI y la del MCER de *Niveles comunes de referencia: escala global*. Repetimos la escala que se ha dado en §2.7.2- A, porque ahora identificamos separadamente cada descriptor para poder referirnos a él en los comentarios; y acompañamos ya cada nivel con la adscripción al nivel o niveles en las *guidelines* de evaluación oral de ACTFL, según nuestra opinión.

Tabla 5. *Niveles comunes de referencia: escala global*

C2 AM (AH)	<ul style="list-style-type: none"> a. Es capaz de comprender con facilidad prácticamente todo lo que oye. b. Sabe reconstruir la información y los argumentos procedentes de diversas fuentes ya sean en lengua hablada o escrita, y presentarlos de manera coherente y resumida. c. Puede expresarse espontáneamente, con gran fluidez y con un grado de precisión que le permite diferenciar pequeños matices de significado incluso en situaciones de gran complejidad. (AH).
---------------------------------------	---

¹⁹² A partir del cual se ha establecido el cuadro presentado al inicio de este trabajo (§1.3).

¹⁹³ En este análisis comparativo entre las escalas del MCER y ACTFL vamos a plantear de manera teórica las ambigüedades o defectos que se encuentran en las escalas. Posteriormente -al hacer las calificaciones de los discursos de los estudiantes, en el capítulo 4- veremos si las escalas utilizadas y las ambigüedades y limitaciones que ya estamos encontrando son un problema para la clasificación o no; y si, de cualquier forma, permiten hacer diferencias útiles entre los niveles de los candidatos.

<p>C1 AM</p>	<p>a. Es capaz de comprender una amplia variedad de textos extensos y con cierto nivel de exigencia, así como reconocer en ellos sentidos implícitos. b. Sabe expresarse de forma fluida y espontánea sin muestras muy evidentes de esfuerzo para encontrar la expresión adecuada. c. Puede hacer un uso flexible y efectivo del idioma para fines sociales, académicos y profesionales. (AM). d. Puede producir textos claros, bien estructurados y detallados sobre temas de cierta complejidad, mostrando un uso correcto de los mecanismos de organización, articulación y cohesión del texto.</p>
<p>B2 AL (AM)</p>	<p>a. Es capaz de entender las ideas principales de textos complejos que traten de temas tanto concretos como abstractos incluso si son de carácter técnico siempre que estén dentro de su campo de especialización. b. Puede relacionarse con hablantes nativos con un grado suficiente de fluidez y naturalidad de modo que la comunicación se realice sin esfuerzo por parte de ninguno de los interlocutores. (AM). c. Puede producir textos claros y detallados sobre temas diversos así como defender un punto de vista sobre temas generales indicando los pros y los contras de las distintas opciones. (AL).</p>
<p>B1 IM (IH)</p>	<p>a. Es capaz de comprender los puntos principales de textos claros y en lengua estándar si tratan sobre cuestiones que le son conocidas, ya sea en situaciones de trabajo, de estudio o de ocio. b. Sabe desenvolverse en la mayor parte de situaciones que pueden surgir durante un viaje por zonas donde se utiliza la lengua. c. Es capaz de producir textos sencillos y coherentes sobre temas que le son familiares o en los que tiene un interés personal. d. Puede describir experiencias, acontecimientos, deseos y aspiraciones así como justificar brevemente sus opiniones o explicar sus planes.</p>
<p>A2 IM</p>	<p>a. Es capaz de comprender frases y expresiones de uso frecuente relacionadas con áreas de experiencia que le son especialmente relevantes (información básica sobre sí mismo y su familia, compras, lugares de interés, ocupaciones, etc.). b. Sabe comunicarse a la hora de llevar a cabo tareas simples y cotidianas que no requieran más que intercambios sencillos y directos de información sobre cuestiones que le son conocidas o habituales. c. Sabe describir en términos sencillos aspectos de su pasado y su entorno así como cuestiones relacionadas con sus necesidades inmediatas.</p>
<p>A1 NM NH</p>	<p>a. Es capaz de comprender y utilizar expresiones cotidianas de uso muy frecuente así como frases sencillas destinadas a satisfacer necesidades de tipo inmediato. b. Puede presentarse a sí mismo y a otros, pedir y dar información personal básica sobre su domicilio, sus pertenencias y las personas que conoce. c. Puede relacionarse de forma elemental siempre que su interlocutor hable despacio y con claridad y esté dispuesto a cooperar.</p>

(Consejo de Europa, 2002: 26).

Análisis de A1: sería inicial medio o alto de ACTFL

No nos parecen claros los descriptores (a) y (b). Nos parece que sus definiciones son muy simples y, ante la necesidad de no dar descripciones negativas, el MCER fuerza demasiado las formas y cae en el fallo de no definir el nivel efectivamente y que haya que interpretarlo por ideas preconcebidas sobre el nivel.

El primero dice, “comprender y utilizar expresiones cotidianas de uso muy frecuente” (a). Sólo si estamos inmersos en un mismo sistema y metodología educativa o si nos referimos a nuestra propia vida sabemos a qué se refiere. No está claro cuáles van a ser esas expresiones cotidianas de uso muy frecuente: ¿“este tamal está muy rico”, “pásame la llave del 3”, “te cae bien esa falda”? Incluso podrían ser enunciados con palabras más técnicas, referidos a personas cuyo único contacto con la lengua sea en el trabajo.

Mejor nos parece la definición de ACTFL que explicita cuáles son los temas básicos: intercambiar saludos, presentarse y nombrar unos cuantos objetos de su entorno familiar (inicial bajo), limitándose al contexto concreto en que han aprendido la lengua (inicial medio), pero limitándose al uso de palabras o sintagmas memorizados o en forma de eco.

El siguiente descriptor, (b) “(poder) presentarse a sí mismo y a otros, pedir y dar información personal básica sobre su domicilio, sus pertenencias y las personas que conoce” creemos que tampoco aclara suficientemente el alcance de los temas. Se dice “información básica”, pero ello no nos parece suficiente para que esta quede delimitada.

Comparándolo con la escala de ACTFL se podría considerar que abarca desde el inicial medio al intermedio bajo. En concreto, creemos que cada descriptor sería un nivel diferente en ACTFL.

Así:

El descriptor (a) sería inicial medio: “Es capaz de comprender y utilizar expresiones cotidianas de uso muy frecuente así como frases sencillas destinadas a satisfacer necesidades de tipo inmediato”. Suponiendo que el nivel de expresión sea tan reducido que se limite a poder satisfacer las necesidades más inmediatas, y que en cuanto nos salimos de un uso frecuente de expresiones cotidianas será imposible toda comunicación. Lo que en ACTFL se expresa al decir que solamente reconoce un

limitado número de expresiones de uso cotidiano que usa de forma memorística sin posibilidades de transformación alguna.

El descriptor (b) sería inicial alto: “Puede presentarse a sí mismo y a otros, pedir y dar información personal básica sobre su domicilio, sus pertenencias y las personas que conoce”, aunque no exprese claramente el tipo y extensión del discurso.

Y el descriptor (c) llegaría hasta intermedio bajo: “Puede relacionarse de forma elemental siempre que su interlocutor hable despacio y con claridad y esté dispuesto a cooperar”.

Análisis de A2: sería intermedio bajo o medio de ACTFL

Su primer descriptor (a) se ocupa solamente de comprensión y es el mismo que en A1, solo que con una cantidad de temas mayor, que abarca “áreas de experiencia que le son especialmente relevantes, información básica sobre sí mismo y su familia, compras, lugares de interés, ocupaciones, etc.”. Creemos que el nivel de comprensión de una persona puede ser muy alto, sin que eso suponga capacidad de expresión alguna. Por lo tanto, no aceptamos esta diferenciación entre A1 y A2 basada en la comprensión. Mejor sería haber definido un nivel de comprensión menor para A1, por ejemplo, decir que el hablante apenas comprende nada porque carece de toda habilidad y conocimiento acerca de la lengua; así se podría dar este descriptor en A2, definiendo a un hablante que sí tiene estas capacidades.

El descriptor (b.), “Sabe comunicarse a la hora de llevar a cabo tareas simples y cotidianas que no requieran más que intercambios sencillos y directos de información sobre cuestiones que le son conocidas o habituales”, se parece en extensión y contenido al A1. No vemos por qué todo lo que se explica no puede estar implícito en lo descrito en A1. Solamente se diferencian en un aspecto, que tomamos con pinzas, el de referirse en A1 a “puede relacionarse” (para lo que se necesitan pocas palabras) y “sabe comunicarse” en A2 (que ya sí necesita más habilidad).

El (c) nos sorprende en su definición: creemos que se mezcla lo más sencillo del nivel anterior, ser capaz de desenvolverse en los contextos habituales y solucionar necesidades inmediatas, junto con la posibilidad de hablar del pasado: “describir (...) aspectos de su pasado”.

Pero no se aclara qué significa hablar del pasado, formalización lingüística, extensión, temática, etc. Piénsese en que una cosa es entender a un hablante si dice “yo

ayer voy al campo y como con la familia”, y otra es aceptar que eso sea “describir en términos sencillos aspectos de su pasado”. La función puede preceder a la capacidad morfológica pero esto no hace que haya más nivel de formulación lingüística todavía, ni por lo tanto más nivel en el uso de la lengua.

Y por el contrario, si un candidato logra hacerse entender en pasado, medianamente bien, calificarlo como A2 es un salto de nivel excesivo: primero, es solamente un nivel más que el nivel inmediatamente anterior, en el que apenas se puede expresar nada, y en segundo lugar, esta definición no mantiene relación con el resto de capacidades limitadas descritas en este mismo nivel en los descriptores a. y b.

Comparado con ACTFL, este nivel podría ser el intermedio bajo o medio, ya que el candidato puede tratar temas cotidianos con cierta tranquilidad y fluidez e incluso lograría iniciar ciertas narraciones sencillas. Raramente se logra una narración en intermedio bajo, pero en determinados candidatos es posible en intermedio medio.

Todo lo que estamos diciendo sobre ambos niveles coincide con los análisis de las escalas anteriores de las que sacábamos la conclusión de que ambos -A1 y A2- marcan un mismo nivel ambiguamente diferenciado por elementos de cantidad y calidad; y, en este caso, diferenciados por aspectos funcionales, el uso de la narración en pasado, que no se dice claramente a qué se refiere.

Análisis de B1: sería intermedio medio o alto de ACTFL

Este nivel, de acuerdo con las conclusiones establecidas en los análisis de las escalas anteriores, marca una diferencia con respecto al A1 y A2. Parece diferente de los niveles inmediatamente inferiores, así como del inmediatamente superior, el B2.

Se refiere a lo que en ACTFL es intermedio medio: al hecho de describir cosas de un ámbito cercano y directo al hablante, sus propias experiencias, acontecimientos en los que él haya tenido participación directa y aspiraciones personales o planes, además de opiniones breves. A esto se refiere su descriptor de habilidad escrita que dice “es capaz de producir textos sencillos y coherentes sobre *temas que le son familiares o en los que tiene un interés personal*”.

No obstante, sus descriptores son bastante ambiguos.

El descriptor (b), “Sabe desenvolverse en la mayor parte de situaciones que pueden surgir durante un viaje por zonas donde se utiliza la lengua”, nos parece demasiado pobremente resuelto y muy ambiguo. No se aclara qué tipo de situaciones pueden surgir

durante un viaje y es evidente que puede haberlas de tipo sencillo o complicado (e incluso de otros tipos). Probablemente se refiera a las sencillas, ya que las complicadas requerirán mayores habilidades lingüísticas. Pero esto hay que deducirlo pensando en que se trata del tercer nivel y que en el anterior apenas existía habilidad suficiente como para desenvolverse en situaciones cotidianas; o sea, mediante un prejuicio sobre sus habilidades.

En cuanto al otro descriptor del nivel, el (d), “Puede describir experiencias, acontecimientos, deseos y aspiraciones así como justificar brevemente sus opiniones o explicar sus planes”, tenemos que decir lo mismo: no se aclara a qué nivel o de qué manera se pueden describir esas cosas. Se limita a dar temas (experiencias, acontecimientos y aspiraciones) y funciones (describir, opinar) pero sin explicar su grado de alcance, salvo en cuanto a opiniones (brevemente). Porque a cada uno de esos temas les pueden corresponder diferentes formas y grados de descripción y explicación, y todo ello no puede estar en el mismo nivel B1.

Nos parece que ACTFL lo soluciona mejor al describir el intermedio medio como propio de hablantes que pueden contar acontecimientos, planes, etc. y tienen capacidad para narrar incipientemente hechos en los que ellos sean protagonistas, con calidad y cantidad suficiente.

En la ambigüedad del MCER, este nivel podría llegar también al intermedio alto de ACTFL, que permite describir, explicar, opinar y comparar a un nivel de organización en forma de párrafos y frases encadenadas, sobre temas generales en los que el candidato no es el único protagonista, aunque sin lograr mantenerse en esas cotas de calidad y cantidad o soltura durante todo el tiempo. Pero es interpretable si el MCER quiere alcanzar el dominio y habilidades del intermedio alto de ACTFL, ya que pensamos que se refiere exclusivamente a la habilidad en ámbitos de actuación personal del candidato.

Análisis de B2: sería el avanzado de ACTFL

El nivel B2 se llama en el MCER “avanzado”. Y sus descriptores quedan bastante claros y cercanos a lo descrito en las *guidelines* de ACTFL. Es de resaltar que cuando el MCER deja la ambigüedad en la definición de sus descriptores, se acerca enormemente a las definiciones de ACTFL, por lo que su influencia nos parece evidente. Y cuando no

se deja llevar por las descripciones de nivel de ACTFL cae en ambigüedades grandes y definiciones poco claras.

El descriptor (b), “Puede relacionarse con hablantes nativos con un grado suficiente de fluidez y naturalidad de modo que la comunicación se realice sin esfuerzo por parte de ninguno de los interlocutores”, se corresponde enteramente con el concepto de avanzado de ACTFL que se define como la capacidad de comunicarse con hablantes nativos no acostumbrados a hablar con extranjeros, sin necesidad de ayuda y sobre casi cualquier tema.

En cuanto a las conclusiones establecidas en los análisis de las escalas anteriores, se mantiene la apreciación de que este nivel B2 marca diferencias mayores con B1, de más calado que con los siguientes C1 y C2. Por ello queda separado de B1 y en cambio se podría agrupar con los “C” en un mismo macronivel.

Análisis de C1: sería avanzado medio de ACTFL

En este nivel el MCER vuelve a ser claro y muy cercano al descrito en ACTFL al definir sus habilidades con parecidas características al avanzado medio: utilizar el lenguaje para temas sociales, generales o profesionales y con una fluidez suficiente y estrategias y soltura tales que no se note la necesidad de buscar el lenguaje necesario en la comunicación.

Estaríamos en un avanzado medio, especialmente por el uso de temas profesionales, por el uso flexible del idioma y por la estructura y organización del discurso. Sin llegar a avanzado alto, que posee otras habilidades propias del superior, que no entran en este descriptor del MCER, todavía.

A nivel de coherencia textual (oral) se define una organización del discurso basada en párrafos y marcadores, suficientemente compleja, propia del mismo avanzado medio de ACTFL. Que viene a ser equivalente al descriptor (d) “Puede producir textos claros, bien estructurados y detallados sobre temas de cierta complejidad, mostrando un uso correcto de los mecanismos de organización, articulación y cohesión del texto”. Con ello, se define la tipología textual mejor que en B2, pero no vemos claro que en aquél no quedara implícita esta misma coherencia textual, al decir que hay que defender los puntos de vista e indicar los pros y contras.

Análisis del C2: sería avanzado medio o alto de ACTFL

Este nivel, comparándolo con la escala de ACTFL, está fundamentalmente en un avanzado medio o en un avanzado alto, porque se hace referencia a que el hablante se puede comunicar perfectamente en situaciones de gran complejidad: “diferenciar pequeños matices de significado incluso en situaciones de gran complejidad”. Sin llegar a superior porque no muestra otras habilidades de discurso abstracto.

3.2.2 Comparación con la escala de *Niveles cualitativos de referencia*

Comparamos la escala del MCER de *Niveles cualitativos de referencia* (ya analizada en §3.1.6), numerada como tabla 4, y la de ACTFL. En esa escala del MCER, cada descriptor está numerado para poder referirnos a él en los comentarios, y se acompaña de la apreciación del nivel que creemos que le correspondería en la escala de ACTFL.

Análisis de A1: sería inicial medio a alto de ACTFL

Estos descriptores muestran una habilidad para hablar en situaciones sencillas o cotidianas, propias de inicial alto e intermedio bajo de ACTFL. Pero contrariamente a lo descrito en otras escalas anterior y posteriormente comentadas, en estos descriptores se describe el alcance, la corrección y la fluidez, con mayores limitaciones de habilidad, pues se restringe a frases memorizadas y bien preparadas, “repertorio básico de palabras y frases sencillas” (1), “repertorio memorizado” (2), “preparadas de antemano” (3) y no a un uso creativo del material aprendido.

Por ello el nivel iría en una horquilla desde el inicial medio, que solamente puede manejar material aprendido, al inicial alto, que hace un uso creativo de lo aprendido, hablando sobre temas cercanos o en los que el hablante es el centro, aunque no siempre pueda mantener el discurso.

En esto, coincide con lo comentado hasta ahora para el A1: horquilla entre inicial medio y alto, con la salvedad de que en otras escalas no se habían descrito tales limitaciones de alcance, corrección y fluidez para A1. Limitaciones que ahora rebajan habilidad oral en el MCER a la descrita en ACTFL en inicial medio.

Análisis de A2: sería desde inicial medio hasta intermedio bajo de ACTFL

Estos descriptores continúan la definición de la habilidad para hablar en situaciones sencillas o cotidianas, propias de inicial medio, alto o intermedio bajo de ACTFL.

La escala sigue planteando fuertes limitaciones orales, debidas a la dependencia de frases memorizadas, “expresiones, grupos de unas pocas palabras y fórmulas memorizadas” (6), continuas pausas, dudas y reformulaciones, “expresiones muy breves, aunque resultan muy evidentes las pausas, las dudas iniciales y la reformulación” (8) y grandes problemas para mantener una conversación, “apenas comprende lo suficiente como para mantener una conversación” (9).

En el campo de fluidez y coherencia, las posibilidades se abren hasta ser capaz de “hacerse entender con expresiones muy breves” (10) y “enlazar grupos de palabras con conectores sencillos” (8).

Análisis de B1: sería intermedio bajo o intermedio medio de ACTFL

Este nivel, tanto en esta como en otras escalas, presenta unas capacidades que claramente corresponden al intermedio de ACTFL, sin llegar al intermedio alto en ningún momento. La horquilla abarca dos niveles, intermedio bajo e intermedio medio.

Ya hemos visto como los descriptores de los niveles A1 y A2 del MCER también entran en el intermedio. Esto hace que un candidato que sea capaz de mostrar algunas habilidades de A2, puede ser evaluado como intermedio o incluso como inicial dependiendo de las características que se consideren de su discurso.

Análisis de B2: sería avanzado de ACTFL

En general, este nivel corresponde con el avanzado de ACTFL, sin precisar más, por sus habilidades: ser capaz de hablar con nativos sin problemas, uso de oraciones complejas, participación en debates, tratamiento de temas generales y capacidad para tomar la iniciativa y los turnos de palabra sin depender del interlocutor.

Solo hay dos palabras que permiten una interpretación más baja o más alta de las habilidades: en el campo de la interacción se habla de temas cotidianos, lo que bajaría el nivel de ACTFL hasta intermedio alto; y en el campo de la coherencia se habla de

discursos, aunque con ciertas restricciones, pues se cita “número limitado de mecanismos de cohesión”, lo que subiría el nivel a un avanzado medio.

Análisis de C1: sería avanzado medio de ACTFL

Los descriptores mantienen la coherencia en todos los campos, salvo lo ya apuntado respecto a la ambigüedad de los términos “académico” y “discurso”.

Aunque podría pensarse que esto lleva las capacidades hasta generar discursos extendidos sobre temas de alto nivel y abstractos, cercanos al superior de ACTFL, las limitaciones que se presentan en el campo de la fluidez, ya comentadas, lo rebajan hasta un avanzado medio.

Lo mismo podemos defender desde otro punto de vista no considerado en el MCER, pero importante para ACTFL: el no tratar temas de nivel superior, “defender opiniones” y “hacer hipótesis”, da a sus habilidades un techo avanzado medio, ya que por muy correcto e incluso abstracto que sea el discurso que el hablante produzca, si este no se genera al tratar temas de complejidad superior, el nivel final del discurso y de sus habilidades queda en avanzado medio.

Es por estas razones por lo que interpretamos que estos descriptores no hacen referencia a un nivel mayor que avanzado medio. A pesar de que en el ámbito académico se pueda llegar a un uso totalmente abstracto y discursivo de la lengua como el que ACTFL define para el nivel superior, y de que el descriptor (22) habla de ausencia de errores, otra característica del nivel superior en ACTFL.

Análisis de C2: sería también avanzado medio de ACTFL

El descriptor que dice “mantiene un consistente dominio gramatical de lengua complejo, aunque su atención esté pendiente de otros aspectos (por ejemplo, de planificación o del seguimiento de las reacciones de otros”, (22) manifiesta capacidades altas para utilizar el lenguaje sin cortapisas, lo que en ACTFL se considera propio de cualquier avanzado medio.

Volvemos a decir lo apuntado en el anterior descriptor: en ACTFL, para llegar al avanzado alto hay que mantener esas características del discurso pero tratando temas de mayor complejidad.

Este descriptor considera un elemento nuevo, el uso del lenguaje en situaciones coloquiales con un adecuado uso de expresiones propias del contexto. En ACTFL esto no se tiene en cuenta directamente. No obstante, si queremos tratar de ver en qué nivel se podrían incluir, esta capacidad nativa bien podría existir desde el avanzado medio, ya que ese es el nivel de los nativos definido por ACTFL¹⁹⁴. El resto de descriptores no ofrecen más explicaciones, sobre los temas, las funciones o la estructura textual como para equiparlo al nivel superior de ACTFL. Por ello nos quedamos en lo que es un avanzado medio.

3.2.3 Comparación con la escala de *Autoevaluación: interacción oral*

Comparamos y analizamos la escala de ACTFL con la escala del MCER de *Autoevaluación: interacción oral* (Tabla 6). Junto a la nomenclatura de niveles del MCER se da la comparación -entre paréntesis- del nivel que estimamos que le correspondería en ACTFL.

Tabla 6. *Autoevaluación: Interacción oral*

AUTO EVALUACIÓN	A1 (NH/ IL)	A2 (IL)	B1 (IM)
Interacción oral	Puedo participar en una conversación de forma sencilla siempre que la otra persona esté dispuesta a repetir lo que ha dicho o a decirlo con otras palabras y a una velocidad más lenta, y me ayude a formular lo que intento decir. Planteo y contesto a preguntas sencillas sobre temas de necesidad inmediata o asuntos muy habituales.	Puedo comunicarme en tareas sencillas y habituales que requieren un intercambio simple y directo de información sobre actividades y asuntos cotidianos. Soy capaz de realizar intercambios sociales muy breves, aunque, por lo general, no puedo comprender lo suficiente como para mantener la conversación por mi mismo.	Sé desenvolverse en casi todas las situaciones que se me presentan cuando viajo. Puedo participar espontáneamente en una conversación que trate temas cotidianos de interés personal o que sean pertinentes para la vida diaria (por ejemplo, familia, aficiones, trabajo, viajes y acontecimientos actuales).

¹⁹⁴ Si vemos la escala de ACTFL, comprobamos que esta considera en el nivel superior unas capacidades mayores que las habituales en los nativos, habilidades fundadas en el uso de discurso extendido al tratar temas abstractos, defendiendo opiniones, entre otras.

AUTO EVALUACIÓN	B2(IH-AL)	C1 (AM)	C2 (AM)
Interacción oral	Puedo participar en una conversación con cierta fluidez o espontaneidad, lo que posibilita la conversación normal con hablantes nativos. Puedo tomar parte activa en debates desarrollados en situaciones cotidianas, explicando y defendiendo mis puntos de vista.	Me expreso con espontaneidad sin tener que buscar de forma muy evidente las expresiones adecuadas. Utilizo lenguaje con flexibilidad y eficacia para fines sociales y profesionales. Formulo ideas y opiniones con precisión y relaciono mis intervenciones hábilmente con las de otros hablantes.	Tomo parte sin esfuerzo en cualquier conversación o debate, y conozco bien modismos, frases hechas y expresiones coloquiales. Me expreso con fluidez y transmito matices sutiles de sentido con precisión. Si tengo un problema, sorteo la dificultad con tanta discreción, que los demás apenas se dan cuenta.

(Consejo de Europa, 2002: 30-31).

Análisis de A1: sería inicial alto o incluso intermedio bajo de ACTFL

Igual que hemos dicho en la tabla 5, *Escala global*, este descriptor manifiesta suficiente nivel (más que el A1 de la escala anterior, *Niveles cualitativos de referencia*) como para que el hablante sea capaz de mantener una conversación sobre temas sencillos de necesidad inmediata, con ayuda y colaboración de los interlocutores.

En ACTFL sería ya nivel inicial alto o incluso intermedio bajo.

El MCER no ofrece descriptores de menor nivel, necesarios para cuando las dificultades de comunicación son mayores o las habilidades menores, propias de personas que no pueden mantener ni siquiera esta mínima conversación.

Análisis de A2: sería intermedio bajo de ACTFL

Creemos que en la escala de ACTFL este A2 sería intermedio bajo¹⁹⁵, un hablante capaz de comunicarse sobre temas sencillos.

La principal diferencia entre el A2 y el A1 está en que ahora no existe ninguna necesidad de ser ayudado para participar en una conversación.

¹⁹⁵ Al igual que hemos dicho al analizar la Tabla 5, en §3.2.1.

En cuanto a temas y calidad del discurso, no varían sustancialmente respecto al A1, salvo que los descriptores se esfuerzan por ser más extensos en la definición de los temas e intercambios, aunque esto no deriva automáticamente en una concepción de mayor habilidad. Por ejemplo, ahora se dice “comunicarme en tareas sencillas y habituales que requieren un intercambio simple y directo”, especificando los temas como “información sobre actividades y asuntos cotidianos”, mientras antes se decía simplemente “participar en una conversación de forma sencilla sobre temas de necesidad inmediata o asuntos muy habituales”.

No vemos que eso suponga más nivel, salvo en el cuantificador “muy” expresado en el A1 junto a temas habituales. Parece que con esto se pretende interpretar que los temas “muy habituales” serían más sencillos de tratar que los temas “cotidianos”; y lo aceptamos. Pero, no se ve una especificación de grado de nivel suficiente como para determinar si una persona es A1 o A2.

Análisis de B1: sería intermedio medio de ACTFL

Este descriptor es bastante pobre pues se limita a dar temas: todo lo que pueda surgir en un viaje¹⁹⁶ o “temas cotidianos de interés personal o que sean pertinentes para la vida diaria, por ejemplo, familia, aficiones, trabajo, viajes y acontecimientos actuales”. Por lo tanto no es un descriptor de nivel sino de temas. Las únicas palabras que se centran en el grado de habilidad son (subrayadas) “sé desenvolverme en casi todas las situaciones” y “pedo participar espontáneamente”.

No obstante, esto ya supone un grado de habilidad mayor respecto al A2, aunque sea definido tan imprecisamente. Otras escalas han tenido más éxito en la diferenciación de las habilidades (Tablas 1, 2, 3 p. ej.).

Comparado con ACTFL seguimos en el intermedio medio, sin describir nada que parezca de mayor nivel.

Análisis de B2: sería intermedio alto de ACTFL

Este descriptor vuelve a dar definiciones de grado (“fluidez o espontaneidad”) además de seguir en su línea de equiparlo con la capacidad de establecer

¹⁹⁶ Ya hemos comentado antes que esto no es suficiente información de nivel ya que en un viaje pueden surgir temas de difícil solución que requieran habilidades de discusión, convencer a gente, etc.

conversaciones con nativos sin problemas, “lo que posibilita la conversación normal con hablantes nativos”, lo que marca una gran diferencia respecto a los anteriores niveles. Además establece nuevas funciones: “debates desarrollados en situaciones cotidianas, explicando y defendiendo mis puntos de vista”.

Creemos que el nivel marcado aquí sube ligeramente respecto al de B1, al ser capaces los hablantes de desarrollar esos debates de que se habla y defender los puntos de vista, para lo que hay que explicar las opiniones. No obstante, el descriptor explica que se trata de temas “cotidianos” y la defensa de esas opiniones se establecerá, lógicamente, dentro de estos campos de acción y no de otros más sociales o generales.

Por ello, comparado con ACTFL, entra limitadamente en el terreno del avanzado pero sin ser completamente avanzado. Para serlo, debe poseerse habilidad para tratar más funciones y más desarrolladas, no solamente a nivel textual y organizativo, sino temático: un avanzado de ACTFL puede tratar todo tipo de temas sociales y no solo los cotidianos.

Así, esta escala¹⁹⁷, se queda en el nivel intermedio alto.

Por otra parte, nos parece que el MCER no suele diferenciar claramente un intermedio alto de un avanzado. En las OPI estos niveles comparten muchas características y se diferencian en grado solamente. En el MCER, estas diferencias de grado no suelen aparecer y se puede dar por avanzado del MCER, o B2, a alguien que, siendo efectivamente B2, quizás sea solamente intermedio alto en ACTFL.

Análisis de C1: sería avanzado de ACTFL

Este descriptor define una habilidad en la que el hablante se expresa sin dificultad sobre temas sociales y profesionales y es capaz de expresar sus ideas y opiniones. Evidentemente, entra en lo que en ACTFL se llama avanzado. Sin diferenciar si se trataría de un avanzado medio o bajo.

¹⁹⁷ Al contrario que otras anteriormente analizadas donde lo equiparábamos con avanzado bajo porque se hablaba de un desarrollo de temas mayor.

Análisis de C2: sería avanzado medio de ACTFL

Este descriptor sube claramente el nivel hasta mostrar una habilidad en la que el hablante puede expresarse sin problemas en cualquier debate transmitiendo matices sutiles y con el uso de los modismos y expresiones necesarios.

Aun suponiendo más nivel que C1, no se llega al siguiente nivel de ACTFL y permanece dentro de lo que puede ser un avanzado medio. Aunque no podemos descartar el avanzado alto en los casos de discursos en los que se dé un alto grado de abstracción y defensa de opiniones ya que el MCER lo define para contextos en los que se pueda tratar “cualquier conversación o debate”¹⁹⁸.

3.2.4 Conclusión de las comparaciones entre las *guidelines* de ACTFL y las escalas del MCER

Ahora, con el fin de proceder a una comparación más cercana y sencilla entre ambos sistemas de evaluar, vamos a agrupar los componentes de los niveles del MCER según los tres criterios de evaluación de ACTFL:

- a. Tipo de discurso.
- b. Contexto y contenido.
- c. Funciones.

Primero presentamos las escalas de cómo quedaría el MCER si se utilizaran los criterios de ACTFL¹⁹⁹ (§3.2.4- A), luego los resultados de compararlo con la escala de ACTFL (§3.2.4- B).

¹⁹⁸ El tipo de definiciones de nivel, habilidad y dominio que el MCER emplea son diferentes a los de ACTFL y aunque definan un nivel después de C1 esto no representa automáticamente un nivel más en ACTFL. Para ser avanzado alto de ACTFL se necesita un dominio de funciones más complejas, como defender opiniones y hacer hipótesis utilizando un lenguaje abstracto en forma de discurso extendido sobre todo tipo de temas, incluso nunca antes pensados, aunque esto no se mantenga todo el tiempo.

¹⁹⁹ Estos criterios de evaluación han sido explicados en §2.6.2- 2.

3.2.4- A) Contenidos de las escalas del MCER según los criterios de evaluación oral de ACTFL

El MCER quedaría delimitado de la siguiente manera.

a. Tipo de discurso en las escalas del MCER

A1- El hablante es capaz de hablar de forma elemental mediante expresiones y frases sencillas y aisladas. Enlaza palabras o grupos de palabras con conectores muy básicos como “y” y “pero”.

A2- El hablante es capaz de hablar mediante una breve lista de frases y oraciones sencillas enlazando grupos de palabras con conectores sencillos tales como “y”, “pero” y “porque”, sobre temas habituales que requieren un intercambio simple, breve y directo.

B1- El hablante es capaz de hablar enlazando una serie de elementos breves, diferenciados y sencillos formando una secuencia lineal de ideas relacionadas con coherencia y razonable fluidez como para que le puedan comprender sin dificultad.

B2- El hablante es capaz de manejar un número limitado de mecanismos de cohesión en un discurso detallado y sistemáticamente desarrollado con cierta fluidez y espontaneidad. Así puede, con sus frases, componer un discurso claro y coherente, ampliando con cierta extensión, dando el énfasis apropiado, exponiendo las ventajas e inconvenientes de un tema. No obstante, puede mostrar cierto nerviosismo si la intervención es larga.

C1- El hablante es capaz de hacer un uso efectivo y flexible del idioma, mediante un uso correcto de los mecanismos de cohesión, articulación y organización del texto para componer discursos fluidos y espontáneos, estructurados y detallados con precisión, con conclusiones apropiadas y sin muestras evidentes de esfuerzo.

C2- El hablante es capaz de hacer un discurso fluido y bien estructurado, haciendo uso completo y apropiado de las estructuras organizativas y variadas y una amplia serie

de conectores y mecanismos de cohesión; y con un estilo adecuado al contexto y una estructura lógica, eficaz y coherente.

b. Contexto y contenido en las escalas del MCER

A1- El hablante trata temas de necesidad inmediata o asuntos muy habituales, como dar datos personales, presentarse a sí mismo y a otros, pedir y dar información personal básica sobre su domicilio o el lugar donde vive, sus pertenencias y sobre las personas que conoce.

A2- El hablante puede hablar sobre actividades, asuntos cotidianos y cuestiones relacionadas con sus necesidades inmediatas y su entorno, como su familia y otras personas, actividades diarias, cosas que le gustan o no, condiciones de vida, origen educativo, condiciones de trabajo, trabajo actual o anterior y aspectos de su pasado.

B1- El hablante puede hablar sobre una variedad de temas cotidianos que sean de interés personal, como su familia, aficiones, trabajo y viajes o que sean pertinentes para la vida diaria, hechos y acontecimientos actuales, experiencias, sueños, esperanzas, deseos, aspiraciones y ambiciones. Puede resolver situaciones que puedan surgir durante un viaje y hablar de sus opiniones y proyectos, describir sus relaciones y narrar una historia o relato, la trama de un libro o película.

B2- El hablante puede tratar temas relacionados con su especialidad y hablar sobre temas generales, o participar en debates desarrollados en situaciones cotidianas.

C1- El hablante puede tratar temas complejos, sociales, académicos y profesionales para fines sociales o laborales.

C2- El hablante puede hablar, tratar temas y participar en cualquier conversación o debate.

c. Funciones en las escalas del MCER

A1- Presentarse y dar información personal.

A2- Breves y sencillas descripciones y presentaciones.

B1- Narrar, describir, explicar y justificar brevemente sus opiniones y argumentos y razonar brevemente.

B2- Hacer descripciones, explicar las ventajas e inconvenientes de un tema, defendiendo su punto de vista e ideas y elaborar cadenas argumentales.

C1- Hacer todo tipo de descripciones y presentaciones.

C2- Argumentar.

3.2.4- B) Comparación de los contenidos en las escalas del MCER según los criterios de evaluación oral de ACTFL

Comparación entre la tipología de texto en las escalas del MCER y de ACTFL

A1 es inicial alto ya que es capaz de hablar de forma elemental mediante expresiones y frases sencillas y aisladas.

A2 es igualmente inicial alto, ya que es capaz de usar frases y oraciones sencillas enlazando grupos de palabras sobre temas sencillos y habituales, pero todavía en intercambios breves y sencillos o directos.

B1 es intermedio bajo, ya que es capaz de, “enlazando una serie de elementos breves”, formar una secuencia lineal de ideas relacionadas con coherencia y razonable fluidez.

B2 es avanzado bajo y medio porque es capaz de utilizar mecanismos de cohesión en un discurso detallado y sistemáticamente desarrollado con cierta fluidez y, con sus frases, puede componer un discurso claro y coherente, ampliando con cierta extensión, exponiendo las ventajas e inconvenientes de un tema.

C1 quizás sea, dependiendo de si puede mantener el discurso en temas abstractos, avanzado alto o solamente avanzado medio, al ser capaz de un uso efectivo y flexible del idioma, mediante un uso correcto de los mecanismos de cohesión, articulación y organización del texto para componer discursos fluidos y espontáneos, estructurados y detallados con precisión.

C2 es igualmente avanzado alto o medio, dependiendo de las mismas características añadidas para C1.

Comparación entre el contexto y contenido del discurso en las escalas del MCER y de ACTFL

A1 sería similar a inicial medio (temas de necesidad inmediata o asuntos muy habituales).

A2 sería similar a inicial alto (actividades, asuntos cotidianos y cuestiones relacionadas con sus necesidades inmediatas).

B1 sería similar a intermedio bajo o medio (una variedad de temas cotidianos).

B2 sería similar a avanzado bajo o medio (temas relacionados con su especialidad y temas generales).

C1 y C2 también serían similares a avanzado bajo o medio (temas complejos, sociales, académicos y profesionales; y participar en cualquier conversación).

Comparación entre las funciones en las escalas del MCER y de ACTFL

Aparentemente, no hay equivalencias entre ambos sistemas. ACTFL gradúa las funciones por niveles, definiéndolas de una manera amplia en conjunción con los temas que tratan y el tipo de discurso que elaboran. El MCER solo utiliza esta diferenciación de funciones entre A2 y B1, nivel donde aparece desplegada ya toda la complejidad funcional posible de las escalas. El resto de niveles se diferencian principalmente en el grado y complejidad de su uso.

Sin embargo, tras los análisis hechos con entrevistas reales en el capítulo 4, y mediante la comparación de los resultados con los de las entrevistas de ACTFL, se puede intuir una diferenciación que subyace a los diferentes niveles del MCER en relación con las funciones que se desplegarían en cada nivel. Remitimos al capítulo 5 para su explicación concreta, (§5.2.1 y cuadro 5.2.1).

3.2.5 Conclusión general de las comparaciones entre las escalas de habilidad oral del MCER y las *guidelines* de evaluación oral de ACTFL

De las comparaciones hechas entre las escalas del MCER y la de las OPI de ACTFL se deduce que sus descripciones de niveles no son claramente transferibles. Esto último es lo que queremos comprobar o rechazar empíricamente mediante nuestra investigación.

Lo más destacable es que los niveles entre ambos sistemas son fluctuantes y quizás sea complicado establecer una comparación efectiva en una evaluación con discurso real de estudiantes extranjeros.

También es importante comprobar que las escalas del MCER no dan apenas datos para niveles bajos (inicial medio o bajo), ni altos (avanzado alto o superior):

- A1 iría desde inicial medio a intermedio bajo, sin poder identificar discursos con características de menor nivel.

- A2 puede alcanzar desde inicial medio hasta intermedio.

- B1 se correspondería con intermedio, bajo, medio o alto.
- B2 se corresponde con que el avanzado de ACTFL, comenzando en el intermedio alto (casi avanzado) y abarcando el avanzado bajo y medio.
- En C1 las habilidades siguen siendo las de avanzado medio, sin llegar a desplegar las habilidades de un hablante superior de ACTFL.
- C2 se corresponde con avanzado medio o alto, sin llegar a desplegar las habilidades de un hablante superior de ACTFL.

Claro que esto es solamente un análisis cualitativo. Para comprobarlo hemos realizado pruebas reales con un test y entrevistas que han sido calificadas por un equipo de evaluadores entrenados para ello. Los resultados empíricos y estadísticos definitivos de todo esto que mantienen o matizan estas afirmaciones se dan en el capítulo 4 y en las conclusiones en el capítulo 5.

Presentamos los resultados de las comparaciones en dos cuadros (3.2.5 a./ 3.2.5 a.a y 3.2.5 b./ 3.2.5 b.), mostrando la horquilla que abarca cada nivel de un sistema respecto al otro.

Cuadro 3.2.5 a. Resultados de las comparaciones de las tablas 4, 5 y 6, entre el MCER y ACTFL

Lo que en el MCER es:	En ACTFL puede ser:		
	Tabla 5	Tabla 4	Tabla 6
C2	AM-AH	AM	AM
C1	AM	AM	AM
B2	AL-AM	AL	IH
B1	IM- (IH)	IL-IM	IM
A2	IM	NM -IL	IL
A1	NM-NH	NM-NH	NH- IL

Cuadro 3.2.5 a.a. Resultados conjuntos de las comparaciones de las tablas 4, 5 y 6, entre el MCER y ACTFL

MCER	--	A1	A2	B1	B2	C1	C2
ACTFL	NL	NM- NH- IL	NM IL - IM	IL -IM- IH	IH AL - AM	AM	AM -AH

Cuadro 3.2.5 b. Resultados de las comparaciones de las tablas 4, 5 y 6, entre ACTFL y el MCER

Lo que en ACTFL es:	En el MCER puede ser:		
	Tabla 5	Tabla 4	Tabla 6
SUPERIOR	No hay	No hay	No hay
ADV. HIGH	C2	No hay	No hay
ADV. MID	C1 C2	C1 -C2	C2
ADV. LOW	B2	B2	C1
INTERM. HIGH	B1	B1-B2	B2
INTERM. MID	A2 B1	B1	B1
INTERM. LOW	A2	A2 B1	A1 A2
NOVICE HIGH	A1	A1 A2	A1
NOVICE MID	A1	A1 A2	No hay
NOVICE LOW	No hay	No hay	No hay

NL: inicial bajo
 NM: inicial medio
 NH: inicial alto
 IL: intermedio bajo
 IM: intermedio medio
 IH: intermedio alto
 AL: avanzado bajo
 AM: avanzado medio
 AH: avanzado alto
 S: superior

Cuadro 3.2.5 b.b. Resultados de las comparaciones de las tablas 4, 5 y 6, entre ACTFL y el MCER

ACTFL	NL	NM	NH	IL	IM	IH	AL	AM	AH	S
MCER	--	A1 A2	A1 A2	A1 A2 B1	A2 B1	B1 B2	B2 C1	C1 C2	C2	--

Acabamos de ver las ambigüedades en las escalas del MCER y la dificultad de comparaciones entre estas y la de ACTFL. Según ello, una persona certificada dentro de un nivel de habilidad con un sistema puede encajar en otro nivel o incluso en varios niveles en el otro con el que lo comparamos.

Así, por ejemplo, decir de una persona que es B1 con el MCER no es suficiente para saber si es intermedio bajo, medio o alto según la OPI de ACTFL. Y al revés, una persona certificada como intermedio alto en una OPI podría poseer según el MCER unas habilidades propias del nivel B2, o de B1.

Según esta ambigüedad de homologaciones, un extranjero que posee las habilidades descritas B2 puede ser intermedio alto o avanzado de ACTFL. Pero otro extranjero evaluado como intermedio alto en ACTFL, puede ser en la escala del MCER un hablante de nivel B1. Ambos tendrían las mismas habilidades según un sistema de evaluación y diferentes habilidades según otro.

¿Qué significa entonces ser B2 o cualquier otra etiqueta más allá de la escala ambigua que define su actuación? Véase el siguiente cuadro (3.2.5 c.) con una escalera de las comparaciones entre los resultados del nivel de estudiantes, explicativo de estas incongruencias. En él se da una circularidad desde el estudiante 1 (Laura) al 4 (Peter), todos ellos ficticios, que muestra la falta de equivalencias²⁰⁰:

Cuadro 3.2.5 c. Escalera de comparaciones entre estudiantes

				3.	Anne	4.	Peter
						ACTFL	Av. bajo
1.	Laura	2.	John	MCER	B2	MCER	B2
		ACTFL	Int. alto	ACTFL	Int. alto		
ACTFL	Int. medio						
MCER	B1	MCER	B1				

Laura (1) y John (2) son B1, pero Laura es solo intermedio medio, mientras que John es intermedio alto en ACTFL.

Anne (3) y John (2) son intermedio alto en ACTFL, pero Anna es B2, mientras que John es solo B1.

²⁰⁰ Este esta basado en las conclusiones del análisis cualitativo entre las escalas realizado en este capítulo. En el capítulo 5 se discutirán las correlaciones entre escalas basadas en la investigación empírica y en los resultados estadísticos.

Finalmente, Peter (4) y Anne (3) son del mismo nivel en el MCER, B2. Pero mientras que Anne es solo intermedio alto, Peter es avanzado en ACTFL.

Claro que ACTFL dice que no se pueden comparar estudiantes entre sí.

Es necesario tener claras estas dificultades de comparación de niveles, para no caer en homologaciones superficiales y errores ingenuos. Para ser conscientes de ello, así como para poder dar pasos tendentes a solventar esas simplificaciones, tanto en la evaluación como en las expectativas que la evaluación produce en todas las personas a las que afecta, desde profesionales o candidatos, hasta sus usuarios últimos en instituciones o mercado de trabajo. Debemos conocer las limitaciones y saber de qué hablamos cuando calificamos a una persona dentro de estas escalas que estamos comparando. Para ello hemos realizado esta investigación²⁰¹.

Evaluar adecuadamente es saber en qué son iguales dos tests y en qué son diferentes. Para no decidir a la ligera que B2 es avanzado, u otras correspondencias sin matizar, que nos pueden llevar a errores.

Esta investigación trata de captar estos matices al describir niveles según el MCER y ACTFL.

En los próximos apartados presentamos el test diseñado para llevar a cabo este trabajo.

En el capítulo 4 se explican las calificaciones de las entrevistas hechas con candidatos voluntarios para comprobar si estas premisas se cumplen. Allí también se explican las ambigüedades surgidas en la práctica al enfrentarse realmente al discurso de los candidatos y a las posibilidades de interpretación y calificación por parte de los evaluadores. Al final, también se presentan las conclusiones estadísticas y analíticas del uso de los tests, con una interpretación que aclara las correspondencias ente las dos escalas utilizadas.

²⁰¹ En el proceso de calificación y evaluación de las entrevistas de candidatos de diferentes niveles, mostrado en el capítulo 4, pueden surgir más ambigüedades y problemas respecto a la suficiente o insuficiente cantidad de datos necesarios para llevar a cabo las calificaciones y valoraciones de sus discursos, así como con el hecho contradictorio de que con los mismos datos y ejemplos de discurso se pueda decidir que se pertenece a un nivel según un sistema (ACTFL o MCER) y quizás a otro nivel según otro sistema (MCER o ACTFL).

3.3 Explicación del trabajo de campo realizado con el test y los candidatos voluntarios

El trabajo de investigación se hizo mediante dos entrevistas, la OPI de ACTFL y una entrevista oral nueva formada por una selección de tareas diseñadas según el MCER, para tal efecto.

Los candidatos voluntarios fueron universitarios de edad comprendida entre 18 y 30 años principalmente, de diversas nacionalidades y lenguas, que estudiaban en los cursos de español para extranjeros del CLM. Solamente dos son mayores de 50 años.

Quedaron para realizar las entrevistas orales en el tiempo libre de que disponían al terminar sus clases de lengua en el CLM, o por las tardes. Las entrevistas se desarrollaron en un aula del CLM en la que solamente se encontraban el evaluador y el candidato, sentados en dos sillas, uno frente al otro y con una tercera silla donde se colocaba una grabadora digital de pequeño formato y unas hojas con las tareas que el entrevistador iba dando al candidato durante la segunda entrevista, la del test del MCER.

Cada sesión duró entre 30 y 60 minutos según el nivel del candidato. En ese tiempo se hacían las dos entrevistas cuyos resultados se han comparado en esta investigación: en primer lugar, la OPI de ACTFL, con una duración de entre 10 y 30 minutos; y después, la selección de tareas diseñadas según el MCER, con una duración de unos 20 a 30 minutos.

Los candidatos sabían el trabajo que iban a realizar. Se les dijo que se trataba de hablar en español y contestar a una serie de preguntas para ser evaluados según dos sistemas diferentes de calificar el dominio y habilidad de uso de la lengua: la OPI de ACTFL (sistema americano) y el MCER (sistema europeo). Se les había dicho previamente que eran dos sistemas diferentes, el primero con una serie de preguntas, sobre temas que irían saliendo de la misma conversación; el segundo, teniendo que realizar una serie de tareas, una tras otra, sobre diversos temas y actividades.

Todos ellos lo realizaron voluntariamente, tras una selección de candidatos filtrada por sus profesores y hecha de forma que los estudiantes voluntarios no fueran los mejores ni los peores de los cursos. Quedaron muy satisfechos con sus entrevistas y conversaciones, y contentos con el transcurso del doble test.

El entrevistador comenzaba explicándoles nuevamente, en resumen, todo el proceso al que se iban a someter, que ya conocían porque se había explicado al seleccionarlos y aceptar ser candidatos; luego encendía la grabadora y comenzaba las entrevistas.

Todas las entrevistas las realizó un solo entrevistador, yo mismo.

3.3.1 Entrevistas utilizadas

La doble entrevista a la que se les sometió constaba de la OPI, en primer lugar, y de la entrevista del MCER para la que se usaron tareas nuevas que se les dieron según el nivel inicialmente previsto para cada candidato, por el curso que estaba estudiando.

Esta segunda entrevista conforma un test que se compone de diferentes tareas según el curso que estudiaban los candidatos en el CLM. Todos hacen las tareas generales de *explicar una receta de comida* (Tarea 2), *explicar el viaje o experiencia más enriquecedores* (Tarea 6), *explicar las ventajas de estudiar la lengua en el país en el que la hablan* (Tarea 7), *argumentar a favor o en contra de las guerras* (Tarea 5), y *contar un cómic* (Tarea 8).

Los de nivel bajo hacen además la tarea de *trabajo de canguro* (Tarea 1). Los de niveles medios hacen la de *opinar sobre una actividad cultural* (Tarea 3). Y en avanzados y superiores hacen la de *explicar ciertos aspectos de su trabajo para un colega que va a ayudarle o sustituirle durante unos días* (Tarea 4).

Así pues, todos realizaban las tareas:

- *Receta de comida.*
- *Explicar el viaje o experiencia más enriquecedora.*
- *Explicar las ventajas de estudiar la lengua en el país en el que la hablan.*
- *Argumentar a favor o en contra de la guerra de Líbano o Irak.*
- *Contar un cómic.*

Los de niveles bajos comenzaban por la del *trabajo de canguro*, tarea que no necesitaban realizar los candidatos de los demás niveles.

Los candidatos de niveles altos realizaban, además, las tareas de:

- *Opinión sobre una actividad cultural.*

- Explica ciertos aspectos de tu trabajo.

El cuadro expone el desarrollo de la entrevista con las 8 tareas nuevas, con el título de cada una, los niveles del MCER para los que se utilizó, así como la escala del MCER de la que se partió para su diseño y las escalas con las que se va a calificar cada tarea principalmente.

Cuadro 3.3.1. Niveles de las tareas y escalas utilizadas en su corrección

	NOMBRE DE LA TAREA	NIVEL	ESCALA USADA EN SU DISEÑO	ESCALAS USADAS PARA CALIFICAR
1	<i>Trabajo de canguro.</i>	A1 A2 B1	- Intercambio de información.	- Intercambio de información.
2	<i>Receta de comida.</i>	B1 B2, C1 y C2	- Intercambio de información.	- Intercambio de información.
3	<i>Opinión sobre actividad cultural.</i>	B1 B2, C1 y C2	- Intercambio de información. - Autoevaluación.	- Intercambio de información. - Autoevaluación. - Monólogo sostenido: argumentación.
4	<i>Explicar ciertos aspectos de tu trabajo...</i>	B2, C1 y C2	- Intercambio de información.	- Intercambio de información. - Monólogo sostenido: argumentación.
5	<i>Argumentar a favor o en contra de la actual guerra...</i>	B1 B2, C1 y C2	- Monólogo sostenido: argumentación. - Autoevaluación.	- Monólogo sostenido: argumentación. - Autoevaluación.
6	<i>Explicar el viaje o experiencia más enriquecedora...</i>	Todos	- Fluidez oral. - Coherencia y cohesión. - Desarrollo de descripciones y narraciones.	- Fluidez oral. - Coherencia y cohesión. - Desarrollo de descripciones y narraciones. - Monólogo sostenido: argumentación.

7	<i>Explicar las ventajas de estudiar la lengua en el país en el que la hablan.</i>	Todos	- Fluidez oral. - Coherencia y cohesión. - Desarrollo de descripciones y narraciones.	- Fluidez oral. - Coherencia y cohesión. - Desarrollo de descripciones y narraciones. - Monólogo sostenido: argumentación.
8	<i>Contar un cómic.</i>	Todos	- Fluidez oral. - Coherencia y cohesión. - Desarrollo de descripciones y narraciones.	- Fluidez oral. - Coherencia y cohesión. - Desarrollo de descripciones y narraciones. - Monólogo sostenido: argumentación.

Además de lo expresado en el cuadro en la columna de la derecha, todas las tareas se evalúan por el conjunto de estas otras escalas: *Flexibilidad, Corrección gramatical, Control y corrección, Compensación, Dominio de vocabulario, Riqueza de vocabulario, Expresión oral en general y Competencia lingüística general.*

3.3.2 Selección de candidatos

La selección de candidatos con los que se hizo la investigación se realizó durante el verano de 2006 en el Centro de Lenguas Modernas de la Universidad de Granada. Se eligió a 35 candidatos de todos sus cursos de lengua y de los cursos especiales para formación de profesores extranjeros de español, pues se necesitaba que pudieran ser calificados en todos los niveles tanto con la escala de ACTFL, que divide el uso de la lengua en 10 niveles, como con las del MCER, que lo hacen en 6.

Los candidatos de los niveles más altos procedían de cursos especiales que se imparten para reciclaje y formación de profesores extranjeros de español. En ellos hay, por lo tanto, personas que llevan años hablando y enseñando nuestra lengua y se supone que pueden tener un nivel alto (avanzado alto o superior de ACTFL y C1 o C2 del MCER).

El resto de los candidatos procedían de los cursos intensivos de lengua española (CILE) que se organizan en nueve niveles (inicial -bajo y alto-, intermedio -bajo y alto-, avanzado -bajo y alto-, superior -bajo y alto- y perfeccionamiento).

La selección total de voluntarios constó de 35 estudiantes, 8 de nivel inicial, 6 de nivel intermedio, 8 de nivel avanzado, 6 de nivel superior, 4 de perfeccionamiento y 3 de talleres de reciclaje de profesores, todos ellos de entre 18 y 65 años y de variadas nacionalidades: Alemania, Chequia, Estados Unidos, Francia, Holanda, Inglaterra, Italia, Japón, Líbano, Noruega, Polonia, Suecia, Suiza, Túnez, Turquía, Zimbawe, y tres nativos de Méjico y Panamá.

El total de estudiantes que hicieron la prueba es el del cuadro siguiente en el que se citan numerados, con un nombre inventado, su nacionalidad y su nivel de procedencia de los cursos del CLM. En negrita están los que finalmente se utilizaron en el corpus de la investigación. De ellos, el candidato número 28 (Amaya) resultó no evaluable para ACTFL y se ha eliminado. El corpus final se compone de 28 candidatos (1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 14, 15, 16, 17, 18, 19, 20, 21, 22, 25, 26, 29, 30, 31, 32, 33, 34 y 35).

1	Biancé	Zimbawe	1 A
2	Stive	Alemania	1 A
3	Johanes	Noruega	1 A
4	María	Holanda	1 B
5	Francesca	Italia	1 B
6	John	Alemania	1 B
7	Carmen	Túnez	2 A
8	Juan	Polonia	2 A
9	Boris	Suecia	2 A
10	Lucía	Francia	2 B
11	Laura	Holanda	2 B
12	Ana	Alemania	2 B
13	María	Estados Unidos	3 A
14	Michelle	Francia	3 A
15	María	Estados unidos	3 A
16	Ana	Estados unidos	3 B
17	Jorge	Estados unidos	3 B
18	Lea	Suiza	3 B
19	Maya	Líbano	4 A
20	Nina	América	4 A
21	Magda	Polonia	4 A
22	Ana	Estados unidos	4 B
23	Verónica	Estados unidos	4 B
24	Laura	Estados unidos	3 A
25	Cuqui	Turquía	5
26	Patricia	Estados unidos	5
27	Ana	Inglaterra	3 A
28	Amaya	Francia	Taller

29	María	Francia	Taller
30	Marta	Chequia	Taller
31	María	Panamá	5
32	Fátima	Méjico	4 B
33	Camila	Méjico	5
34	Sachiyo	Japón	1 A
35	Kayo	Japón	1 A

3.3.3 Preevaluación de los candidatos

Se tenía una primera apreciación sobre el nivel de lengua de los estudiantes porque estaban estudiando en los cursos del CLM y habían sido asignados a diferentes grupos mediante la prueba de clasificación que realizan todos los estudiantes al llegar al centro e iniciar sus cursos²⁰². Pero además, y para una mejor selección de candidatos, se pidió a los profesores de cada curso que seleccionaran, de entre los estudiantes que estuvieran dispuestos voluntariamente a hacer los tests, a los que creyeran que respondían a las características más estándares del nivel del curso, esto es, sin sobresalir por rápidos, inteligentes o habilidosos, ni por lo contrario. Con esto nos aseguramos de no tener ni los mejores, ni los peores de cada nivel y conseguíamos que, dentro de la voluntariedad de los candidatos, el conjunto final no estuviera sesgado por estudiantes con un interés especial en aprender y por una motivación mayor²⁰³.

Tras estas dos características de selección, prueba previa de nivel para asistir a los cursos del CILE y opinión del profesor respecto a su verdadero nivel correspondiente con ese curso concreto, se formó el grupo principal. A este grupo se le sumaron luego los profesores extranjeros, igualmente seleccionados de forma voluntaria de entre los cursos de especialización, más otro grupo de tres estudiantes nativos de similares características y perfil que el resto de los estudiantes de los cursos del CLM. Estos últimos no se tomaron de la población española general, sino que se seleccionaron del Curso de Estudio Hispánicos que se imparte en el CLM entre cuyos estudiantes se

²⁰² Podemos considerar que con esta clasificación ya sabemos un nivel de conocimientos y habilidades de los candidatos, pues la prueba que se utiliza para organizar los cursos tiene un nivel de fiabilidad alto comprobado con análisis estadísticos y a lo largo del transcurso de los años y una validez predictiva igualmente alta. El número de estudiantes que pide un cambio de nivel por creer -tanto él como sus profesores- que debe cambiarse a un grupo de mayor o menos nivel, es mínimo, rondando entre un 4 y un 7%.

²⁰³ Ya que los estudiantes más interesados en participar en la investigación podrían estar más motivados y ser más hábiles que los otros del grupo.

encuentran nativos de español pero procedentes de Méjico, Estados Unidos u otros países de habla hispana, de semejantes características educativas, culturales y sociales que el resto. Así asegurábamos que fueran de similares características que el resto de candidatos no nativos, tanto en relación a su formación universitaria como a posibles conocimientos culturales. Haberlos elegido españoles podría influir a su favor o en contra en sus resultados a la hora de contestar a algunas de las tareas del test, por ejemplo la de hablar de una experiencia cultural interesante en España o la de opinar respecto a las ventajas y desventajas de estudiar en España.

Con este dato del nivel de habilidad lingüística aproximado de cada candidato, procedente del curso que estaban estudiando en el CLM, pudimos decidir qué tareas darles en su evaluación con el test del MCER. Comenzando por las sencillas, para los de niveles bajos, o por las de nivel medio, para los demás.

3.3.4 Grabación y almacenamiento de las entrevistas

Tus labios están calientes todavía
(Shakespeare).

Todas las entrevistas realizadas fueron grabadas y archivadas con una etiqueta que recoge para cada candidato un número, su nombre ficticio y el nivel del curso del CLM, de esta manera:

1. Biancé, 1 A.
6. John, 1B.

Posteriormente se procedió a escuchar, calificar y evaluar cada entrevista por separado por varios profesionales o profesores de lengua formados al efecto. Los resultados de estas evaluaciones, comentarios y consideraciones sobre la capacidad de las escalas para ofrecer resultados claros y válidos se ofrecen en posteriores capítulos de este trabajo.

3.3.5 Calificación y evaluación de las entrevistas

Las entrevistas fueron grabadas para su posterior corrección y almacenamiento²⁰⁴. La evaluación se hizo con dos equipos de correctores: uno para la OPI, formado por evaluadores miembros de ACTFL, de la misma manera que se hace en esta organización²⁰⁵, y otro para nuestro test del MCER, entrenado para lograr homogeneidad de criterios en cuanto a los descriptores del MCER y los sistemas de calificar las muestras de discurso de los candidatos, según los criterios de formación de evaluadores expuestos en §2.4.3- 4) y según las escalas señaladas para el diseño y evaluación de las tareas. Este equipo lo formaban 25 personas en total; 23 trabajando en grupos de dos o tres personas, corrigieron 2 entrevistas cada equipo; y otras dos personas, formadas con los mismos criterios, corrigieron todas las entrevistas por separado.

A todos ellos se les dio un curso de 10 horas sobre evaluación oral, en el que se les instruyó en las escalas del MCER y su uso para la evaluación oral. La formación se centró en el análisis del MCER: familiarización con sus objetivos, niveles y escalas; interpretación de sus descriptores; y prácticas con ejemplos de discursos de diferentes estudiantes, respondiendo a preguntas como las de los tests que se debían corregir, así como de casos raros o de difícil interpretación, por su acento o por dificultades de fluidez.

Se estudiaron todas las escalas que se iban a utilizar en la evaluación. Primero, las que se habían utilizado para diseñar cada tarea, posteriormente las demás que se iban a utilizar para dar una calificación definitiva. Y se tuvo en cuenta, especialmente, el hecho de que cada profesor o estudiante tiene, en principio, unos criterios personales que deben evitarse y dejar aparte. Para esto, en todo momento se exigió que los argumentos

²⁰⁴ Las entrevistas se presentan en un DVD como apéndice a este trabajo.

²⁰⁵ Las entrevistas de ACTFL se hicieron igual que se hacen en esta organización y por miembros de ACTFL, pero en algunos casos se llevó a los candidatos a pruebas de más nivel que el necesario, para poder comparar sus resultados con los obtenidos en el test del MCER, en el que se pretendían obtener muestras de estrés. No son una evaluación oficial de la organización. No podíamos hacer un test oficial ya que ello exigiría haber realizado la investigación exclusivamente con candidatos que hubieran hecho oficialmente el test de la OPI, lo que además supone un costo económico, y a los que luego se hubiera pedido voluntariamente realizar el otro test, el del MCER. Todo esto habría dificultado enormemente la selección de candidatos de todos los niveles. Las entrevistas, realizadas por un miembro de ACTFL, se enviaron a otro evaluador certificado de ACTFL, M. Dolores Delgado, para que diera su calificación en una doble corrección, de la misma manera que se hace en las entrevistas oficiales llevadas a cabo por la organización; a este evaluador oficial se le avisó de que algunas entrevistas tendrían mayor duración de lo esperado, muestras de estrés y preguntas de mayor nivel que lo habitualmente necesario, y que hiciera la calificación siguiendo los criterios habituales, salvando esos excesos innecesarios.

esgrimidos para dar una calificación procedieran, exclusivamente, de las escalas del MCER; nunca de otras opiniones o experiencias profesionales.

Se hicieron sesiones en grupos, con trabajo de calificación de entrevistas y discusiones sobre los resultados, con el objetivo de acercar posturas para que se llegara a unos criterios comunes.

Cada entrevista del MCER obtuvo tres calificaciones:

- 1 y 2. Las calificaciones dadas por las dos personas de forma independiente.
- 3. La dada por el grupo formado por dos o tres personas, escuchando y calificando la entrevista, en una doble sesión. Primero, para decidir el nivel de forma independiente; después, para llegar a un acuerdo conjuntamente.

Tenemos tres notas de cada candidato y con ellas se ha obtenido una sola utilizando el criterio de la mayoría como nota final, de forma global. Esto es, como cada entrevista del test del MCER se compone de 5 o 6 tareas y cada tarea puede ser evaluada según una lista de varias escalas del MCER, se pidió que los calificadores dieran una calificación a cada tarea y una global para toda la entrevista.

Los problemas surgidos para calificar según las escalas se explican en §4.2.1 –A). Los surgidos para aclarar cómo ponderar esa calificación final teniendo en cuenta las diversas subcalificaciones obtenidas en cada tarea, y a través de diferentes escalas, se explican en §4.2.1- C).

Los análisis estadísticos que se han hecho con los resultados de todas las entrevistas, y se han utilizado todos los datos disponibles: la nota general para las entrevistas de cada candidato, la nota general de cada uno de los tres evaluadores, y las notas de cada evaluador respecto a cada una de las tareas de la entrevista del MCER. Con ello se dispone de 588 datos con los que se ha hecho también un análisis factorial que se explica en §4.5. Con todos los resultados finales se hizo un análisis de consistencia entre evaluadores, una regresión factorial y el coeficiente de correlación de Spearman con la OPI. Estos datos de las calificaciones y de la comparación entre los resultados de ambas formas de evaluación, se dan en los capítulos 4 y 5 y en los Apéndices 9 y 10.

3.3.6 Planteamiento sobre el tipo de test a utilizar

Para medir la lengua de los candidatos podíamos optar entre estos dos modelos de administración de tests:

- a. Hacer exámenes de diferente nivel y que cada candidato se presentara al del nivel que aspirase a superar.
- b. Hacer un examen que pudiera servir para todos los candidatos, con independencia de su nivel.

En nuestro test hemos optado por un modelo híbrido consistente en diseñar unas tareas para unos niveles específicos y otras aptas para varios niveles o para todos los niveles al mismo tiempo.

A cada candidato, además de la OPI, se le pasaron tres tipos de tareas:

- Pensadas para su nivel.
- Pensadas para todos los niveles.
- Pensadas para niveles más elevados del que se le suponía al candidato.

Así, por ejemplo, a un candidato al que se le supone un nivel de conocimientos y expresión bajo, porque estaba haciendo un curso de nivel inicial en el CLM, se le pasaban las tareas del A1 y A2, las demás tareas aptas para evaluar cualquier nivel, y tareas de nivel superior, en este caso del nivel B1 y algunas de nivel superior.

El objeto de esto es poder saber el nivel más alto en el que pueda desenvolverse cada candidato. Pero, además, queríamos ver cómo actuaba en caso de que se le propusieran actividades más difíciles que pudieran colapsar sus habilidades. Esto lo hicimos así para poder comparar el test con el sistema de la OPI, donde se evalúa a los candidatos a lo largo de una serie de actividades de nivel creciente hasta demostrar tanto sus habilidades como sus limitaciones o el nivel donde ya no pueden desenvolverse. En este proceso se puede producir lo que aquí hemos llamado *estrés comunicativo*, cuando se colapsa su lenguaje porque ya no puede actuar adecuadamente ante determinado nivel que excede sus habilidades. La OPI no penaliza este colapso lingüístico. Nosotros,

en nuestro interés por comparar ambos sistemas, quisimos averiguar cómo se interpreta con las escalas del MCER la actuación de los candidatos cuando eso sucede.

3.4 Descripción de las tareas de que se compone la segunda entrevista, diseñadas según el Marco común europeo de referencia.

La descripción de estas 8 tareas es la siguiente.

3.4.1 Tarea 1: “Trabajo de canguero”

Planeada para niveles A1 y A2.

Trabajo de canguero.
(Para el alumno) <i>Vas a trabajar en una casa ayudando a la familia con su hijo pequeño y enseñándole un poco de tu lengua (inglés, francés, etc.). El padre de la familia te va a hacer unas preguntas antes de empezar el trabajo. Después si quieres puedes hacerle preguntas sobre su hijo o familia, trabajo, etc.</i>
(Algunas ideas para el entrevistador) ¿Cómo te llamas? ¿De dónde eres? ¿Cuántos años tienes? ¿Qué estudias? ¿Dónde estudias? ¿Trabajas? ¿Es la primera vez que haces este trabajo? ¿Qué información necesitas sobre el trabajo, los niños, la casa, etc.?

Esta tarea trata de que el candidato intercambie información con el interlocutor para intentar obtener un trabajo de canguero. Está pensada para medir la capacidad de interacción de candidatos de poco nivel, A1 y A2, hablando sobre temas cotidianos y respecto a cosas habituales: preguntar sobre actividades que hacen las personas en su vida diaria, horarios de trabajo, salario, nombre, edad y características personales de la familia de la persona con la que se está hablando, etc.

Para ello hemos seleccionado una tarea en la que se espera que el candidato pregunte cosas sencillas como las siguientes: ¿cómo se llama su hija?, ¿cuántos años

tiene?, ¿qué le gusta hacer?, ¿qué cosas hace normalmente en casa?, ¿podemos ir al parque?, ¿le gusta?, ¿cuántas horas necesito trabajar?, ¿dónde está su casa, la escuela?

El interlocutor, para dar realismo a la situación, debe preguntar al candidato cosas como si ha trabajado antes con niños o si le gustan los niños, etc.

Los descriptores del nivel A1 y A2 en que nos hemos fijado fundamentalmente para su diseño son estos de la escala de *Intercambiar información* (Consejo de Europa, 2002: 80).

A2	<ul style="list-style-type: none"> - Plantea y sabe cómo responder a preguntas relativas a lo que se hace en el trabajo y en el tiempo libre. - Pide y da indicaciones que se refieren a un mapa o a un plano. - Plantea y contesta preguntas sobre costumbres y acciones de la vida cotidiana.
A1	<ul style="list-style-type: none"> - Plantea y contesta preguntas sencillas, realiza afirmaciones... - Plantea y contesta preguntas sobre sí mismo y sobre otras personas: dónde vive, personas a las que conoce, cosas que tiene

3.4.2 Tarea 2: “Receta de comida”

Planeada para B1, B2, C1 y C2.

<p>Receta de comida²⁰⁶.</p>
<p><i>Explica a tu profesor cómo se hace una comida que te guste y sepas hacer. Y contesta a sus preguntas para aclararle las cosas que no entienda bien cómo hacer.</i></p>

Esta tarea trata de que el candidato describa la forma de hacer algo dando instrucciones detalladas y para ello se ha seleccionado que cuente alguna comida que él sepa hacer, o cómo se hace una comida de su país o de su familia. No obstante, se pretende que no sea un monólogo sino una charla con su interlocutor, en la que tenga

²⁰⁶ Esta tarea es habitual en muchas actividades de enseñanza y aprendizaje de lenguas, así como en otros tests como por ejemplo el de Bordón (Bordón, 2006: 230).

que ir explicando los pasos y atendiendo a sus demandas de aclaraciones, como suele ser habitual en la vida normal.

De los descriptores del B1 de la escala de *Intercambiar información* (Consejo de Europa, 2002: 80) se pensó fundamentalmente en estos:

<p>B1.2</p> <p>B1.1</p>	<ul style="list-style-type: none"> - Describe la forma de hacer algo dando instrucciones detalladas. - Sabe cómo encontrar y comunicar información concreta y sencilla.
---------------------------------------	---

El tipo de lenguaje que buscamos con ella es el de una descripción no muy complicada pero con suficiente detalle, porque para explicar una comida, por sencilla que sea, se necesitan describir pasos concretos y ordenados y nombres de utensilios y productos.

3.4.3 Tarea 3: “Opinión sobre actividad cultural”

Planeada para B1, B2, C1 y C2.

<p>Opinión sobre actividad cultural.</p>
<p>(Para el alumno)</p> <p><i>Has estado un mes en Granada viviendo y estudiando español. Seguro que has ido al cine o a algún espectáculo, o has leído algo interesante sobre su cultura.</i></p> <p><i>¿Puedes decir a tu profesor qué cosas te han gustado más y charlar con él sobre una de ellas expresando tus opiniones y contestando a sus preguntas si él necesita más información o detalle?</i></p>
<p>(Algunas ideas para el entrevistador)</p> <p>¿Qué cosas interesantes culturales has hecho o visto este mes? ¿Cuál te parece más interesante? ¿Me podrías explicar por qué y darme tu opinión sobre ella? ¿Por qué te parece especialmente relevante o interesante culturalmente? Otras preguntas para pedir detalle.</p>

Esta tarea está diseñada para evaluar a candidatos de nivel B1. Con ella se trata de conseguir que el candidato opine brevemente sobre alguna actividad cultural -sea del tipo que sea, culto, popular, etc.- y responda a preguntas para explicar algunos detalles.

Para su diseño se buscó en el descriptor de la escala de *Intercambio de información*, de B1: “Resume y da su opinión sobre relatos, artículos, charlas, discusiones, entrevistas o documentales breves y responde a preguntas complementarias que requieren detalles” (Consejo de Europa, 2002: 80).

Asimismo, en la escala de *Autoevaluación* se dice que los candidatos del nivel B1 “pueden participar en conversaciones que traten temas cotidianos de interés o que sean pertinentes para la vida diaria, como acontecimientos culturales y que los del nivel B2 puedan explicar y defender sus puntos de vista” (Consejo de Europa, 2002: 30-31).

3.4.4 Tarea 4: “Explicar ciertos aspectos del trabajo”

Planeada para B2, C1 y C2.

Explicar ciertos aspectos de tu trabajo para un colega que va a ayudarte o sustituirte durante unos días.
(Para el alumno) <i>Te vas de vacaciones una semana y tienes que explicarle a un colega cómo es tu trabajo, porque él lo va a hacer. Responder a sus preguntas.</i>
(Algunas ideas para el entrevistador) Vas a trabajar durante una semana en el puesto de tu colega. Pregúntale todo lo necesario para saber hacer bien ese trabajo (ten en cuenta que es posible que cuando termines- si lo has hecho bien- consigas un puesto en esa empresa).

Esta tarea está pensada para los tres niveles más altos, B2, C1 y C2, y trata de que el candidato comunique con total certeza información detallada y realice descripciones claras y detalladas acerca de la forma de llevar a cabo un procedimiento o intercambie información compleja y consejos sobre todos los asuntos relacionados con su profesión.

Fundamentalmente pensamos en estos elementos de la escala de *Intercambiar información* (Consejo de Europa, 2002: 80):

C1 y C2	Como en B2
B2.2 B2.1	<ul style="list-style-type: none"> - Comprende e intercambia información compleja y consejos sobre todos los asuntos relacionados con su profesión. - Comunica con total certeza información detallada. - Realiza descripciones claras y detalladas acerca de la forma de llevar a cabo un procedimiento. - Sintetiza y comunica información y argumentos procedentes de varias fuentes.

La diferencia de habilidad entre B1 y B2 estriba solamente en el grado de detalle y certeza mostrados al hacer las descripciones y en el hecho de hacerlo al transmitir un procedimiento. En niveles más bajos que B2 los descriptores de la escala de *Intercambiar información* mencionan simplemente la capacidad de comunicar información concreta y sencilla o describir la forma de realizar algo dando instrucciones detalladas.

Se pensó en esta tarea porque permite elaborar una explicación meticulosa, en primer lugar, para que la persona a la que se le explique sea capaz de hacer bien el trabajo y sustituir durante una semana al que lo cuenta; y en segundo lugar, pensando en que la explicación debe estar lo suficientemente bien dada como para que el jefe no se preocupe ni quede mal la persona sustituida.

El tipo de lenguaje que se pretende elicitar es, por una parte, el de una información clara, detallada, pausada y organizada y, por otra, una interacción entre los hablantes con aclaraciones interesantes, significativas e importantes.

3.4.5 Tarea 5: “Argumentar sobre la guerra...”

Dos opciones a elegir sobre el mismo tema y con iguales objetivos, planeadas para todos los niveles.

Argumentar a favor o en contra de la actual guerra de Israel contra Líbano.

Como quizás sepas, estos días Israel está atacando al Líbano y ha matado a varios cientos de personas inocentes civiles, unas 300, y está destruyendo importantes servicios de uso civil, porque las milicias de Hezbolá secuestraron a dos soldados en la frontera norte de Israel. ¿Te parece que esto es justo, que Israel tiene el derecho a defenderse, como ha dicho Estados Unidos, o crees que en realidad y aunque tiene el derecho a defenderse, debe usar otros medios o tener más cuidado con sus acciones?

Argumentar a favor o en contra de la guerra de Estados Unidos en Irak.

¿Crees que la guerra de Irak ha sido justa y necesaria? Estás de acuerdo con las afirmaciones de EE.UU. de que ahora el mundo es más seguro y era necesario terminar con un dictador y asesino como Sadam Hussein?

Esta tarea cambia la habilidad a evaluar. Las anteriores pretenden evaluar la capacidad de intercambiar información. Ahora se quiere evaluar la habilidad para argumentar según los siguientes descriptores de la escala de *Monólogo sostenido: argumentación* (Consejo de Europa, 2002: 63).

C1 y C2	- No hay descriptor disponible.
B2	Desarrolla argumentos sistemáticamente, dando un énfasis apropiado a los aspectos importantes y apoyándose en detalles adecuados. Desarrolla argumentos con claridad, ampliando con cierta extensión y defendiendo sus puntos de vista con ideas complementarias y ejemplos adecuados. Construye cadenas argumentales razonadas. Explica puntos de vista sobre un tema, proponiendo las ventajas y desventajas de ciertas opciones.
B1	Desarrolla argumentos lo bastante bien como para que se puedan comprender sin dificultad la mayor parte del tiempo. Ofrece breves razonamientos y explicaciones de opiniones, planes y acciones.
A1 y A2	- No hay descriptor disponible.

Igualmente en la escala de *Autoevaluación* se dice que un candidato del nivel B1 puede justificar brevemente sus opiniones y uno de B2 “puede explicar un punto de vista sobre un tema exponiendo las ventajas e inconvenientes y explicando y defendiendo sus puntos de vista” (Consejo de Europa, 2002: 30-31).

La tarea 5, Argumentar y opinar sobre la guerra, está definida según los descriptores de la escala *Monólogo sostenido: argumentación*. Además está pensada por ser fácilmente comparable con las funciones evaluadas en una OPI que diferencian a un candidato entre si es avanzado o superior:

- Opinar sobre un tema social o general.
- Organizar la información en una secuencia coherente de párrafos encadenados hasta -si el nivel del candidato lo permite- elaborar un discurso abstracto.
- Defender opiniones.
- Hacer hipótesis sobre acontecimientos.

Esta tarea, por ello, se diseñó con el fin de generar en los candidatos un discurso que permitiera medir todas estas diferencias de estilo y tipo de argumentación:

- La expresión de la mera opinión (breves razonamientos y explicaciones de opiniones -B1.1-).
- El desarrollo de argumentos inteligibles sin casi problemas (B1.2),
- El desarrollo extenso de argumentos y cadenas argumentales razonadas ampliadas con diferentes puntos de vista e ideas complementarias (B2.1),
- La capacidad (ambiguamente expuesta en los descriptores, a nuestro juicio) de desarrollar argumentos sistemáticamente, dando un énfasis apropiado a los aspectos importantes y apoyándose en detalles adecuados.

A pesar de definirse a partir de los descriptores de B1 y B2, la tarea se utilizó en todos los niveles porque se pensó que cualquier persona es capaz de tener ideas y opiniones sobre un tema tan general como una guerra y porque se quería analizar además en qué medida candidatos de menor nivel (A1, A2) también eran capaces de realizar tareas de argumentación aunque el MCER no ofreciera descriptor para esos niveles bajos.

El lenguaje que se pretendía elicitar depende del nivel del candidato, desde meras frases sueltas expresando su idea y opinión respecto a las guerras, ya con frases sueltas y sintagmas breves, hasta frases enlazadas o todo tipo de recursos lingüísticos de alto nivel y organización textual como lo definido en los descriptores del B2.

Las tareas 6, 7 y 8 se enuncian seguidas y después se explican conjuntamente.

3.4.6 Tarea 6: “Explicar el viaje o experiencia...”

Planeada para todos los niveles.

Explicar el viaje o experiencia más enriquecedora o interesante de los realizados durante la estancia en España durante este mes.
(Para los alumnos de niveles bajos: A1, A2 y B1) <i>¿Puedes contarme algún viaje que has hecho durante este mes?</i> Si no da suficiente información, usar las preguntas normales sobre viajes (dónde, con quién, cuándo, cómo, cuántos días, qué tal, experiencias interesantes, visitas, alojamiento, incidentes, vuelta a Granada...).
(Para los alumnos de niveles altos: B2, C1 y C2) <i>Explícame el viaje o experiencia más enriquecedora o interesante de los realizados durante la estancia en España durante este mes.</i>

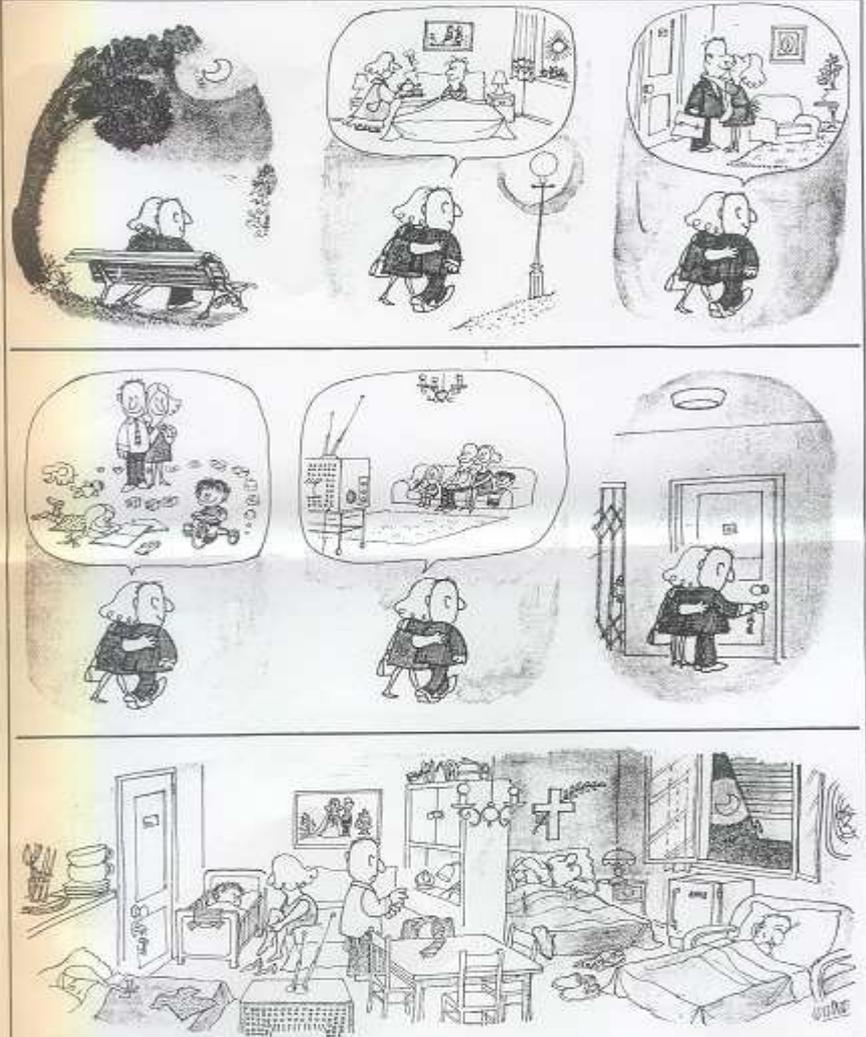
3.4.7 Tarea 7: “Explicar las ventajas de estudiar...”

Planeada para todos los niveles.

Explicar las ventajas de estudiar la lengua en el país en el que la hablan.
(Para los alumnos de niveles bajos: A1, A2 y B1) <i>Explícame las ventajas de vivir en un país extranjero con una familia, frente a vivir en tu propio país, para aprender una lengua. ¿Qué es lo mejor de vivir en un país extranjero para estudiar una lengua? Y vivir con una familia, ¿está bien? ¿Por qué?</i>
(Para los alumnos de niveles altos: B2, C1 y C2) <i>Explícame las ventajas de vivir y estudiar en el país con una familia frente a vivir con otros estudiantes de tu misma nacionalidad.</i>

3.4.8 Tarea 8, “Explicar un cómic”

Planeada para todos los niveles.

Explicar un cómic.
<i>Después de mirar unos minutos este cómic, ¿puedes explicarlo o contarme qué pasa o qué es lo principal de esta historia?</i>


(Miquel L. / Baulenas N. S., 1985: 93).

Explicación de las tareas 6, 7 y 8

Las tareas 6, 7 y 8 se diseñaron pensando en evaluar a los candidatos en relación al dominio general de la lengua a través de diferentes temas y según una serie de escalas en las que se consideran diversos elementos del uso de la lengua como la fluidez oral, la

coherencia y cohesión, y el desarrollo de descripciones y narraciones, principalmente. Estas tareas se diseñaron partiendo de estas 3 escalas:

Desarrollo de descripciones y narraciones.	
C2	Como en C1
C1	Realiza descripciones y narraciones detalladas integrando varios temas, desarrollando aspectos concretos y terminando con una conclusión apropiada.
B2	Realiza descripciones y narraciones claras ampliando y apoyando sus puntos de vista sobre los aspectos principales, con detalles y ejemplos adecuados.
B1	Realiza, con razonable fluidez, narraciones y descripciones sencillas, mediante una secuencia lineal de elementos.
A2	Cuenta historias o describe algo con la ayuda de una lista sencilla de elementos.
A1	No hay descriptor posible.

(Consejo de Europa, 2002: 122).

Coherencia y cohesión.	
C2	Crea textos coherentes y cohesionados haciendo uso completo y apropiado de una variedad de criterios de organización y de una gran diversidad de mecanismos de cohesión.
C1	Produce un discurso claro, fluido y bien estructurado, mostrando un uso adecuado de criterios de organización, conectores y mecanismos de cohesión.
B2.2	Utiliza con eficacia una variedad de palabras de enlace para señalar con claridad las relaciones que existen entre las ideas.
B2.1	Utiliza un número limitado de mecanismos de cohesión para enlazar frases y crear un discurso claro y coherente, aunque puede mostrar cierto “nerviosismo” en una intervención larga.
B1	Enlaza una serie de elementos breves, concretos y sencillos para crear una secuencia cohesionada y lineal.
A2	Enlaza grupos de palabras con conectores sencillos, como, por ejemplo, “y”, “pero” y “porque”.
A1	Enlaza palabras o grupos de palabras con conectores muy básicos, como “y” o “entonces”.

(Consejo de Europa, 2002: 122)

Fluidez oral.	
C2	Se explica mediante un discurso amplio y seguido con una fluidez natural, sin esforzarse ni dudar. Se detiene sólo para considerar las palabras más adecuadas con las que expresar sus pensamientos o para buscar un ejemplo o explicación adecuada.
C1	Se expresa con fluidez y espontaneidad, casi sin esfuerzo. Sólo un tema conceptualmente difícil puede obstaculizar la fluidez natural del discurso.
B2.2	Se comunica espontáneamente, a menudo mostrando una fluidez y una facilidad de expresión notables incluso en períodos más largos y complejos.
B2.1	Produce discursos con un ritmo bastante regular, aunque puede dudar mientras busca estructuras y expresiones; provoca pocas pausas largas. Participa en la conversación con un grado de fluidez y espontaneidad que hace posible la interacción habitual con hablantes nativos sin producir tensión en ninguno de los interlocutores.
B1.2	Se expresa con relativa facilidad. A pesar de algunos problemas al formular su

	discurso, que dan como resultado pausas y “callejones sin salida”, es capaz de seguir adelante con eficacia y sin ayuda.
B1.1	Es capaz de mantener su discurso, aunque sean muy evidentes las pausas para ordenar la gramática y el léxico, y de corregirse; sobre todo, en períodos largos de producción libre.
A2.2	Se hace entender en intervenciones breves, aunque resulten muy evidentes las pausas, las dudas iniciales y la reformulación.
A2.1	Construye frases sobre temas cotidianos con la facilidad suficiente como para desenvolverse en breves intercambios, a pesar de tener dudas muy evidentes y tropiezos en el comienzo.
A1	Se desenvuelve con enunciados muy breves, aislados y preparados, y con muchas pausas para buscar expresiones, articular palabras menos habituales y salvar la comunicación.

(Consejo de Europa, 2002:126)

Para calificar el resultado se pensó en utilizar también otras escalas de dominio, como la de *Flexibilidad*, la de *Corrección gramatical*, la de *Control y corrección*, la de *Compensación*, la de *Dominio de vocabulario*, la de *Riqueza de vocabulario* y la de *Competencia lingüística general*.

El objeto de las tareas 6 y 7 es averiguar la capacidad para hablar sobre experiencias personales, pero con la dificultad intrínseca de que sea una narración a la que se puedan añadir todo tipo de descripciones u opiniones y respecto a la que se puedan analizar los elementos ya señalados presentes en las escalas (vocabulario, coherencia y cohesión textual, etc.). A su vez, y como estas escalas presentan diferentes descriptores según los 6 niveles del MCER, se ha buscado que las tareas sean suficientemente abiertas y posibilitadoras como para elicitarse desde las muestras más sencillas del lenguaje hasta las más complicadas. Por ello se dan instrucciones diferentes para el evaluador, ya se trate de un candidato de niveles bajos o altos.

La tarea 6 (explicar un viaje o experiencia enriquecedora), pensada para todos los niveles, se extrajo de la escala de *Desarrollo de descripciones y narraciones*, intentando poder hacer una comparación entre el lenguaje que elicite y el de la OPI -donde estas funciones de narrar, describir, explicar y opinar son primordiales para establecer si un candidato es intermedio alto o avanzado bajo o medio-.

La tarea 7 (explicar las ventajas de vivir y estudiar en el país), pensada para todos los niveles, se puso por ser esta una de las funciones que se califican en las OPI, una comparación en la que se expresan los pros y contras de un tema. Esta se puede evaluar

según las escalas de *Coherencia y cohesión* y de *Expresión oral en general* del MCER entre otras.

La tarea 8 (explicar un cómic), se pensó para todos los niveles, usando la escala de *Fluidez oral*. Con ella se pretende elicitarse y calificar el lenguaje en general y graduar al candidato en uno de los 6 niveles establecidos según sus descriptores. Se trata de averiguar la capacidad para explicar, usar vocabulario o estrategias de compensación, organizar información, narrar o describir, así como el resto de capacidades respecto al dominio de la lengua señaladas en las otras escalas apuntadas.

Para la calificación de la actuación de los candidatos se usan las escalas con las que se diseñó cada tarea, además de otras consideradas interesantes. Estas otras escalas son las siguientes²⁰⁷:

- *Competencia lingüística general.*
- *Riqueza de vocabulario.*
- *Dominio de vocabulario.*
- *Compensación.*
- *Control y corrección.*
- *Corrección gramatical.*
- *Flexibilidad.*
- *Expresión oral en general.*

3.5 Precauciones tomadas para que los resultados de la segunda entrevista sean comparables con los de una OPI de ACTFL

Para sortear algunas de las diferencias existentes entre los dos sistemas (ACTFL y MCER) y hacerlos más fácil y objetivamente comparables, al diseñar el nuevo test se han intentado acercar los sistemas y técnicas de elicitación de discurso.

Se han diseñado tareas similares a las que se persiguen en las OPI, para que las actuaciones del candidato, mediante un sistema u otro, sean similares y puedan someterse a una calificación comparativa. La selección de actividades se ha hecho

²⁰⁷ Se pueden ver en el Apéndice 6.

teniendo en cuenta que en una OPI se buscan una serie de funciones como la habilidad de descripción y de narración o la habilidad para dar explicaciones y opiniones y hacer comparaciones con detalle. Y se ha pretendido que esto también surja en el nuevo test.

A continuación citamos varios de estos elementos tenidos en cuenta, que ofrecen garantías de una evaluación adecuada y ayudan a hacer comparaciones entre ambos sistemas:

1.- Todas las entrevistas se grabaron en casete digital para que su calificación se pudiera hacer como se hace con cada una de las OPI de ACTFL, con sucesivas escuchas y análisis, incluso por un segundo o tercer evaluador. Esto obliga, asimismo, al evaluador a mantener un protocolo en su comportamiento y técnicas de elicitación del discurso que sea justo, objetivo y verificable. Con ello se pueden evitar los problemas ya discutidos (§2.4) respecto a la variabilidad y falta de autocontrol respecto al comportamiento del evaluador y las técnicas de elicitación, producidos cuando el test no se graba.

2.- Todas las entrevistas, las del test del MCER y las de la OPI, tuvieron al mismo entrevistador, una misma persona formada por ACTFL y conocedora del método de la OPI²⁰⁸.

Para evitar la variable de subjetividad derivada de una falta de homogeneización entre entrevistadores o de la natural disparidad de criterios, todas las entrevistas las ha realizado un solo evaluador formado con los mismos criterios, protocolos y objetivos²⁰⁹.

No obstante, para la calificación de las grabaciones una vez realizadas, sí se ha dispuesto de varios equipos de personas formadas de acuerdo con el mismo sistema y objetivos, como se ha señalado en §3.3.5.

3.- El diseño de las actividades del test se ha hecho teniendo en cuenta los elementos que se evalúan en una OPI de ACTFL, porque cuanto más similares sean las tareas, más fácil es su comparación.

²⁰⁸ Esta persona es el autor de esta tesis y miembro de ACTFL que superó los cursos de formación de esta asociación en 1999.

²⁰⁹ Un equipo de entrevistadores que pusiera en práctica las nuevas tareas diseñadas, podría dar resultados discrepantes y estar influido por variables indeseables, salvo que antes se procediera a una adecuada formación. Este lujo no era necesario para la investigación, ya que el objeto de esta no es ver cómo entrevistan sino ver si las calificaciones dadas a las grabaciones de la OPI son comparables con las dadas a las grabaciones de nuestro test del MCER.

La OPI analiza, en el nivel intermedio, la capacidad de los candidatos para elaborar discursos siendo creativos en el uso del lenguaje, combinando material aprendido para producir todo tipo de información personal, utilizando frases sencillas. Por eso se han diseñado nuevas tareas en las que se deba utilizar el lenguaje en este sentido y que van a ser calificadas según la escala de *Intercambio de información oral* del MCER u otras escalas que contengan similares características. Entre ellas una que existe en la OPI de ACTFL (la de hacer preguntas para solicitar un trabajo de canguro).

Igualmente, en la OPI, en el nivel avanzado se busca calificar la habilidad del candidato para realizar discursos más complejos en los que la coherencia textual y organizativa sea mayor y se forme mediante frases complejas encadenadas formando párrafos bien organizados; y las funciones que se califican son las de describir, explicar, comparar, narrar y opinar, en forma detallada y coherente y sobre temáticas sociales o generales, más allá de la propia explicación de experiencias del candidato. Por ello, en el nuevo test se han diseñado también tareas en las que el candidato deba hacer estas mismas cosas, describir, explicar con detalle, comparar narrar y opinar.

Se eligieron las escalas y funciones con objeto de poder elicitar un lenguaje que pudiera ser semejante al que se elicitaba mediante una OPI: el intercambio de información (como en las tareas 1, 2, 3 y 4), la argumentación (como en la tarea 5), las descripciones y narraciones (como en la tarea 6) y la coherencia y cohesión (como en la tarea 7), entre otras cosas.

Algunas tareas tienen un correlato directo en las actividades que se realizan mediante una entrevista de OPI. La tarea 1, trabajo de canguro, es una tarea que existe en la OPI como *role play* para determinar si los candidatos son capaces de realizar preguntas sobre temas sencillos y cotidianos y establecer si dominan el nivel intermedio o son todavía iniciales. También la tarea 3, opinar sobre una actividad cultural, presenta una fácil comparación con otras actividades realizadas en la OPI, ya que muy a menudo se trata de hablar de aspectos culturales interesantes y dar opiniones al respecto. Igualmente las tareas 5, 6 y 7 existen de forma similar en una OPI, en la que en nivel avanzado se pide que los candidatos hagan comparaciones o análisis de ventajas y desventajas sobre un tema, o que argumenten sus ideas y expliquen viajes o experiencias enriquecedoras.

4.- Las habilidades de cada estudiante se han evaluado en conjunto, siendo la calificación final la encargada de decidir su nivel.

En cuanto a la estructura del test, para asemejarlo a las OPI, hemos decidido que cada estudiante no se presente a un nivel determinado preseleccionado por él mismo, por su profesor o por una institución en la que va a estudiar o para la que quiere trabajar (como suele ser habitual en otros exámenes donde el candidato se matricula para evaluarse en un nivel que él aspira a probar que domina y sobre el que se le examina). En nuestro test hemos propuesto a cada candidato tareas de diferentes niveles, comenzando por una sencilla y ascendiendo en complejidad y dificultad; así, cada candidato ha tenido que realizar tareas de su nivel junto a otras tareas de mayor o menor nivel en una progresión de dificultad y complejidad. Con ello, si superaba las primeras pruebas obteníamos otros datos respecto a sus habilidades en otras tareas de nivel inmediatamente superior, lo que nos permitía juzgar y calificar otros elementos necesarios para averiguar todo el rango de posibilidades de uso de la lengua.

Estas tareas de mayor dificultad nos permiten, además, ver en qué medida el estrés comunicativo se puede controlar en un test como el que estamos utilizando y cómo este estrés se localiza, interpreta y juzga al utilizar las escalas de medición del MCER. Además, al ir utilizando un escalonamiento de tareas en progresión de dificultad creciente, se puede decidir cuál es el nivel más alto que posee cada candidato y no solo decidir si un candidato posee un nivel predeterminado o no. Con ello, logramos saber cuál es el nivel más alto de cada candidato, lo que permite una mejor comparación entre ambos sistemas de evaluación.

5.- El método de la entrevista, la actuación del evaluador y las técnicas de elicitación se han acercado a las técnicas de elicitación y las directrices del modelo de ACTFL, teniendo especial cuidado en evitar crear estrés en los candidatos.

Por esto hemos intentado atender a tres elementos importantes, como son:

a. Perseguir el objetivo de cada tarea con preguntas adecuadas; aclaratorias o motivadoras cuando el candidato parezca no lograrlo, para así facilitar las respuestas; buscando pruebas de que esas habilidades realmente existen o no existen²¹⁰.

b. Atender a lo que se llama “el momento de la entrevista”, que consiste en ser consciente en todo momento de lo que está haciendo o pretendiendo hacer el candidato,

²¹⁰ Para más detalles sobre este sistema de preguntar y la tipología de preguntas motivadoras en cada momento, remitimos al apartado que explica el método de entrevistas OPI de ACTFL, §2.6.2.

con independencia de lo que se le haya sugerido que haga²¹¹ (Lazaraton, 2002: 37). Esto permite reorientar el discurso mediante nuevas preguntas que guíen las respuestas hacia los objetivos deseados; en vez de permitir al candidato que siga por su camino alejándose de lo que se pretendía, por error de apreciación o por carencia de otras posibilidades.

c. Evitar una progresión de temas o preguntas que hagan surgir estrés e impidan al candidato mostrar su mejor nivel de lengua²¹².

En los siguientes apartados describimos: a. el diseño del test; b. la validación de las tareas; c. el pilotaje a que se sometieron con expertos y candidatos voluntarios y d. la validación del test que conforman.

3.6 Validación del test

El test de ACTFL ya ha sido definido en el apartado correspondiente (§2.6). Queda ahora dejar constancia de que el nuevo test del MCER no penaliza o premia a los candidatos. En el capítulo 4 se pueden ver los resultados y cómo los evaluadores muestran una alta correlación y homogeneidad en sus calificaciones, por lo que el sistema de corrección utilizado y los resultados obtenidos son válidos. Además, y puesto que pretendemos que el test esté diseñado de acuerdo con el MCER, se debe cumplir la serie de requisitos de formación de evaluadores, diseño de tareas y validación, expuesta en el manual para relacionar exámenes con el MCER (Consejo de Europa, 2002 b). De este manual se ha tratado en §2.7.5. Los formularios se encuentran en el Apéndice 4²¹³. No obstante, para nuestra tesis se ha realizado una validación independiente que se presenta en este capítulo en los siguientes apartados.

²¹¹ De este tema de las intenciones del candidato se habla en §2.4.3- 2.

²¹² En nuestra investigación las hemos podido manejar porque el evaluador ha sido entrenado para poder hacerlo y era una sola persona. El objeto de la investigación no era ver en qué medida otras personas van a ser capaces de hacer una entrevista objetiva y homogénea siguiendo esta técnica sino elicitare las mejores muestras de lengua para proceder a comparar la calificación que sobre estas se haga utilizando el MCER con la calificación obtenida en la OPI.

²¹³ Los formularios del Apéndice 4 son los que oficialmente se deben presentar para solicitar la aceptación del nuevo test como test relacionado con el MCER de acuerdo con los protocolos del Consejo de Europa. A pesar de que aquí se ha seguido un sistema de validación independiente, los presento cumplimentados, para mostrar que se cumplen los requisitos.

Como nuestro empeño está en encontrar si ante un mismo candidato dos formas de calificar su actuación pueden ser homologables (la de ACTFL y la realizada mediante nuestro test del MCER), es importante establecer de antemano las características de las tareas elegidas y demostrar que las diferencias de calificación no se deben a la influencia de ellas sino a las escalas utilizadas al calificar, o a lo sumo, al modo del test, o al modo de elicitar las respuestas y provocar el comportamiento lingüístico²¹⁴.

Las tareas tienen la importante función de provocar que en la respuesta de los candidatos surja el tipo de lengua que queremos evaluar y no un tipo de lengua artificial producido solamente por o para la evaluación. En la medida en que el test ofrezca oportunidades al candidato para que este actúe a su más alto nivel de competencia, logrando los objetivos lingüísticos definidos, este será un buen test.

Por ello, en la definición de las tareas debemos determinar adecuadamente las características de la lengua real y las condiciones en que se da en su ambiente natural, para luego compararlas con las condiciones y características de la lengua elicitada mediante las tareas del test. Eso es lo que vamos a hacer en este apartado, mediante el pilotaje y validación de las tareas y el test, de forma que se pueda tener una clara referencia de su alcance o de sus posibles limitaciones.

A continuación damos los resultados de la validación de las tareas, donde se analiza su naturalidad, su contenido, el lenguaje que elicitan, el nivel para el que pueden servir y el sistema de calificación que requieren. Primero (§3.6.1), los resultados obtenidos mediante encuestas hechas con profesores expertos y (§3.6.2) con estudiantes voluntarios a los que, asimismo, se pidió que realizaran las tareas. Y luego (§3.6.3) se explica la validación general del test.

3.6.1 Validación de las tareas y su tipo de lenguaje con expertos

Los análisis de cada tarea se realizaron con expertos que contestaron y reflexionaron sobre determinados aspectos de cada tarea.

²¹⁴ Los formularios del Apéndice 4 son los que oficialmente se deben presentar para solicitar la aceptación del nuevo test como test relacionado con el MCER de acuerdo con los protocolos del Consejo de Europa. A pesar de que aquí se ha seguido un sistema de validación independiente, los presentamos cumplimentados, para mostrar que se cumplen los requisitos.

Se trabajó con 10 profesores de español como lengua extranjera del Centro de Lenguas Modernas de la Universidad de Granada, varios de ellos doctores y todos con amplia experiencia en la enseñanza de español como lengua extranjera, que contestaron a un cuestionario de 6 preguntas²¹⁵, para averiguar:

- a. Si las tareas sirven para evaluar la habilidad oral general de una persona y conforman un constructo apropiado (pregunta 1).
- b. Si las tareas son auténticas (pregunta 2).
- c. Si el lenguaje que provocan las tareas sirve para medir las habilidades planeadas (pregunta 3).
- d. El nivel para el que sirve cada tarea (pregunta 4).
- e. Si las escalas del MCER seleccionadas son aptas para evaluar la lengua del candidato (pregunta 5).
- f. El tipo de lenguaje que producen las tareas (pregunta 6).

Estas preguntas cubrían un amplio marco de características necesarias para validar el test en conjunto: aspectos relativos a la validez de constructo, la validez aparente y autenticidad de las pruebas, la validez de respuesta e interactividad de las tareas, y la validez de contenido. De todo ello se ha tratado en §2.3.

En la realización de las 8 tareas que finalmente se han aceptado como válidas, se cumplieron casi en su totalidad las expectativas respecto a los niveles para los que se podían utilizar, las escalas con las que se debían calificar y el tipo de lenguaje que generaban. No se aprecian grandes diferencias entre el lenguaje usado en las actividades en la vida real y en el usado en el test, ni hay elementos en el tipo de input dado para la realización de las tareas que le resten fiabilidad; así mismo, cumplen con el tipo de lenguaje que se esperaba obtener en las respuestas.

Tras este proceso de análisis con expertos y candidatos se incluyeron algunos cambios en el test. Estos son los siguientes:

²¹⁵ Véase el cuestionario con sus preguntas concretas en el Apéndice 7.

- Se flexibilizó la forma de dar las instrucciones de las tareas para adaptarlas a cada participante.

- Se decidió utilizar más escalas en el proceso de calificar los resultados de ciertas tareas, en concreto la escala de *Monólogo sostenido* para calificar las tareas sobre la actividad cultural (tarea 3), explicar ciertos aspectos de tu trabajo... (tarea 4), explicar un viaje enriquecedor..., (tarea 6), explicar las ventajas de estudiar la lengua en el país..., (tarea 7) y contar el cómic, (tarea 8). Y las escalas de *Fluidez* y de *Monólogo sostenido* para la tarea de contar el cómic (tarea 8).

- Se amplió el número de escalas utilizadas para la calificación general de la actuación de los candidatos.

- Se vio que algunas tareas servían para evaluar las habilidades y elicitar suficiente información en candidatos de diferente nivel de lo que se había pensado al diseñarlas: la tarea de contar un cómic (tarea 8) parece apta para candidatos de todos los niveles, la de explicar las ventajas de estudiar la lengua en el país en que se habla (tarea 7), también parece apta para niveles medios, como B1, y la de indicar el camino (tarea 9) en un plano, parece apta para candidatos de menos nivel (aunque esta finalmente hubo que desecharla porque no generaba el lenguaje esperado).

Las tareas que presentan variaciones mayores respecto a su realización como actividad en el test y fuera del test son la 1 (trabajo de canguro) y la 4 (explicación del puesto de trabajo) ya que en la vida real requieren una relación especial entre los interlocutores y se deben realizar en un lugar específico que difiere del aula de lenguas en la que, además, hay que simular una ficticia y especial relación entre el candidato y el evaluador. No obstante, como se verá en su validación, esto no las dificulta ni invalida.

En cuanto a otras diferencias producidas en el tipo de lenguaje esperado o en la actuación de los candidatos, por las diferentes condiciones físicas entre los lugares de realización de las actividades en la vida real y durante el test o por la forma de realización del test o el tipo de instrucciones dadas, estas solamente se aprecian en las tareas 1, 2 y 4 y no son decisivas.

Se señalan en el siguiente cuadro:

Elementos	Lengua real	Lengua de la tarea del test
<p>Características del escenario Características físicas</p>	<p>Tarea 1. Casa de familia, lugar de trabajo, una cafetería, etc. Sentado o de pie e incluso tomando un café y visitando las diferentes partes de la casa o la niña o niño que hay que cuidar.</p> <p>Tarea 2. Casa de familia, lugar de trabajo, una cafetería, etc. Sentado o de pie.</p> <p>Tarea 4. En su lugar de trabajo.</p>	<p>Tarea 1. Aula de clase de lenguas. Sentado en una silla ante el evaluador que hace las veces de padre o madre de familia que necesita un canguro.</p> <p>Tarea 2. Aula de clase de lenguas. Sentado en una silla ante el evaluador que hace las veces de colega o amigo interesado.</p> <p>Tarea 4. Aula de clase de lenguas. Sentado en una silla ante el evaluador que hace las veces de colega o amigo interesado.</p>
<p>Participantes</p>	<p>Tarea 1. Dos personas o más. Canguro que necesita un trabajo. Padre o madre de la familia.</p> <p>Tarea 4. Colega que va a sustituirte en tu trabajo durante unos días.</p>	<p>Tarea 1. Candidato que simula ser una persona que necesita un trabajo de canguro. Y evaluador que simula ser el padre de familia que tiene una hija a la que hay que cuidar.</p> <p>Tarea 4. Candidato que simula que explica su trabajo al evaluador, que simula ser un colega que va a sustituir en el trabajo al candidato.</p>
<p>Hora de la tarea</p>	<p>No concreta. No es relevante.</p>	<p>No concreta. No es relevante.</p>

Respecto al test en su conjunto, formado por todas estas tareas, los expertos manifiestan en su casi totalidad- 80%- que sirve adecuadamente para medir el nivel general de habilidad lingüística oral de los candidatos, con comentarios como que “es bastante completo”, “es redondo” o “varía suficientemente las tareas como para producir diferentes y variadas muestras de lengua” que ayudan a tener una visión global de la habilidad de los candidatos.

Un 20% dijo que para servir como test de habilidad general de uso de la lengua debería incluirse alguna otra tarea que ayudara a calificar aspectos como la interacción oral y la fluidez, alguna tarea con una planificación previa por parte del estudiante que sirviera para considerar aspectos académicos o profesionales usuales en ámbitos laborales y educativos, alguna tarea que sirviera realmente para diferenciar en los niveles altos a quien es B2, C1 y sobre todo C2 y tareas que elicitaran un lenguaje basado en suposiciones e hipótesis argumentando sobre temas nunca antes planteados.

En respuesta a estos datos tenemos que decir, como se ha expuesto en §3.4, que las muestras de lengua obtenidas con estas tareas se van a calificar siguiendo la escala de la que se diseñaron, juntamente con otras escalas de dominio y fluidez. Efectivamente, reconocemos en todo test oral de este tipo una carencia de aspectos de interacción más abierta, pero tanto este como otros aspectos se intentan solventar mediante el protocolo de comportamiento del entrevistador, que se acerca a las maneras de elicitar las muestras de lengua utilizadas en las pruebas orales de ACTFL (ya explicadas en §2.6.2 y en este capítulo en §3.5).

3.6.1- a) Validación de la tarea 1: “Trabajo de canguro”

Los datos confirman nuestras intenciones respecto al nivel en que se iba a usar la tarea, A2, al tipo de lenguaje generado; a la escala utilizada para su calificación, *Intercambio de información*.

Los expertos opinaron que la naturalidad de la tarea era alta y que el lenguaje que se podía esperar que produjera era el adecuado a la escala de *Intercambio de información*, en los niveles bajos. Respecto al tipo de lenguaje que elicitaba²¹⁶, la mayoría dijo que era sencillo, sin mucha necesidad de organización retórica, ni de uso de elementos pragmáticos, ni de conocimientos culturales especiales.

Manifestaron:

- Que consideraban bastante natural la tarea: el 70% le da un 5 en una escala de 5 puntos (un 30% le da un 4).

²¹⁶ A los expertos se les pidió opinión sobre 4 aspectos (complejidad, organización retórica, saberes culturales de que depende la realización de la tarea y elementos pragmáticos que se pueden producir en el discurso generado). Véanse las preguntas 3 y 6 de la encuesta en el Apéndice 7.

- Que sirve efectivamente para elicitación y medir las habilidades de intercambio de información que habíamos pretendido con su diseño (el 100% está de acuerdo en esto).

- Que la escala con la que se diseñó la tarea sirve para la calificación de la lengua generada con la tarea.

En este sentido, identifican claramente algunos descriptores de la escala como útiles para definir los aspectos de realización de dicha tarea; en concreto para el nivel A2 (70% de expertos) todos sus descriptores, pero sobre todo los 13, 16 y 17, y para el nivel B1 (50% de expertos) los descriptores 5, 9 y 10²¹⁷.

Respecto al tipo de lenguaje, sus respuestas fueron:

- Complejidad: la mayoría estima que el lenguaje generado es sencillo (el 60%).

- Organización retórica: la mayoría estima que no se necesita mucha organización retórica (el 80% estima que el nivel de organización retórica generado es bastante sencillo o sencillo).

- Saberes culturales: la mayoría cree que no es necesaria ni poca ni mucha formación cultural (el 60%).

- Elementos pragmáticos generados: aquí hay división de opiniones y la mitad de los expertos consideran que no son necesarios demasiados elementos pragmáticos para realizar bien esta tarea (el 20 % cree que no necesita ni produce muchos elementos pragmáticos, el 10% que un poco, otro 10% que ni pocos ni muchos) pero el 50% cree que sí (un 20% que bastantes y un 30% que muchos).

3.6.1- b) Validación de la tarea 2: “Receta de comida”

Los datos confirman nuestras intenciones respecto al nivel B1 de la tarea y tipo de lenguaje generado, así como a la escala utilizada para su calificación, *Intercambio de información*. Asimismo los expertos opinaron que la naturalidad de la tarea era alta y que el lenguaje que se podía esperar que produjera era el adecuado.

²¹⁷ La escala de *Intercambio de información*, con sus descriptores numerados, puede verse en el Apéndice 6.

Respecto al tipo de lenguaje que elicitaba, la mayoría dijo que era sencillo, sin mucha necesidad de organización retórica, ni de uso de elementos pragmáticos, ni de conocimientos culturales especiales.

Los expertos manifestaron:

- Que consideraban bastante natural la tarea (el 70% le da 4 o 5 puntos).
- Que sirve para elicitación y medir las habilidades de *Intercambio de información* que habíamos pretendido (el 100% de los expertos).
- Que la escala con la que se diseñó la tarea sirve para la calificación de la lengua generada con la tarea. Identifican claramente algunos descriptores de la escala como útiles para definir los aspectos de realización de dicha tarea, para el nivel B1 especialmente (90%).

Respecto al tipo de lenguaje, sus respuestas fueron:

- Complejidad: el 70% estima que el nivel de lenguaje generado es bastante sencillo.
- Organización retórica: el 80% estima que el nivel de organización retórica generado o necesario es sencillo.
- Saberes culturales: el 60% estima que no son necesarios saberes culturales especiales.
- Elementos pragmáticos generados: el 60% estima que esta tarea no necesita o produce muchos elementos pragmáticos.

3.6.1- c) Validación de la tarea 3: “Opinión sobre actividad cultural”

Los datos confirman nuestras intenciones respecto a los niveles de la tarea y tipo de lenguaje generado, pero nos hicieron considerar también la escala de *Monólogo sostenido: argumentación* para realizar las calificaciones del lenguaje que genera, pues el 90% de los expertos manifestó que podía medirse con esta, además de con la de *Intercambio de información*.

Asimismo, los expertos opinaron que la naturalidad de la tarea era alta y que el lenguaje que se podía esperar que produjera era el adecuado.

Respecto al tipo de lenguaje que elicitaba, la mayoría dijo que era sencillo, sin mucha necesidad de organización retórica, pero que sí son necesarios conocimientos culturales especiales así como el uso de elementos pragmáticos.

Los expertos manifestaron:

- Que consideraban bastante natural la tarea (el 90% le da un 5 o 4 puntos).
- Que el lenguaje que produce sirve para medir la habilidad de *Intercambio de información* (70%) aunque apuntan en mayor medida que se puede medir la habilidad de *Monólogo sostenido: argumentación* (90 %). En este sentido identifican claramente algunos descriptores de las escalas como útiles para definir los aspectos de realización de dicha tarea, en concreto para el nivel B1 (80%) -los descriptores 5 y 7 de *Intercambio de información* y 5 y 6 de *Monólogo sostenido*-.

Respecto al tipo de lenguaje que elicitaban, las respuestas fueron:

- Complejidad: la mayoría (50%) estima que el nivel de lenguaje generado o necesario no es ni sencillo ni complicado, frente a solo un 20% que cree que se necesita un lenguaje un poco complicado.
- Organización retórica: el 70% estima que el nivel de organización retórica generado o necesario no es complicado.
- Saberes culturales: el 60% estima que son necesarios bastantes, lógicamente para contar la experiencia cultural vivida.
- Elementos pragmáticos: en este caso hay división de opiniones: el 40% estima que esta tarea no necesita, pero otro 40% cree que sí.

3.6.1- d) Validación de la tarea 4: “Explicar ciertos aspectos del trabajo”

Los datos confirman nuestras intenciones respecto a los niveles de la tarea, B2 o más, tipo de lenguaje generado y escala de calificación, *Intercambio de información* (70%) y en menor medida (40%) también de *Monólogo sostenido: argumentación*. Asimismo, los expertos opinaron que la naturalidad de la tarea era alta y que el lenguaje que se podía esperar que produjera era el adecuado.

Respecto al tipo de lenguaje que elicitaba, hubo bastante diversidad de opiniones tanto sobre su dificultad como sobre la necesidad de conocimientos, uso de elementos pragmáticos y organización retórica.

Las respuestas fueron:

- Complejidad: el 50% estima que el nivel de lenguaje generado o necesario es complicado, frente a un 20% que dice que es sencillo, y un 30% que dice que no es sencillo ni complicado.

- Organización retórica: el 40% estima que el nivel de organización retórica generado o necesario es complicado, pero el 30% estima que es sencillo.

- Saberes culturales: el 40% estima que son necesarios pero el 30% que no.

- Elementos pragmáticos generados: el 50% estima que esta tarea necesita y produce bastantes elementos pragmáticos, pero un 30% cree que no.

3.6.1- e) Validación de la tarea 5: “Argumentar sobre la guerra...”

Los datos confirman nuestras intenciones respecto al nivel de la tarea -B2 o más- y al tipo de lenguaje generado y escala de calificación, *Monólogo sostenido: argumentación*. Asimismo los expertos opinaron que la naturalidad de la tarea era alta (70%) y que el lenguaje que se podía esperar que produjera era el adecuado (100%).

Respecto al tipo de lenguaje que elicitaba, la mayoría dijo que era difícil, que requería conocimientos culturales especiales así como el uso de elementos pragmáticos, pero que la organización retórica no es ni sencilla ni complicada.

Los expertos manifestaron:

- Complejidad: el 80% estima que el nivel de lenguaje generado o necesario es complicado o muy complicado.

- Organización retórica: el 50% estima que el nivel de organización retórica generado o necesario es un poco complicado o muy complicado, pero el 40% estima que no es complicado.

- Saberes culturales: el 60% estima que son necesarios bastantes.

- Elementos pragmáticos: el 60% estima que esta tarea necesita y produce bastantes elementos pragmáticos.

3.6.1- f) Validación de la tarea 6: “Explicar el viaje o experiencia...”

Esta tarea se diseñó tanto para niveles bajos como altos y los datos confirman nuestras intenciones respecto al uso en todos los niveles, al tipo de lenguaje generado y escalas de calificación, *Intercambio de información* y *Monólogo sostenido*.

Asimismo, los expertos opinaron que la naturalidad de la tarea era alta (90%) y que el lenguaje que se podía esperar que produjera era el adecuado (100%).

Respecto al tipo de lenguaje que elicitaba, la mayoría dijo que era difícil, que requería conocimientos culturales especiales, así como el uso de elementos pragmáticos, pero que la organización retórica no es ni sencilla ni complicada.

Por último, respecto al tipo de lenguaje que elicitaban, las respuestas fueron:

- Complejidad: en líneas generales manifestaron que era necesaria poca o muy poca.
- Organización retórica: hay un reparto de opiniones entre quienes piensan que no es necesaria organización retórica o muy poca, quienes opinan que es necesaria bastante y quienes opinan que ni poca ni mucha.
- Saberes culturales: en niveles bajos el 40% cree que apenas se necesitan, y en niveles altos, el 40% cree que sí son bastante o muy necesarios.
- Elementos pragmáticos generados: en general hay una profunda división de opiniones pues el 40% estima que necesita y produce bastantes elementos pragmáticos frente a otro 40% que cree que no son necesarios.

3.6.1- g) Validación de la tarea 7: “Explicar las ventajas de estudiar...”

Los datos confirman nuestras intenciones respecto a los niveles de la tarea, B1 y B2, al tipo de lenguaje generado y a la escala de calificación, *Monólogo sostenido: argumentación*.

Asimismo, los expertos opinaron que la naturalidad de la tarea era alta (60%) y que el lenguaje que se podía esperar que produjera era el adecuado.

Respecto al tipo de lenguaje que elicitaba, la mayoría dijo que no es fácil ni difícil, que requería organización retórica y conocimientos culturales especiales pero no hay acuerdo respecto a la necesidad del uso de elementos pragmáticos.

Las respuestas fueron:

- Complejidad: el 60% estima que tiene una complejidad media.
- Organización retórica: el 60% estima que tiene una organización retórica media.
- Saberes culturales: también el 60% de expertos estima que son necesarios bastantes o muchos.
- Elementos pragmáticos generados: hay división de opiniones entre quienes creen - 40% de expertos -que sí van a producirse o necesitarse y los que creen -otro 40%- que estos no van a ser demasiados.

3.6.1- h) Validación de la tarea 8: “Explicar un cómic”

Los datos confirman nuestras intenciones respecto a la posibilidad de utilizar esta tarea en casi todos los niveles, de evaluarla con varias escalas de habilidad y dominio y respecto a la variedad de lenguaje generado. Ya preveíamos una falta de naturalidad en el hecho de contar un cómic y no fue necesario introducir ningún cambio en su planteamiento.

Asimismo, los expertos opinaron que la naturalidad de la tarea era baja (80%) y que el lenguaje que se podía esperar que produjera era el adecuado.

Respecto al tipo de lenguaje que elicitaba, hubo opiniones de todo tipo (fácil, difícil y ni fácil ni difícil), igualmente respecto a la organización retórica requerida o generada y a la necesidad del uso de elementos pragmáticos; respecto a la necesidad de conocimientos culturales especiales, la mayoría cree que no son necesarios.

Las respuestas fueron:

- Complejidad: los expertos se dividen en cuatro franjas diferentes a proporciones iguales del 20% cada una; un 20% no responde a esta cuestión, un 20% opina que la complejidad de la lengua va a ser baja, otro 20% que mediana, otro 20% que alta y otro que muy alta.
- Organización retórica: el 30% no contestó a esta pregunta, el 20% estima que el nivel de organización retórica generado o necesario era sencillo, el 10% que no es sencillo ni complicado, el 30% que es un poco complicado y un 10% que muy complicado, (es decir, un 40% estima que es o complicado o muy complicado, frente a un 20% que estima que es muy sencillo),

- Saberes culturales: el 40% estima que son necesarios pocos saberes culturales.
- Elementos pragmáticos generados: el 30% estima que esta tarea no necesita ni produce ni pocos ni muchos, el resto de opiniones discrepan completamente una de otra con porcentajes del 10%: un 10% estima que se necesitan o produce muy pocos, otro 10% que pocos, otro 10% que bastantes y otro 10% que muchos (el 30% no contestó).

3.6.1- i) Validación de la tarea 9: “Indicar el camino en un plano”

Los datos no confirman nuestras intenciones respecto al nivel B1 de la tarea sino que lo estiman más sencillo (un 50% cree que es apta en el nivel A2) aunque sí respecto al tipo de lenguaje generado, sencillo en general.

Como se verá tras los análisis con candidatos voluntarios (apartado 3.6.2 i), esta tarea se desechó por la imposibilidad de realizarla al no estar familiarizados con las calles de la ciudad ni con el plano, ni utilizar elementos de referencia para explicar el camino. No obstante, los expertos no apreciaron la necesidad de conocimientos culturales, es decir, el conocimiento de las calles de la ciudad y de elementos de referencia necesarios para explicar el camino a un lugar en una ciudad.

En las pruebas con expertos, manifestaron que consideraban bastante natural la tarea (el 100% le da un 5 o 4 puntos); que sirve efectivamente para elicitación y medir las habilidades de *Intercambio de información* que habíamos pretendido. Todos los expertos consultados están de acuerdo en esto y, asimismo, dan por buena la relación entre esta tarea y la escala de la que se ha partido para su diseño, la escala de *Intercambio de información*.

Y respecto al tipo de lenguaje que eliciten, las respuestas fueron:

- Complejidad: el 70% estima que el nivel de lenguaje generado o necesario es bastante sencillo o sencillo.
- Organización retórica: el 70% estima que el nivel de organización retórica generado o necesario es bastante sencillo o sencillo.
- Saberes culturales: el 80% estima que no son necesarios.
- Elementos pragmáticos generados: el 50% estima que esta tarea no necesita o no produce apenas elementos pragmáticos, y el 40% que bastantes o muchos.

3.6.2 Validación de las tareas y su lenguaje con estudiantes

El pilotaje y validación con estudiantes se hizo con 37 estudiantes del Curso intensivo de lengua española (CILE), del Curso de Lengua y Cultura Españolas (CLCE) y del Curso de Estudios Hispánicos (CEH) del Centro de Lenguas Modernas de la Universidad de Granada, a los que se plantearon una serie de preguntas relativas a la autenticidad, viabilidad y validez de respuesta, para averiguar:

- a. En qué medida las tareas y su sistema de administración era efectivo.
- b. El grado de naturalidad de las tareas requeridas.
- c. Las reacciones que les provocaban.
- d. El tipo de lenguaje que generaban.

Unos cuantos, además, realizaron pruebas con los tests.

a. En cuanto a la administración de las tareas y el test completo, no hubo apenas comentarios sobre el lugar u hora de realización, contenido, etc. Pero sí sobre las instrucciones dadas, por lo que hubo que flexibilizarlas para adaptarlas a cada candidato, asegurándose de que las entendían antes de comenzar a hacerlas. Ya que leídas tal y como se habían diseñado inicialmente, algunos candidatos voluntarios manifestaron que a veces eran insuficientes y otras demasiado formalmente expresadas. Por ello, se pensó en dejarlas como ideas para el evaluador y darle permiso para que el planteamiento de cada tarea fuera abierto y se adaptara al estilo de cada candidato, pudiendo ser explicadas de una manera más coloquial.

Algunos estudiantes manifestaron que el test completo (con las dos partes, la de la OPI de ACTFL y las tareas de nuestro test del MCER) parecía largo, otros que el tiempo se les pasaba rápidamente.

b. En cuanto a la naturalidad de las tareas, la mayoría de los 37 estudiantes a los que se preguntó al respecto manifestó que eran bastante o muy naturales, solían realizar cosas similares en su lengua o no les parecía raro tener que hacerlas en alguna ocasión.

Las tareas que se salen de esta consideración de ser tareas naturales son:

- La tarea 2 (explicar una receta de comida).
- La 5 (opinar y argumentar sobre una guerra).
- La 8 (contar la historia que se muestra en un cómic).

La tarea 2, porque había un 15% de candidatos que manifestaba su total incapacidad para explicar una receta ya que no sabía hacer ninguna comida, salvo una hamburguesa o pasta.

La tarea 5 a, porque había otros candidatos que manifestaban que no sabían nada respecto de la guerra del Líbano e Israel. Y la 5 b porque había algunos candidatos que manifestaron que en su país era un tema muy delicado del que no solían hablar casi nunca. Estos eran todos americanos, pero no todos los americanos opinaban igual.

Y la tarea 8 se vio como algo muy artificial, obteniendo solamente entre 0 y 2 puntos de naturalidad (en una escala sobre 10). Solo una persona dijo estar muy familiarizado con la lectura y explicación de cómics entre sus amigos y le otorgó un 10.

Sin embargo, sí dijeron que estaban muy familiarizados con tareas de este tipo en el aula de lengua y por esa razón se mantuvo la tarea en el test final.

c. En cuanto al grado de emoción y reacciones que les provocaban y si en alguna medida eso iba a ser un problema para la realización de la tarea, contestaron que no lo consideraban duro, ni siquiera ante la tarea de hablar de la guerra. Aunque esto luego en la realidad no fue así, ya que algunos no pudieron hablar por temor o estrés.

En la realización del test, ya una vez terminada la fase de pilotaje con voluntarios, efectivamente, las tareas 2 y 5 b presentaron algún problema (no la 8 ni la 5 a). Hubo candidatos que manifestaron no poder realizar la 2 por carecer de conocimientos de cocina. Y algún candidato que, aunque manifestó estar dispuesto a hablar y opinar sobre la guerra de Irak sin problemas, luego se emocionó mucho y no pudo terminar adecuadamente (expresó que tenía amigos en Irak y se echó a llorar).

La tarea 9 sí planteó problemas²¹⁸ y se decidió eliminarla, ya que hubo muchos candidatos que mostraron ser incapaces de hacerla, incluso en su lengua materna.

²¹⁸ En el Apéndice 8 se explica esta tarea.

d. En cuanto al tipo de lenguaje que generaban, por ser uno de los elementos más importantes del constructo del test, vamos a explicarlo en cada tarea por separado, en los siguientes subapartados.

3.6.2- a) Validación de la tarea 1: “Trabajo de canguro”

No hubo problemas ni opiniones negativas respecto a la tarea. Preguntados sobre la naturalidad de la tarea, la puntuación obtenida fue de 9 puntos sobre 10 y todos ellos creían que era un trabajo natural al que podían enfrentarse sin problemas en su lengua materna.

Tal y como era de esperar, la tarea produce la interacción requerida para la calificación con la escala de *Intercambio de información* y logra muestras de lengua similares a las que produce la actividad en la vida real, con limitaciones de vocabulario pero sin que nada de ello invalide el diseño ni los objetivos propuestos.

En las prácticas que se realizaron con los estudiantes, el grado de compromiso con la tarea, la riqueza y complicación de los objetivos y el lenguaje utilizado fue incrementándose en niveles altos. En los niveles bajos los candidatos se limitaron a una negociación básica y a preguntar las características del trabajo, las personas y el entorno del trabajo sobre el que estaban hablando, así como otras cosas normales de ese tipo de interacción: los horarios y actividades que se requieren para realizar el trabajo, precios mínimos, etc. Y en niveles más altos (la tarea no está pensada para niveles altos pero se probó con todo tipo de candidatos), se llegó a una interacción con características cercanas a las de un nativo, negociando y discutiendo todo tipo de condiciones e incluso intentando imponer condiciones propias.

En la vida real -y entre nativos- se puede producir una mayor complejidad de conceptos y temas llegándose hasta la negociación total de las condiciones de trabajo, cosa que en el test no se espera ni se pretende, salvo en candidatos de niveles altos (como efectivamente -en las pruebas- se llegó a producir).

3.6.2- b) Validación de la tarea 2: “Receta de comida”

No hubo problemas ni opiniones negativas al respecto. La puntuación obtenida fue de 9 puntos y todos opinaron que era una tarea natural a la que podían enfrentarse sin problemas en su lengua materna. Únicamente hubo dos personas que dijeron que eran incapaces de hablar de comidas, incluso en su lengua materna, porque solamente sabían cocinar pasta.

Respecto al lenguaje que provoca en las pruebas con candidatos de varios niveles, se aprecia una interacción menor de la esperada en esta situación en la vida real, en cuanto a los procedimientos para explicar cómo hacer una comida y a la necesidad de cooperar entre los interlocutores para entender y aclarar los conceptos y asegurarse de que se transmiten adecuadamente.

En el cumplimiento de la tarea durante el test, la interacción puede limitarse a la mínima información y, en efecto, se nota que los estudiantes que la intentan realizar carecen del vocabulario relativo a las instrucciones (verbos específicos para cada paso que se realiza), a los aparatos que se deben usar o a los nombres de las comidas. Y eso hace que las explicaciones carezcan de detalle, lo que es más evidente en niveles bajos.

3.6.2- c) Validación de la tarea 3: “Opinión sobre actividad cultural”

Esta tarea presenta una gran variedad de posibilidades de realización tanto en la vida real como en la actividad del test, pero no se encontraron diferencias significativas entre el lenguaje usado en la vida real y el usado en el test.

A veces los estudiantes tardaban en centrarse en una actividad cultural concreta o intentaban evitar el asunto por interpretarlo como algo demasiado culto o complejo. El evaluador debía insistir y ayudar a encontrar el tema sobre el que charlar abriendo el campo a todo tipo de manifestaciones culturales o de diferencias en los modos de vida y civilización: no solamente actos cultos, como teatro o espectáculos, sino cine, conciertos u otros aspectos de la vida diaria que pudieran ser interesantes desde el punto de vista cultural, incluyendo actividades informales u otros elementos festivos y cotidianos diferentes entre las culturas.

3.6.2- d) Validación de la tarea 4: “Explicar ciertos aspectos del trabajo”

No se plantearon problemas al respecto sobre la naturalidad (8 puntos sobre una escala de 10) ni sobre la posibilidad o dificultad de realización de la tarea.

Es fácil entender que la responsabilidad de explicar el trabajo a un colega para que te sustituya es grande en la vida real y requiere un esfuerzo y unos objetivos que son difícilmente conseguibles en un test. Dado el carácter imaginario de la situación se aprecian diferencias entre la vida real y el test basadas en la mayor cantidad de elementos de contenido pragmático que surgen en la vida real ante la necesidad de asegurarse de que se entienda todo bien por la importancia del tema requerido. No obstante el objetivo marcado en la escala de habilidad correspondiente se puede cumplir sin necesidad de tal compromiso y la tarea puede funcionar perfectamente en el test.

3.6.2- e) Validación de la tarea 5: “Argumentar sobre la guerra...”

La tarea tiene dos opciones -como hemos visto, sobre Líbano o sobre Irak- porque se preveían algunas dificultades desde el primer momento. En las opiniones pedidas a los estudiantes, preguntados sobre la naturalidad, le dieron 7 puntos. Y solamente hubo 4 estudiantes, de 37, que manifestaron que tendrían dificultad para realizarla por ser un tema sensible y porque decían que en su cultura no estaban acostumbrados a tener conversaciones sobre este tema. Entre los europeos no se planteó ningún problema al respecto. Sin embargo, algunos americanos que no pensaban que tendrían problemas para hablar del tema (Irak), cuando hicimos pruebas con el test se mostraron incapaces de hablar.

El tipo de lenguaje utilizado en el test y en la vida real puede ser similar, salvo en el grado de discusión -bajo o a veces inexistente- dado en el test. No obstante, para cumplir la tarea es suficiente con expresar opiniones y ciertos argumentos sin esperar una discusión ni una confrontación de ideas.

3.6.2- f) Validación de la tarea 6: “Explicar el viaje o experiencia...”

No hubo ningún comentario problemático al respecto de esta tarea en las pruebas con estudiantes extranjeros. Respecto a su naturalidad las marcaron con 10 puntos sin discusión.

El tipo de lenguaje utilizado en el test y en la vida real puede ser igual, salvo el interés que en la vida real puede surgir -con determinados interlocutores- por parecer cultos, divertidos, etc. interés que no se genera en el test.

3.6.2- g) Validación de la tarea 7: “Explicar las ventajas de estudiar...”

El tipo de lenguaje utilizado en el test y en la vida real puede ser igual, salvo que en la vida real quizás se dé la necesidad de convencer realmente al interlocutor, necesidad inexistente en el test.

No hubo ningún comentario problemático respecto a esta tarea.

3.6.2- h) Validación de la tarea 8: “Explicar un cómic”

Casi todos los estudiantes manifestaron que les parecía una tarea poco natural y no veían en su vida situaciones en las que realizaran nada parecido. Y muchos creían que tendrían problemas para expresar adecuadamente todo lo que se les ocurriera y al tener que usar el vocabulario específico que en él aparece.

La puntuación sobre naturalidad fue de solamente 2 puntos, excepto para un estudiante.

Sin embargo, el lenguaje surgido en las prácticas se adecuaba a lo esperado: descripciones de las situaciones, opiniones, comentarios, detalles más o menos interesantes sobre los dibujos, con una variedad bastante grande, dependiendo de lo que el candidato veía en la historieta y lo que se le ocurría al respecto.

3.6.2- i) Validación de la tarea 9: “Indicar el camino en un plano”

Los estudiantes a los que se pidió su opinión sobre la naturalidad de las tareas dijeron que era natural y que se hacía sin problemas aparentes, con 8 puntos sobre 10, excepto dos que afirmaron ser incapaces incluso en su propia lengua materna.

Sin embargo, en las prácticas que se hicieron con la tarea, mostraron un gran desconocimiento de las calles y direcciones de la ciudad y eran incapaces de realizar la tarea o cuando lo lograban era utilizando un lenguaje muy simple y poco informativo, limitándose al uso de verbos y marcadores de dirección básicos, sin otros elementos de referencia ni informaciones adicionales.

Por ello y por considerarla difícil para determinados estudiantes, y más al vivir en una ciudad nueva, finalmente se desechó incluirla en el test²¹⁹.

3.6.3 Validación del test en su conjunto y de su fiabilidad

Como ya se ha visto en §2.3, la validación de una prueba es necesaria para averiguar si esta sirve para lo que se pretendía (*validez de la prueba*), si repetida daría los mismos resultados (*fiabilidad*) y si se puede llevar a efecto y administrar entre candidatos sin problemas de variación de resultados (*viabilidad*). Entonces ya se comentaron los distintos tipos de validez existentes así como los métodos de averiguarlas.

Vamos a terminar la explicación del nuevo test con la explicación de los principales elementos de validación general empleados y ya comentados en el capítulo 2. Para ello damos a continuación -en forma de esquema y en cuadros para cada componente- los datos sobre fiabilidad de la prueba, validez de constructo, autenticidad, interactividad, impacto y uso práctico, tomados de las listas de validación y fiabilidad de Bachman / Palmer²²⁰.

De sus análisis se puede ver que:

- El test es fiable en su uso²²¹.

²¹⁹ Véase la tarea en el Apéndice 8.

²²⁰ Seguimos, en este proceso de validación, las preguntas y conceptos ya explicados en §2.3, tomadas de Bachman / Palmer (1996: capítulo 7).

²²¹ Este dato también se obtuvo de los análisis estadísticos a que se sometieron los resultados de las entrevistas y de la correlación entre sus resultados y los de la OPI. Véase el capítulo 4 y Apéndice 10.

- Ha sido descrito de forma que no penalice la actuación de determinados candidatos.

- Su constructo está bien definido y es suficiente como para evaluar las habilidades que se pretende.

- Las escalas que se han usado para definir la lengua usada en las tareas sirven para calificar la actuación de los candidatos y hacer las inferencias suficientes como para dar una adecuada descripción de sus habilidades.

- El tipo de lenguaje que provocan las tareas es auténtico.

- Las tareas no están sesgadas hacia ninguna característica especial, de los candidatos, ni hacia otros aspectos como conocimientos culturales requeridos o necesidad de habilidades particulares.

- En su diseño se han tenido en cuenta las habilidades y características de los candidatos para que todos puedan hacer el test en igualdad de condiciones.

Nosotros hemos realizado nuestra propia validación del test, cuyos datos en fase de pilotaje acabamos de exponer y cuyos datos generales van a continuación. El Consejo de Europa, en el documento para relacionar exámenes con el MCER del que se trató en §2.7.5, propone una serie de pasos para demostrar que un test además de ser válido se relaciona con el Marco. Estos también los hemos rellenado y figuran en el Apéndice 4.

3.6.3- A) Fiabilidad

En nuestra investigación todas las entrevistas las ha realizado un solo entrevistador entrenado, por lo que las posibilidades de variación de resultados en función de la falta de homogeneización de criterios entre evaluadores es inexistente.

El cuadro siguiente apunta lo principal en cuanto a otros elementos que pueden hacer que los resultados del test varíen, porque este se haga de forma diferente con cada candidato.

Remitimos a §4.5 para los datos concretos de su uso efectivo y los análisis estadísticos efectuados.

Análisis de la fiabilidad en la forma de hacer el test (Bachman / Palmer, 1996: 150):

1. ¿Cambian las características del lugar en que se realiza el test de una vez a otra?	No en aspectos fundamentales. Calidad totalmente adecuada.	Pueden usarse diferentes aulas, aunque todas ellas son similares y disponen de los elementos necesarios para realizar la entrevista. En algunas el nivel de sonido externo es mayor que en otras pero no se han observado elementos distorsionantes.
2. ¿Pueden cambiar las características del test o de sus partes o hay diferentes formatos del test?	Calidad totalmente adecuada. Los cambios son los necesarios según el diseño y características del test.	Los cambios entre las partes y tareas cambian de acuerdo con las características definidas en los planteamientos del test y las tareas. Cada estudiante -según su nivel inicialmente considerado y el obtenido en la OPI- realiza las tareas predeterminadas para ese nivel. Los cambios son los apropiados según el diseño del test.
3. ¿En qué medida cambia el input dado de una tarea a otra, de una parte del test a otra o en diversas administraciones del test?	No en aspectos fundamentales. Calidad totalmente adecuada.	Las instrucciones son similares pero se adecuan al nivel de comprensión del candidato asegurándose en todo momento de que las entiende. Se presentan de forma personal a cada uno. Si el candidato en su respuesta muestra no haber entendido o no responde adecuadamente, el evaluador clarifica la tarea para que el candidato la realice si tiene el nivel suficiente. Esto que es habitual en la OPI de ACTFL se ha trasladado también a cada tarea del test del MCER
4. ¿En qué medida varían las características de la respuesta esperada de una tarea a otra en diversas administraciones del test?	Calidad totalmente adecuada.	La respuesta esperada es acorde con las propuestas y planteamientos del test y cuando el candidato no presenta tales características el evaluador tiene el deber de plantear las preguntas para lograr que surja el discurso esperado o demostrar que su carencia es por falta de nivel, habilidad lingüística o dominio.
5. ¿En qué medida cambia la relación entre el input y la respuesta de una forma injustificada en diversas administraciones del test?	Calidad totalmente adecuada.	La respuesta es calibrada a medida que se obtiene. Se busca que se adecue a lo esperado o se persigue su cumplimiento o la búsqueda de pruebas de que este es imposible debido a falta de dominio.

3.6.3- B) Validez de constructo

En el pilotaje realizado con expertos expuesto en §3.6.1, ya se obtuvieron datos sobre este aspecto. Se detallan a continuación otros aspectos de la validez de constructo.

Análisis de la validez de contenido y constructo del test:

6. ¿Está la habilidad lingüística definida claramente y sin ambigüedades?	Calidad totalmente adecuada.	En la definición del constructo se han establecido varias partes diferenciadas según la tipología de discurso y funciones que se han planificado y según el grado de dominio de la lengua utilizada al realizarlas. Esto se usa como base para el sistema de calificaciones. Las pruebas con expertos atestiguan estos extremos.
7. ¿Es la habilidad lingüística medida en el test relevante y adecuada para el propósito del test?	Calidad totalmente adecuada.	El constructo se ha consultado con profesores expertos y responde a lo esperado de acuerdo con las escalas de dominio y habilidad que se propone medir. Cada tarea se plantea obtener un tipo de discurso determinado y las pruebas realizadas en pre-test así lo muestran suficientemente.
8. ¿En qué medida las tareas reflejan la definición del constructo?	Calidad totalmente adecuada.	Las tareas están diseñadas para obtener un tipo de discurso determinado y en cantidad suficiente como para probar su existencia o imposibilidad. Y esto se ha demostrado ya en las pruebas con candidatos en fase de pre-test.
9. ¿En qué medida el sistema de calificaciones refleja la definición del constructo?	Calidad totalmente adecuada.	Cada tarea se mide de acuerdo con su definición de habilidad lingüística y de dominio establecido en el constructo. Cada tarea se diseñó teniendo en cuenta unas habilidades concretas que debía provocar - extraídas de las escalas del MCER- y cuya existencia en el discurso de cada candidato es calificada mediante esas mismas escalas.
10. ¿Sirve el sistema de notas y calificaciones para hacer las interpretaciones sobre la habilidad de los candidatos?	Calidad totalmente adecuada.	Las habilidades pueden extraerse directamente del sistema de calificaciones puesto que estas son la muestra de lo que puede hacer cada candidato en su nivel.
11. ¿Qué características del lugar pueden llevar a los candidatos a actuar de forma diferente?	Ninguna.	Todos los candidatos están cómodos en el lugar y familiarizados con él y con el trabajo a realizar. No hay ningún elemento especial que interfiera.
12. ¿Hay características en la forma del test que pueden llevar a los candidatos a actuar de formas diferentes?	Ninguna.	Las instrucciones, partes, tareas y actitud del evaluador están diseñadas al nivel necesario para cada candidato. Ninguno sale penalizado de ellas. Ningún candidato es más hábil que otro en relación a este aspecto ni obtiene una ventaja o desventaja. El tiempo, la estructura del test o el método de calificaciones no favorecen de forma obvia a nadie.

13. ¿Qué elementos del input pueden llevar a los candidatos a actuar de formas diferentes?	Ninguno sustancialmente.	Las instrucciones que se dan en las tareas se adaptan al nivel y necesidades de los candidatos. Si cada uno actúa de forma diferente se debe a sus carencias de habilidad, no por la forma de actuar el entrevistador.
14. ¿Qué elementos de la respuesta esperada pueden causar diferencias en la actuación de los candidatos.	Ninguno sustancialmente.	Los únicos son los necesarios y considerados en el constructo que se va a medir: en el grado de dominio y habilidad demostrados en la realización de las tareas y funciones.
15. ¿Qué elementos de relación entre el input y la respuesta esperada pueden hacer variar la actuación de los candidatos?	Posiblemente la experiencia personal de los candidatos.	Si la experiencia de los candidatos al responder a las tareas es insuficiente, esta puede lastrar sus respuestas inicialmente pero el evaluador debe perseguir mediante otras preguntas el objetivo a evaluar y el constructo definido.

3.6.3- C) Autenticidad

Hemos tratado este tema al explicar las pruebas de validación del lenguaje que provocan las tareas (§3.6.1) junto a otras opiniones que se pidió a expertos y candidatos voluntarios (§3.6.2). Se detallan a continuación otros aspectos relativos a la autenticidad²²².

Análisis de la validez aparente y autenticidad del test:

16. ¿En qué medida la descripción de los objetivos de las tareas en su ámbito de actuación real incluye información sobre localización, input, respuesta esperada, y relación entre input y respuesta?	Completa.	La descripción de las tareas en su uso en contextos reales se encuentra en §3.4 e incluye suficiente información sobre la localización, input, respuesta esperada y relación entre input y respuesta.
17. ¿En qué medida las características de las tareas en su lengua real se corresponden con las de las tareas del test?	Corresponden bastante fielmente.	El análisis de las correspondencias se encuentra en §3.6.1 y §3.6.2. De ellos se deduce que hay una suficiente correspondencia salvo en el

²²² El concepto se ha explicado en su forma teórica en §2.2, desde el punto de vista de Bachman / Palmer, y §2.3.1 en general.

		lugar en que se realizan (un aula de lenguas) y en que se fingen unos papeles entre el evaluador y el candidato.
--	--	--

3.6.3- D) Interactividad

En §2.3.4, al hablar de la validez de respuesta, se presenta este aspecto de forma teórica; y en § 3.6.2, al explicar el pilotaje realizado con candidatos voluntarios sobre la validez de respuesta de cada tarea del test, de forma práctica.

A continuación detallo otros aspectos de la validez de respuesta y de la interactividad del test en su conjunto.

Análisis de la validez de respuesta e interactividad del test:

18. ¿En qué medida las tareas presuponen un conocimiento del tema y en qué medida suponemos que los candidatos lo poseen?	Aceptable para los propósitos de la prueba.	Las tareas se basan en su propia experiencia y habitualidad y naturalidad para hablar de los temas propuestos. Solamente en las tareas en que hay que hablar de recetas de comidas se plantearon problemas, las de viajes y explicación de actos culturales, pueden llevar a determinados candidatos a no hablar por falta de experiencias pero no en los casos de los estudiantes usados en las pruebas.
19. ¿En qué medida se ha considerado al candidato y sus características para diseñar las tareas?	Aceptable para los propósitos de la prueba.	Como ya se ha explicado en §3.6.2 todos los candidatos responden a unas características que permiten ser usuarios de la prueba en igualdad de condiciones.
20. ¿En qué medida las características de las tareas las pueden realizar los candidatos?	Aceptable para los propósitos de la prueba.	Salvo lo apuntado en la pregunta 18, no hay aspectos que impidan a ningún candidato la realización de las tareas.
21. ¿En qué medida se debe utilizar el conocimiento lingüístico para desenvolverse en las tareas?	Amplia gama.	Las tareas implican el uso de muchos elementos lingüísticos. Aunque cada una está diseñada con un objetivo determinado, el resultado final es útil para calificar las capacidades lingüísticas de los candidatos en la medida en que la escala del MCER lo sea.

22. ¿En qué medida las tareas se emplean para demostrar el conocimiento de funciones diferentes a la mera demostración del conocimiento y posesión de una habilidad lingüística?	Intrascendente.	La realización de las tareas implica el uso de conceptos culturales. No obstante el sistema de calificación se centra en los aspectos lingüísticos de la habilidad y no en los conocimientos del candidato respecto al contenido.
23. ¿En qué medida son las tareas interdependientes?	Son independientes.	Las tareas son independientes totalmente. Cada tarea se evalúa de acuerdo a sus objetivos concretos y con su escala determinada.
24. ¿Cuánta tiempo se da para el uso de estrategias de planificación?	Aceptable.	El habitual que existe ante estas situaciones y tareas en ámbitos naturales. Las estrategias están incluidas entre los elementos a evaluar en la OPI y en parte se consideran en las demás tareas.
25. ¿Contienen las tareas algo que pudiera provocar una respuesta afectiva positiva o negativa que afectara al desarrollo de las tareas?	Muy poco.	En el diseño de las tareas se ha tenido en cuenta la opinión de usuarios similares. No obstante una tarea presenta posibilidades afectivas negativas y se ha preparado una tarea alternativa similar como opción alternativa. (Esto se explica en §3.6.2- e).

3.6.3- E) Impacto

Aunque este extremo no es relevante para nuestra investigación pues el test diseñado se usa para corroborar o matizar nuestra tesis, comentamos algunos aspectos de su posible impacto porque es interesante para la comprensión del constructo de todo test.

Análisis del impacto del test:

26. ¿En qué medida cambiará el test las concepciones que el candidato tiene sobre la lengua, la sociedad, el uso de estrategias o el conocimiento?	Razonablemente.	Este extremo no es relevante para nuestra investigación. El tipo de tareas y el feedback recibido le pueden orientar en una manera de aprendizaje y concepción de la lengua más funcional que la tradicional. Además puede orientar el desarrollo de las motivaciones de aprendizaje y de producción oral.
--	-----------------	---

27. ¿En qué medida se tuvieron en cuenta las opiniones y sugerencias de los candidatos en el diseño de las tareas?	Razonablemente.	En el pilotaje al que se sometieron las tareas se analizaron y discutieron con estudiantes similares y se atendió a sus demandas para redefinir las tareas.
28. ¿En qué medida se da a los candidatos un feedback claro, significativo, relevante e inteligible?	Razonablemente.	Este extremo no es relevante para nuestra investigación. No obstante a algunos candidatos se les dio el resultado según ambos tipos de test, con una fotocopia de los niveles de la OPI y una valoración de su habilidad según el MCER. Si el test fuera usado normalmente se darían estos resultados de forma independiente y explicando el alcance de cada destreza y las limitaciones de cada candidato en los aspectos más débiles de su actuación.
29. ¿Se aplican los criterios y procedimientos igualmente a todos los candidatos?	Totalmente.	Este test no se utiliza para otras cosas (promoción, etc.) y sus resultados y procedimientos son los mismos para todos.
30. ¿El sistema de calificaciones es apropiado para tomar las decisiones que se van a tomar con el test?	Totalmente.	En su uso normal este test se usaría para averiguar el nivel de los candidatos. No hay otros usos previstos. El sistema de calificaciones es relevante para el propósito de la investigación: averiguar el nivel de acuerdo con las escalas del MCER y compararlo con el nivel de una OPI.
31. ¿Están los candidatos informados del sistema utilizado para tomar las decisiones?	Razonablemente.	Al seleccionar a los candidatos se les informó. En las instrucciones se indica cuál es el objetivo del test y de las tareas suficientemente, como para que su realización sea encaminada en ese sentido.
32. ¿Se siguen efectivamente los protocolos y criterios?	Totalmente.	Sólo se usan los sistemas descritos para el test y las tareas.
33. ¿Hay relación entre la lengua evaluada y el tipo de lengua que se enseña en las clases de lengua?	Innecesaria.	La lengua evaluada pretende ser reflejo de la lengua real usada por el estudiante con independencia de lo que estudie o donde lo estudie. En principio no pretende tener relación ni impacto en otras tareas pedagógicas.
34. ¿Qué relación tiene las tareas y las características de las lecciones de lengua?	Independiente.	Inicialmente esta relación no existe ya que se mide la lengua con independencia de cómo se aprenda esta. No obstante puede tener un efecto de que los profesores realicen en el aula tareas similares de tipo real que se aproximen a las que se evalúan en el test si este les parece interesante.

<p>35. ¿Qué relación tiene el test con los valores y objetivos de los profesores o de las instituciones?</p>	<p>Indirecta.</p>	<p>El MCER pretende servir de incentivo para una discusión sobre muchos aspectos de la lengua, su enseñanza y evaluación pero es muy abierto respecto a ellos y sus metodologías y objetivos. En la medida en que este test contribuyera a ello -cosa que no es objetivo directo de esta investigación- tiene una incidencia relativa porque trata de sus temas y procedimientos. En la medida en que no interese directamente a los profesores o instituciones su incidencia será baja.</p>
<p>36. ¿Qué relación tiene el test con los valores y objetivos de la sociedad o el sistema educativo?</p>	<p>Grande.</p>	<p>Si aceptamos que el MCER es la referencia que se está utilizando realmente y de forma práctica en nuestra sociedad actualmente, su incidencia es total y muy importante. En realidad la incidencia depende del grado en que el MCER esté teniendo realmente en nuestra sociedad actual. Y del grado en que la sociedad y candidatos acepten estos valores, probablemente más que las propias instituciones y profesores.</p>
<p>37. ¿En qué medida los valores y objetivos del test coinciden o entran en conflicto con los de la sociedad actual?</p>	<p>De completo acuerdo.</p>	<p>No hay discrepancias ideológicas ni sociales al respecto.</p>
<p>38. ¿De qué manera afecta el uso del test a la sociedad o al sistema educativo?</p>	<p>Positivamente.</p>	<p>Hay buenas consecuencias potenciales debidas al concepto de lengua manejado y a su utilidad en consideraciones laborales y sociales. E igualmente para el sistema educativo y sus objetivos de aprendizaje de la lengua. No obstante este no es un objetivo directo de nuestra investigación.</p>
<p>39. ¿Cuál es la consecuencia positiva más importante o lo mejor que podría suceder a consecuencia del uso de este test?</p>		<p>Que profesores y estudiantes se plantearan el uso de la lengua real como objetivo de aprendizaje teniendo en cuenta la comunicación y la funcionalidad como lo principal.</p>
<p>40. ¿Qué es lo peor que podría desearse que pasara como consecuencia del uso de este test?</p>		<p>Que en el intento de ser comunicativo y funcional se descuidaran otros elementos de la lengua que son necesarios para llegar a este objetivo deseable y ello supusiera un lastre en el proceso de aprendizaje y enseñanza de las lenguas que impidiera logra los primeros objetivos.</p>

3.6.3- F) Uso práctico

Al igual que su impacto, este extremo no es relevante para nuestra investigación pero lo comentamos porque es interesante para el test en conjunto. De ello se ha tratado de forma teórica en §2.3.7.

Análisis del uso del test:

41. ¿Qué cantidad de materiales se necesitan en los procesos de diseño, pruebas y administración?	Están calculados y previstos.	todos y	Todo lo necesario es de fácil acceso y uso y no plantea problemas.
42. ¿Están los recursos disponibles durante los procesos de diseño, pruebas y administración?	Están calculados y previstos.	todos y	La elaboración, pruebas y administración de estas tareas no presentan ninguna dificultad de recursos.

3.7 Ficha del nuevo test del MCER utilizado en esta investigación

El test final utilizado en esta investigación se describe en la siguiente ficha siguiendo a Bordón (2006: 218)²²³:

	Ficha de características del test nuestro
Tipo de examen	Examen de dominio de habilidad oral
Nivel que define	Todos, según los 6 niveles del MCER
Características de los candidatos	Adultos extranjeros universitarios de cultura general media
Formato del examen	Entrevista de interacción con tareas de expresión e interacción, entre el candidato y el evaluador.
Áreas temáticas	Lo descrito en la explicación de cada tarea (§3.4):

²²³ En el Apéndice 1 hay otra ficha que explica este test siguiendo las características expuestas por Bachman / Palmer (1996).

incluidas	<ol style="list-style-type: none"> 1. Trabajo de canguro. 2. Receta de comida. 3. Opinión sobre actividad cultural. 4. Explicar ciertos aspectos de su trabajo para un colega que va a ayudarlo o sustituirle durante unos días. 5. Argumentar a favor o en contra de la guerra. 6. Explicar el viaje o experiencia más enriquecedora o interesante. 7. Explicar las ventajas de vivir y estudiar en el país. 8. Explicar un cómic.
Funciones comunicativas	Las descritas al explicar las tareas (§3.4): intercambio de información, argumentación en monólogo sostenido, desarrollo de narraciones y descripciones, entre otras.
Fases del examen	<ol style="list-style-type: none"> 1. Saludos al candidato y explicación de los objetivos de examen y de que se va a grabar en un casete digital. 2. una serie de tareas que se dan una seguida de otra, con una breve pausa para seleccionar la siguiente al terminar cada una. Entre una y otra tarea el entrevistador se limita a decir “Bueno, esto ya está. Ahora vamos a hacer... (explicación de la siguiente tarea)”. 3. Final. Se despide el entrevistador y queda con el candidato para ofrecerle los resultados de la evaluación.
Formato de las pruebas y partes	<p>Hablado, en español, con input oral salvo en un caso (fotocopia de un cómic).</p> <p>Interacción oral y expresión oral basada en actividades propuestas a los candidatos (ya descritas en §3.4). Duración de cada tarea: entre 5 y 10 minutos. Duración total 20 o 30 minutos aproximadamente.</p>
Evaluación	Mediante las escalas de habilidad y dominio del MCER ya señaladas y explicadas en sus apartados. Con evaluadores de lengua formados para conocer, interpretar y aplicar las diversas escalas que se usaron (<i>Interacción oral en general, Monólogo sostenido: argumentación, Desarrollo de descripciones y narraciones, Coherencia y cohesión, Fluidez oral, Competencia lingüística general, Riqueza de vocabulario, Dominio de vocabulario, Compensación, Control y corrección, Corrección gramatical</i>). De ello se ha tratado en §3.3.5.
Experimentación	El examen se experimentó con candidatos similares y con expertos, como se ha señalado en §3.6 para comprobar la validez de las pruebas
Retroalimentación	Tras la fase de pilotaje se subsanaron los fallos encontrados y se desecharon las actividades que no daban resultados efectivos

CAPÍTULO 4

ANÁLISIS DE LOS RESULTADOS DE LAS ENTREVISTAS ORALES

Este capítulo comienza explicando la formación del corpus de entrevistas utilizadas en la investigación (§4.1).

Luego presenta tres partes bien diferenciadas:

Primero (§4.2) se exponen las carencias y problemas que han encontrado los evaluadores que utilizaban las escalas del MCER al calificar las entrevistas, con sus ambigüedades y el sistema de ponderación de resultados utilizado.

Después (§4.3 y §4.4) se presentan los resultados de las entrevistas hechas a los candidatos con los dos modelos de exámenes utilizados: el nuestro, diseñado según el MCER, y el de ACTFL; y se comentan algunas entrevistas. Damos los resultados y comentamos las calificaciones dadas por los equipos de evaluadores; en el caso del MCER, según sus propios argumentos. Los casos raros, como el de ser calificado como B2 siendo solamente intermedio medio, cuando normalmente en ese nivel en se han situado candidatos de nivel intermedio alto o avanzado, los explicamos en §4.4.

En tercer lugar (§4.5) se presentan los análisis estadísticos realizados con los resultados de las entrevistas. Todos estos análisis y comentarios, permiten, finalmente, extraer las conclusiones generales de equiparación entre los niveles de la OPI y los niveles del MCER, que, junto a otras, se exponen en el capítulo 5.

En este capítulo 4 se seleccionan candidatos y muestras de sus discursos pero no se hacen las transcripciones de sus entrevistas. Todas las entrevistas, las de los 28 candidatos, se acompañan en un DVD.

4.1 Corpus de entrevistas utilizadas en esta investigación

Las personas entrevistadas han sido 35, pero sólo se pudo utilizar un corpus de 28 personas con 56 entrevistas, dos por persona. Las restantes no cumplían algún requisito para su calificación. Estaban incompletas o con problemas para ser evaluadas.

Para componer el corpus de investigación se realizó una preevaluación de las entrevistas del modelo OPI, según los criterios de ACTFL. Así, primero se procedió a decidir qué entrevistas de ACTFL eran calificables, ya que los candidatos que no eran evaluables según ACTFL no podían entrar en el corpus final para comparar sus resultados con los de nuestro test del MCER. En el MCER no existen estos criterios estrictos y el test que se diseñó para evaluar a los candidatos no invalida, en principio, a ninguna entrevista.

Para poder obtener muestras de lengua que abarcaran todo el campo de las escalas del MCER y de ACTFL se había seleccionado a estudiantes de todos los niveles del CLM. Pero en esa primera preevaluación no oficial de las OPI no se obtuvo ninguna entrevista del nivel inicial bajo y sí, en cambio, bastantes de otros niveles.

Una vez determinado el corpus de candidatos válidos para la calificación con la OPI, 29 en ese momento, se procedió a calificar a los candidatos de ese corpus en sus dos entrevistas: nuestra entrevista del MCER, por un lado, con un equipo que se formó al efecto; y las OPI, que se enviaron a Estados Unidos para que se calificaran por evaluadores oficiales de la organización. En total se enviaron 29, pero una la desecharon por no evaluable. Por eso el corpus final es de 28 candidatos.

Hay, de ACTFL, tres grabaciones de inicial medio, tres de inicial alto, tres de intermedio bajo, cinco de intermedio medio, tres de intermedio alto, cuatro de avanzado bajo, dos de avanzado medio, dos de avanzado alto y tres de superior; y del MCER, dos de menos de A1, cuatro de A1, dos de A2, cuatro de B1, once de B2, una de C1 y cuatro de C2.

El corpus de 28 entrevistas es el siguiente (damos el número de la entrevista y el nombre ficticio que el candidato usó durante las entrevistas, el nivel obtenido con el

MCER y su nivel en la OPI. En la columna de la derecha damos el número de orden usado para los análisis estadísticos²²⁴).

En el capítulo 5, en las conclusiones, se dan estos mismos resultados, explicados cuantitativamente, con análisis estadísticos.

Cuadro 4.1 a. Corpus de entrevistas y resultados de las calificaciones²²⁵

Número y nombre del alumno	Nivel según MCER	Nivel * según ACTFL	Numero en el análisis estadístico
34. Sachiyo	-A1	NM	1
35. Kayo	-A1	NM	2
4. María	A1	NM	5
1. Biancé	A1	NH	3
3. Johanes	A1	NH	4
6. John	A1	NH	7
2. Stive	A2	IL	6
7. Carmen	A2	IL	8
5. Francesca	B1	IL	9
8. Juan	B1	IM	10
9. Boris	B1	IM	11
10. Lucia	B1	IL	12
15. María	B2	IM	13
19. Maya	B2	IM	14
14. Michelle	B2	IH	15
17. Jorge	B2	IH	16
22. Ana	B2	IH	17
16. Ana	B2	AL	18
18. Lea	B2	AL	19
20. Nina	B2	AL	20
21. Magda	B2	AL	26
25. Cuqui	B2	AM	21
30. Marta	B2	AM	27
29. María	C1	S	24
26. Patricia	C2	AH	22
Nat ²²⁶ 32. Fátima	C2	AH	28
Nat 31. María	C2	S	23
Nat 33. Camila	C2	S	25

*Siglas para los niveles de ACTFL: NM (*Novice Mid*): inicial medio. NH (*Novice High*): inicial alto. IL (*Intermediate Low*): intermedio bajo. IM (*Intermediate Mid*): intermedio medio. IH (*Intermediate High*): intermedio alto. AL (*Advanced Low*): avanzado bajo. AM (*Advanced Mid*): avanzado medio. AH (*Advanced High*): avanzado alto. S (*Superior*): superior.

²²⁴ Las tablas con los análisis estadísticos sobre estas calificaciones se encuentran en el Apéndice 10.

²²⁵ El Apéndice 11, el DVD con las grabaciones de las entrevistas, sigue la numeración de este cuadro, con el nombre ficticio y número de cada alumno-candidato.

²²⁶ Los candidatos 31, 32 y 33 son nativos de lengua española pero no españoles.

Disponemos de comentarios de los calificadores de nuestro test del MCER, pues se les pidió una calificación general del candidato, pero a su vez se les pidió que la argumentaran explicando el nivel obtenido en las tareas de acuerdo con las diferentes escalas que se tomaban en consideración para la evaluación.

Cada tarea se ha calificado mediante las escalas del MCER con las que se diseñó y con otras escalas que abarcaban otros aspectos importantes y a las que los evaluadores debían referirse en todo momento para razonar las calificaciones dadas. En §3.4 se han explicado las tareas utilizadas, sus objetivos y las escalas de las que se partió para su diseño y con las que se procedió a su calificación. Estas escalas se encuentran en el Apéndice 6.

No tenemos comentarios de la evaluación de ACTFL, salvo la calificación final. No obstante, una vez determinado el nivel, hemos intentado extraer ejemplos que muestren cómo es el discurso y las habilidades de los candidatos.

Los resultados de las calificaciones de ambas entrevistas muestran que los candidatos obtienen una calificación generalmente convergente con las calificaciones obtenidas por otros candidatos similares, pero no siempre. Por ejemplo, todos los candidatos que han obtenido inicial alto en ACTFL son A1 en el MCER, pero a su vez hay candidatos que son A1 en el MCER y en ACTFL son un nivel menos, inicial medio; o todos los candidatos que han obtenido avanzado bajo de ACTFL son B2 en el MCER, pero al mismo tiempo otros candidatos han obtenido en el MCER B2 y son en ACTFL un nivel menos, es decir, intermedio alto o un nivel más, avanzado medio (incluso hay dos casos que son solamente intermedio medio), por lo que el B2 alcanza desde el intermedio alto (y en algunos casos desde intermedio medio) hasta el avanzado medio.

La casuística de candidatos y niveles según la entrevista del MCER y de ACTFL es la siguiente:

Cuadro 4.1 b. Casuística de niveles.

MCER	ACTFL	Candidatos	Número de casos
Menos que A1	Inicial medio	34. Sachiyo 35. Kayo	2
A1	Inicial medio	4. María	1
A1	Inicial alto	1. Biance	3

		3. Johanes 6. John	
A2	Intermedio bajo	2. Stive 7. Carmen	2
B1	Intermedio bajo	5. Francesca 10. Lucia	2
B1	Intermedio medio	8. Juan 9. Boris	2
B2	Intermedio medio	15. María 19. Maya	2
B2	Intermedio alto	14. Michelle 17. Jorge 22. Ana	3
B2	Avanzado bajo	16. Ana 18. Lea 20. Nina 21. Magda	4
B2	Avanzado medio	25. Cuqui 30. Marta	2
C1	Superior	29. María	1
C2	Avanzado alto	26. Patricia N32. Fátima	2
C2	Superior	N31. María N33. Camila	2

En §4.5 se analizan estos datos para ver si son estadísticamente significativos, se explican y se sacan las conclusiones de todo ello.

4.2 Carencias y problemas detectados al calificar utilizando las escalas del MCER

Los problemas de la calificación con la escala de ACTFL, respecto a su fiabilidad, o consistencia entre evaluadores, se han señalado en §2.6.3. Ahora, queremos explicar los problemas con los que se encontraron nuestros evaluadores al proceder a calificar las entrevistas utilizando las escalas del MCER.

Ya hemos hecho referencia a la ambigüedad de sus escalas y a si es real la diferenciación en seis niveles o no (véase §3.1). Ahora vamos a ver si tales críticas persisten en la práctica, al proceder a evaluar las entrevistas y discursos de los candidatos; si, por el contrario, desaparecen los problemas y ambigüedades; o si surgen otros no esperados.

De los comentarios de los calificadores podemos sacar conclusiones respecto a estos y otros problemas que queremos agrupar y estudiar en dos bloques:

- A. La ambigüedad de interpretación de los descriptores (§4.2.1- A).
- B. El surgimiento de criterios formalistas en las calificaciones (§4.2.1- B).
- C. La ponderación de las notas (§4.2.1- C).

4.2.1- A) Ambigüedades de interpretación

La ambigüedad de interpretación de los descriptores de la que hablábamos en §2.7.4 y §3.1 es evidente y hace, a veces, difícil calificar el discurso de los candidatos, pero no lo hace imposible. Esto puede llevar a que, a veces, las calificaciones padezcan de argumentaciones débiles o poco consistentes; pero, en conjunto, estas son bastante homogéneas. Ya que gracias a la formación dada a los evaluadores y a la propia estructura del diseño del test y del sistema de ponderación (véase siguiente punto §4.5.1- B), todos logran dar una calificación similar²²⁷.

A pesar de esa consistencia interna de los resultados, los evaluadores manifiestan haber tenido problemas para decidir si un discurso o el resultado de una tarea, debe ser calificado como A2 o como B1, como B1 o B2, como C1 o C2; y las diferencias existentes entre los descriptores de nivel de C1 y C2, más que ninguna otra, parecen mínimas²²⁸. A estas dificultades y ambigüedades se une la carencia de descriptores de C1 y C2 en muchas escalas, lo que es un aprieto al intentar evaluar a un candidato diferenciando si es B2, C1 o C2: si un candidato muestra características de discurso de nivel B2, muchas escalas determinan que ya posee las mismas habilidades que otros candidatos de nivel C1 o C2, pues son las mismas.

En las escalas de *Intercambiar información* y de *Monólogo sostenido: argumentación*, sólo se dan descriptores de B2; en las de *Desarrollo de descripciones* y

²²⁷ La homologación mediante una correlación entre el test del MCER y el de ACTFL, también con un índice alto, del .95 (véase §4.5 para estos datos) da a entender que entre ambos tests se está midiendo algo similar, con lo que el constructo del test utilizado para evaluar mediante las escalas del MCER queda validado, o hablando estrictamente, que las inferencias que se deducen de él pueden ser válidas. La fiabilidad de los correctores es muy alta (véase capítulo 5 para estos datos estadísticos), pero la existencia de una alta fiabilidad no significa que el test mida el dominio de lengua que deseábamos; otros evaluadores formados en otra parte y de otra manera pueden llegar a conclusiones diferentes a estas y obtener igualmente una alta fiabilidad.

²²⁸ Esto es lo que preveíamos y así lo hemos señalado en el capítulo 1, §1.4

narraciones, faltan los de C2 y en la de *Compensación y Flexibilidad* los de C1. Además, en muchas ocasiones los descriptores de C1 y C2 son una mínima cuestión de matiz difícilmente identificable en la mayoría de los discursos. Esto lo acusan varias veces los calificadores al utilizar las escalas de *Fluidez oral*, *Coherencia y Cohesión*, *Competencia lingüística general* y *Expresión oral en general*.

Vamos a ver algún ejemplo de estos comentarios sobre la ambigüedad, dados por los propios evaluadores al calificar a los candidatos 32 (Fátima), 29 (María), 20 (Nina) y 5 (Francesca).

- Al evaluar al candidato número 32 (Fátima), quien es C2, se dice:

- *Según la escala de Desarrollo de descripciones y narraciones, es difícil saber si la alumna tiene un C1 o un C2 ya que el descriptor de C2 es el mismo que para C1. No tengo muy claro si se trata de uno o de otro, ya que la chica a veces tiene errores gramaticales típicos de los nativos. Cierto es que realiza descripciones y narraciones detalladas integrando varios temas, desarrollando aspectos concretos y terminando con una conclusión apropiada.*

- *La escala de Intercambiar información tampoco nos indica cómo saber si se trata de un C1 o un C2. Está claro que se trata de un nivel superior pero parece difícil discernir el C1 del C2.*

- *Su gramática es propia de un nativo, pero tiene errores que apenas se notan: dequeísmo, colocación de los elementos de la oración, etc.*

- Al intentar evaluar al candidato número 29 (María), que es C1, se dice:

- *Me han resultado las escalas muy ambiguas y no sólo he dudado durante las audiciones sino que también he seguido dudando después. Pero había que ajustarse a algo.*

- *En cuanto a la diferencia entre C1 y C2, aunque también hay ambigüedad, creo que el MCER es todavía menos claro en niveles como el B1, B2 y C1, porque se supone que el C2 ya tiene que parecer casi nativo, ¿no?; pero también considero que, como la ambigüedad pura empieza desde los descriptores de B1, pues ésta ya está presente hasta el final. Además en algunos descriptores de C1 te dice "igual que B2 pero un poco más", y con los de C2 pasa lo mismo ("igual que C1"), pero ¿se tendrán que diferenciar en algo, no?*

- Al evaluar al candidato número 20 (Nina), quien es B2, se dice:

- Según la escala de Argumentación no veo clara la diferencia entre B2.1 y B2.2: ambos descriptores son similares y la actuación de la candidata podría servir para ambos descriptores.

- Podría estar entre B2/C1 pues se detectan leves errores en la expresión de las frases, aunque son leves y esporádicos y no producen ambigüedad (“lo inconveniente”, “estoy en la misma dirección”).

- En general, hay un buen control gramatical, sin malentendidos. Dudo entre el B2 y el C1. El B2, al ser más descriptivo, ofrece más datos de análisis. El C1 detecta también un alto grado de control gramatical y dice que los errores son escasos, como en nuestro caso.

- Este tipo de comentarios no es exclusivo de los niveles altos. También al evaluar al candidato número 5 (Fancesca), quien es finalmente B1, se dice:

- Al calificar su discurso según la escala de Riqueza de vocabulario dudo entre A2 y B1.

- Al utilizar la escala de Compensación podríamos calificarla como B1, ya que “adapta” una palabra de su lengua materna y pide confirmación, pero realmente no pienso que “adapte”, sino que realmente la sustituye por la palabra en su lengua.

- Al utilizar la de Control y corrección tengo las mismas dudas que en el caso anterior, realmente no creo que realice un control y corrección, tampoco creo que sea consciente de la sustitución de palabras españolas por italianas.

- Al usar la de Corrección gramatical la situaría entre A2 y B1, inclinándome al primero.

4.2.1- B) La preponderancia de criterios formalistas en ciertas calificaciones

Hemos visto que todas las escalas señaladas han quedado mermadas para poder apreciar diferencias claras entre el discurso de C1 y C2. Aunque se pueden utilizar como guía, no sirven para decidir el nivel ya que carecen de descriptores o estos son muy ambiguos. Las escalas con las que finalmente se ha podido proceder a una diferenciación clara entre un discurso de C1 y C2 han sido:

- Riqueza de vocabulario.

- Dominio de vocabulario.

- Control y corrección.

- Corrección gramatical.

- Autoevaluación.

En estas escalas, las diferencias fundamentales que se muestran entre un discurso de nivel C1 y otro de nivel C2 se basan en cosas como:

- Si *“hay deficiencias” aunque “se logren superar con circunloquios” (C1 de Riqueza de vocabulario), o ya se han eliminado esas deficiencias y “se muestra buen dominio del repertorio léxico” (C2 de Riqueza de vocabulario).*
- Si *“el vocabulario es correcto” (C2 de Dominio de vocabulario), o todavía “hay errores aunque no importantes” (C1 de Dominio de vocabulario).*
- Si *“se sortean las dificultades sin que el interlocutor se dé cuenta” (C2 de Control y corrección) o se hace “volviendo a formular lo que se quería decir, aunque sin perder fluidez en el discurso” (C1 de Control y corrección).*
- O si *“los errores gramaticales son escasos” (C1 de Corrección gramatical) o “inexistentes” (C2 de Corrección gramatical).*

Por ello, su uso como argumentos de peso frente a la ambigüedad del uso de las otras escalas señaladas provoca una calificación con criterios mucho más formalistas que en una OPI. Esto se ha notado fundamentalmente en los niveles altos, ya que en su calificación adquieren mucho peso los criterios tradicionales de gramática y vocabulario, en detrimento de otras habilidades y funciones de expresión. Así, para observar diferencias entre los discursos de nivel B2, C1 o C2 se emplean, en muchas ocasiones, meros matices de corrección, nada sustancial que suponga la existencia de nuevas habilidades lingüísticas entre los diferentes niveles; solo cuestiones de grado.

Por ello, ante la falta de criterios claros surge uno nuevo y externo a las propias escalas, intuitivo: el recurso a la lengua nativa como comparación con la del candidato. Con esto se intenta dilucidar si su discurso es de nivel C1 o C2. Pero ello incurre en un error de criterio, que es peligroso pues puede invalidar la escala utilizada, convirtiéndola en una escala basada en normas y no en criterios, con la consiguiente variación que supone del concepto de lengua a evaluar (de esto se ha tratado en §2.5.1 y más específicamente en §2.6.3- E).

Esto se ve en que muchos de los comentarios que los evaluadores dan respecto a los niveles B2, C1 o C2 se basan en si tiene deslices o errores, más corrección o menos, o deficiencias léxicas; y sólo en menor medida se fijan en aspectos funcionales o de estructura del discurso. No es que no se utilicen otros criterios sino que los formales adquieren una preponderancia muy grande: una vez que se ha establecido mediante

otras escalas que un discurso es de más nivel que B2, el definitivo paso para decidir si es C1 o C2 es el formal.

Por ejemplo (a continuación presento cuatro ejemplos):

- La candidata 29

La candidata 29 (María), es calificada como C1 porque, según un evaluador:

- *Carece de suficiente corrección gramatical: la duda que nos cabe para concederle un nivel alto es la cierta confusión que observamos en el uso del pasado que en ocasiones hace:*

“Y ella también quería aprender tango y le he dicho...” (en vez de “le dije”).

“Y bueno, empecé a pensar por qué no vendría a Francia y se lo he dicho” (en vez de “y se lo dije”).

- *En la fluidez oral sólo se detiene para considerar las palabras más adecuadas.*

- *Por otra parte, teniendo en cuenta la escala de Competencia lingüística, también admitimos que el nivel de esta alumna está más cerca de un C1 que de un B2, puesto que se expresa con claridad y sin limitar lo que quiere decir.*

O porque, según otro evaluador:

- *Tiene un gran control en cuanto al vocabulario que pone en juego, pero a veces se ayuda de léxico que no es tan exacto en español para lo que ella pretende decir aunque es totalmente válido: “tragar todas las costumbres” o “pasar exámenes”.*

- *Normalmente suple sus carencias: primero dice “mirando la televisión” pero rápidamente repite la expresión de una forma más precisa: “viendo la televisión” o reestructura con mucha sutileza: “y resulta que el sueño, no sé, no es exactamente así”. Sin embargo, al final de la actividad no sustituye “roncos” por otra palabra o circunloquio a pesar de que ella sabe que no es correcto.*

- *Creo que se le daría el certificado del nivel C1, aunque en determinadas ocasiones no logra superar todos los descriptores de todas las escalas; sin embargo considero que presenta una gran fluidez oral.*

- La candidata 31

La candidata 31 (María), es calificada como C2, fundamentalmente porque:

- *Muestra un dominio amplio y fiable de un completo repertorio de elementos lingüísticos para formular sus ideas con precisión y sin limitaciones (Competencia lingüística general); presenta un amplio repertorio léxico relacionado con el trabajo a realizar (Riqueza de vocabulario); empleado con total consistencia (Dominio de*

vocabulario); *no muestra ningún tipo de duda a la hora de utilizar el léxico y las estructuras lingüísticas (Control y corrección); no comete errores gramaticales. Su monólogo tiene los rasgos del lenguaje oral de cualquier nativo, por lo que lo valoro como un C2 (Corrección gramatical); se muestra natural y sin dudas o esfuerzos (Fluidez).*

- La candidata 32

La candidata 32 (Fátima), también C2, lo es fundamentalmente con argumentos formales. Los comentarios de uno de sus evaluadores manifiestan tarea a tarea las carencias de las escalas para dar una buena calificación y se utilizan otras escalas que son sobre todo estructurales:

- Para evaluar la tarea de “explicar su trabajo”, la escala de Intercambiar información tiene descriptor sólo hasta B2.2, así que no tenemos referencia real sobre su nivel.

- Al calificar la tarea de “explicar una actividad o un viaje”, la escala llega sólo a C1, que cumple la alumna (realiza una descripción detallada integrando varios temas).

Finalmente, el evaluador tiene que echar mano de otros argumentos y explica que mantiene un consistente control gramatical.

Al intentar calificar la tarea de “explicar las ventajas y desventajas de estudiar en España”, sí utiliza un criterio más funcional:

- El estudiante desarrolla argumentos dando un énfasis apropiado a los aspectos importantes y apoyándose en detalles adecuados; se explica mediante un discurso amplio con fluidez natural, y se detiene sólo para buscar una explicación adecuada.

Y al calificar la tarea de “opinar sobre la guerra” dice: “En este caso también el alumno muestra tener un léxico muy apropiado relacionado con el tema, y se expresa sin tener que limitar lo que quiere decir”. En general, concluye:

- Al subir el nivel de las tareas no muestra dificultad en adaptarse al nivel requerido; ha sido siempre capaz de seleccionar las palabras más adecuadas al tema del que tenía que hablar, sin tener faltas gramaticales. Por esta razón lo calificamos con un C2.

- La candidata 33

Igualmente a la candidata 33 (Camila), se la califica según elementos gramaticales o léxicos, sin tener en cuenta lo que dicen realmente las escalas del MCER:

- El discurso de la alumna es totalmente natural y propio de un hablante nativo, ni siquiera en el acento o el tono observamos características que no son propias de un hablante que tiene el español como lengua materna. La competencia lingüística de la estudiante es perfecta (o al menos muy próximo de serlo); por ello encontramos

estructuras gramaticales construidas correctamente: concordancia y tiempos verbales adecuados (perfecta diferenciación entre el pretérito perfecto y el indefinido), un léxico variado (estoicas, matriarcal, aderezos) y adaptado a las necesidades comunicativas e incluso la entonación marca el énfasis necesario para transmitir lo que quiere. En cuanto al uso del español, queda bastante claro que estamos ante una muestra de español de América, confirmado mediante la utilización de expresiones propias de Latinoamérica (tomar la bebida, tomar fotos, ahorita...) y más concretamente, mejicanismos (mesero, ponchar).

Acabamos de ver cómo los evaluadores recurren a la norma nativa como criterio (se supone que el C2 ya tiene que parecer casi nativo, ¿no? Tiene errores gramaticales típicos de los nativos. Su gramática es propia de un nativo) olvidando los criterios de las escalas del MCER. Es un recurso que han comentado por escrito y oralmente muchos de los evaluadores. Cuando encuentran una soltura suficiente y una destreza de vocabulario y organización de ideas alta, argumentan con ese criterio. Esto es un error porque lleva a desentenderse de los verdaderos criterios de las escalas y a dar una evaluación diferente a la pretendida en el examen, sin perseguir las habilidades descritas en las escalas. Y no es un criterio que sirva para evaluar, no solo por no utilizar el criterio descrito en el MCER sino por las razones ya explicadas en §2.4.1. Podemos decir, además, que hay muchos modelos de lengua nativa y fácilmente podría ser que un hablante nativo no cumpliera con los criterios del C2 en determinadas escalas, quedándose en una calificación de su nivel de uso de la lengua de C1.

Este criterio, además, puede llevar a engaños: una vez que se ha decidido que el nivel se acerca al de un nativo porque no comete errores y tiene mucha fluidez, se decide calificarlo como de nivel C2 y ya es difícil sustraerse a esa apreciación ni aún en el caso de que en otras tareas el nivel sea bajo comunicativamente, deficiente en la organización del discursos o débil en el sistema de descripción, argumentación, etc. Si el candidato ya no se esfuerza en otras tareas y baja el nivel de su lengua, su calificación difícilmente se va a rebajar, salvo que aparezcan errores de gramática o vocabulario. El criterio se convierte entonces en la formalidad y corrección antes que en el dominio de las demás habilidades lingüísticas.

El efecto es similar a lo que ocurre cuando un evaluador se deja llevar porque conoce otras notas altas de un candidato y eso le influye para dar notas altas a las nuevas tareas aunque objetivamente no se aprecie tanto nivel en ellas.

Con ACTFL no pasa esto. Porque, aunque un candidato sea capaz de hablar con soltura y fluidez, e incluso sin errores de léxico o gramática, no se tiene una idea del nivel y se deben continuar las pruebas en niveles y preguntas de mayor complejidad para demostrar que posee más habilidades. La estructura del test es importante y determinante del nivel y la calificación final. Esta no viene determinada por la soltura y corrección de una tarea sino por el proceso del test completo y su estructura de dificultad creciente. Es, al ir subiendo el nivel de las preguntas y de las exigencias de dominio y habilidad requeridas para cumplirlas, cuando se comprueba el uso de determinados mecanismos y elementos de la lengua; no se puede tomar una decisión antes de pasar por todas las pruebas necesarias, en las que se debe demostrar la posesión de determinadas características o habilidades y el uso adecuado de cada elemento que se enjuicia.

Por ello, aunque un estudiante no cometa errores y parezca “nativo” (si queremos utilizar este argumento) en determinadas tareas, debe demostrar que mantiene esa habilidad en otras tareas más complejas que son, en avanzado, describir, narrar, explicar, comparar y opinar con detalle respecto a temas sociales en los que el protagonista sea ajeno a la experiencia del candidato; y, en superior, ser capaz de elaborar discursos abstractos, defender ideas y elaborar hipótesis. Ninguna ausencia de errores supone el superior por sí mismo, aunque sea necesario no mostrar errores sistemáticos para ser calificado como superior.

Este nos parece un sistema mucho más fiable y realista que impide que el evaluador se deje influir en las calificaciones, ya que en cada momento la fase de la entrevista tiene un sentido y un objetivo determinado y el entrevistador sabe esto y sabe qué está haciendo y lo que pretende del candidato con el manejo de las preguntas y de la interacción. Cada elemento, aunque la evaluación sea holística, se juzga independientemente y con objetividad.

4.2.1- C) Ponderación

El MCER tiene muchas posibilidades de uso en el aula, como diagnóstico de problemas o como presentación de resultados, bien en sí mismos, bien a través del portfolio. Pero nosotros hemos compuesto un test independiente del aula, un test de nivel global de lengua oral para certificación. Por ello necesitamos dar una nota global

de las habilidades generales que se miden mediante sus tareas. Para ello, ya se ha dicho al comienzo de este apartado, se pidió a los evaluadores que, tras escuchar y analizar las diferentes tareas y calificar el dominio de habilidades mostradas en ellas, se diera una nota final del candidato, de forma global.

Acordar esto y llegar a entender qué representaba la nota que se daba finalmente precisó de la explicación de algunos de los posibles aspectos de su interpretación y llegar a un acuerdo respecto a la forma de darla y su significado final²²⁹. Primero hubo que aclarar que se trata de una nota del dominio general que se posee efectivamente, es decir, que se trata de una nota para certificar un dominio y no para decidir a qué curso debería pasar el candidato si deseara estudiar más cosas de la lengua. Posteriormente, hubo que aclarar el sentido de los marcadores y submarcadores de nivel, A2.1 y A2.2, B1.1, B1.2, B2.1, B2.2 y C1.1 y C2.2. Y aún, en varias ocasiones, una vez obtenidos los resultados, hubo que aclarar y asegurarse del sentido que los evaluadores estaban dando con la calificación que ellos mismos notificaban de cada candidato.

Las preguntas, problemas o dudas que surgieron eran del tipo de que si se decide que un candidato es B1, ¿significa que debe asistir a cursos de lengua donde se trabajen y practiquen habilidades para lograr ser finalmente B1 o que ha superado las habilidades de este nivel y puede dedicarse a aprender las habilidades del nivel siguiente, B2? El MCER permite subdividir cada escala en nuevos niveles como B1.1 y B1.2, etc. y en algunas escalas él mismo ofrece estas subdivisiones. Por ello, si decimos que un candidato es B1.1 estamos diciendo que no es B1.2 y que, por lo tanto, no ha superado el B1. Entonces, con el mismo razonamiento, si decimos que es B1, sin más subdivisiones, quedaba la duda de si había superado el B1 o no.

Nos encontramos ante este problema en la forma en que los diferentes evaluadores se referían al nivel que tenía cada candidato, tanto durante las prácticas como, a veces, en las notas finales.

También es importante el problema de ponderar entre las diferentes notas obtenidas con las escalas utilizadas en cada calificación, para dar una nota final: si un candidato muestra una habilidad en unas tareas o con unas escalas (por ejemplo ser en *Dominio de vocabulario*, B1) y al calificar la misma tarea con otra escala obtiene otra nota (por

²²⁹ La ventaja de dar notas sobre destrezas aisladas es evidente pero no nos sirve para los objetivos de nuestro test. Al respecto podemos apuntar las palabras de Bordón (2006: 101): “al presentar las destrezas aisladas el evaluador puede comprobar exactamente el dominio de cada una en particular, ya que la combinación de destrezas puede acarrear resultados que no sean fácilmente interpretables”.

ejemplo en *Corrección gramatical*, B2), ¿cuál es el nivel finalmente, si necesitamos dar una nota global final, coherente e inteligible? ¿O cuál es su nivel final si en unas tareas obtiene una nota y en otras, otra?

El MCER tiene una virtud pedagógica, que ya hemos señalado en §2.7.1 y §2.5.3, pues permite la horizontalidad de las habilidades, pudiendo superarse un nivel en una determinada habilidad mientras no en otra; y parece muy posible poseer diferentes niveles en tareas diferentes o en determinados aspectos de ellas. Esto está bien para ámbitos educativos donde el profesor puede desmenuzar las habilidades y realizar actividades diferentes, viendo el desarrollo de sus estudiantes y dándoles un feedback apropiado a cada situación, para poder trabajar en las clases de lengua los diferentes aspectos de su aprendizaje de forma separada. Igualmente es positivo si se trata de hacer un portfolio y se pueden dar por separado diferentes notas y comentar aspectos diversos. El MCER en este aspecto tiene elementos fuertes e interesantes.

Pero esta horizontalidad tan rica del MCER se vuelve en su contra a la hora de dar una nota global en un examen. ACTFL soluciona el problema prohibiendo que se compensen las características del discurso y que unas destrezas de alto nivel sirvan para certificar ese nivel aunque en otras situaciones aún no se logre: ACTFL deja claro que para obtener una calificación de avanzado, por ejemplo, se deben cumplir al 100% todos y cada uno de los requisitos del avanzado. Queda la posibilidad de cumplirlos casi todos y entonces se obtiene la calificación del nivel inferior, en este caso intermedio alto (casi avanzado). Pero no hay dudas (o no suele haberlas) entre los evaluadores respecto al significado de cada nota o calificación: se sabe qué representa ser avanzado y qué intermedio alto, como el resto de sus 10 niveles. Ya se ha hablado de esto en §2.6.2- 5

El MCER, en cambio, no tiene directrices e instrucciones al respecto, por lo que es necesario explicarlas claramente y describirlas en las especificaciones del test. Y a ello debe dedicarse tiempo y trabajo para que entre todos los usuarios puedan tener claro el sentido de la nota que obtengan y para notificar adecuada y claramente el nivel y que este sea bien interpretado.

Así que hubo que llegar a acuerdos con todos los equipos de evaluadores y con los profesores que calificaron las entrevistas, sobre la ponderación y el sentido de los

términos usados en las calificaciones y notas finales. Y los acuerdos y directrices respecto al test global de dominio de la lengua oral fueron estos²³⁰:

- Decir que un candidato es o tiene el nivel B1, por ejemplo, supone que posee sus habilidades completamente y que si el candidato quisiera mejorar la lengua debería estudiar un curso de más nivel, en este caso de B2. Por ello, y para facilitar la calificación, ante las ambigüedades detectadas, se desecharon todas las submarcas de nivel B1.1, B1.2, etc. y se decidió que cada nivel era un conjunto completo que se poseía completamente o no se poseía todavía; esto es, si solamente se dominan ciertas habilidades dentro de un nivel, pero no las demás, significa que el nivel que posee el candidato es el anterior. Así, no se da una nota de B1.1 sino de A2 y ello significa que el candidato ya posee las habilidades de A2 y, si es el caso, puede estar en el camino de aprendizaje y uso de las habilidades del siguiente nivel, B1, que todavía no se dominan, por lo que puede encontrarse en el B1.1. Pero se le da un certificado de A2.

- Si el discurso de un candidato muestra un nivel más alto en determinadas tareas, porque la tarea es sencilla y le permite mayor fluidez o mejor lucimiento, pero en otras tareas no logra esos mismos resultados, se califica al candidato como dentro del nivel más representativo de su habilidad a lo largo de toda la entrevista. Así, hay candidatos que muestran una habilidad de A2 en tareas como contar una receta de comida o argumentar sobre la guerra, pero se desenvuelven dentro del B1 en el resto de las tareas y de la entrevista: su calificación final se considera como B1. Igualmente, otros candidatos que parecen de nivel B2 en determinadas tareas pero no se mantienen en ese nivel, son B1 si es este el nivel que mantienen como forma de expresión más general durante la entrevista. Haberlos calificado como B2 sería artificial o falso. Y utilizar subdivisiones de nivel como B1.1 o B1.2 parecía poco práctico, dados los problemas de ambigüedad ya existentes para encontrar suficientes diferencias de nivel en las escalas.

Hay muchos ejemplos de este tipo de ponderación del que venimos hablando aquí. Damos tres:

²³⁰ Con esto no queremos decir que deba ser así siempre. Dependerá de los objetivos del test y del contexto en que se utilice el MCER y sus escalas. Por ejemplo, seguimos pensando que, para contextos educativos y académicos, desmenuzar las habilidades de un estudiante en todas las posibilidades que permite el MCER a través de sus casi 60 escalas, y en los subniveles que ofrece, es muy interesante.

- Los comentarios sobre el candidato 2 (Stive), del que un evaluador dice que puede tener nivel A2 o B1 en diferentes escalas y tareas.

Hablando de la tarea de buscar un trabajo de canguro, dice:

- Se le puede aplicar el nivel A2.2 porque “comprende lo suficiente como para desenvolverse sin esfuerzo en intercambios sencillos y habituales” (...)

- Según las otras escalas de dominio lo situaríamos en un nivel B1, aunque también es cierto que hay escalas como la de Fluidez en las que presenta un nivel A2. En cuanto a la corrección gramatical tiene errores pero aunque comete errores, queda claro lo que intenta decir; por eso, en ésta escala, puede ser de nivel B1. En la escala de Control y corrección también presenta un B1 porque pide confirmación de que la forma utilizada es la correcta. (...)

Y respecto a la tarea de explicar una actividad cultural:

- En esta actividad se identifican parámetros de dos niveles: el A2.2 y el B1. El primero porque “plantea y contesta preguntas sobre pasatiempos y actividades pasadas”. Pero, cuando se le pregunta su opinión sobre Córdoba, presenta un nivel B1 porque “sabe como encontrar y comunicar información concreta y sencilla”. Hablar de las características arquitectónicas de Córdoba es un tema relativamente complejo y muy alejado de lo cotidiano y por eso finalmente hemos decidido situarlo en un B1.

A2 es el nivel que finalmente se le asigna a este candidato. Y el que, efectivamente obtuvo por el resto de evaluadores como nota final global. Es decir, domina el A2, el B1, no.

En este sentido, debemos decir que, de acuerdo con estas instrucciones y la labor de formación de evaluadores, el estrés comunicativo que producen determinadas tareas difíciles, haciendo que la lengua se colapse, ha sido bien interpretado por los evaluadores y no han penalizado a los candidatos ante muestras de discurso de poco nivel que no representan en realidad sus habilidades generales.

El caso de Stive es uno de ellos. En él se puede ver cómo a pesar de ser de nivel A2 muestra un colapso lingüístico en una tarea que le impide hablar. Sin embargo, su calificación final no ha sido perjudicada por ello.

Aquí está la transcripción²³¹ de esta situación de estrés:

²³¹ Transcripción hecha por Karina Warren según adaptación de las normas explicadas en Lazaraton, 2002: 62-72. Significado de cada símbolo: (.3): pausas, con el tiempo en décimas de segundo, en este caso 3; .hhh: inhalación; .hhh!: inhalación profunda; hhh: exhalación; hhh!: exhalación profunda; (hhh):

- E.- ¿Me puedes explicar un poco (.) eh: <[¿por qué
C.- [Porque
E.- ---> es mejor estudiar, vivir y estudiar aquí en España para aprender
Español? ¿Qué cosas buenas y qué cosas malas tiene (.2) venir aquí a España
para:
C.- mmhmm
E.- ---> aprender una lengua?>
C.- mmhmm. E.:m (.2) creo que::m:, e:m Granada es una ^ciudad e:m, .hhh
e:m, *e*:m, e:m, *e*h creo %k-e:m que:% estudiar español es: em,.hhh
eh: muy bien en Granada,.hhh em porque: e:m (1) es una ciudad %he:m%
(1.2) muy eh: interesante: y e:m,.hhh e:m (1) ha:y e:m *e*-¿escuelas
diferentes para estudiar español aquí?
E.- [mm
C.- ---> [y: .hhh e:m, e:m español es una lengua que me gusta mucho y: s em
.hhh (.5) muy importante: para mi también. Yo: .hhh es difícil a
trabajar e:[m en Alemania=
E.- [Sí
E.- =Ah ¿sí?=
C.- =Je, je, je.=
E.- =[Y ¿qué cosas], qué cosas malas tiene
C.- ---> [de arquitect-
E.- ---> venir a España para estudiar? <¿Lo económico:??> >¿La, la, la< la
gente:?
C.- E:m= No sé qué cosas...
C.- [Niget=
E.- =[Sí, negativas puedes-=
C.- =E:m me:m (.8) <%no gusta:%> e:m .hhh e:m ?la comida: con aceite?
E.- A ha [Sí.
C.- [Todo es con aceite.
E.- A[ha
C.- [E:m y: e:m (.8) y: e:m (3) ?y hace mucho calor=
E.- =[Sí.
C.- ---> [(,[%sí%.)
E.- [Je, je. .hhh >Bueno, pero no es un problema para=
C.- =%No.%
E.- ---> aprender la len[guá ¿o] sí?<=
C.- [%>No,no,no,no]<%
C.- =Je, je. E:m (1.2) No, no tengo problemas.
E.- ¿No? ¿Qué bien! ¿No?=
C.- =Je, je, je.
E.- Y ¿tus amigos también van a venir a estudiar español aquí? ¿Tu crees que
^sería bueno?
C.- .hhh Ah, sí. Em:-=
E.- =¿Mejor que en Aleman[ia]?=
C.- =[(%sí%%Es posible, sí.=
E.- =¿Sí?=
C.- =E:m, sí.

palabra llena de aspiraciones; (.) : entonación descendente; =: palabras enganchadas sin intervalo; []: solapamientos; : sonido que se ha alargado; -: corte, normalmente una oclusión glotal; % %: volumen muy bajo; * *: sonido rasgado; ^: marcada subida de tono en una palabra; > <: habla veloz; < >: habla lenta; psk: chasquido con los labios; tch: chasquido con la lengua; Subrayado: énfasis; () : no se entiende lo que se dice.

Como se ve, baja su nivel notablemente, apenas tiene nada que decir. Pero no le penalizan globalmente. Mantiene su calificación de las anteriores intervenciones: A2. O sea, que hacen como en la OPI, consideran que no tiene más nivel, pero no le bajan el suyo. Porque existe esa posibilidad global de calificar. Si hubiera que poner notas a cada intervención, no sabemos si se mantendría la nota final, pues al hacer una media una nota baja penalizaría el resultado final.

- Los comentarios sobre el candidato 8 (Juan), finalmente B1, del que uno de los evaluadores finalmente dice:

- Teniendo en cuenta todos los comentarios realizados, observamos que el alumno realiza sus tareas mayoritariamente en un nivel B1, aunque haya algún caso de nivel superior e incluso inferior.

- Los comentarios sobre el candidato 29 (María), finalmente C1, de quien, hablando sobre la tarea de “explicar una actividad cultural”, se dice:

- Si nos atenemos a esta escala, Intercambio de información, pensamos que no podríamos evaluar correctamente el nivel de la alumna. Nos quedaríamos muy por debajo del que, en función de otras escalas de dominio, creemos que tiene.

- Así pues, si nos fijamos en otras escalas de dominio, su nivel se corresponde con un nivel elevado. Si nos detenemos en la escala de Compensación, por ejemplo, es verdad que la alumna demuestra un C2, pues en ocasiones “sustituye una palabra que no recuerda por un término equivalente de una forma tan sutil que apenas se nota”:

“No, bueno, es que, cómo decir, los versos de Lorca no pasaron... era como si hubiera bailado los versos de Lorca” (el subrayado es del evaluador).

- Del mismo modo, según las escalas de Fluidez oral o de Riqueza de vocabulario pensamos que el nivel logrado sería un C1, igual que sucede con la de Corrección gramatical, ya que, ciertamente, “los errores son escasos y apenas se notan”.

En definitiva, este sistema de hacer la ponderación y la calificación final del nivel de cada candidato, unido a la formación de los evaluadores y al diseño del test acercándose a las funciones que se elicitaban en la OPI, ha permitido que se evite lo que en §2.4.3- 3 llamábamos estrés comunicativo, y que se penalice a los candidatos que presentan un discurso que se colapsa por la dificultad de la tarea o por una situación difícil provocada por las preguntas del entrevistador. De estos problemas, considerando la dificultad de la tarea, de la situación, y del contexto generado en la interacción a lo largo y dentro del test, se ha tratado en §2.4.

4.3 Análisis de las entrevistas

En este apartado ofrecemos un análisis cualitativo de las entrevistas, seleccionando los principales ejemplos de la casuística, con comentarios extraídos de los propios evaluadores.

Estos datos los presentamos según este guión:

- Una introducción para el nivel obtenido en el test del MCER (desde menos que A1 hasta C2) por cada candidato, en la que se recuerdan los objetivos para ese nivel.
- Un cuadro con los resultados de todas las entrevistas que encajan en ese nivel; cuadro que muestra la calificación obtenida según el MCER y según la OPI y el nivel en que estudiaron los candidatos en el CLM.
- El comentario de los resultados del candidato, con ejemplos de ambas entrevistas, la realizada según nuestro test del MCER y la OPI.

Del test del MCER presentamos la calificación obtenida por cada candidato y las explicaciones en las que se basa la decisión del nivel, según algunos de los razonamientos que nos han dado sus calificadores, con ejemplos extraídos de las grabaciones. A veces, breves extractos, otras veces diálogos completos sobre un tema o una tarea.

De la entrevista realizada según ACTFL, damos el resultado acompañado de extractos seleccionados por nosotros mismos con los que intentamos mostrar algunas de las características de la habilidad del candidato.

4.3.1 Entrevistas de menor nivel que A1 en el MCER

Tras calificar todas las entrevistas, hubo dos que dieron el resultado de ser de menor nivel que A1 y ambas fueron calificadas como inicial medio en la escala de ACTFL.

Estos son sus resultados (en negrita el que va a ser comentado):

Numero y nombre	Calificación de la entrevista hecha para el MCER	Calificación de la OPI de ACTFL	Curso estudiado en el CLM
35. Kayo.	Menos que A1	Inicial medio	Inicial A
34. Sachiyo.	Menos que A1	Inicial medio	Inicial A

Estas entrevistas corresponden a dos candidatos que habían estudiado durante un mes en el CLM en un curso de principiantes, pero que todavía no poseen ni dominan las características del A1. Ambos candidatos son, además, del mismo país, Japón. Esto nos hace pensar que el desarrollo del aprendizaje del español por estudiantes japoneses, tradicionalmente lento por numerosos factores, salvo que los estudiantes ya conozcan otras lenguas y estén familiarizados con nuestro alfabeto y sistema cultural, no se ve reflejado en las escalas del MCER. Quizás debería replantearse, puesto que al no mostrar un avance explícito en el aprendizaje del español para estudiantes de estas nacionalidades, se puede ver frustrada su motivación si se utilizan evaluaciones acordes con el MCER que no muestren el esfuerzo y trabajo desarrollado a lo largo de un curso²³².

Los candidatos son menos que A1 pero son, sin embargo, el segundo nivel de la escala de ACTFL, inicial medio.

Tener un nivel inferior al A1 significa no dominar las habilidades que se muestran en las diferentes escalas del MCER respecto a este nivel, que damos a continuación:

Es capaz de comprender y utilizar expresiones cotidianas de uso muy frecuente así como frases sencillas destinadas a satisfacer necesidades de tipo inmediato. Puede presentarse a sí mismo y a otros, pedir y dar información personal básica sobre su domicilio, sus pertenencias y las personas que conoce. Puede relacionarse de forma elemental siempre que su interlocutor hable despacio y con claridad y esté dispuesto a cooperar.

Niveles comunes de referencia: escala global (A1) (Consejo de Europa, 2002: 26).

²³² North / Moyer (2004), en el documento «Exemples de productions orales illustrant, pour le français, les niveaux du Cadre Européen commun de reference pour les langues», en el que dan ejemplos de 23 estudiantes, evaluados en parejas, reconocen esta carencia, ya que tras tres semanas de curso de lengua a 15 horas diarias, una estudiante americana, Jessica, no llega al A1 y la califican como de nivel 0. Se puede ver el documento en la página web del Consejo de Europa: www.coe.int/T/DG4/Portfolio/?L=E&M=/main_pages/illustrationse.html 8; una vez en la página dirigida a: Material illustrating the CEFR levels, página 7 (consulta: 21 de enero de 2008).

Interacción oral	Puedo participar en una conversación de forma sencilla siempre que la otra persona esté dispuesta a repetir lo que ha dicho o a decirlo con otras palabras y a una velocidad más lenta y me ayude a formular lo que intento decir. Planteo y contesto preguntas sencillas sobre temas de necesidad inmediata o asuntos muy habituales.
Expresión oral	Utilizo expresiones y frases sencillas para describir el lugar donde vivo y las personas que conozco.

Niveles comunes de referencia: cuadro de autoevaluación (Consejo de Europa, 2002: 30-31).

ALCANCE	CORRECCIÓN	FLUIDEZ	INTERACCIÓN	COHERENCIA
Dispone de un repertorio básico de palabras y frases sencillas relativas a sus datos personales y a situaciones concretas.	Muestra un control limitado de unas pocas estructuras gramaticales sencillas y de modelos de oraciones dentro un repertorio memorizado.	Sólo maneja expresiones muy breves, aisladas y preparadas de antemano, utilizando muchas pausas para buscar expresiones, articular palabras menos habituales y corregir la comunicación.	Plantea y contesta preguntas relativas a datos personales. Participa en una conversación de forma sencilla, pero la comunicación se basa totalmente en la repetición, reformulación y corrección de frases.	Es capaz de enlazar palabras o grupos de palabras con conectores muy básicos y lineales como <i>y entonces</i> .

Niveles comunes de referencia: aspectos cualitativos del uso de la lengua hablada (A1)(Consejo de Europa, 2002: 32-33).

<p>Plantea y contesta preguntas sobre sí mismo y otras personas, dónde vive, personas a las que conoce, cosas que tiene.</p> <p>Plantea y contesta preguntas sencillas, realiza afirmaciones sencillas sobre temas de necesidad inmediata o muy cotidianos y responde a este tipo de afirmaciones.</p> <p>Comprende preguntas e instrucciones que se le formulan despacio y con cuidado.</p>
--

Intercambio de información (A1) (Consejo de Europa, 2002: 80)

Ser inicial medio en las *guidelines* de ACTFL supone, como hemos dicho en §2.6:

- *Comunicarse mínimamente y con dificultad usando un número de palabras aisladas y locuciones memorizadas, ciñéndose al contexto concreto en que han aprendido la lengua. Al responder a preguntas concretas solamente se pueden emitir dos o tres palabras cada vez u ocasionalmente alguna respuesta de repertorio. Se hacen pausas constantemente buscando palabras sencillas o se repiten y reciclan las suyas y las de su*

interlocutor. A causa de las dudas, falta de vocabulario, incorrección, o fallos en responder adecuadamente, estos candidatos no logran hacerse entender fácilmente ni ante interlocutores acostumbrados a tratar con extranjeros. Cuando se les pide que manejen temas y realicen actividades de nivel intermedio suelen recurrir a repeticiones, palabras de su lengua materna o se callan (Breiner-Sanders, 2000: 17).

Las candidatas no mostraron habilidades suficientes como para calificarlas como A1, en ninguna tarea. Sin embargo, en la entrevista de ACTFL son inicial medio. Por otra parte, si calificamos estas entrevistas del MCER, calificadas como menos que A1, con los criterios de ACTFL, sus resultados serían, igualmente, inicial medio. De donde se deduce que el problema para no poder darles nivel según el MCER (que no posee descriptores para su poco nivel) no es de las muestras de lenguaje obtenidas en el test sino de las escalas utilizadas para calificar.

Comentamos una de ellas, la número 35.

4.3.1- A) Entrevista con el MCER del candidato 35 (Kayo): menos que A1

Vamos a verlo en algunas tareas:

- Trabajo de canguro (0:34).

Esta tarea está pensada para los niveles bajos y se ha diseñado con el objeto de probar si el candidato:

A2	- Plantea y sabe cómo responder a preguntas relativas a lo que se hace en el trabajo y en el tiempo libre. - Pide y da indicaciones que se refieren a un mapa o a un plano. - Plantea y contesta preguntas sobre costumbres y acciones de la vida cotidiana.
A1	- Plantea y contesta preguntas sencillas, realiza afirmaciones... - Plantea y contesta preguntas sobre sí mismo y sobre otras personas: dónde vive, personas a las que conoce, cosas que tiene.

El candidato apenas puede hablar, responder ni preguntar y el examinador tiene que repetir todas las preguntas (C. = candidato; E. = entrevistador):

E.- *Tú ¿quieres trabajar con mi hija aquí en casa?*

C.- *Ah, no, ah, eh?*

E.- *¿No quieres trabajar con mi hija aquí en casa? ¿Sí o no?*

- C.- *No, ¿Trabajar?, eh, no, mi estudiante.*
 E.- *Ya, ya, tú necesitas trabajar, ¿sí?*
 C.- *mm*
 E.- *Vale, yo tengo una hija, entonces tú me has dicho que tú quieres ser el baby sitter, el canguro de mi hija, ¿sí?*
 C.- *Sí*
 E.- *Y tú, ¿Te gustan los niños a ti?*
 C.- *¿Niños?*
 E.- *¿Los niños pequeños, te gustan?*
 C.- *¡Ah!, sí.*

El candidato apenas es capaz de plantear preguntas sencillas:

- E.- *¿Tienes alguna pregunta sobre mi hija, el trabajo o cualquier cosa?*
 C.- *Ah, ¿tú - casa - aquí—cerca?*
 C.- *Ah, ¿qué? ...¿cuánto?, ¿cuánto?, ¿cuánto?, ¿qué her-mano?, ¿cuánto her-mana?, ¿qué?, ¿quién?*
 E.- *No, sólo tienes que trabajar con mi hija, mi hija pequeña, sólo tengo una hija.*
 C.- *Una.*
 E.- *Sí, solamente. Nada más, ¿sí?*
 C.- *Sí, eh, más, ah, ah, trabajar cuanto profesor... ¿divertido?*

Se confirma que ni siquiera las escalas de dominio que tienen descriptor posible para el nivel A1 se ajustan a la capacidad del candidato, cuyo nivel es claramente inferior. Los ejemplos que hemos seleccionado, que son prácticamente toda la producción oral que se consigue, demuestran que tiene algunas nociones mínimas que no son suficientes para resolver una tarea básica propia del A1.

- Explicar una receta de comida (4:37).

Esta tarea, como ya se ha explicado, está pensada para nivel intermedio. Lógicamente no llega a cumplir con las habilidades del descriptor del nivel.

- E.- *¿Puedes explicarme una comida que sea buena de China o de Japón?*
 C.- *Ah... China, Japón...*
 E.- *Sí, es diferente, a ver, una comida, una comida para mí.*
 C.- *Atoo... Cocina... ¿eh?*
 E.- *La comida, ¿Cómo se llama?*
 C.- *Atoo, frecuentemente comida, japonés, cocina ehhhhhh Sopa de miso atoo... purrr, purrr*
 E.- *¿Qué es?*
 C.- *Mmmmmmm*

En cuanto a la *Riqueza de vocabulario*, conoce y reconoce alguna palabra, pero no llega a tener un repertorio ni a crear frases aisladas, como se dice en la escala (*Tiene un repertorio básico de palabras y frases aisladas relativas a situaciones concretas*).

- E.- *¿Qué cosas, qué ingredientes?*
C.- *Miso, ah, salsa... agua. Me gusta chino, mi japonés*
E.- *¿Es fácil hacer una sopa de miso?*
C.- *Muy fácil*
E.- *¿Cuánto tiempo?*
C.- *Ehh, un poco*
E.- *¿Es buena?*
C.- *Sí. Ah... comida*
E.- *¿Qué comida te gusta?*
C.- *Me gusta chino, mi japonés.*

La conclusión es que tampoco llega al nivel A1.

Las siguientes tres tareas, explicar una actividad o viaje interesante, comentar las ventajas de estudiar en el país y contar un cómic, se diseñaron para todos los niveles según las escalas de *Fluidez oral*, *Coherencia y cohesión* y *Desarrollo de descripciones y narraciones*. No vamos a detallar su resultado. En ninguna de ellas llega al nivel A1. Apenas habla ni contesta a las preguntas ni describe las situaciones, ni tiene suficiente dominio de vocabulario o corrección gramatical.

- Explicar una actividad o viaje interesante (7:20).

- C.- *Pasado, pasado, han ido pasado, atooo, ito ita italia.*
E.- *¿Has hecho algún viaje aquí?*
C.- *España, atoo, interior, Granada.*
C.- *No, no... viaje. Siempre aquí estudia.*
C.- *Si, eh... Italia... renacimiento, me gusta renacimiento.*
E.- *¿Por qué?*
C.- *Primero, Duo...católico...*
E.- *Ah, si Milán, el Duomo ¿Es bonito?*
C.- *Si es bonito ato más ato Firenze ato Duomo.*

- Comentar las ventajas de estudiar español en España (10:27).

- E.- *¿Por qué es mejor en España y no en Japón?*
C.- *Ahhh, estudiar español, mmm ¿Por qué?*
E.- *¿Por qué en España es mejor que en Japón?*
C.- *Ahh. mm.. volar, mmm, trabajar, volar... Español atto quiero usar...*

- Contar el cómic (14:05).

Apenas puede dar nombres *novio, novia*, y algún lugar, *parque, cama*. El uso de verbos es difícil y necesita pensarlo mucho, como se ve en la penúltima línea.

- E.- *A ver, explícame quiénes son esas personas, donde están, qué pasa...*
 C.- *Ah, novio y novia.*
 C.- *Ato parque.*
 E.- *¿Y qué pasa?*
 C.- *Noche.*
 E.- *¿Dónde están?*
 C.- *Ato parque.*
 E.- *¿Y después, que pasa?*
 C.- *Él pens recorda mañana ato él es, es, estoy estés está ato cama.*
 C.- *Novia ahhh ahh.*

4.3.1- B) OPI del candidato 35 (Kayo): inicial medio

El resultado de la calificación de la OPI es de inicial medio. Cuando contesta lo hace con palabras sueltas y en un par de ocasiones puede hacer alguna frasecita.

Al hablar sobre su ciudad (0:20):

- C- *Tokio am hay todos am, pero am todos am un poco caro pero am todos diferente, persona diferente... muy am persona.*
 E.- *¿Te gusta vivir en Tokio? ¿Por qué?*
 C- *Hay ah muy... musi, museo, ah, ... mmmm, muy ah hay no.... discoteca,(risas) ah... museo, ah, ah, (risa) ah... mmm.*

De su familia (1:48):

- C- *Familia es ah vive Japón de norte, Tokio de norte, si. Ah, sí. Familia es ato padre y madre y to mayor her mano, pero mayor ato elmano Tokio vivir vivo.*

También intenta hacer alguna pregunta, aunque se ríe y grita cuando se le pide que lo haga y da muestras de incomodidad. Lo único que pudo preguntar fue (6:34):

- C- *¿Qué gusto? ¿Qué gusto Granada?*
 C- *Ah. Mmm (risas) ¿Cuánto años? ¿Cuánto tiene años?*
 C- *Ah. Mmm. (risa) mmm ¿tú, familia, ah, anque a to padre trabaja?*

Todo esto es lo correspondiente a un inicial medio de ACTFL, que conoce un número limitado de frases y palabras memorizadas, con las que apenas puede hacer mensajes y responder a preguntas sobre temas preparados y cotidianos.

Con ello ha sido evaluada con las *guidelines* de ACTFL como inicial medio.

Y en el test del MCER como menos que A1.

4.3.1 C) Conclusión

Como hemos comprobado, las entrevistas de nivel inicial medio de la OPI no tienen cabida en el MCER, pues requieren descriptores de menor nivel que los más bajos que se describen. De esto, así como de sus consecuencias pedagógicas, hablaremos en el capítulo 5.

4.3.2 Entrevistas de nivel A1

Las entrevistas que dieron este nivel son cuatro: la 1 (Biancé), 3 (Johanes), 4 (María) y 6 (John). De ellos, tres dieron inicial alto en ACTFL y uno de ellos solamente inicial medio.

Los dos primeros estudiantes estudiaron el curso *inicial A* durante un mes, los otros dos estudiaron un nivel más, el *inicial B*.

Estos son sus resultados (en negrita la que vamos a comentar):

Numero y nombre	Calificación de la entrevista hecha para el MCER	Calificación de la OPI	Curso estudiado en el CLM
4. María	A1	Inicial medio	Inicial B
1. Biance	A1	Inicial alto	Inicial A
3. Johanes	A1	Inicial alto	Inicial A
6. John	A1	Inicial alto	Inicial B

Lograr una calificación de A1 con el MCER supone tener el dominio que hemos señalado al comienzo del apartado anterior (§4.1). La candidata 4 (María) logra ese

dominio pero es calificado de inicial medio en la OPI, por lo que logra ese nivel mediante una serie de elementos memorizados con los que apenas puede formar un discurso creativo sobre sus actividades habituales (ya que estas son algunas de las características de inicial medio en ACTFL). En cambio, los otros tres candidatos obtienen en la OPI inicial alto, por lo que su nivel A1 significaría que pueden realizar un discurso creativo sobre sus actividades diarias y habituales aunque no lo mantengan en todo momento.

En la OPI estas son diferencias importantes. Sin embargo todas ellas quedan en el test del MCER dentro del A1.

Vamos a comentar uno de los tres que son A1 y obtienen inicial alto en la entrevista de ACTFL, el 3 (Johanes).

4.3.2- A) Entrevista del MCER del candidato número 3 (Johanes): A1

Los evaluadores dicen que su nivel es A1. Y ponen como ejemplo la receta de comida en *donde la interacción del alumno con el evaluador se ha limitado a la mínima información, con detalles muy escasos y vocabulario muy básico, sin hacer uso de ningún verbo en imperativo* (evaluador).

Vamos a ver algunos de sus comentarios:

- Trabajo de canguro (1:21).

Comprende las preguntas. Plantea y contesta preguntas sencillas como: *¿Cuánto tiempo en la semana?* Y tiene un repertorio básico de palabras y frases aisladas propio de A1.

C- ¿Cuánto años tiene?/ ¿Está tranquila?/ ¿Cuánto dineros... eh / ¿Cuánto tiempo ah, en, en la semana?/ ¿El fin de semana?/ ¿Cuánta, cuántas horas?/ ¿Vacaciones?/ ¿Es posible trabajar más?

- Explicar una receta de comida (4:17).

Esta tarea, como ya se ha explicado, está pensada para nivel intermedio. Lógicamente, el candidato es A1 y no llega a cumplir con las habilidades del descriptor del nivel: “comprende las preguntas pero se expresa con palabras y frases aisladas y

entrecortadas”. Hay un uso casi exclusivo del verbo ser: *Es patatas, pero no fritas, eh, caliente agua...*

C- Es patatas, eh, pero no fritas, eh en... caliente agua. Y eh, es carne eh... eh... (pausa) en... (pausa) bol, bollo carne, no sé, no... no file para ah... no... partida. Y una salsa... moreno.

- Explicar una actividad o viaje interesante (6:00).

Apenas enlaza las palabras y solo hace uso del conector “y”. No es capaz de contar historias o describir algo con la ayuda de una lista sencilla de elementos. Usa expresiones sencillas relativas a situaciones concretas. Intercambia información limitada a través de la descripción.

Por ejemplo, cuando trata de contar el calor que pasaron por la tarde en Sevilla: *Sevilla fue muy caliente. Sí. Es 45. Horrible. Entre 2 y 8 en la tarde necesitamos...*

C- Fue aun fiesta en otro barrio, no recuerdo dónde.

C- Fue una concierto de rok, de celtic y rok, un grupo muy conocida en español, en España, y mi profesora y un fere, fere y mucho bebidas.

C- Porque ene noruega es muy frío, es solamente los, la... las mesa más caliente es posible eh posible una un fiesta.

- Comentar las ventajas de estudiar español en España (11:40).

Su repertorio básico es muy limitado. Solo es capaz de emitir enunciados aislados, breves, con gran dificultad y se desenvuelve con muchas pausas para buscar expresiones.

C- Las cosas buenas, eh...eh puedo practicar español puedo todos los días. Y... en... está más eh, más barato vivo en español, vide en español que en Oslo. Y eh, eh...y eh, es más, más caliente que en Noruega, es bien... y... yo puedo ver eh las chentes y la país, es bueno. ¿Mal? no sé mal.

- Contar el cómic (13:50).

En esta tarea dos calificadores, de los cuatro que lo han evaluado, le dan A2 porque produce expresiones breves de acciones habituales y queda claro lo que intenta decir. *Él piensa que ella hace el desayuno para los dos... Y después...*

Dicen que se ve más fluidez y su discurso es más organizado, concluyendo con la explicación breve de la moraleja del cómic: *Es posible soñar, eh, mejor días.*

C- Novia y novio, eh, es un noche romántico en la parque. Y él piensa piensa eh que eh ella eh hacer desayuno para los dos, ella hace desayuno para él está en la ... la cama y después el va a trabajo cuando ella está en la casa Y después eh, en la mismo tiempo eh ella piensas que los dos de... eh... ven los hijos jugar y ve el televisión juntas y cuando los dos eh llegan a la casa eh eh viveis con con con eh los padres y eh eh los eh ...vives con sus padres y sus abuelos y es muy pequeño.

La conclusión general, de todas formas, es que su habilidad es de nivel A1.

4.3.2- B OPI del candidato número 3 (Johanes): inicial alto

A este candidato de A1 en la OPI se le ha calificado como inicial alto. Ser inicial alto en ACTFL supone, según dijimos en §2.6:

- Poder manejar una variedad de temas y actividades de nivel Intermedio pero estos hablantes son incapaces de mantener el discurso en ese nivel. Pueden desenvolverse con éxito en un número de actividades comunicativas sencillas propias de situaciones sociales simples. Su conversación se limita a unos pocos temas predecibles necesarios para la supervivencia en la cultura de la lengua tales como información personal básica, objetos básicos y un número reducido de actividades, necesidades inmediatas y preferencias. Estos hablantes iniciales altos responden a preguntas simples y directas o piden información; cuando se les pide que hagan preguntas sólo son capaces de hacer unas pocas y memorizadas.

El candidato es creativo con el lenguaje que conoce y contesta a las preguntas aunque con problemas. Muestra un uso creativo del lenguaje aprendido para funciones personales y situaciones básicas predecibles. A veces parece intermedio pero no se mantiene en ese nivel todo el tiempo. Otras veces ni siquiera son frases lo que logra componer.

El resultado es que no hay cantidad de información, ni de frases, ni calidad suficiente como para considerar que es un intermedio bajo.

Preguntado por su ciudad, da información con una funcionalidad cercana a un intermedio, creativa y en español (00:16):

- E.- ¿Cómo es Oslo?*
- C.- Oslo es frío, es pequeño, más pequeño que que Granada. No. Es un poco más grande que Granada pero es un pequeño.*
- E.- ¿Y te gusta Oslo?*
- C.- Sí, porque tiene los cines, teatros, las cosas neces nece sita en el ciudad, la ciudad.*

- E.- *¿Es la capital, no, de Noruega?*
 C.- *Sí.*
 E.- *¿Es una ciudad turística, industrial o cómo es?*
 C.- *No, no industrial no porque... no industria en Noruega, porque... eh... no industria y no turístico.*

Al hablar de su familia, parece dar algo más de información y mantenerse en un intermedio (1:20):

- E.- *¿Y tú vives allí siempre, vives con tu familia, estudias, trabajas?*
 C.- *Vive con mi familia de cero a 21 años y después yo he estudiado en Trodheim.*
 E.- *¿Cómo es tu familia?*
 C.- *Mi madre y mi padre tienen cuarenta y ocho años más o menos y tengo dos hermanos eh... Eric ah el menor y eh..., los dos son mi menores pero Eric tiene diecisiete años y Seti, el otro, tiene diecinueve años.*
 E.- *Tú ¿eres el mayor? ¿Solo sois chicos?*
 C.- *Solo chicos.*
 E.- *¿Cómo es la vida con chicos?*
 C.- *Somos tranquilo.*
 E.- *Y tu madre ¿está contenta con tres chicos, no quiere una chic?*
 C.- *Quieres un chico pero es no posible, tres veces pero no chica.*

Pero al preguntarle si fue difícil vivir solo, por sus estudios, su lengua se colapsa (4:20):

- C- *Vivo en..., ah..., vivo con dos, dos otros estudiantes, no sé palabra en español pero es un piso, un piso con tres habitaciones, una para ver tele, una sí, ah, y eh una estudiante eh, por un año y nuevo estudiante so, más, más personas.*

El nivel dado a esta entrevista según la escala de ACTFL es de inicial alto.

Y el de la entrevista del MCER, A1.

4.3.2- C) Conclusión

Como vemos, los candidatos de nivel inicial alto de la OPI (nivel 3) son en el test del MCER A1, su nivel más bajo (nivel 1), aunque todavía hay un caso que es A1 e inicial medio, que se comentará en §4.4.

4.3.3 Entrevistas de nivel A2

Dos entrevistas dieron A2 en el test del MCER, la 2 (Stive) y la 7 (Carmen). Ambos obtienen la calificación de intermedio bajo según ACTFL. El primero estudió un curso de principiante bajo durante un mes (inicial A), el otro uno de intermedio bajo (intermedio A).

Estos son sus resultados (en negrita el que vamos a comentar):

Numero y nombre	Calificación de la entrevista hecha para el MCER	Calificación de la OPI	Curso estudiado en el CLM
2. Stive	A2	Intermedio bajo	Inicial A
7. Carmen	A2	Intermedio bajo	Intermedio A

Que los candidatos tengan el nivel A2 significa que dominan las habilidades:

Es capaz de comprender frases y expresiones de uso frecuente relacionadas con áreas de experiencia que le son especialmente relevantes (información básica sobre sí mismo y su familia, compras, lugares de interés, ocupaciones, etc.) Sabe comunicarse a la hora de llevar a cabo tareas simples y cotidianas que no requieran más que intercambios sencillos y directos de información sobre cuestiones que le son conocidas o habituales. Sabe describir en términos sencillos aspectos de su pasado y su entorno así como cuestiones relacionadas con sus necesidades inmediatas.

Niveles comunes de referencia: escala global (A2) (Consejo de Europa, 2002: 26).

HABLAR	A2
Interacción oral	Puedo comunicarme en tareas sencillas y habituales que requieren un intercambio simple y directo de información sobre actividades y asuntos cotidianos. Soy capaz de realizar intercambios sociales muy breves, aunque, por lo general, no puedo comprender lo suficiente como para mantener la conversación por mí mismo.
Expresión oral	Utilizo una serie de expresiones y frases para describir con términos sencillos a mi familia y otras personas, mis condiciones de vida, mi origen educativo y mi trabajo actual o el último que tuve.

Niveles comunes de referencia: cuadro de autoevaluación (Consejo de Europa, 2002: 30-31)

ALCANCE	CORRECCIÓN	FLUIDEZ	INTERACCIÓN	COHERENCIA
Utiliza estructuras compuestas por oraciones básicas con expresiones, grupos de unas pocas palabras y fórmulas memorizadas con el fin de comunicar una información limitada en situaciones sencillas y cotidianas.	Utiliza algunas estructuras sencillas correctamente, pero todavía comete sistemáticamente errores básicos.	Se hace entender con expresiones muy breves, aunque resultan muy evidentes las pausas, las dudas iniciales y la reformulación.	Sabe contestar preguntas y responder a afirmaciones sencillas. Sabe indicar cuándo comprende una conversación, pero apenas comprende lo suficiente para mantener una conversación por decisión propia.	Es capaz de enlazar grupos de palabras con conectores sencillos tales como «y», «pero» y «porque».

Niveles comunes de referencia: aspectos cualitativos del uso de la lengua hablada
(A2) (Consejo de Europa, 2002: 32-33).

Ser intermedio bajo en las *guidelines* de ACTFL significa cumplir, aunque de forma mínima, todas las características del nivel: llevar una conversación cara a cara, haciendo y respondiendo preguntas simples, y creando y combinando el material aprendido, para hacer sus mensajes, pedir información, preguntar horarios. Y ser capaz de hacerlo en escenarios informales habituales y limitados a intercambios básicos; con temas relacionados con su medio personal e inmediato: trabajo, familia, ciudad, amigos, etc. Todo ello mediante la organización del discurso con frases sueltas.

Vamos a comentar las entrevistas de uno de los candidatos, el número 7 (Carmen), según las calificaciones obtenidas con el sistema del MCER y las *guidelines* de ACTFL.

4.3.3- A) Entrevista del MCER del candidato 7 (Carmen): A2

Esta candidata recibió la calificación de A2. Estos son algunos de los comentarios de sus evaluadores:

- La entrevistada se hace entender mediante expresiones sencillas de manera general y se hacen patentes las dudas iniciales y los silencios, por lo que no es capaz de mantener un discurso fluido de producción libre.

- Adapta frases o fórmulas sencillas aprendidas; confunde tiempos verbales (eres un poco frío), olvida mantener la concordancia (no tiene mucho palabras), confunde los artículos (la padre); sin embargo suele quedar bastante claro lo que pretende decir por lo que el nivel de intercambio comunicativo es aceptable. El dominio del léxico es suficiente para mantener un intercambio comunicativo básico y cotidiano y no consigue manejar una secuencia lineal de elementos.

Vamos a verlo en algunas de las tareas:

- Trabajo de canguro (0:30).

Se desenvuelve de acuerdo con los descriptores de A2.2 de *Intercambiar información*: *Comprende lo suficiente como para desenvolverse sin esfuerzo en intercambios sencillos y habituales. Plantea y contesta preguntas sobre costumbres y acciones de la vida cotidiana*, etc. Responde y reacciona adecuadamente en la primera parte de las preguntas que hace el evaluador:

E.- ¿Vives en España ahora?

C.- Vivo en España y yo necesito trabajo.

E.- ¿Tú estudias o no?

C.- Sí. Después de la clase yo necesito trabajo y yo puedo venir a tu hija por la tarde.

E.- ¿Es la primera vez que cuidas niños, que haces este trabajo?

C.- Sí.

E.- ¿Te gustan los niños?

C.- Me gustan mucho.

E.- ¿Tienes alguna pregunta?

Y estas son sus preguntas, que se mantienen en el nivel de *Intercambio de información* que estamos describiendo:

C- Tu mujer ¿es a la casa por la tarde?/ ¿Qué te gusta que yo hego por ta por te por tu hija?/ ¿Yo puedo salir también, por ejemplo?/ ¿Necesita ella necesita por ejemplo una comida por la tarde?

- Explicar una receta de comida (3:10).

Explica como se hace el cuscús. Comienza admitiendo que no sabe muchas palabras y cuando el entrevistador le dice que lo intente con otras palabras y lo explique, se lanza. Durante la explicación pide ayuda al entrevistador para ver si le entiende y para

explicar ciertas palabras que desconoce. Se da, entre paréntesis, la ayuda que el entrevistador le ofrece.

Ya hemos advertido de que esta tarea se pensó para niveles más altos, como B1, para ver en qué medida se era capaz de explicar un procedimiento. Lo que esta estudiante logra hacer no alcanza el nivel B1. Pero su discurso continúa dentro de lo descrito para A2:

- C.- *La primero... meta el poco de tomate, el poco de sel, eh, no sé, (el entrevistador se lo dice: sí cebolla)*
- C.- *cebolla, bien, y un poco al la fuego por tres minutos, dos minutos e siempre...Un poco de agua, un poco de agua, e después con esta comida, como se dice ¿fuego?*
- E.- *(¿más alto?)*
- C.- *No mas alto, más bajo. Y tú corta la verdura y después todos juntos y siempre la fuego más se bajo. Y otra comeda, ¿como dice? Come dro, por hace la comida*
- E.- *(La cazuela o sartén)*
- C.- *Una cazuela y un poc de agua...*

- Explicar una actividad o viaje interesante.

Al preguntarle por un viaje no hay una secuencia lineal de elementos, que sería el nivel B1, sino una serie de frases y oraciones sencillas para contestar a las preguntas. Su discurso fue el siguiente (5:57):

- C- *Este mes... mucho viajes... yo fui a las Canarias*
- C - *La primera semana, semana de julio, yo fui a Canarias.*
- C - *Sí, a Las palmas e una puerto que se llama puerto... la semana después la semana pasada yo fui a Barcelona*
- C - *Sí, me gusto mucho, me gusta eh pero yo me gusta eh yo so baña pero estaba ah eres es un poco frió y no puedo, puedo.*
- C - *No porque aquí yo tengo una familia, la familia de mi buela, esta familia tenia una trabajo a Canarias y yo..., eh... yo...*
- C - *Sí, no siempre porque vive lejos un poco de aquí de Granada eh yo esta familia me das un apartamento enfrente del corte inglés.*

- Comentar las ventajas de estudiar español en España (8:15).

Igual que en la anterior tarea, la estudiante crea un mensaje que no logra articular con claridad y se pierde.

- C.- *Bien la primer cosa a Túnez no tengo tiempo para estudiar a centro cultural de España quel yo trabajo por la mañana por la tarde y siempre. Otra cosa, siempre aquí Granada en España yo puedo practicar la lengua muy bien a la calle con la gente. Practicar muy bien. No olvido que solo un mes el año pasado este año una mes yo puedo hablar.*

- Contar el cómic (9:26).

Se muestra incapaz de explicar el cómic, necesita buscar las palabras y no las encuentra, hace constantes pausas y finalmente se pierde y llega a un callejón sin salida:

C.- Este comic qui ah... tenemos dos o cupe en jardín por la noche eh juntos eh después ah eh andando el hombre pienso piense pienso que espero que por la noche espero espero eh que la mañana sa mujer ah eh pone pone el café el café a la cama si eh con ah y vien de la..., vine de (se para).

Ante la pregunta del entrevistador, continúa (10:46):

C.- Que pasa que cet hombre vive con la familia... con la familia de sa mujer... con la pa... con la hijos, con la padre, la madre y todo eh la padre todos masho. Yo creo que tenemos todo 60 años sí más está hombre por la noche no puedo gusta que pasa la noche a... que pasa la noche a la calle mejor que a la casa eh bueno este hombre andando con sa mujer pienso que espero que vive ton el solo sin la familia sin la persona y que la mañana sa mujer viene venía poner con el café a la cama y cuando viene al trabajo el tenía un beso y espero que tener un salón por ver la"tivi" con la hijos y sa mujer y espero que tener habitaciones para la niños.

La conclusión es que su habilidad es de nivel A2 y que en varias tareas la dificultad le supera impidiéndole expresarse.

4.3.3- B) OPI del candidato 7 (Carmen): intermedio bajo

El nivel en la entrevista OPI es intermedio bajo. Realiza mínimamente las funciones del intermedio y utiliza su tipo de lenguaje durante todo el tiempo, incluso cuando se le hacen preguntas más complicadas y propias de un nivel avanzado, pues mantiene el mismo tipo de discurso. En general, da poca cantidad de información pero a veces de calidad suficiente.

Al preguntarle por su ciudad y la playa empieza a hablar y se ve que está utilizando el lenguaje aprendido de forma creativa, para hacer mensajes relacionados con su propia vida, lo que es propio del intermedio. Aunque no hay ni mucha cantidad ni calidad (0:14).

C.- Yo vivo a Túnez en centro de Túnez en una cuidad cerca de la playa, turística.
E.- ¿Cómo es?

- C.- *Es muy grande la playa, mucho extranjeros, mucho españoles, que visto este esta ciudad.*
- E.- *¿Te gusta? ¿Por qué?*
- C.- *No necesito que yo voy a otra actividad pa por ejemplo en verano por la playa e todo es cerca*
- E.- *¿Te gusta ir a la playa?*
- C.- *Sí, me gusta mucho si después el trabajo después todo el año por mucho trabajo si necesitas necesito vacaciones.*

Al preguntarle por su familia, sus frases son de calidad, cortas pero están más o menos bien (1:20):

- C.- *Yo tengo dos hermanas e uno hermano. La hermana en quince de agosto va se casa. Lotra hermana estodia historia in la universita. Y mi hermana estodio a Francia, a París.*
- E.- *¿Vivis todos juntos?*
- C.- *No. Yo vivo sola porque yo trabajao lejos de la familia y lotra hermana que va a se casar eh también sola porque todos a si trabajar trabajan lejos da familia.*

Y al hacer preguntas también se defiende aunque mínimamente, son pocas y de calidad limitada, pero entran en lo esperado en un intermedio bajo de ACTFL (6:16):

C-¿Tú tienes hijos?/ ¿Cuánto años tiene?/ ¿E dónde eres?/ Ese ciudad ¿es muy lejos?

Ha sido evaluada en la OPI de ACTFL como intermedio bajo. Y en nuestro test del MCER como A2.

4.3.3- C) Conclusión

Los candidatos de nivel intermedio bajo de la OPI son en el test del MCER A2.

4.3.4 Entrevistas de nivel B1

Las entrevistas que dieron este nivel son cuatro, la 5 (Francesca), la 8 (Jorge), la 9 (Boris) y la 10 (Lucía). En la entrevista de ACTFL dos de estos candidatos han obtenido la calificación de intermedio medio y dos de intermedio bajo.

Su procedencia es de diversos cursos de distinto nivel. El primero estudió un curso de inicial alto (inicial B), el segundo y tercero uno de intermedio bajo (intermedio A), el último uno de intermedio alto (intermedio B).

Estos son sus resultados (en negrita la que va a ser comentada):

Numero y nombre	Calificación de la entrevista hecha para el MCER	Calificación de la OPI	Curso estudiado en el CLM
5. Francesca	B1	Intermedio bajo	Inicial B
8. Juan	B1	Intermedio medio	Intermedio A
9. Boris	B1	Intermedio medio	Intermedio A
10. Lucia	B1	Intermedio bajo	Intermedio B

Queremos hacer notar que hay otros candidatos que han obtenido también intermedio bajo e intermedio medio en la entrevista de ACTFL, con una calificación diferente en el test del MCER. Dos candidatos tienen una calificación de A2 en el test del MCER e intermedio bajo en ACTFL. Han sido comentados en el punto anterior. Igualmente, hay dos candidatos que obtienen intermedio medio en la entrevista de ACTFL y en la del MCER tienen una calificación de B2, el candidato 15 (María) y el 19 (Maya) y de los que se hablará más adelante (§4.4.3).

Tener un dominio de nivel B1 significa, por ejemplo:

Sabe desenvolverse en la mayor parte de las situaciones que pueden surgir durante un viaje por zonas donde se utiliza la lengua. Es capaz de producir textos sencillos y coherentes sobre temas que le son familiares o en los que tiene un interés personal. Puede describir experiencias, acontecimientos, deseos y aspiraciones, así como justificar brevemente sus opiniones o explicar sus planes.

Niveles comunes de referencia: escala global (B1) (Consejo de Europa, 2002: 26).

Interacción oral	Sé desenvolverme en casi todas las situaciones que se me presentan cuando viajo donde se habla esa lengua. Puedo participar espontáneamente en una conversación que trate temas cotidianos de interés personal o que sean pertinentes para la vida diaria (por ejemplo, familia, aficiones, trabajo, viajes y acontecimientos actuales).
Expresión oral	Sé enlazar frases de forma sencilla con el fin de describir experiencias y hechos, mis sueños, esperanzas y ambiciones. Puedo explicar y justificar

	brevemente mis opiniones y proyectos. Sé narrar una historia o relato, la trama de un libro o película y puedo describir mis reacciones.
--	--

Niveles comunes de referencia: cuadro de autoevaluación (Consejo de Europa, 2002: 30-31).

<i>ALCANCE</i>	<i>CORRECCIÓN</i>	<i>FLUIDEZ</i>	<i>INTERACCIÓN</i>	<i>COHERENCIA</i>
Tiene un repertorio lingüístico lo bastante amplio como para desenvolverse y un vocabulario adecuado para expresarse, aunque un tanto dubitativamente y con circunloquios, sobre temas tales como su familia, sus aficiones e intereses, su trabajo, sus viajes y acontecimientos actuales.	Utiliza con razonable corrección un repertorio de fórmulas y estructuras de uso habitual y asociadas a situaciones predecibles.	Puede continuar hablando de forma comprensible, aunque sean evidentes sus pausas para realizar una planificación gramatical y léxica y una corrección, sobre todo en largos periodos de expresión libre.	Es capaz de iniciar, mantener y terminar conversaciones sencillas cara a cara sobre temas cotidianos de interés personal. Puede repetir parte de lo que alguien ha dicho para confirmar la comprensión mutua.	Es capaz de enlazar una serie de elementos breves, diferenciados y sencillos para formar una secuencia lineal de ideas relacionadas.

Niveles comunes de referencia: aspectos cualitativos del uso de la lengua hablada (B1) (Consejo de Europa, 2002: 32-33).

Comentamos ahora un ejemplo de B1 que en la entrevista de ACTFL es intermedio medio, el candidato 8 (Juan). Los candidatos que son B1 e intermedio bajo los consideramos casos raros y se comentan en §4.4.

4.3.4- A) Entrevista del MCER del candidato 8 (Juan): B1

El nivel determinado por los evaluadores es B1.

Vamos a verlo con alguno de sus comentarios:

- Trabajo de canguro (1:00):

Muestra un nivel B1 en la realización de esta tarea, pues, con un pequeño abanico de recursos plantea y sabe cómo responder a preguntas sencillas relativas al trabajo y cuestiones sobre sí mismo o sobre otras personas. Aunque el entrevistador tiene que estar continuamente provocando la reacción del candidato, como cuando le pregunta si

le gustan los niños, a lo que éste contesta *Sí, mucho. Tengo una educación de pedagógica y tengo una experiencia... de dos años... con niños*, que es la respuesta más larga que da a lo largo de toda la tarea. Pero cuando habla lo hace adecuadamente, preguntando por el número de horas que tendrá que trabajar, el horario, lo que podrá ganar, la edad de su hija, etc. Y en todo momento hace comentarios a las respuestas y preguntas de clarificación como muestra de que está comprendiendo, repitiendo las respuestas del entrevistador y utilizándolas para enlazarlas con una nueva pregunta adecuada:

- C.- *Eh, ¿Cuánto horas tengo?*
 E.- *Necesitamos dos días de tres horas cada día, por la tarde.*
 C.- *¿Por la tarde? Está bien para mí, no hay problema. Entonces, eh, ¿puedo empezar mañana?*
 E.- *Mañana no, el lunes.*
 C.- *El lunes. Eh, mm, ¿Cuánto puedo ganar?*
 E.- *Aquí pagamos más o menos 9 euros por hora,*
 C.- *9 euros, está bien*
 E.- *Dos días, seis horas.*
 C.- *Seis horas. Dos días está bien. Y ¿todos los días de la semana?*
 E.- *No, dos días.*
 C.- *Dos días.*
 E.- *El lunes y el miércoles ¿puedes?*
 C.- *¿Por la tarde?, ¿a qué hora?*
 E.- *Bueno, de 5 a 8...*
 C.- *Muy bien, trabajo por la mañana. Muchas gracias.*
 E.- *¿No quieres saber algo de mi hija o otras cosas?*
 C.- *Tu hija, pienso que conocemos el lunes y ahí... ¿cuánto años tiene?*
 E.- *Tiene 10 años.*
 C.- *Tu hija 10 años, ah risas.*
 E.- *Ah, pienso que tu hija tiene una niña (risas).*

- Dar la opinión sobre una actividad cultural (4:00):

El descriptor de la escala de *Intercambio de información* del nivel B1 dice que “Resume y da su opinión sobre relatos, artículos, charlas, discusiones, entrevistas o documentales breves y responde a preguntas complementarias que requieren detalles”, así como “Intercambia, comprueba y confirma con cierta confianza información concreta sobre asuntos, cotidianos o no”. Y según los calificadores que han evaluado a este candidato esto se cumple, aunque quizá mínimamente. Algunos ejemplos que justifican esta opinión son:

- C.- *El domingo fue a un concierto en el Generalife, el Romancero Gitano. Flamenco, baila, canta y también multimedia muy interesante.*
 E.- *¿Qué te pareció?*

C.- *Es maravilloso. Es muy bonito, todas la combinación de música, de baila, colores,...es muy bueno. A mí me gusta el flamenco mucho y música de folk. Toco guitarra un poco.*

Igualmente, respecto a las escalas de *Dominio de vocabulario* y *Riqueza de vocabulario*: dispone de una serie de elementos lingüísticos para explicar los principales puntos de una idea y para expresar pensamientos sobre temas abstractos o culturales, como el que demanda la actividad. Comienza dando una valoración *Más fácil es para mí recordar el día antes de mañana, fue al Generalife...*, en referencia a la pregunta en la que se le pide que cuente una actividad cultural, y quiere decir que va a hablar sobre una reciente; y todo el tiempo acompaña los sustantivos con adjetivación adecuada, como en *Es maravilloso, Es muy bonito, A mí me gusta..., Fue muy artístico y alegre, Un poco turístico, Pequeñas películas muy modernas, Es muy natural*, etc. Y utiliza léxico como *flamenco, música, baile, colores, artístico, alegre, películas modernas, poemas de Lorca, guitarra*, etc.

Preguntado sobre su opinión dice:

E.- *¿Y el espectáculo que era flamenco, flamenco puro, fue flamenco o era turístico, qué te pareció?*

C.- *No turístico. Pensaba que es un poco turístico, es un poco kitch, pero después fue... Artístico level es muy alto, con poemas de Lorca recitados, pequeñas películas, muy ah modernas... combinación de moderna y viejo flamenco es...*

E.- *¿Te pareció interesante?*

C.- *Mucho.*

E.- *¿No fue demasiado comercial?*

C.- *Muy natural. No, no tienes... piensas que... es... a mi parece era moderno.*

- Argumentar a favor o en contra de una guerra.

El candidato formula un discurso con ciertas dificultades, del que sacamos la conclusión de que no apoya la guerra de Irak. Lo calificaremos como B1. Expone sus ideas sobre un tema complejo, usando varios argumentos y explicaciones, y llegando a realizar una comparación entre el sistema político de su país y el de los Estados Unidos.

Vamos a ver el principio de sus argumentaciones (14.35):

C- *Para mi es una cosa muy mal, no pienso que es una buena idea solver conflictos con la guerra. Especialmente es un conflicto cultural también, es no solo sobre los resorces*

de cómo gasóleo, es una cosa cultural también como isla. Es muy peligroso eh tener un conflicto con otra grande cultura y pienso que Bush es un poco estúpido si no piensa en cosas como diferencias culturales y... todo el mundo países arábigos es muy juntos, cuando uno país tiene problema es solidario, entonces no soy de acuerdo con la política internacional de Estados Unidos. Pienso que gente de Estados Unidos también no son feliz con eso; hay mucho gente de universidades que piensan el mismo. Es un problema...

- Explicar una actividad o viaje interesante.

El candidato cuenta un viaje enriquecedor culturalmente, que es una tarea bastante más compleja que la señalada por el descriptor del B1, por lo que, aunque la fluidez le falle, es ése el nivel en el que debemos encuadrarlo, pues un nivel B2 ya sería excesivo y un A2 poco: cuenta historias o describe algo empleando únicamente una lista sencilla de elementos.

Si nos fijamos en la escala de *Intercambiar información* lo colocaríamos también en el B1, porque para un A2 se dice que *plantea y contesta preguntas sobre pasatiempos y actividades pasadas*, y lo que hace el alumno es más que un simple intercambio de información, es la descripción de un viaje, algo que sobrepasa el nivel A2.

No obstante y debido a la dificultad de la tarea, no hay cohesión entre las partes del texto, partes que más bien son pequeñas frases que va enlazando con la conjunción ‘y’. Quizá pudiera haberlo hecho mejor empleando otras estructuras más sencillas que debe conocer, pero ha elevado su tono y eso le ha hecho equivocarse en muchos casos: *parti, parada, paradi, paradamos en Salamanca; Nosotros teníamos un coche pero la camina es cerca de la calle todos los tiempo, puedes ver caminar.*

Vamos a ver un extracto de su narración (6:30):

C- Fue a Santiago de Compostela también pero parti, parada, paradi, paradamos en Salamanca. Es muy precioso Salamanca, con la marcha, Salamanca por la noche, ¡guau! con las luces en el las iglesias la gente de universitarios y podría probar un vino de ribera de Duero,... es rico y también la catedral en Santiago de Compostela en el Camino de Santiago. Nosotros teníamos un coche pero la camina es cerca de la calle todos los tiempo no puedes ver el caminar...

El nivel con el que le han calificado es B1.

4.3.4- B OPI del candidato 8 (Juan): intermedio medio

Tener una calificación de intermedio medio de ACTFL supone ser capaz de realizar todo lo del nivel intermedio con suficiente soltura, calidad y cantidad y a veces algunas cosas del avanzado, aunque con problemas.

Según la escala de ACTFL dada en §2.6:

- Los hablantes del nivel intermedio medio pueden manejar con acierto una variedad de actividades comunicativas sencillas en situaciones sociales básicas. Su conversación se limita a algunos de los intercambios establecidos y predecibles necesarios para la supervivencia en la cultura de la lengua. Estos comprenden información personal relacionada consigo mismo, la familia, la casa, actividades diarias, preferencias e intereses personales, así como necesidades físicas y sociales como alimentación, compras, viajes y alojamiento. Tienden a reaccionar al discurso por ejemplo contestando a preguntas directas y pidiendo información. Sin embargo son capaces de hacer una variedad de preguntas cuando necesitan obtener información simple para satisfacer sus necesidades básicas, como direcciones, precios y servicios. Cuando se les pide que desarrollen funciones o manejen temas de nivel avanzado, dan cierta información pero muestran dificultades para conectar las ideas, manipular el tiempo y aspecto y usar estrategias comunicativas como los circunloquios (Breiner-Sanders, 2000: 16).

Este extracto intenta mostrar en este candidato estas características del nivel (1:45):

- C.- Soy de Cracovia.
C.- Es una bonita ciudad al sur de Polonia.
C.- Es muy turística, es tiene un millón habitans y es muy antigua, fue una capital de Polonia. Es ciudad de reyes, entonces tenemos castillo, muchas iglesias y muchos bares. Trescientos bares.
E.- ¿Cómo España?
C.- Como España, es una ciudad con una grande universidad. Es más vieja que Universidad de Granada.
E.- Y tú ¿estudias allí?
C.- Yo estu-dió ahí. He terminado. Y ahora...
E.- ¿Qué has estudiado?
C.- He estudiado sicología.
E.- ¿Es una buena universidad?
C.- Sí. Muchas ciudades, también con Granada, copera, también con universidades de Estados Unidos...
E.- ¿Y trabajas?
C.- No, no tiene trabajo, porque he terminado mi estudios doctorales en junio y ahora yo voy a una post doc, posdoctorado estudios, y quería estar en la Universidad de Granada porque conozco a un profesor y tengo amigos y amigas aquí, entonce...
E.- ¿Quieres vivir en Granada?
C.- Por un año. Después no sé.
(...)

Habla de su ciudad, estudios, trabajo, amigos. Siempre es creativo y con suficiente calidad. Pero no puede realizar otras tareas y responder a otras preguntas del nivel avanzado, para narrar, describir, explicar u opinar. O cuando lo hace mantiene su discurso en las mismas características, sin mejorar su calidad.

Cuando se le pide que explique la situación económica de su país (9:04):

C- Es muy mal con trabajo. Es muy mal, gente no tiene trabajo y especialmente con trabajo... cuando estes (no se entiende) es no posibilidad; también si tienes doctorado hay no dinero para investigación... y el gobierno cortó dinero para investigación. Es dinero solo para cosas más importantes para ellos como agricultura y otras cosas. So es no fácil y también trabajar eh físicamente las gente tiene poquito dinero.

Con ello ha sido evaluada en la OPI de ACTFL como intermedio medio.

Y en nuestro test del MCER como B1.

4.3.4- C) Conclusión

Como vemos, los candidatos de nivel B1 (nivel 3) son intermedio medio de la OPI (nivel 5), aunque todavía hay un caso que es B1 e intermedio bajo que se comentará en §4.4, para ver sus características especiales, que le hacen alcanzar el nivel B1 con un discurso en la OPI de, aparentemente, menos nivel que el del resto de candidatos.

4.3.5 Entrevistas de nivel B2

Las entrevistas que dieron este nivel son 11. En la entrevista de ACTFL, nueve están dentro de la funcionalidad del avanzado, siendo tres de nivel intermedio alto (casi avanzado), cuatro de avanzado bajo y dos de avanzado medio. Hay dos más extrañas: simplemente intermedio medio, cuya funcionalidad es normalmente diferente a la de un avanzado.

Queremos hacer notar que los demás candidatos que han obtenido intermedio medio en la entrevista de ACTFL han tenido una calificación en la del MCER de B1. Han sido comentados en los puntos anteriores. Por ello, consideramos como casos raros a estos candidatos, núm. 15 y núm. 19, que son B2 pero al mismo tiempo intermedio medio.

Estos son sus resultados (en negrita los que van a ser comentados):

Numero y nombre	Calificación de la entrevista hecha para el MCER	Calificación de la OPI	Curso estudiado en el CLM
15. María	B2	Intermedio medio	Avanzado A
19. Maya	B2	Intermedio medio	Superior A
14. Michelle	B2	Intermedio alto	Avanzado A
17. Jorge	B2	Intermedio alto	Avanzado B
22. Ana	B2	Intermedio alto	Superior B
21. Magda	B2	Avanzado bajo	Avanzado A
16. Ana	B2	Avanzado bajo	Avanzado B
18. Lea	B2	Avanzado bajo	Avanzado B
20. Nina	B2	Avanzado bajo	Superior A
25. Cuqui	B2	Avanzado medio	Perfeccionamiento
30. Marta	B2	Avanzado medio	Reciclaje profesores

Comentamos ahora dos ejemplos de B2 que mantienen la funcionalidad lingüística dentro del avanzado: intermedio alto, candidato 14 (Michelle) y avanzado bajo, candidato 20 (Nina).

Los casos que obtienen solamente el nivel intermedio bajo los consideramos raros y se comentan en §4.4.

Tener un dominio de nivel B2 significa, por ejemplo:

Es capaz de entender las ideas principales de textos complejos que traten de temas tanto concretos como abstractos, incluso si son de carácter técnico siempre que estén dentro de su campo de especialización. Puede relacionarse con hablantes nativos con un grado suficiente de fluidez y naturalidad de modo que la comunicación se realice sin esfuerzo por parte de ninguno de los interlocutores. Puede producir textos claros y detallados sobre temas diversos así como defender un punto de vista sobre temas generales indicando los pros y los contras de las distintas opciones.

Niveles comunes de referencia: escala global (B2) (Consejo de Europa, 2002: 26).

Interacción oral	Puedo participar en una conversación con cierta fluidez y espontaneidad, lo que posibilita la comunicación normal con hablantes nativos. Puedo tomar parte activa en debates desarrollados en situaciones cotidianas explicando y defendiendo mis puntos de vista.
Expresión oral	Presento descripciones claras y detalladas de una amplia serie de temas relacionados con mi especialidad. Sé explicar un punto de vista sobre un tema exponiendo las ventajas y los inconvenientes de varias opciones.

Niveles comunes de referencia: cuadro de autoevaluación (B2) (Consejo de Europa, 2002: 30-31).

ALCANCE	CORRECCIÓN	FLUIDEZ	INTERACCIÓN	COHERENCIA
Tiene un nivel de lengua lo bastante amplio como para poder ofrecer descripciones claras y expresar puntos de vista sobre temas generales sin evidenciar la búsqueda de palabras y sabe utilizar oraciones complejas para conseguirlo.	Demuestra un control gramatical relativamente alto. No comete errores que provoquen la incomprensión y corrige casi todas sus incorrecciones.	Es capaz de producir fragmentos de discurso con un ritmo bastante uniforme; aunque puede dudar mientras busca estructuras o expresiones. Se observan pocas pausas largas.	Inicia el discurso, toma su turno de palabra en el momento adecuado y finaliza una conversación cuando tiene que hacerlo, aunque puede que no lo haga siempre con elegancia. Colabora en debates que traten temas cotidianos confirmando su comprensión, invitando a los demás a participar, etc.	Utiliza un número limitado de mecanismos de cohesión para convertir sus frases en un discurso claro y coherente, aunque puede mostrar cierto «nerviosismo» si la intervención es larga.

Niveles comunes de referencia: aspectos cualitativos del uso de la lengua hablada
(B2) (Consejo de Europa, 2002: 32-33).

Tener un nivel intermedio alto o avanzado supone estar dentro de la funcionalidad del avanzado. En intermedio alto, esta no se mantiene en todo momento pero sí en más del 50% de los casos. En avanzado, al 100%. Esta funcionalidad es, según lo apuntado en §2.6, *llevar una conversación social o laboral y tomar la iniciativa, describir, narrar, explicar y opinar en la mayor parte de franjas de aspecto y tiempo, utilizar la lengua en la mayoría de contextos informales y formales de trabajo y relaciones sociales; sobre temas y hechos de interés general y otros no relacionados directamente con ellos: viajes, modos de vida, sistema de transporte, sistema educativo, etc.; organizando el discurso en forma de párrafos.*

Vamos a comentarlas según las calificaciones obtenidas con las escalas del MCER y las *guidelines* de ACTFL.

Comentamos los candidatos 14 y 20.

Candidato 14. Michelle.	Entrevista del MCER	Entrevista de OPI
Nivel	B2	Intermedio alto
Candidato 20. Nina.	Entrevista del MCER	Entrevista de OPI
Nivel	B2	Avanzado bajo

4.3.5 -1) Entrevistas de candidatos de B2 que son intermedio alto en ACTFL

De los tres candidatos que son B2 en el test del MCER e intermedio alto en la entrevista de ACTFL, escogemos uno para mostrar y comentar sus habilidades, Michelle.

Candidato 14. Michelle.	Entrevista del MCER	Entrevista de OPI
Nivel	B2	Intermedio alto

4.3.5 -1 A Entrevista del MCER del candidato 14 (Michelle): B2

Los calificadores la califican como B2. Los comentarios de uno de ellos son así de claros, utilizando los descriptores del B2 a modo de preguntas:

¿Manifiesta ostensiblemente que debe limitar lo que quiere decir? No.
¿Hace descripciones claras? Sí.
¿Usa oraciones complejas? Sí
¿Se le nota mucho que está buscando las palabras que necesita? Un poco a veces, cuando le faltan las palabras puesta en escena y screen, (pantalla) nada más.
A mi me parece que puede ser B2.

Vamos a ver algunos ejemplos:

- Explicar una actividad o viaje interesante (6.33).

Según la escala de *Desarrollo de descripciones y narraciones*, en el nivel B2, “Realiza descripciones y narraciones claras ampliando y apoyando sus puntos de vista sobre los aspectos principales, con detalles y ejemplos adecuados” o, según la de *Compensación*, también en el nivel B2, “Utiliza circunloquios y paráfrasis para suplir

carencias de vocabulario y de estructura”, se observa que la candidata es capaz de ampliar y dar detalles en su descripción, cumpliendo los descriptores:

- C.- *Fui a Italia con mi escuela cuando tenía 16 años. Y hemos visitado, visitamos Roma y Siena y era muy bien porque hacía en ese tiempo latín, porque es que en Francia es mejor hacer latín para estudiar carreras literarias, y me ha encantado Italia porque es un país culturalmente rico.*
- E.- *¿Y tú entendiste la cultura italiana? ¿Es diferente que la francesa o que la europea...hay cosas que te llamaron la atención?*
- C.- *Pienso que sí, sí. Pienso que hay cosas que son diferentes. Por ejemplo hay más cosas de la época antigua. En Francia no hay muchas cosas de esta época y podemos decir que, no es malo para Italia, pero podemos decir que en Italia hay muchas ruinas, que son buenas visitar, pero es un país de ruina. Y he visto el Vatican (...)*

- Comentar las ventajas de estudiar español en España (8:15).

Según los comentarios de los calificadores, el candidato entra en las descripciones de las escalas de *Monólogo sostenido: argumentación* en el nivel B2, “Explica puntos de vista sobre un tema, proponiendo las ventajas y desventajas de ciertas opciones”; de *Coherencia y cohesión*, “Utiliza con eficacia una variedad de palabras de enlace para señalar con claridad las relaciones que existen entre las ideas”; así como de *Competencia lingüística general*, “Se expresa con claridad y sin manifestar ostensiblemente que tenga que limitar lo que quiere decir. Dispone de suficientes elementos lingüísticos como para hacer descripciones claras, expresar puntos de vista y desarrollar argumentos, utilizando para ello algunas oraciones complejas y sin que se le note mucho que está buscando las palabras que necesita”; y de *Flexibilidad*, “Adapta lo que dice y los medios de expresarlo a la situación y al receptor, y adopta un nivel de formalidad adecuado a las circunstancias”.

- E.- *Vivir aquí en España, ¿tú crees que es bueno para aprender y practicar español, ¿qué ventajas y qué inconvenientes tiene estar aquí?*
- C.- *Pienso que estar aquí nos hace aprender el español de todos los días y nos hace hablar y entender porque en Francia aprendemos un español de los libros. Entonces podemos conocer una lengua muy bien en Francia, ir al país y no entender casi nada. Porque los cursos no están hechos para hablar y comunicar. Es para aprender una lengua con la gramática y su cultura con los libros y las obras importantes del país, pero no es hecho para comunicar.*
- E.- *Y qué inconvenientes puede tener.*
- C.- *Sí, porque no hay el otro lado de la cultura, aquí no he estudiado, pero estoy contenta, no he estudiado Lorca y todo eso. Y es un poco, pienso que hay una diferencia mayor. Es que cuando aprendemos una lengua con una persona francesa que conoces esta lengua es más fácil para decirnos es difícil para*

vosotros aprender eso porque en francés pensamos así. Y aquí no saben porqué hacemos faltas. Entonces hay ventajas e inconvenientes. Pienso que lo mejor sería aprender una lengua con un profesor español y un profesor francés, los dos. Así podríamos aprender la lengua realmente que se habla, que podemos, para comunicar y la otra para entender a hablar como francés.

- Contar el cómic.

Igualmente al contar el cómic. Sería B2 según unas escalas:

(Fluidez oral): “Produce discursos con un ritmo bastante regular, aunque puede dudar mientras busca estructuras y expresiones; provocas pocas pausas ligeras”.

(Coherencia y cohesión): “Utiliza con eficacia una variedad de palabras de enlace para señalar con claridad las relaciones que existen entre las ideas”.

(Riqueza de vocabulario): “Varía la formulación para evitar la frecuente repetición, pero las deficiencias léxicas todavía pueden provocar vacilación y circunloquios”.

(Dominio de vocabulario): “Su precisión léxica es generalmente alta, aunque tenga alguna confusión o cometa alguna incorrección al seleccionar las palabras, sin que ello obstaculice la comunicación”.

En cuanto a su grado de *Corrección gramatical*, un corrector manifiesta que

- Tiene buen control gramatical; todavía puede cometer “deslices” esporádicos, errores no sistemáticos y pequeños fallos en la estructura de la frase, pero son escasos y a menudo puede corregirlos retrospectivamente. Pero que diferenciar entre “errores escasos” y “deslices esporádicos” resulta imposible. Finalmente decide que es B2: nos apoyamos en la segunda parte del descriptor de C1 “errores que apenas se notan” para situar a la candidata en el B2, porque sí se notan.

Vamos a ver parte de la transcripción de esta intervención del candidato describiendo el cómic (10:48):

C.- *Es una pareja... que está en un sitio romántico, con la luna y... sobre un banco...y eso, los tópicos románticos. Y ah, después se van a casa, y durante el camino para ir a casa hablan, pienso que hablan, sí hablan... Pienso que es una idealización*

C.- *Es difícil...*

E.- *Cuéntame.*

C.- *Es difícil. Hablan de al principio, si no miramos al fin hablan de la vida, de la vida de todos los días, de la vida cotidiana. Y primero vemos el hombre en la cama y su mujer trayéndole su desayuno, que es muy, eso es muy, no romántico pero con mucho cuidado. Y parece una pareja, como se dice, feliz. Después podemos ver el hombre que está apunto de ir a trabajar y la mujer le da un beso. Es una, un pareja que parece muy unido. Después podemos ver a la pareja mirar a los chicos jugar. Dos chicos, una niña y un hijo, es... como decir, la familia típica, la familia ideal, que se puede ver presenta padres y dos hijos, una mujer un chico, una niña un niño. Y después podemos ver a la familia toda junta mirar la televisión, todos juntos y parecen muy feliz. Y después*

podemos ver la pareja entrar en su casa, a la puerta principal, y al fin ah, em, no sé cómo decir, es... em, es una familia un poco extraña. No es lo que hemos visto antes es una familia con viejos, con los abuelos pienso y hay, no sé cómo decirlo...

- E.- *¿Por qué al final está eso, qué pasa en la historia, hay algo con esa familia?*
C.- *Pienso que la familia está escapando, se escapa de la vida cotidiana y cuando llegan hay todo lo que hace su cotidiano, es decir, abuelos, los abuelos y los niños y una pequeña, muy pequeña casa.*
E.- *¡Qué pena, la vida no es como nos gustaría.*
C.- *Sí claro, al principio es una vida idealizada y todo lo que podemos ver puede tenerlo, pero con los otros inconvenientes de la vida real...*

4.3.5- 1 B) OPI del candidato 14 (Michelle): intermedio alto

Si un candidato muestra una habilidad considerada de nivel intermedio alto en ACTFL, significa que puede hacer las cosas del avanzado casi todo el tiempo pero que no mantiene ni la funcionalidad ni la habilidad todo el tiempo. Es por lo tanto, casi avanzado pero no avanzado. Para mostrarlo vamos a seleccionar algún extracto donde se comprueba que es avanzado y otro donde se ve que no mantiene ese nivel. Probar el nivel en estricto exige realizar varias pruebas en el avanzado para ver en qué casos y en qué cantidad mantiene el nivel. Nosotros aquí solo damos un par de muestras. En la grabación se encuentran todas.

Se muestra como un avanzado, hablando con párrafos sobre un tema general y explicando o describiendo con detalle, cuando explica los estudios en una universidad privada en Francia; pero al mismo tiempo tiene dificultades y limitaciones que le impiden serlo efectivamente. Reduce el nivel de la comunicación a ejemplos de su propia vida o a situaciones concretas, porque le falta la soltura y habilidad para mantenerse en un contexto general (4:01):

- C.- *Sí, todo es fantástico, porque cuando entramos en estas escuelas somos funcionarios del estado y entonces después tenemos que trabajar para el estado diez años, es por eso que nos pagan antes y después debemos diez años...*
E.- *No he oído yo nunca...*
C.- *Sí, es muy francés.*
E.- *Y todos los funcionarios ¿tienen que entrar en esas escuelas?*
C.- *No, no, Hay los concursos abiertos para todos. Eso es para la elite.*
E.- *Entonces tu eres una elite o quieres ser una elite.*
C.- *No quiero. Pienso que no voy a entrar, pero son, las clases preparatorias son muy difíciles y yo quiero una buena preparación.*

- E.- *Vale y ¿por qué quieres entrar ahí y no en una universidad normal?*
 C.- *Porque en Francia voy hay una universidad que prepara bien, es muy difícil o muy laisisto, cómo se dice laisisto, no, cómo se... es francés. Cómo decir, por ejemplo este año durante tres, no, dos meses, casi tres meses no había clase y es casi siempre así. No es... La universidad es un lugar para la gente que no quieren trabajar.*
 E.- *¡Ah sí!*
 C.- *Sí, pero, es un poco complicado pero el nivel de las universidades es bajo.*
 E.- *Tú crees que la universidad francesa no funciona bien, que no tiene buen nivel, la gente no quiere trabajar...*
 C.- *Sí hay, hay gente que está muy buenos, pero en Francia cuando decimos que estamos en la universidad, mm. Pero si decimos Ah, estoy en clases preparatorias, ¡Aah!*
 E.- *Ah, ¿sí?*
 C.- *¡Aah!*

Las últimas líneas de la transcripción, respondiendo a la pregunta “¿Tú crees que la universidad francesa no funciona bien, que no tiene buen nivel, la gente no quiere trabajar?” manifiestan estas dificultades claramente (5: 35).

Tampoco informa suficientemente, le cuesta trabajo buscar las expresiones y baja el nivel con ejemplos concretos, cuando le pedimos que hable de las diferencias entre Malta y Granada (8:03):

C- No, no he podido encontrar a ventajas en Malta. Pienso que mejor venir en España. Eh, eh, es más para los pequeños y eso... Y España es más para la gente que ha viajado y.... Es muy, más, más interesante estar aquí con gente que ya ha viajado que en Malta con chicos que vienen para ser, salir de marcha todo el tiempo. Aquí sa, sa, salgo de marcha pero no es lo mismo. Am, no es hacer, salir de marcha para salir de marcha. Es salir de marcha aquí para discutir, encontrar a otras personas hacer intercambios...

Se ha calificado como intermedio alto en la OPI de ACTFL.

Y en nuestro test del MCER, B2.

4.3.5- 2) Entrevistas de candidatos de B2 que es avanzado bajo en ACTFL

De los cuatro candidatos que son B2 en el test del MCER y avanzado bajo en la entrevista de ACTFL, escogemos uno para mostrar y comentar sus habilidades, Nina.

Candidato 20. Nina.	Entrevista del MCER	Entrevista de OPI
Nivel	B2	Avanzado bajo

4.3.5- 2 A) Entrevista del MCER del candidato 20 (Nina)

Esta entrevista está calificada con nivel B2. No obstante hay algunas tareas en las que el nivel del discurso es C1, como las de “Contar un viaje” o “Explicar las ventajas de estudiar español en España”. Y una en la que la decisión parece muy difícil: “Argumentar sobre la guerra de Irak”.

En las escalas de *Corrección gramatical* y de *Competencia lingüística general* es un B2. Pero en otras, los evaluadores expresan la dificultad de tomar una decisión sobre si el nivel es B2 o C1. A veces en las escalas la diferencia conceptual entre lo que se dice en B2 y C1 es mínima o es casi lo mismo pero con otras palabras, o insuficiente como para tomar una decisión basada en el discurso concreto ante una tarea.

Así los evaluadores los califican entre B2.2 y C1 en la escala de *Coherencia y cohesión*:

- B2.2: “Utiliza con eficacia una variedad de palabras de enlace para señalar con claridad las relaciones que existen entre las ideas”.

- C1: “Produce un discurso claro, fluido y bien estructurado, mostrando un uso adecuado de criterios de organización, conectores y mecanismos de cohesión”.

O en la de *Expresión oral en general*:

- B2.2: “Realiza descripciones y presentaciones claras y sistemáticamente desarrolladas, resaltando adecuadamente los aspectos significativos y los detalles relevantes que sirvan de apoyo”.

- C1: “Realiza descripciones y presentaciones claras y detalladas sobre temas complejos, integrando otros temas, desarrollando ideas concretas y terminando con una conclusión adecuada”.

Y cuando se acude a otras escalas para solucionar la ambigüedad, los evaluadores se encuentran con que en varias de ellas el descriptor de B2 es igual que los de C1 o C2.

Vamos a ver algunos ejemplos del discurso generado por las tareas de la entrevista:

- Explicar ciertos aspectos de tu trabajo (00:40).

Esta tarea se pensó para comprobar la capacidad de explicar un procedimiento mediante descripciones claras y detalladas (B2, de la escala de *Intercambio de*

información). Algo que la estudiante es capaz de hacer, por ejemplo al explicar las normas de la piscina (4:02):

C- No se puede correr, pos supuesto, porque hay mucha agua en el área de la piscina y se puede caer, y también no se puede fumar, no se puede hacer algo para sacar la atención del salvavidas de la piscina y también no se puede ah, ... no se puede comer en el área alrededor de la piscina pero hay un área donde hay una tienda que se vende pizza, helado y refrescos y allí se puede comer pero no se puede traer la comida de allí a la piscina.

- Explicar una actividad o viaje interesante (5:20).

El viaje a Marruecos lo explica con bastante precisión y detalle e incluso combina o integra varios temas al explicar al mismo tiempo las diferencias culturales y religiosas.

Le dieron una calificación de C1:

(6:06)

C- También ahora tengo un respeto nuevo para la cultura musulmanes, muy interesante, son muy ah muy parecidos al al cristianismo, muy parecidos a los judíos, todos son, todos se usan el mismo dios pero los musulmanes creen que no necesitan, no necesita algo entre dios y tú, es como una lucha directa entre dios y tú. Es muy interesante para mi porque soy cristiana, soy católica y en la religión católica siempre tienes los obispos, tienes los curas, los curas sí, y siempre hay alguien entre tú y dios, y la iglesia entre tú y dios y la iglesia es la ruta entre tú y dios.

- Comentar las ventajas de estudiar español en España (8:28).

También en esta tarea se expresa con claridad y sin parecer que se limita lo que tiene que decir, más propio de C1:

(8:28)

C- Para mí en general no había tantas cosas malas, me gus..., me encanta Granada. Quiero regresar aquí, quiero vivir aquí algún día, sí me encanta Granada. Pero lo inconveniente era que viviendo con una familia, con una señora, ella nos trata más como, no son como familia, para ella es como negocios, estrictos negocios. Y es como un hostel, nuestra casa es como eso. No almorzamos juntos, no cenamos juntos. La comida está allí en la mesa cuando quieres, sí. Y la hija de mi señora no e muy amable, no e muy abierta. Pero he notado que la mayoría de la gente aquí en Granada son muy abiertas, muy amables, muy chistosos, ayudantes y todo. Por ejemplo, uno de mis primeros días cuando estaba perdida, una mujer pues pidió, pidió a una mujer direcciones y ella “Oh, ven conmigo, estoy en la misma dirección, ven conmigo”.

- Contar el cómic.

En esta tarea su discurso es lineal y con un vocabulario sencillo (10:32):

C- Pues los dos son una pareja y ella tiene... ah sueños de una vida perfecta con su marido y ella puede ah llevarse el desayuno en la cama y cada día él sale para el trabajo y ella cuida de los niños y los niños son perfectos y son angelitos y todos pueden ver la tele juntos...

Pero al final este candidato se ve limitado al terminar la descripción, cuando apenas puede describir la situación real de la pareja que vive en casa con los abuelos o los suegros (11:08):

C- Y, (risas) triste eso, (ja, ja) tienen un vida llena de problema con su piso, está completo (risa) con todos de, con los, con los padres del marido, con el padre de ella y también su niños, sus niños, los dos son mismos niños que en el sueño pero los dos no tienen, no son angelitos, ja ja, aunque su piso, es limpio más o menos, pero es lleno, muy completo, y no tienen la vida perfecto que quieren.

- Argumentar a favor o en contra de una guerra (14:00).

En esta tarea el discurso estaría entre B2 y C1, con la dificultad de que las escalas de *Desarrollo de descripciones y narraciones, Coherencia y cohesión, Fluidez oral*, etc. en sí no ayudan mucho. Si se decide, como es el caso, que es B2.2, es antes por el uso de otras escalas más formales, de *Dominio de vocabulario* o de *Corrección gramatical*, por la falta de vocabulario, porque los errores no son esporádicos deslices o porque los errores no son escasos y se notan.

C- He notado que los Estados Unidos cuando quieren hacer algo no siempre terminan con su causa, no siempre se quedarse con la causa al final. Como en Afganistán había la guerra y Bush dijo que “vamos a traer la democracia a los afga, a Afganistán y a construir un democracia donde todos pueden vivir en libre con libertad”. Y ahora no estábamos en Afganistán más. ¿Sí? Todo está en Irak. Y ¿quién va a soportar la democracia que es tan joven?, ¿sí? y no hay mucho apoyo ahí. Por eso las acciones no, eso es el problema con las acciones del gobierno americano, que no se quedan al final. Y en Irak ahora no tenemos mucho apoyo de los otros países y ¿cómo se puede tener una guerra sin apoyo? si solo es tú y otro ¿qué va a pasar? Nada. Necesita apoyo de muchos, de mucha gente, de muchos país, de muchos gobiernos y con sus acciones Bush está haciendo más lejos el resto del mundo, ¿sí? y los Naciones Unidos y todo. Y... pienso que sí, pienso que la democracia buena idea, maravillosa idea, pero no es para todos y es algo que necesita hacer, ah pasar más lento y más poco a poco, se puede traer la democracia a otro, no se puede usar la fuerza para “aquí su democracia, ¡tómalo! ¡No se puede hacerlo como así, él está haciendo!

Por todos estos argumentos la calificación definitiva del candidato es de B2.

4.3.5- 2 B) OPI del candidato 20 (Nina): avanzado bajo

El candidato muestra una charla fluida en la que usa párrafos todo el tiempo y realiza todas las funciones juntas con lenguaje social hablando de temas generales. Explica su experiencia ante el huracán Katrina en Alabama, donde se encontraba y explica y opina sobre la situación de los pobres y sus oportunidades educativas, temas generales propios de la evaluación en la OPI de niveles avanzados.

Veamos algún extracto que muestra este nivel propio de un avanzado bajo:

(6:00):

- E.- *Cuando tú fuiste ya había pasado el huracán, ¿no?*
C.- *Sí. No. Pues..., estaba allí. Dos días antes del huracán.*
E.- *Y ¿te fuiste?*
C.- *Sí. Porque.*
E.- *Cuéntame. Yo he visto las noticias pero tú estabas allí. ¿Había mucha información...?*
C.- *La espera no iba a empezar hasta miércoles, después del huracán. Y el huracán pasó por la ciudad lunes en la mañana. Por eso, ah, fui allí el viernes antes con mi mamá, porque tenía una casa nueva en que iba a vivir durante el año. Alquilada, sí, de una amiga, una amiga que estaba en Roma durante el semestre de otoño. Y mi compañera de cuarto estaba allí ya con sus padres. Y todos estábamos comprando los muebles, limpiando la casa..., para hacerlo, para hacerlo lista para el año para nosotros poder vivir. Y en el avión... en el aeropuerto en Chicago cuando estaba mi madre y yo estábamos esperando, había las noticias del huracán. Ah, un huracán en...cerca de Florida, cerca de Georgia, Alabama. ¡Oh!, no pasa nada. Porque los dos años pasados, eh salió de, salí de Nueva Orleans a causa de huracanes y nunca pasó nada. Eso tiempo el sábado cuando nos levantamos en la mañana, los padres de mi amiga dicen, "Salen ahora, necesitan salir a causa del huracán". "¡Qué!, ¿qué pasó? Sólo estamos aquí para un día. ¿Qué pasó?" Y después de eso a las nueve de la mañana fuimos a la casa de nuestros amigos (...)*

(15:23)

- E.- *Porqué hay tantos pobres en Nueva Orleans. ¿Es que no hay trabajo o no quieren trabajar?*
C.- *Pienso que es los dos. Es los dos. No quieren trabajar y no hay tanto trabajo. Hay... Pero cuando vas a los bares, cuando vas a los restaurantes, siempre las personas que trabajan en comida rápida, como Mc Donalds, Burger King, todo eso, siempre son los pobres, con los que trabajan en los para lavar los platos, limpiar los cuartos en los hoteles y todo. Siempre son los pobres. Yo cuando estoy en la escuela, cuando estoy en la universidad siento como... estoy racista, soy arcaísta a causa que todos que trabajen para mi son negros. Y es difícil vivir en un ambiente como esos, sí. Porque se siente mal. Pero tienen trabajos. Son trabajos buenos, son trabajos con seguridad, de la universidad, de nuestra... Nuestra universidad es el empleador más grande de la ciudad. El empleador privado más grande.*

- E.- *¡Ah! ¿Sí?*
- C.- *Nuestra ciudad tiene, ah nuestra presidente de la universidad está, no sé, en la conferencia de personas que van a ayudar a la ciudad, con... (dice nombres de personas en inglés, algo así como “Ronald Beige” y otro, apenas comprensibles).*
- E.- *Y esas personas que tiene trabajo, realmente ¿han tenido oportunidades para estudiar o no?*
- C.- *Eso es el problema, porque las escuelas privadas en Nueva Orleans son muy buenas pero las escuelas públicas son peores del país, son peores del país. En Louisiana y Misisipi, los dos estados son al lado y las escuelas públicas en los dos son los peores en todos los Estados Unidos. Y es el sistema de educación. Porque yo he hecho trabajo voluntario con los chicos en las escuelas públicas y es como, no tienen ganas de tener éxito, no tienen, no piensan que pueden tener éxito.*

El candidato es evaluado como intermedio alto según las *guidelines* de ACTFL.

Y B2 en nuestro test del MCER.

4.3.5- 3) Conclusión

Como vemos, los candidatos de nivel intermedio alto, avanzado bajo y avanzado medio, se califican en el test del MCER como de nivel B2 (nivel 4).

Consideramos raro el caso de obtener esa calificación de B2 siendo solamente intermedio bajo, pues todos los demás candidatos de ese nivel eran B1. Se comentará en §4.4 para ver sus características especiales, que les hacen alcanzar el nivel B2 con un discurso en la OPI de, aparentemente, menos nivel que el del resto de candidatos.

4.3.6 Entrevistas de nivel C1

Sólo una entrevista dio nivel C1. Es la 29 (María). Esta estudiante hizo un curso de reciclaje de profesores y en la entrevista de ACTFL ha sido calificada como de nivel superior. Se comentará en el apartado siguiente (§4.4) con los casos raros, porque no encuadra con los resultados normales para C1.

4.3.7 Entrevistas de nivel C2

Las entrevistas que dieron este nivel son cuatro, la 26 (Patricia), la 32 (Fátima), la 31 (María) y la 33 (Camila). Las candidatas 26 y 32, calificadas como avanzado alto en la entrevista de ACTFL realizaron un curso de lengua de nivel de perfeccionamiento y superior alto, respectivamente; las otras dos tienen la calificación de superior en la entrevista de ACTFL y realizaron un curso de lengua de nivel de perfeccionamiento.

Estos son los resultados de las entrevistas (en negrita las que se van a comentar):

Numero y nombre	Calificación de la entrevista hecha para el MCER	Calificación de la OPI	Curso estudiado en el CLM
32. Fátima	C2	Avanzado alto	Superior B
26. Patricia	C2	Avanzado alto	Perfeccionamiento
31. María	C2	Superior	Perfeccionamiento
33. Camila	C2	Superior	Perfeccionamiento

Tener un dominio de nivel C2 significa, por ejemplo:

Es capaz de comprender con facilidad prácticamente todo lo que oye o lee. Sabe reconstruir la información y los argumentos procedentes de diversas fuentes, ya sean en lengua hablada o escrita, y presentarlos de manera coherente y resumida. Puede expresarse espontáneamente, con gran fluidez y con un grado de precisión que le permite diferenciar pequeños matices de significado incluso en situaciones de mayor complejidad.

Niveles comunes de referencia: escala global (C2) (Consejo de Europa, 2002: 26).

Interacción oral	Tomo parte sin esfuerzo en cualquier conversación o debate y conozco bien modismos, frases hechas y expresiones coloquiales. Me expreso con fluidez y transmito matices sutiles de sentido con precisión. Si tengo un problema, sorteo la dificultad con tanta discreción que los demás apenas se dan cuenta.
Expresión oral	Presento descripciones o argumentos de forma clara y fluida y con un estilo que es adecuado al contexto y con una estructura lógica y eficaz que ayuda al oyente a fijarse en las ideas importantes y a recordarlas.

Niveles comunes de referencia: cuadro de autoevaluación (C2) (Consejo de Europa, 2002: 30-31).

ALCANCE	CORRECCIÓN	FLUIDEZ	INTERACCIÓN	COHERENCIA
Muestra una gran flexibilidad al reformular ideas diferenciando formas lingüísticas para transmitir con precisión matices de sentido, enfatizar, diferenciar y eliminar la ambigüedad. También tiene un buen dominio de expresiones idiomáticas y coloquiales.	Mantiene un consistente dominio gramatical de un nivel de lengua complejo, aunque su atención esté pendiente de otros aspectos (por ejemplo, de la planificación o del seguimiento de las reacciones de otros).	Se expresa espontánea y detalladamente con fluidez natural y coloquial, evitando o sorteando la dificultad con tanta discreción que los demás apenas se dan cuenta.	Participa en una conversación con facilidad y destreza, captando y utilizando claves no verbales y de entonación sin esfuerzo aparente. Interviene en la conversación esperando su turno, dando referencias, haciendo alusiones, etc., de forma natural.	Creación de un discurso coherente y cohesionado, haciendo un uso completo y apropiado de estructuras organizativas variadas y de una amplia serie de conectores y de otros mecanismos de cohesión.

Niveles comunes de referencia: aspectos cualitativos del uso de la lengua hablada, (C2) (Consejo de Europa, 2002: 32-33).

A pesar de lo que estas escalas y las otras que se utilizan para la calificación describen respecto del C2, un argumento utilizado por los evaluadores para calificarlos como C2 es que el discurso parece el de un español o alguien que sabe perfectamente el español. Sin prestar atención a si hace funcionalmente la serie de cosas que se describen en las escalas. Es decir, sin atender a los criterios que están establecidos. De este punto trataremos enseguida, en el apartado dedicado a los problemas y carencias del MCER en la evaluación (§4.3.1).

Ser superior en ACTFL equivale, como hemos apuntado en el punto anterior, a que se puede discutir extensamente cualquier tema defendiendo la opinión y haciendo conjeturas e hipótesis; en la mayoría de situaciones formales e informales; sobre una amplia gama de temas de interés general concretos y abstractos y algunos campos de su propia especialización y experiencia; y organizando el discurso en forma de párrafos organizados hasta componer un discurso extendido.

Ser avanzado alto es mantener esa funcionalidad casi todo el tiempo, un 50% del tiempo, pero sin ser capaz de utilizar ese tipo de discurso y habilidad en todo momento.

Vamos a comentar las entrevistas del candidato 33, según las calificaciones obtenidas en nuestro test del MCER y la OPI de ACTFL.

Candidato 33. Camila.	Entrevista del MCER	Entrevista de OPI
Nivel	C2	Superior

4.3.7- A) Entrevista del MCER del candidato 33 (Camila): C2

Los evaluadores califican a este candidato como C2:

- El discurso de la alumna es totalmente natural y propio de un hablante nativo, ni siquiera en el acento o el tono observamos características que no son propias de un hablante que tiene el español como lengua materna. La coherencia a lo largo de todas las tareas es de manera lineal, gramática perfecta, variedad léxica y capacidad para crear e imaginar respuestas completamente contextualizadas, en las que ejerce, además, una crítica a la altura de un estudiante totalmente formado, con opiniones convincentes con su manera de pensar.

- Encaja en el nivel C2 y como tal, puede desarrollar unas tareas de perfeccionamiento de español que pudieran serle de utilidad en campos de español más especializados como cultura, literatura, ciencia o cualquier otra rama.

Veamos alguna parte de su discurso. Con un ejemplo es suficiente porque se ve enseguida su soltura.

- Explicar una actividad o viaje interesante (5:21):

E.- Cuando fui a Oaxaca. Porque había leído mucho acerca de Oaxaca. Y más que nada de las mujeres oaxaqueñas y que Oaxaca es un lugar matriarcal. Entonces también a mí eso me llamó mucho la atención porque en la cultura tan machista de Méjico para mí que me hubieran dicho que Oaxaca es un lugar matriarcal se me hizo...

C.- Y ¿era... como...?

E.- Es un poblado al sur de Méjico que se llama, bueno, dentro del estado de Oaxaca, que se llama Cuchitán (...) que incluso (nombre ininteligible) ha escrito mucho acerca de ella, de ese estado. Y...para mí fue un lugar mágico porque... sí, en verdad las mujeres son grandes, son estoicas, son fuertes, son ellas las que mandan los hombres igual pasan ... a pescar, pero al regresar la que vende el pescado, la que vende la fruta, la que manda, la que administra es la mujer (...).

4.3.7- B) OPI del candidato 33 (Camila): superior

Igual que en la entrevista anterior, de nivel superior, vamos a mostrar unos extractos donde se vea que el candidato puede hacer discursos abstractos defendiendo opiniones. Algo más que un nativo, un nativo educado, requisito necesario para ser considerado de nivel superior en ACTFL.

(13:43)

- E.- *¿Crees que tienen razones los americanos para negarles esa visa, ese derecho de trabajo a los mejicanos, o no?*
- C.- *Al hacerlo perderían muchísimo dinero, porque muchísimo del dinero de la mano de obra de los mejicanos no se reporta, entonces hay muchísimo (no se entiende). Entonces al estado no le conviene. Y también yo no lo haría porque, de alguna manera se debe de parar esta ola de emigración porque Méjico tiene que hacer algo para cambiar todo es. Algo que no me gusta es que el ingreso económico más grande de Méjico son las remesas que mandan. Entonces, siento que eso hace que Méjico sea un país holgazán y dependiente, que nada más espera a que le llegue el dinerito de acá. Que es el ingreso número uno. De alguna manera se debería parar. Entonces haciendo eso, negar eso, o poner la barrera, lo que quería hacer para parar la ola de emigración. Bueno, (no se entiende), yo estoy de acuerdo, porque la gente debe de, o sea, porque incluso mucha gente muere al estar pasando, al estar intentando, entonces se debe de parar con eso. Si se les niega eso es para que se queden en Méjico y el gobierno haga algo.*
- E.- *Pero ¿tú crees que lo niegan porque piensan que...en Méjico?, están pensando en ellos mismos...*
- C.- *Claro que no piensan en Méjico, ellos piensan en su conveniencia.*

Con las *guidelines* orales de ACTFL se ha calificado como superior. Y en el test del MCER como C2.

4.3.7- C) Conclusión

Como vemos, los candidatos de nivel superior del MCER, C2 (nivel 6), son en la OPI avanzado alto o superior.

4.4 Casos raros

Vamos a analizar y comentar los casos que parecen raros, que hemos ido señalando. Estos parecen raros, en principio, porque obtienen calificaciones diferentes en una y otra entrevista, que se salen de lo que es la mayoría de casos. Esto es, por ejemplo, haber obtenido B2 siendo solamente intermedio medio, cuando lo normal es tener un nivel de intermedio alto o avanzado, o haber obtenido A1 siendo inicial medio, cuando los otros A1 son inicial alto.

Estos casos raros son los señalados en negrita en el cuadro:

- El candidato 4 (María), pues los demás candidatos calificados como inicial medio de ACTFL son calificados como de menos nivel que A1 con el MCER, y este es A1.
- El candidato 5 (Francesca), pues los demás candidatos calificados como intermedio bajo de ACTFL son calificados como de nivel A2 en el test del MCER, y este es B1.
- Los candidatos 15 (María) y 19 (Maya), pues los demás candidatos calificados como intermedio medio de ACTFL son calificados como de nivel B1 en el test del MCER, y estos son B2.
- El candidato 29 (María), pues es C1 y además superior, cuando lo normal parece ser superior y C2.

Cuadro 4.4. Cuadro donde se señalan los casos raros, en negrita.

34. Sachiyo	-A1	NM
35. Kayo	-A1	NM
4. María	A1	NM
1. Biance	A1	NH
3. Johanes	A1	NH
6. John	A1	NH
2. Stive	A2	IL
7. Carmen	A2	IL
5. Francesca	B1	IL
8. Juan	B1	IM
9. Boris	B1	IM
10. Lucia	B1	IL
15. María	B2	IM
19. Maya	B2	IM
14. Michelle	B2	IH

17. Jorge	B2	IH
22. Ana	B2	IH
16. Ana	B2	AL
18. Lea	B2	AL
20. Nina	B2	AL
21. Magda	B2	AL
25. Cuqui	B2	AM
30. Marta	B2	AM
29. María	C1	S
26. Patricia	C2	AH
Nat 32. Fátima	C2	AH
Nat 31. María	C2	S
Nat 33. Camila	C2	S

4.4.1 Entrevistas del candidato inicial medio en ACTFL, que es A1

El candidato núm. 4 (María), obtuvo en el test del MCER la calificación de A1; sin embargo en ACTFL se le dio menos calificación que al resto de candidatos que obtienen A1, solamente inicial medio. O, al revés, con solamente inicial medio de ACTFL ya se la calificó como A1 del MCER.

34. Sachiyo	-A1	NM
35. Kayo	-A1	NM
4. María	A1	NM
1. Biance	A1	NH
3. Johanes	A1	NH
6. John	A1	NH

El resto de los candidatos que obtienen A1 son inicial alto, casi intermedio, en la calificación de la OPI.

Veamos las características de su discurso:

Candidato 4. María.	Entrevista del MCER	Entrevista de OPI
Nivel	A1	Inicial medio

4.4.1- A) Entrevista del MCER del candidato número 4 (María): A1

Los evaluadores dicen que su nivel es A1, *en todas las tareas: sabe bastante vocabulario pero es incapaz de hacer frases.*

Vamos a ver algún ejemplo de su habilidad:

- Trabajo de canguro (0:25).

Aunque su primera pregunta parece muy buena, enseguida se aprecian las limitaciones que exigen que el entrevistador haga un esfuerzo por comprender:

C.- ¿Cómo se llama su hija?/ ¿Cómo uh tiene años?/ ¿Amos e amos por jugar, juga?/ Mmm... ¿la lengua el lengua e lengua es ingles y francés tu hija?/ Mnnn, ¿el trabajo de tu familia?/ ¿Cuándo días de semana? / ¿Y tu mujer?/ La hija ¿es permitir permito sus amigas a la casa?

- Explicar una actividad o viaje interesante (3:15).

C.- La Málaga es muy bueno, la ciudad muy grande y la mar muy bonito. Eh, pasear eh en la Alpujarras. Eh, también la Capileira.

E.- ¿Qué te gusta más?

C.- Me gusta más Capileira. Más tranquilo...

E.- ¿Cuándo fuiste?

C.- Tres meses pasado.

E.- ¿En coche o en autobús?

C.- Con coche y eh eh pasea eh cuatro horas con mi hermano y mi marido.

E.- ¿Qué hiciste?

C.- Cuatro horas pasa, pasea, le río, y eh, uh, las cortillos.

E.- ¿Por qué te gustó?

C.- Muy bueno la casa y la río y la...

E.- ¿Es más bonito que Holanda?

C.- (Risas) Sí.

- Comentar las ventajas de estudiar español en España (5:25).

C.- No, porque estudiar español en España es mejor.

C.- Para eh más importante por habla lacento del pueblo o en Holanda también importante hablar francés, inglés, alemán. Porque aquí es español la lengua importante so es importante socializad más fácil en la gente de España.

- Contar el cómic (7:20).

Al describir el cómic se limita a una secuencia de palabras y verbos y hay que esforzarse para entender lo que quiere decir:

C.- Dos personas eh es eh eh amor, en amorse en la las la sofa, en la parte en la luna es romántica y eh la la calle e habla hablando de desayuno de mañana y después que hablan de situación de la mañana y de la hombre va a la trabaja y la mujer es eh tiene un (no se entiende la palabra) y después de eh piensa eh de tiene dos hijos, un hija un hijo más, después eh la familia es en la sofa en la sofa y via y .. y ... no sé, la televisión... volvo vuelve, volvemos ala casa es otro situación porque la dos hijas es los eh los dos hijos pero también un tío y la las los abuelos es dormir en la cama, en la casa, so. Eh no no muy romántico.

La conclusión de los calificadores es que su habilidad general en nuestro test del MCER es de nivel A1.

4.4.1- B) OPI del candidato número 4 (María): inicial medio

Esta entrevista ha sido calificada con las *guidelines* de ACTFL como inicial medio. Esto es que la estudiante no muestra la suficiente habilidad para utilizar el material aprendido de forma creativa y formar sus propios mensajes y, cuando lo intenta, solo lo hace con grandes dificultades y no durante suficiente cantidad de tiempo, intentos o funciones insuficientes como para decir que es casi un intermedio (por lo menos el 50% del tiempo) con lo que lograría un inicial alto. No es así, y queda en inicial medio.

Damos un extracto donde se ve cómo se sirve de lo que sabe para contestar mínimamente a algunas preguntas y explicar su vida con cierto sentido, pero sin llegar a ser casi intermedio, sin mantenerse en intermedio casi todo el tiempo. Por ello se queda en inicial medio.

- E.- *Hola María, ¿De dónde eres?*
C.- *Soy de holandesa.*
E.- *¿Dónde vives en Holanda?*
C.- *No vive en Holanda ahora porque emigras-sión yo en España*
E.- *Tu familia ¿dónde vive?*
C.- *En Lecrín el valle, Valle de Lecrín.*
E.- *¿Con quien vives en España?*
C.- *Con mi marido y dos hijas.*
E.- *¿Cómo son tus hijas?*
C.- *Anna es diesiete años, tiene diecisiete años y Sara tiene tres años, trece, treche.*
E.- *¿Estudian aquí en Granada?*
C.- *Sí.*
E.- *¿Qué tal en Granada?*
C.- *Bien.*
E.- *¿Te gusta?*
C.- *Sí.*

- E.- ¿Y a tus hijas?
 C.- Sí, también, se gusta, son gusta.
 E.- Ellas ¿hablan español?
 C.- Sí pero practicar aquí es bien y mi hija Anna estudia un año la español aquí.
 E.- ¿En el CLM?
 C.- Sí en un clase de CLM.
 E.- Y tú y tu marido ¿trabajáis?
 C.- No trabajar, ahora puedo construir un hotel de eh Lecrín, le Vale de Lecrín.
 E.- Tu marido ¿es el jefe?
 C.- Otro personas construir y eh pasado año próximo año eh empieza hotel.
 E.- ¿Vais a trabajar en el hotel o volvéis a Holanda?
 C.- No, no, trabajo en hotel.
 E.- ¿Por qué queréis vivir en Granada?
 C.- Porque vive en Holanda es muy rápido y aquí es tranquilo.

Más adelante al preguntarle sobre su vida en Granada dice:

C.- Bueno, la montaña es bueno. Visitar mucho veces pero para viven es no más difícil pero es es bien, es la gente es muy amable y la restaurantes es bien, la comer es bien.

Con este tipo de habilidades ha sido evaluada en la OPI de ACTFL como inicial medio.

Y en nuestro test del MCER como A1.

4.4.2 Entrevistas de candidatos de nivel B1 que son intermedio bajo en ACTFL

Los candidatos núm. 5 (Francesca) y núm. 10 (Lucía) obtuvieron en el test del MCER la calificación de B1; sin embargo en ACTFL se les dio menos calificación que al resto de candidatos que obtienen B1, solamente intermedio bajo. O, al revés, con solamente intermedio bajo de ACTFL ya se la calificó como B1 del MCER.

2. Stive	A2	IL
7. Carmen	A2	IL
5. Francesca	B1	IL
8. Juan	B1	IM
9. Boris	B1	IM
10. Lucia	B1	IL

¿Es que estos candidatos de intermedio bajo son más intermedio bajo que los otros que se quedaron en A2?, ¿se parecen más a los que son intermedio medio, por lo que obtienen la misma calificación en el MCER que ellos, B1?

Vamos a ver las características del discurso de uno de ellos, el candidato núm. 5.

Candidato 5 Francesca.	Entrevista del MCER	Entrevista de OPI
Nivel	B1	Intermedio bajo

4.4.2- A) Entrevista del MCER del candidato 5 (Francesca): B1

El nivel de este candidato es B1. Los evaluadores dicen que necesita recurrir constantemente al italiano y carece de vocabulario suficiente pero que es capaz de cumplir con las características del nivel.

Estos son algunos ejemplos de sus tareas:

- Receta de comida (0:20).

Los evaluadores manifiestan que *describe la forma de hacer algo dando instrucciones pero usa mucho el vocabulario italiano en la descripción y enlaza series de elementos breves, sencillos para crear una secuencia con cohesión y lineal, pero de una forma muy básica.*

C- Similar a la tostada del desayuno aquí en España. Pane diferente, más mórvido. Necesita de pan, tomates y basílica. La sal eh el ajo. Tu tomar el pane y cortarlo bastante espeso. Por lo mel forno. Después el tomato sobra el pan eh tomato cortado finipino, sale, basílico, orégano, del aceite (...)

- Explicar una actividad o viaje interesante (4:20).

Respecto a esta tarea un evaluador dice:

- Cuando se pregunta a la estudiante por un viaje que haya realizado en el último mes y haya supuesto un estímulo cultural para ella no sabe bien de qué hablar exactamente, con lo que pasa a perder soltura y prácticamente habla en italiano a la vez que mucho más rápido. La calificaría como un B1.

Este es un extracto de la entrevista:

C.- A mi me gusta mucho estar en la calle y mirar a la gente, le diferente modalidad de ser de español, le le es tocar mucho la gente es contacto físico, esto creo que e una peculiaridad de la del español, que en Italia he conocido otra otra personas son francés, anglé, alemán, son diferente. Modalidad de aprocho entre personas...

- Comentar las ventajas de estudiar español en España (6:35).

Igual que en los otros casos los argumentos de los evaluadores son: que ofrece explicaciones o respuestas no muy extensas y que siempre echa mano del italiano.

C.- La cosa muy buena e poder hablar siempre en español, eh esperar de hablar español. Eh, es una residencia diferente aquí, porque aquí ya también de turistas. En la residencia donde yo fui en Verona solo estudiante italiano.

E.- ¿Pero estás con españoles, o no?

C.- Aquí sí, la primera semana he vido siempre estudiante universitario. Yo creo que a una misma cultura universitaria, europea donde se puede haber de interés comune eh más o menos culturali o más divertidi, no sé.

- Contar el cómic (8:58).

C.- Esta una historia d'amor donde un chico una chica se encuentra, se enamora, después van a casarse, eh van a tener dos hijos, eh viven en una casa eh entre mujer y dos hijo. Después yo creo que el trabajo de... no va muy bien, eh van a vivir en la casa de sus padres, muy afolada, eh entonces la la vida cambia mucho porque la casa e la familia que esperaban de haber eh no habían previsto más gente.

4.4.2- B) OPI del candidato 5 (Francesca): intermedio bajo

Esta candidato de B1, en la OPI se ha calificado como intermedio bajo. Repetimos, como hemos dado en §4.3, que ser intermedio bajo en ACTFL supone:

- Cumplir aunque de forma mínima todas las características del nivel: llevar una conversación cara a cara, haciendo y respondiendo preguntas simples, y creando y combinando el material aprendido, para hacer sus mensajes, pedir información, preguntar horarios. Y ser capaz de hacerlo en escenarios informales habituales y limitados a intercambios básicos; con temas relacionados con su medio personal e inmediato: trabajo, familia, ciudad, amigos, etc. Todo ello mediante la organización del discurso con frases sueltas.

Aunque en muchos momentos muestra un nivel de intermedio con calidad, no mantiene esta todo el tiempo ya que usa mucho italiano y al pedirle que realice el papel

de buscar trabajo de canguro y hacer preguntas no logra que estas sean buenas preguntas con calidad y cantidad suficiente. Ello la deja en intermedio bajo. Para ser intermedio medio debe mantener la calidad y cantidad en todo momento.

Veamos su forma de hablar normalmente, de intermedio (1:45)

- E.- *¿Qué planes tienes?*
C.- *Ahora no sé, porque espero de terminar el estudio a diciembre y después no sé si ir a otra ciudad o ir a Verona. No sé.*
E.- *¿Qué estudias?*
C.- *Yo ciencias de educación, estoy haciendo una especialización en pedagogía muy muy interesante.*
E.- *Y tu familia dices que vive en Varese, háblame un poco de tu familia. ¿Es una familia grande?*
C.- *Ah, un hermano y una hermana, una familia bastante grande para la típica familia italiana, sí, sí.*
E.- *¿Estudian, trabajan?*
C.- *Mi hermana e hermano, trabaja. Eh, mi padres, mi mama eh trabaja en casa e mi papa ha terminado el trabajo, no trabaja más, eh sesenta, tiene sesenta y cuatro años y en Italia no se trabaja. Yo soy la pequeña de familia, porque mi hermana a treinta años y mi hermano veintisiete. Mi hermano vive en Roma a Roma, trabaja a Roma. Mi hermana vive en Pavía, que e una ciudad cerca de Milano.*

Veamos sus preguntas, donde baja su calidad y su nivel (13.35):

- C.- *¿Cuánto años tienes la niña?*
E.- *Mi hija tien 10 años.*
C.- *¿Una sola, no?*
E.- *Sí, sólo una.*
C.- *¿Por la cena, por la merienda, qué hora empieza el trabajo?*
E.- *Por la tarde, mi mujer y yo estamos fuera lunes y miércoles. Es cuando necesitamos.*
C.- *¿Hasta...el horario?*
E.- *Desde las cinco, de cinco a ocho.*
C.- *¿Debo tomar el autobús antes por ser puntual? Vale.*
C.- *¿Qué debo solo...mm, no sé, mm, a la niña le gusta el parco, ir al parco o hacer de gimnasio, no sé?*
E.- *Pues le gusta pasear, sí, puede salir a dar un paseo. En casa también le gusta dibujar.*
C.- *¿Cuánto, cuánto el, cuánto dinero?*

Con ello ha sido evaluada en la OPI de ACTFL como intermedio bajo.

Y en nuestro test del MCER como B1.

4.4.3 Entrevistas de candidatos de B2 que son intermedio medio en ACTFL

Hay dos candidatos que son B2 en el test del MCER e intermedio medio en la entrevista de ACTFL, el núm. 15 (María) y el núm. 19 (Maya). Lo más normal parece, sin embargo, que los candidatos de intermedio medio sean B1 en el test del MCER. Los evaluadores, sin embargo, les han dado el nivel B2.

8. Juan	B1	IM
9. Boris	B1	IM
10. Lucia	B1	IM
15. María	B2	IM
19. Maya	B2	IM
14. Michelle	B2	IH
17. Jorge	B2	IH
22. Ana	B2	IH
16. Ana	B2	AL
18. Lea	B2	AL
20. Nina	B2	AL
21. Magda	B2	AL
25. Cuqui	B2	AM
30. Marta	B2	AM

Y que los candidatos que son B2 tengan en la OPI un nivel más alto o considerablemente más alto; intermedio alto, avanzado bajo o avanzado medio.

Veamos las características del discurso de estos candidatos:

Candidato 15. María.	Entrevista del MCER	Entrevista de OPI
Candidato 19. Maya.	Entrevista del MCER	Entrevista de OPI
Nivel	B2	Intermedio medio

4.4.3- 1 A) Entrevista del MCER del candidato 15 (María): B2

La calificación final del estudiante según las escalas de evaluación es un B2, porque *se expresa sin manifestar ostensiblemente que tenga que limitar lo que quiere decir (...)* dispone de suficientes elementos como para hacer descripciones claras, expresar puntos de vista y desarrollar argumentos, dos descriptores de B2 de la escala de

Competencia lingüística general. Aunque uno de los calificadores la evaluó como solamente B1, más cercana a las calificaciones obtenidas por los demás candidatos de intermedio medio.

Vamos a ver ejemplos de algunas tareas:

- Dar la opinión sobre una actividad cultural (1:55).

C.- *Muy interesante porque había mucha gente y you know, es tan divertidos, y todas la gente son bebiendo y tenía un tiempo bueno. Y en los Estados Unidos nunca hacer nada como eso. Y es muy interesante que aquí es legal y en los Estados Unidos no es legal porque en los Estados Unidos, necesito, necesitaban o tener veintiuno años para beber, y no puedes beber en las calles excepto en Las Vegas, pero solamente en el...(nombre en inglés que no se entiende) en Las Vegas, donde estén todos los hoteles.*

E.- *¿Te parece bien, qué opinas?*

C.- *Es diferente pero es la vida de Granada, creo que si no tiene clase, o si tiene clases pero pueden ir a clase y no faltas su trabajo o su escuela no es un cosa malo porque es una cosa que en donde pueden ah encontrar con personas nuevas y como ayer fui y no regresó a mi casa a hasta las seis de la mañana y conocí mucha gente. Y no es algo que pienso que puedo hacer cada noche pero si los fines de semana o una vez cada semana no es una mala cosa.*

- Explicar una actividad o viaje interesante (8:50).

C.- *Fui a ver el Sixtin Chapel y fue una experiencia que no puede describir, era muy... no puede creerlo, fue muy bonita y todo es muy magnifico. Y necesitaba esperar tres media horas en la calle y solamente estaba en la Chapel Sextina por veinte minutos pero fue muy, muy divertido.*

E.- *¿Que fue lo que te impresiono más?*

C.- *Las pinturas de Miguel Angel son increíbles y cuando caminas, cuando entró al Chapel no podía hablar, no podía hacer nada, estaba como con mi boca abierta y no tenía palabras por él y mi papa me dice, María, ¿todo es bueno? Yo no podía reaccionar.*

- Comentar las ventajas de estudiar español en España (12:36).

C.- *Creo que es más fácil para aprender español aquí porque hay muchas personas que quieren ayudarte, como encontro a chicos que querían salir y hablo con ellos y me ayudan en mi español y yo puedo ayudarte puedo ayudar en su inglés. Pero en los Estados Unidos si me dicen algo malo, mal no pueden ayudarme, pero dicen ah no eso no es; pero aquí se dice no, no, no dice como eso, pero así. Y son muy simpáticos y la gente son más... más divertidos Y me gusta porque en mi casa ahora no puedo hablar inglés porque mi señora no sabe inglés.*

Y cuando explica que pasó una semana en el hospital (14:04):

C.- También la gente en los hospitales no hablan inglés y fue muy difícil porque yo no sabía mucho español y estaba muy enferma no comí nada por eso no puedo no podía pensar y y fui difícil para or pero mm mi español fui más o menos bien y podía hablar y describir qué necesitaba.

- Contar el cómic (16:30).

C.- Hay un pareje ah creo que son ah esposos y están caminando en la calle ah soñando o pensando de su vida que quieren, la vida que quieren, como la muchacha puede traer a su esposo el desayuno en la cama y los niños jugaban buenos con los otros y será una familia normal pero cuando regresan a su casa es un niños durmiendo en la cama y los padres y todos los personas en el mismo casa y no es como sus sueño.

E.- Irónicamente ¿Qué te parece que es lo que quiere contar?

C.- Para mi es como la vida que vean en la tele y la vida que es real, porque hay muchas personas que viven con sus padres o con otra gente porque no pueden pagar por el pisa pero en la tele se miran todo es bueno todo es fácil, la vida es muy fácil para vivir, peor en la vida real no es, es muy difícil.

- Argumentar a favor o en contra de una guerra (19:17).

C.- Cuando empezó no creía que era necesaria, porque tenía, o tengo muchos amigos que están en Irak ahora y cuando empezó y muchos amigos que me quieren mucho, me amo mucho a los. Y fue muy difícil por mí porque no entiendo. Pero ahora creo que es bueno y creo que si los Estados Unidos salgan voy a ser más daño que bueno, ya más daño a la gente en Irak. Porque con los Estados Unidos Irak, o allí, la gente pueden ir a clase o hacer cosas que ah antes no podían hacer y con la ayuda de los Estados Unidos se pueden hacerlos. Y realmente creo que todos los guerras son tontos, no necesitan hacer nada porque si..., si empiezas un guerra no va a terminar, nunca va a ser paz, nunca.

Como hemos apuntado, la calificación recibida fue de B2, aunque en la OPI, como se verá en el siguiente punto, sólo alcanzó intermedio medio.

4.4.3- 1 B) OPI del candidato 15 (María): intermedio medio

Tener una calificación de intermedio medio de ACTFL supone ser capaz de realizar todo lo del nivel intermedio con suficiente soltura, calidad y cantidad y algunas cosas, aunque con problemas, del avanzado, en unas pocas ocasiones. Remitimos a §4.3.4- B, en este capítulo, o a §2.6, para una explicación más extensa de las habilidades del nivel.

Veamos algún extracto de su discurso, calificado como intermedio medio:

- Al explicar el trabajo de sus padres o el suyo tiene dificultades e incluso recurre a su lengua materna (1:45):

- C- *Ah, mi padre es el jefe de...ah una... ¿parking company? para el aeropuerto. Tiene tres, cuatro ah locks del parking y... mi mamá hace los... ah, no sé cómo se dice ¿pay checks?*
- E- *No sé, explícame.*
- C- *Mi mamá es la persona que... firman, to sign, todos los cheques que recibimos para... cuando las personas trabajan necesitan ganar dinero? y mi mamá firma los cheques.*
- E- *Ah, vale, entonces trabaja también en la compañía con tu padre. Y tú también, ¿trabajas con ellos?*
- C- *Sí. Yo soy un conversiery de los autobuses. Trabajo en los fines de semanas.*
- E- *Y ¿te gusta el trabajo?*
- C- *Me gusta muchísimo. Tengo muchas amigas y amigos en mi trabajo y es muy divertido.*
- E- *¿No es peligroso?*
- C- *No. Eh, a veces pero... es... mm... Todo el tiempo no. A veces sí pero... es peligroso porque necesito levantar los maletas y son muy heavy?*

- Al explicar un incidente no logra narrarlo adecuadamente (3:40), por lo que no alcanza el avanzado, ni siquiera el intermedio alto:

- E- *¿Has tenido algún conflicto alguna vez?*
- C- *No, porque cuando tenía un problema... me dice la persona “siéntate, cállate y no me digas nada porque si me digas algo voy a... a parar el autobús y tú necesitas caminar al aeropuerto”.*
- E- *¿De verdad?, cuéntame, ¿cómo fue, dónde?*
- C- *Estaba... manej... conduciendo por el estar que es otro de los lugares que mi padre tiene, y estaba, cerca del aeropuerto, como dos o tres blocks, pero... un hombre ah... en la atrás del autobús, ah... estaba diciendo palabras malos y hay un otra gente en el autobús y niños también y me dice eh “tú necesitas cállate” y él dice “¿Porqué? No necesito hacer nada”. Y me dice a él “si quisieras que yo parar, tú puedo. Y si tengas un problema tú puedes hablar con mi jefe, porque yo no quiero problemas y las otras personas aquí no quieren que tú hablas como eso”. Y él no me dijo nada. Calló, calló. Y me hablo con mis managers y ellos dicen, si tengas a otro problema como eso dis, dímelo.*

Por lo que su calificación es de intermedio medio, cumpliendo los requisitos de un intermedio con cantidad de información y calidad suficiente, pero no los del avanzado en suficiente medida como para llegar al intermedio alto.

4.4.3- 2 A) Entrevista del MCER del candidato 19 (Maya): B2

Este candidato está calificado como de nivel B2 en casi todas las escalas utilizadas, incluso en alguna ocasión, con las escalas de *Dominio de vocabulario* y *Corrección gramatical*, en determinadas tareas, se aprecia más nivel, C1.

Algunos de los comentarios son que, según la escala de *Fluidez*, *se comunica espontáneamente con fluidez*. Según la de *Coherencia y cohesión*, *realiza un discurso elaborado usando conectores y enlaces*. Atendiendo a la de *Competencia lingüística general*, *describe claramente su opinión sin renunciar a elementos lingüísticos para ello*. Según la escala de *Riqueza de vocabulario*, *desarrolla los argumentos con un vocabulario adecuado a los temas que trata*.

En la tarea de argumentar sobre la guerra llegaría hasta C1 según las escalas de *Dominio de vocabulario*, *porque apenas comete errores de vocabulario, quizá el único sea el usar “oil” en lugar de “petróleo”* y *Corrección gramatical*, *Manifiesta un alto grado de corrección gramatical, los errores son pocos*.

Tarea a tarea podemos ver que en varias ocasiones su nivel es alto aunque no se mantenga todo el tiempo en el mismo grado de dominio.

- Dar la opinión sobre una actividad cultural (2:54)

Respecto a esta tarea, uno de los evaluadores comenta:

-Yo creo que “amplía y apoya sus puntos de vista sobre los aspectos principales, con detalles y ejemplos adecuados” que es B2. No son sólo descripciones sencillas. Habla de varias cosas y las usa para organizar sus opiniones y argumentos dando detalles de ello.

Y otro evaluador dice:

- Según la escala de Competencia lingüística general, la estudiante posee recursos suficientes para hablar de un tema cultural, como es la visita a un museo. En tanto fluyen las ideas va desarrollando argumentos paralelos al hecho de que le ha gustado el museo, como por ejemplo cuando recomienda al entrevistador ir al museo como plan alternativo a las salidas clásicas.

Veamos un extracto:

C.- Hemos visitado el Museo de Picasso.
E.- ¿Dónde?

- C.- *En Málaga*
 E.- *¿Y te pareció interesante?*
 C.- *Muy interesante.*
 E.- *A ver, cuéntame por qué.*
 C.- *Porque en mi país no tenemos ese tipo de museos. Y sobre todo Picasso, es un poco, es muy famoso. En el museo había muchas fotos del artista, de su cuando era joven hasta su últimos días. Y también pinturas de Picasso: mujeres, niños, vasos, pinturas de vasos, muchos colores...*
 E.- *¿Te gustó?*
 C.- *Mucho.*
 E.- *¿Me recomendarías ir al museo?*
Sí, porque... es bien un poco cambiar, hacer algo un poco diferente de salir por la noche y beber o ver una película, es un poco diferente ver fotos de un artista...

- Explicar un viaje (5:54).

No tenemos nada que argumentar contra sus opiniones a la hora de hacer un trabajo objetivo en el que no podemos intervenir. No obstante nosotros pensamos que en esta tarea no se expresa con suficiente fluidez ni coherencia llegando a un callejón sin salida la final de su intervención sobre la playa de Málaga.

Transcribimos un extracto de esta parte de la entrevista.

C.- Este fin de semana fuimos a Málaga y visitamos la catedral y la casa de Picasso, el museo de Picasso y... caminamos un poco en la ciudad, descubríamos la ciudad andando. Y por la tarde fuimos a la playa, playa muy bonita, muy limpia que se llama Malagueta, con la arena muy eh, muy suave, muy...sí, y me gustaba, gusté-me gustó mucho la playa, la playa de Málaga porque... el agua es muy fresc, era muy fresco y limpio y el sol era... a las 5 estaba... el tiempo, estaba mejor que el sol, estaba menos... (silencio).

- Explicar las ventajas o desventajas de estudiar en España (8:15).

C.- En mi país eh descubre la lengua en los libros. La gramática, los vocabularios. Solamente en los libros y en general, pienso que cada persona para descubrir la lengua más debe eh debes visitar al país para descubrirla, eh: en el descubrirla en el, en el país de nacimiento. Y para practicar la lengua porque en mi país no puedo hablar eh Español, ¿no? Hablamos el árabe: es diferente. Y también en la clase no eh los cursos no son suficientes para hablar y para eh practicar la lengua suficientemente. Aquí, eh yo debo hablar español porque si no la gente no va a comprender y para comunicar debo utilizar la lengua, y por eso pienso que era un buen, una buen idea de de ser, de estar aquí en España.

- Argumentar a favor o en contra de una guerra.

El nivel descrito por los evaluadores es B2. Un evaluador argumenta:

- La alumna desarrolla varios argumentos para justificar su opinión acerca de la guerra de Irak, explicando varios puntos de vista complementarios: el de la población de Irak, del Líbano, o Bin Laden, y algunas razones que han motivado la guerra, como el petróleo y la riqueza económica. Sin embargo no acaba de desarrollar dichos argumentos sistemáticamente, por ello considero que no alcanza el nivel B2.2. Los enlaces entre las frases no son demasiado ricos, sino limitados; finalmente sólo conectan ideas yuxtapuestas. (...) mantiene un nivel general de B2. Logra desarrollar de forma exitosa ciertas ideas y expresar puntos de vista sin que se le note demasiado que, en ocasiones, busca las palabras o aplica mecanismos de evitación. En alguna ocasión aparece alguna deficiencia léxica. En este caso utiliza una palabra que la estudiante conoce, como sucede con llamar "oil" al petróleo, o se produce cierta vacilación sin consecuencias negativas para la comunicación. En cuanto a los errores son escasos (algún participio irregular, como "vuelto", "ser en conflicto" en lugar de "estar en conflicto", etc.), pero mantiene un nivel alto de control gramatical. La impresión final es que ha cumplido la tarea y ha respondido a las cuestiones del entrevistador.

Veamos un extracto de la argumentación sobre la guerra (14:58):

C- Pienso que los EE.UU. siempre hacen lo que quieren y hacen la guerra y entran en nos países para aprovechar de la riqueza. Y en el caso de Irak pienso que han aprovechado del oil del país porque es muy rico y pienso que la versión de, o el nivel del terrorismo no va a pasar de esta manera, porque tampoco como nosotros Irak ahora es un país destruido y hay un conflicto entre los dos partidos musulmanes, los sunís y los sumí, son conflicto y los estados, los americanos intentan, intentan, no intentan de acercar los dos o de... Bueno, pero en general han destruido todo Irak. Irak tiene mucha riqueza cultural y una historia muy llena de muchas cosas, como por ejemplo aquí en España hay muchas cosas culturales del pasado. Y ahora el país es totalmente destruido, la gente sufre mucho... Y Bin Laden, no sé, pienso que este hombre tiene el poder de hacer, de hacer lo que quiere también y puede preparara oro ataque terrorista sobre los EE.UU. y como nosotros los iraquíes y los libaneses vivimos en la misma región, podemos un poco, eh, tenemos los mismos problemas y que siempre los extranjeros vienen a nuestro país para molestar un poco porque ellos tienen el poder, tiene el dinero...

- Contar un cómic (11:57).

C.- Sí. Es una: pareja, un hombre y una, una mujer, sentados en, en jardín por la tarde, y están eh pensando en el futuro que van a que van a hacer y... Bueno, eh están eh visualan, mirando qué va a pasar, si por ejemplo van a casarse el uno del otro y imaginan como va a, como sus vida va a funcionar. Por la mañana la mujer va a prepararle desayuno al a su hombre y comer eh, van a comer los dos eh en la cama muy relajar- en la cama?

E.- ¡Qué bien!

C.- Y luego el marido va a ir al trabajo para trabajar y ganar eh dinero para poder vivir y después eh piensan em tener hijos y hijas y para poder pasar el tiempo con ellos y educar a poder educar a los hijos y ver la tele con ellos

E.- Aha.

C.- Y todo esto, todos estos imágenes fue- fueron en el jardín

E.- Sí.

- C.- *Y estos em, del jardín la par, los dos han regresado a la casa y la última es eh es la realidad de sus vida.*

Ha sido calificada como de nivel B2, aunque en la OPI se ha calificado como intermedio medio.

4.4.3- 2 B) OPI del candidato 19 (Maya): intermedio medio

Tener una calificación de intermedio medio de ACTFL supone ser capaz de realizar todo lo del nivel intermedio son suficiente soltura, calidad y cantidad y algunas cosas, aunque con problemas, del avanzado, en unas pocas ocasiones. Remitimos a §4.4.2- B en este capítulo o a §2.6 para una explicación más extensa de las habilidades del nivel.

Veamos algún extracto de su discurso de intermedio medio (1:50):

- E.- *¿A ti te gusta vivir allí en el Líbano? ¿Es una ciudad fácil para la vida, cómoda? ¿Cómo es?*
- C.- *Es difícil encontrar trabajo. Y tenemos muchas deudas por eh los países eh extranjeros. Tenemos problemas de la clase política, es un poco la corrupción, pero se puede vivir ah. Si tienes eh amigos, tu propia vida, la Universidad. Bueno se puede vivir, ah.*
- E.- *Y tú allí, ¿estudias o trabajas?*
- C.- *Todavía estoy estudiante.*
- E.- *¿Sí? ¿Qué estudias?*
- C.- *La traducción.*
- E.- *¿En la universidad?*
- C.- *Sí.*
- E.- *¿Sí? Y ¿por qué estudias traducción? ¿Qué quieres hacer en el futuro? ¿Qué planes tienes?*
- C.- *Porque pienso que la lengua es muy importante para poder comunicar. Y desde, es mi punto de vista Luego yo me gustaría trabajar en una no sé em-embajadora sí embajadora, o em ser una profesora de lengua.*
- E.- *¿Sí?*
- C.- *Sí.*
- E.- *Aha, vale. Y ahora, estás en España. ¿Qué tal?*
- C.- *Muy bien.*
- E.- *¿Sí?*
- C.- *Sí.*
- E.- *¿Es la primera vez que has venido a España?*
- C.- *Es mi primer viaje.*
- E.- *¿Sí? ¿Tu primer viaje?*
- C.- *En toda mi vida. Sí.*
- E.- *En toda tu vida, ¿nunca has salido del Líbano? Ah, bueno y ¿qué tal? ¿Bien?*
- C.- *Bien.*
- E.- *¿Qué te parece España, Granada? ¿Qué- qué te gusta?*

- C.- *Sí. Es un pequeñito pueblo donde podemos andar ir eh a todas las tiendas ah un poco hay muchos sitios turísticos, Albaicín, Alhambra.*
- E.- *Vale. Y ¿Cómo te sientes viviendo en España?
¿Te sientes diferente?*
- C.- *No.*
- E.- *No.*
- C.- *Ahora Muy bien. La- la gente es muy amable, sociable y , no te sientes diferente o...No.*

4.4.4 Entrevistas del candidato de nivel C1 que es superior en ACTFL

El candidato núm. 29 (María), obtuvo en el test del MCER la calificación de C1; sin embargo en ACTFL se le dio superior.

Tener un dominio de nivel C1 significa, por ejemplo:

Es capaz de comprender una amplia variedad de textos extensos y con cierto nivel de exigencia, así como reconocer en ellos sentidos implícitos. Sabe expresarse de forma fluida y espontánea sin muestras muy evidentes de esfuerzo para encontrar la expresión adecuada. Puede hacer un uso flexible y efectivo del idioma para fines sociales, académicos y profesionales. Puede producir textos claros, bien estructurados y detallados sobre temas de cierta complejidad, mostrando un uso correcto de los mecanismos de organización, articulación y cohesión del texto.

Niveles comunes de referencia: escala global (C1) (Consejo de Europa, 2002: 26).

Interacción oral	Me expreso con fluidez y espontaneidad sin tener que buscar de forma muy evidente las expresiones adecuadas. Utilizo el lenguaje con flexibilidad y eficacia para fines sociales y profesionales. Formulo ideas y opiniones con precisión y relaciono mis intervenciones hábilmente con las de otros hablantes.
Expresión oral	Presento descripciones claras y detalladas sobre temas complejos que incluyen otros temas, desarrollando ideas concretas y terminando con una conclusión apropiada.

Niveles comunes de referencia: cuadro de autoevaluación (C1) (Consejo de Europa, 2002: 30-31)

ALCANCE	CORRECCIÓN	FLUIDEZ	INTERACCIÓN	COHERENCIA
Tiene un buen dominio de una amplia serie de aspectos lingüísticos que le	Mantiene con consistencia un alto grado de corrección gramatical; los	Se expresa con fluidez y espontaneidad sin apenas esfuerzo. Sólo	Elige las frases adecuadas de entre una serie disponible de funciones del	Produce un discurso claro, fluido y bien estructurado con el que

<p>permiten elegir una formulación para expresarse con claridad y con un estilo apropiado sobre una gran variedad de temas generales, académicos, profesionales o de ocio sin tener que restringir lo que quiere decir.</p>	<p>errores son escasos, difíciles de localizar y por lo general los corrige cuando aparecen.</p>	<p>un tema conceptualmente difícil puede obstaculizar la fluidez natural de su expresión.</p>	<p>discurso para introducir sus comentarios con el fin de tomar o mantener la palabra y relacionar hábilmente sus propias intervenciones con las de los demás interlocutores.</p>	<p>demuestra un uso controlado de estructuras organizativas, conectores y mecanismos de cohesión.</p>
---	--	---	---	---

Niveles comunes de referencia: aspectos cualitativos del uso de la lengua hablada (C1) (Consejo de Europa, 2002: 32-33).

En ACTFL ser de nivel superior supone que se puede discutir extensamente cualquier tema defendiendo su opinión y haciendo conjeturas e hipótesis; en la mayoría de situaciones formales e informales; sobre una amplia gama de temas de interés general concretos y abstractos y algunos campos de su propia especialización y experiencia; y organizando el discurso en forma párrafos organizados hasta componer un discurso extendido.

Vamos a comentar las entrevistas según las calificaciones obtenidas con nuestro test del MCER y la OPI de ACTFL.

Numero y nombre	Calificación de la entrevista hecha para el MCER	Calificación de la OPI	Curso estudiado en el CLM
29. María	C1	Superior	Reciclaje de profesores

4.4.4- A) Entrevista del MCER del candidato 29 (María): C1

Los evaluadores manifiestan la dificultad de calificar discursos dentro del C1 o C2 ya que son escasas las escalas que muestran diferencias y cuando las describen son de tal matiz que es imposible verlo en la práctica. Dos deciden que es C1 pero otro cree que es

C2, aunque le gustaría tener más elementos entre los descriptores de las escalas para apoyar esta opinión suficientemente, dado que en muchas escalas se equiparan C1 y C2, sin poder encontrar diferencias. Este dice: “de los pocos descriptores que diferencian entre C1 y C2 creo que es C2, porque tiene gran riqueza y dominio de vocabulario, realiza descripciones y narraciones detalladas integrando varios temas y cumple el descriptor de C2 de *Fluidez oral*”.

Por ahora nos quedamos con que finalmente le dan una calificación de C1, basada en los siguientes comentarios y argumentos:

- Explicar una receta de comida.

Un evaluador explica que el nivel de esta tarea en este candidato puede variar según las escalas con que se analice. Así:

- *Según la escala de valoración propia de esta tarea, Intercambio de información, creemos que el nivel correspondiente a esta alumna sería un B1, porque comprobamos que, con la forma de responder al profesor con preguntas, obtiene información más detallada. Es decir, que, aunque su nivel de lengua pueda estar por encima de un B1, sin embargo, no se atiene a los parámetros que exigen los niveles superiores para resolver esta tarea. Ya que, en ningún momento, como cabría esperar, “realiza descripciones claras y detalladas acerca de la forma de llevar a cabo un procedimiento”. Sino que se limita a responder formulando cuestiones.*
- *Pero teniendo en cuenta las demás escalas de Dominio, Fluidez oral, Coherencia y cohesión, Corrección gramatical, Dominio de vocabulario, Riqueza de vocabulario etc. su nivel podría corresponder a un B2, aunque en algunas escalas, como las de Riqueza y Dominio de vocabulario se acercaría más bien a un nivel C1.*

- Dar la opinión sobre una actividad cultural.

Continúan los comentarios sobre la escasez de datos encontrados en determinadas escalas para evaluar estos discursos de nivel alto. Un evaluador, para calificarla como C1, argumenta:

- *Atendiendo a la escala indicada para la valoración de esta tarea, Intercambio información, y teniendo en cuenta el desarrollo del contenido con el que la alumna lleva a cabo dicha tarea, pensamos que la alumna no sólo relata cómo fue esa actividad cultural, en qué consistía etc., sino que, además va respondiendo a las preguntas con las que el profesor la va interrumpiendo en su discurso.*
- *Ahora bien, si nos atenemos a esta escala, pensamos que no podríamos evaluar correctamente el nivel de la alumna ya que se queda en B2. Nos quedaríamos por debajo del que, en función de otras escalas de dominio, creemos que tiene.*
- *Así pues, si nos fijamos en otras escalas de dominio, su nivel se corresponde con un nivel elevado, pues, si nos detenemos en la escala de Fluidez oral o de Riqueza de vocabulario pensamos que el nivel logrado sería un C1, igual que sucede con la de Corrección gramatical, ya que, ciertamente, los errores son escasos y apenas se notan.*

En realidad su nivel de interacción y su forma de hablar es cercano a la de un nativo, y es un gusto escucharle hablar. Veamos un fragmento de su intervención, aunque en este caso la transcripción no da cuenta de su capacidad, de su fluidez y de la búsqueda casi nativa de la negociación y co-construcción del discurso que en este caso se observan. Recomendamos la escucha de esta entrevista.

(1:31)

- C.- *Fui a este ballet de, del Romancero Gitano.*
 E.- *Vale, cuéntame un poco cómo fue eso. Tu opinión sobre esa actividad*
 C.- *Me gustó mucho, me gustó mucho.*
 E.- *Sí, ¿por qué?*
 C.- *Bueno...primero me gustó mucho el lugar donde pasó, los jardines de la... del Generalife, que es un lugar muy bonito, maravilloso, y después vimos a Cristina Hoyos que es una, bueno es una señora que baila y las castañuelas, como, como esta señora toca las castañuelas, es increíble. Y bueno yo he visto, me encantó esto, la manera de, bueno, el baile, también el vestuario, muy...*
 E.- *¿Ella cantaba o bailaba sola o era un espectáculo, qué era?*
 C.- *No... bueno....es que... ¿cómo decir? los versos de Lorca no pasaron... era como si hubiera bailado los versos de Lorca. Porque tomaba algunos fragmentos del Romancero, por ejemplo, no sé, hubo varios pero yo estoy pensando como se llama la historia de este gitano...*
 E.- *Antonio el camborio.*
 C.- *Antonio el camborio. Y, bueno, entonces bailaron sobre la historia de Antonio el camborio. Y encima me gustó mucho porque era muy discreto, sin mucho, sin,...*
 E.- *¿Estaba bien montado, te pareció una buena interpretación de Lorca?*
 C.- *Sí, para mí sí...porque, no sé, al principio pensaba que iba a ver más, no sé, no sé cómo se dice las cosas que se pone alrededor, muy barroco, no, no, nada de eso, todo era muy escueto. Había muy pocas cosas...Había los vestidos... había verde que te quiero verde.*
 (...)
 E.- *¿Cómo era, turístico, muy flamenco para turistas o era culturalmente...?*
 C.- *No, no, yo no lo vi nada, no turístico, no, no creo.*

- Explicar una actividad o viaje interesante (6:05).

Sus evaluadores dicen:

- *Tomando como referente de valoración de esta tarea la escala de Desarrollo de descripciones y narraciones, en nuestra opinión, el nivel en el que situaríamos la realización de la actividad sería un nivel C1, ya que, verdaderamente, la alumna en su discurso “realiza descripciones y narraciones detalladas integrando varios temas, desarrollando aspectos concretos y terminando con una conclusión apropiada”.*

- *En cuanto a la escala de Intercambio de información, el descriptor que se le podría asignar al resultado de la tarea sería también el correspondiente a un nivel C1 ya que, como puede apreciarse en su discurso, la alumna “comunica con total certeza*

información detallada". Porque, como se ve en los ejemplos, trata de narrar con precisión las experiencias que vivió durante su viaje a Buenos Aires, intentando extenderse en todo tipo de detalles, haciendo para ello, en muchas ocasiones, uso del estilo directo.

Veamos unos extractos:

C- *Hace tres años fui a Buenos Aires y esto fue una experiencia para mí extraordinaria y... bueno...mucho...eh...primero porque era un viaje muy largo, porque bueno, ir a Argentina es un viaje muy largo y Buenos Aires no es una ciudad exótica o sea que en Buenos Aires tú ves fragmentos de Europa: fragmentos de Italia, fragmentos de Francia, fragmentos de Inglaterra...Bueno no es una ciudad exótica, pero hay una atmósfera increíble y lo que me gustó tanto es toda esta cultura. Es que haya...no hay un barrio que no tenga su teatro, su... en fin... tú, tú das dos pasos, hay una librería, es increíble toda la cultura que hay y hay el tango.*

(...)

C- *Fui a clases de tango y me encontré con una pareja. Un señor, un chico que tenía 33 años, que hubiera podido ser mi hijo. Y, entonces voy a clase con él, yo me acuerdo que estaba con una chica que tenía 26 ó 27 y ella también quería aprender el tango y he dicho bueno, vente conmigo, vamos a prender el tango y vamos a ver a este chico que era uno de estos argentinos tan guapos, tan guapos que ella me dijo: ah, no, no voy (...)* bueno la diferencia es que yo con mi edad yo no lo veía, pero ella sí. Bueno, al final vino y aprendimos a bailar.

(...)

C- *Y entonces yo lo monté todo así y en noviembre un susto...estaba aterrorizada porque he dicho y a ver qué hacen estos chicos y si no vienen y si vienen y que sea un espectáculo horrible qué hago yo y vinieron y cuando los vi en la estación de Limoges ah...he dicho por lo menos están aquí.*

(...)

C- *Un poco de todo, pero muy bien. Cambiaban de vestido para cada cosa, en fin muy bien hecho y lo hemos pasado fenomenal. Tú te vienes a mi casa, yo te hago la comida, la cama, lo tienes todo, sólo tienes que pagar el viaje y yo te organizo espectáculos en los institutos y en los colegios.*

Parece ser que se la calificó como C1 porque comete ciertos errores, aunque apenas se le notan. He aquí un par de ejemplos de la tarea de hablar de la actividad cultural (subrayados): *no sé cómo se dice las cosas que se pone alrededor, muy barroco, no, no, nada de eso, todo era muy escueto. Había muy pocas cosas... Había los vestidos... había verde que te quiero verde.*

4.4.4- B) OPI del candidato 29 (María): superior

Si un candidato es de nivel superior, quizás sirva cualquier muestra de su discurso para mostrarlo. Pero si vamos a ser rigurosos con los criterios de ACTFL debe poder defender opiniones y hacer hipótesis utilizando discurso extendido.

Vamos a transcribir, por ello, una parte de su entrevista en la que explica los cambios que ha visto en la sociedad española desde que veraneaba de joven hasta ahora que hay una ley que permite el matrimonio homosexual, donde explica y defiende sus opiniones. Es largo así que vamos a hacer diversas catas en él.

(23:40):

- E.- *Pero te parece que va por buen camino, que todas estas cosas son de la cultura occidental en general o son peculiares de España que tiene una forma muy peculiar y muy diferente de ser...*
- C.- *Bueno yo veo que las cosas van rápido, que, que, que los cambios, bueno, fueron muy rápidos. Y a veces tenemos la impresión de que los españoles no sé, no pueden correr tan rápido para tantos cambios. Porque, porque cambiaron las cosas pero de manera muy rápida. Porque yo me acuerdo a partir del momento en que hubo democracia aquí, por ejemplo, esto es un detalle perfecto. Esto no lo hubiéramos visto antes en las tiendas de periódicos. Hubo un, un una hola de de revistas pornográficas. Esto era, bueno, es que esto no lo podía uno ver porque era...*

(24:50)

- E.- *Pero esto no lo podemos asimilar los españoles, crees.*
- C.- *No, no, no, no. Ahora yo no digo que no pero al principio, es que todo fue tan rápido, porque entre las restricciones que había antes y esta ola después de liberalismo...*

(25:20)

- E.- *Pero tú crees que es difícil de asimilar y que no se asimila, que la sociedad español es más franquista de lo que creemos...*
- C.- *No, yo te hablo de, de, de, del momento de esta curva. Ahora ya no. Yo diría más; es que yo os veo muy vanguardistas, ¿no? Porque mira, ¿quién aceptó la ley homosexual aquí? Los franceses, no. Y yo creo que no estamos apunto de ellos.*
- E.- *Y ¿a ti qué te parece? ¿Te parece bien que los homosexuales puedan casarse e incluso adoptar hijos?.*
- C.- *Bueno, a mí, bueno, bueno. A mí me da igual. Bueno, lo que digo es que yo digo cómo puede ser que sea España un país tan, tan, bueno hay que decir, tradicionalista. Porque, ¿sois vosotros los que adoptéis esta lengua, es increíble! Y nosotros no creo que no estamos apunto de hacerlo. Bueno a mí e parece que hay que vivir con su tiempo. Los homosexuales tienen sus tendencias, qué vamos a hacer. Pero ir hasta adoptar niños, creo que hay inconvenientes. Porque, bueno, yo creo que los niños necesitan al papa y a la mamá, dos papás y dos mamás nos sirven (...).*

Este candidato ha sido calificado como superior en la OPI de ACTFL, sin penalizarle los errores que se han considerado no sistemáticos o propios de un nativo, pero C1 en nuestro test del MCER, donde los mínimos errores que apenas se notan, sí se le han tenido en cuenta, siguiendo la escala de *Corrección gramatical*.

4.5 Resultados estadísticos: explicaciones y comentarios

4.5.1 Análisis estadísticos

Presentamos a continuación los resultados estadísticos que se han obtenido al analizar las calificaciones de todas las entrevistas.

El conjunto de datos analizados es de 588 y se encuentran en el Apéndice 9: las calificaciones de ACTFL de cada una de las 28 OPI y las calificaciones del test del MCER de las 28 entrevistas hechas con nuestro test del MCER (Apéndice 9.1), las calificaciones globales de cada uno de los tres evaluadores para cada candidato (Apéndice 9.2) y los juicios de cada evaluador respecto a cada tarea de nuestro test del MCER (Apéndice 9.3, 9.4, y 9.5)²³³.

Los resultados estadísticos a los que aquí vamos a referirnos, se encuentran en el Apéndice 10, en siete apartados.

1. Correlaciones de Spearman entre los resultados de la OPI y de nuestro test del MCER (Apéndice 10. 1).
2. Regresión y fórmula para averiguar la variable de ACTFL dado el nivel en el test del MCER, así como la estimación de su capacidad de predicción (Apéndice 10. 2).
3. Regresión y fórmula para averiguar la variable del MCER dado el nivel en la OPI de ACTFL, así como la estimación de su capacidad de predicción (Apéndice 10. 3).
4. y 5. Correlaciones entre evaluadores según el coeficiente de Pearson y Spearman (Apéndices 10.4 y 10.5).
6. Análisis de las medias de las tareas 5, 6, 7, y 8 y de su fiabilidad según el alfa de Cronbach (Apéndice 10.6).
7. Análisis factorial de las tareas 4, 5, 6, 7, y 8 para averiguar los factores que intervienen en ellas (Apéndice 10.7).

²³³ En el Apéndice 9. 6, se encuentran así mismo los promedios de las calificaciones de cada tarea.

Los datos obtenidos mediante el análisis estadístico muestran que:

A. Existe una alta correlación entre los resultados obtenidos en las 28 entrevistas del MCER y las 28 de la OPI. Esto es así pues el coeficiente de Spearman es de .953 (Apéndice 10. 1).

B. Respecto al nuevo test diseñado para esta investigación, el test del MCER, las tareas analizadas tienen una fiabilidad alta (Alfa de Cronbach de .993, Apéndice 10.6) y todas ellas evalúan el mismo factor o, lo que es igual, un mismo y único componente (Apéndice 10.7).

C. Respecto a los evaluadores, estos presentan una consistencia alta, tanto según el *coeficiente de correlación de Cronbach* (el primer evaluador con el segundo .98, el primero con el tercero .96 y el segundo con el tercero .95; Apéndice 10.4), como según el *coeficiente de correlación de Spearman* (el primer evaluador con el segundo .98, el primero con el tercero .96 y el segundo con el tercero .96; Apéndice 10.5).

Estos datos muestran que ambos tests, el nuestro del MCER tal y como lo hemos diseñado para nuestro trabajo, y la OPI, evalúan los mismos elementos o el mismo constructo de una forma equivalente y, por lo tanto, podrían funcionar como pruebas paralelas. Así mismo, ponen de manifiesto que el test del MCER diseñado para este trabajo es fiable en sus resultados y evalúa un solo componente. Y que los evaluadores han realizado sus calificaciones de forma consistente y objetiva utilizando el mismo conjunto de conceptos y criterios en sus calificaciones.

Como decimos en el capítulo 1, la investigación se fundamenta en obtener dos tests similares que puedan ser calificados y evaluados por escalas diferentes, con el fin de comprobar si esas escalas, de los sistemas de ACTFL y el MCER, pueden ser equiparadas o no. El capítulo 3 (§3.3 y §3.4) explica cómo el diseño del nuevo test se hizo seleccionando tareas con las que se pudiera establecer una comparación con la OPI, para lo que nos fijamos en su contenido, el tipo de lenguaje que elicitaban, así como su complejidad, entre otras cosas. Igualmente, se diseñó el test de forma que, en su concepto y en su forma de realización y calificación, se pareciera lo más posible a la OPI (§3.5). Y este test se validó siguiendo los protocolos habituales en evaluación de lenguas (§3.6).

De los resultados que estamos comentando ahora, se deduce que el trabajo de diseño, validación y preparación de evaluadores ha tenido la consistencia y efectos deseados y ha sido adecuado para nuestros propósitos: se ha conseguido un test válido y fiable y los evaluadores han actuado con objetividad y consistencia. Ambos tests son similares en su constructo y concepto de lengua a evaluar, por lo tanto pueden ser comparados sin problemas, como si fueran dos versiones de un mismo test.

Nos queda ahora explicar y comentar los resultados de esta comparación. Ya hemos señalado en §1.2 que era imposible calificar un test con las escalas del otro o viceversa y que por ello se diseñó el nuevo test de forma similar al primero. Calificando ahora los dos tests (que pueden funcionar como uno mismo) con dos escalas diferentes, cada uno con la suya, comprobaremos en qué medida estas escalas ofrecen resultados similares o diferentes ante situaciones análogas. Si hemos comprobado que los evaluadores actuaron en sus calificaciones con suficiente fiabilidad, objetividad y consistencia, el resultado depende enteramente de las escalas utilizadas para medir los discursos generados con los tests: si las escalas utilizadas coinciden o no, dependerá de sus capacidades para extraer los resultados, dependiendo de sus conceptos de lengua y competencia lingüística, en definitiva, de lo que midan. El instrumento lo tenemos bien elaborado, el patrón con el que se asigne sentido a los datos, como se refiere Ortega al hablar de las escalas (1996: 28), es lo que vamos a juzgar en adelante.

Para esto necesitamos averiguar los valores de equiparación de ambas escalas, es decir, qué nivel obtendría un candidato en un sistema una vez que haya sido evaluado y calificado en el otro, y viceversa.

4.5.2 Equiparaciones entre los resultados obtenidos con las escalas del MCER y ACTFL

Vamos a dar las equivalencias entre ambas escalas, primero conociendo la variable de MCER, cuadro 4.5.2 a., luego conociendo la de ACTFL, cuadro 4.5.2 b. Para averiguarlas realizamos un análisis de regresión con los datos de ambos tests: conociendo los niveles del MCER, averiguamos el nivel asignado para ACTFL (Apéndice 10.2); y conociendo los niveles obtenidos en la OPI de ACTFL, averiguamos el nivel asignado para el test del MCER (Apéndice 10.3).

Las fórmulas resultantes de la regresión calculada son:

$$\text{ACTFL (x)} = 1.317 + 1.335 (\text{MCER}) \quad \text{MCER (x)} = -0.448 + 0.655 (\text{ACTFL})$$

Con estas fórmulas, partiendo de los resultados de una escala, se pueden averiguar los valores para la otra²³⁴.

Estos valores se muestran en las siguientes tablas: 4.5.2 a.: cálculo de niveles para ACTFL (conociendo los del MCER), 4.5.2 b.: cálculo de niveles para el MCER (conociendo los de ACTFL).

Cuadro 4.5.2 a. Valores para ACTFL, conociendo los datos del MCER (en negrita los valores enteros, que representan a los niveles en su punto de inicio del tramo)

MCER conocido	Valores numéricos	Correspondencia en ACTFL	Nivel de ACTFL
- A1	-0.23	1.00	NL
- A1	0.00	1.31	NL + ²³⁵
- A1	0.51	2.00	NM
A1	1.00	2.64	NM ++
A1	1.27	3.00	NH
A2	2.00	3.987 ²³⁶	NH- IL ²³⁷
A2	2.02	4.00	IL
A2	2.77	5.00	IM
B1	3.00	5.3	IM +
B1	3.52	6.00	IH
B2	4.00	6.63	IH ++
B2	4.27	7.00	AL
C1	5.00	7.992 ²³⁸	AL- AM ²³⁹
C1	5.03	8.00	AM
C1	5.78	9.00	AH
C2	6.00	9.29	AH +
C2	6.53	10.00	S

Fórmula: $\text{ACTFL} = 1.31 + 1.33 (\text{MCER})$.

(Ejemplo: si MCER es 1.00, ACTFL es: $1.31 + 1.33 (1.00) = 2.64$).

²³⁴ Para calcular los valores utilizamos solamente dos decimales, excepto para los valores de A2 y C1, como se explica en su momento.

²³⁵ En ACTFL no existe esta gradación de niveles plus, pero la hemos señalado para hacer notar la diferencia entre este (.3) y del inicio del rango. Igualmente señalo el centro del rango (.6) con dos ++.

²³⁶ Los valores cercanos a la unidad se han averiguado utilizando tres dígitos para ver hasta dónde se acercan al siguiente valor. La fórmula completa usada es $\text{ACTFL} = 1.317 + 1.335 (\text{MCER})$, que también está en el Apéndice 10.

²³⁷ Parece posible aproximar el 3.987 al 4. Esto daría la calificación de intermedio bajo.

²³⁸ Como se ha dicho, de los valores cercanos a la unidad se han averiguado tres dígitos.

²³⁹ Parece posible aproximar el 7.99 al 8 y daría la calificación de avanzado medio.

Cuadro 4.5.2 b. Valores para el MCER, conocidos los de ACTFL (en negrita los valores enteros, que representan a los niveles de las escalas en su punto de inicio del tramo)

ACTFL conocido	Valores numéricos	Correspondencia en el MCER	Nivel del MCER
NL	1.00	0.21	- A1
NM	2.00	0.86	- A1
NM	2.15	1.00	A1
NH	3.00	1.51	A1 + ²⁴⁰
NH	3.75	2.00	A2
IL	4.00	2.16	A2
IM	5.00	2.81	A2 ++
IM	5.29	3.00	B1
IH	6.00	3.46	B1 +
IH	6.83	4.00	B2
AL	7.00	4.11	B2
AM	8.00	4.76	B2 ++
AM	8.36	5.00	C1
AH	9.00	5.41	C1 +
AH	9.90	6.00	C2
S	10.00	6.06	C2

Fórmula: $MCER = -0.44 + 0.65 (ACTFL)$

(Ejemplo: si ACTFL es 1.00, MCER es: $-0.44 + 0.65 (1.00) = 0.21$).

Si juntamos ambos cuadros, se aprecia que los rangos no son equivalentes sino consecutivos o, con otras palabras, que los niveles se alcanzan en cada escala con diferente ritmo. Como se ve en el cuadro siguiente (cuadro 4.5.2 c).

Cuadro 4.5.2 c. Progresión de niveles del MCER y ACTFL (En negrita los niveles del MCER y en cursiva los de ACTFL)

<i>NL</i> __ <i>NM</i> __ A1 <i>__</i> <i>NH</i> __ A2 / <i>IL</i> <i>__</i> <i>IM</i> __ B1 <i>__</i> <i>IH</i> __ B2 <i>__</i> <i>AL</i> <i>__</i> <i>AM</i> __ C1 <i>__</i> <i>AH</i> __ C2 <i>__</i> <i>S</i> <i>__</i>
--

(Ejemplo de interpretación: A1 es NM, todavía no llega a NH; B2 es IH, todavía no llega a AL. O, NH es A1, sin llegar A2; IH es B1, todavía no llega a B2).

²⁴⁰ El MCER permite hacer subdivisiones de nivel con grados de .1 y .2 o plus. Pero los que marcamos aquí no se corresponden con los que el MCER señala. Ya que no señala estas subdivisiones para A1 ni C1, sino para A2, A2+, B1, B1+, B2 y B2+. Y tampoco existen las opciones A2 ++ y B2 ++. Nuestra marca es para hacer notar el punto del rango en el que se sitúa el nivel.

Un candidato que obtenga una calificación en el MCER puede trasladar su nota a la escala de ACTFL y viceversa. Pero esto, en la práctica, lleva a que haya diferencias de niveles entre unos candidatos con la misma nota en una escala y distinta nota en la otra, excepto el A2, que es IL y el IL que es A2.

Así:

A1 es en ACTFL NM (inicial medio), pero NM (inicial medio) es en el MCER “menos que A1”.

B1 es en ACTFL IM (intermedio medio), pero IM (intermedio medio) es en el MCER A2.

B2 es en ACTFL IH (intermedio alto), pero IH (intermedio alto) es en el MCER B1.

C1 es en ACTFL AM (avanzado medio), pero AM (avanzado medio) es en el MCER B2.

C2 es en ACTFL AH (avanzado alto), pero AH (avanzado alto) es en el MCER C1.

Por ejemplo, un candidato que en el MCER sea B1, sería en la escala de ACTFL IM (intermedio medio). Pero este candidato no se corresponde en sus calificaciones con otro que tenga IM (intermedio medio) en ACTFL, ya que en el MCER le corresponde solamente A2, sin llegar todavía al B1 del anterior candidato. B1 es IM, pero IM es solo A2.

Esto hará que, utilizando este sistema de equiparaciones, podamos encontrarnos con dos candidatos que sean IM pero uno sea solamente A2 y el otro ya B1. El primero no ha llegado todavía al B1, el segundo sí.

Sigamos aclarando el cuadro 4.5.2 c. Si interpretamos la serie podemos ver que los niveles del MCER pueden entenderse como un *crescendo* de la funcionalidad del discurso de ACTFL que iría: A1 es inicial bajo y medio (sin características de intermedio), y siendo A1 se alcanzaría el inicial alto (casi intermedio); el A2 equivaldría al intermedio bajo y alcanzaría al intermedio medio (ambos con la misma funcionalidad y sin características del avanzado); el B1 comenzaría siendo intermedio medio pero alcanzaría al intermedio alto (casi avanzado), ambos presentando ya características del avanzado, (en intermedio medio, insuficientes como para ser intermedio alto, o sea, menos del 50% del tiempo); B2 comenzaría siendo casi avanzado (intermedio alto) y llegaría al avanzado (avanzado bajo y avanzado medio); el C1 ya sería avanzado medio

y llegaría al avanzado alto (o casi superior, quizás ambos con características de superior, aunque el avanzado medio en menos del 50% del tiempo); y el C2 comenzaría siendo todavía un no superior, un avanzado alto, para alcanzar al superior.

Esto lo muestro en el siguiente cuadro, cuadro 4.5.2. d. donde damos el resultado que se alcanzaría durante el nivel; o sea, los resultados al límite más alto posible para cada nivel.

Cuadro 4.5.2 d. Solapamientos y límite de funcionalidad de los niveles de ACTFL en los niveles del MCER

MCER	Correspondencia funcional en ACTFL
- A1	NL, NM: Inicial
A1	NH: Inicial alto o casi intermedio
A2	IL, IM: Intermedio medio
B1	IM, IH: Intermedio alto o casi avanzado
B2	IH, AL: Avanzado
C1	AM, AH: Avanzado medio, alto o casi superior
C2	AH, S: avanzado alto o casi superior; y superior

Ahora, si analizamos los resultados en la dirección contraria, de ACTFL al MCER, encontramos con que los niveles se solapan (cuadro 4.5.2 e.).

Cuadro 4.5.2 e. Cuadro de solapamientos y correspondencias de los niveles del MCER en los niveles de ACTFL

ACTFL	Niveles máximos en el MCER	
Inicial bajo (NL)	“menos que A1”	- A1
Inicial medio (NM)	“menos que A1” y alcanzaría el A1	- A1 -> A1
Inicial alto (NH)	A1 y alcanzaría el A2	A1+ -> A2
Intermedio bajo (IL)	A2	A2+
Intermedio medio (IM)	A2 y alcanzaría el B1	A2 ++ -> B1
Intermedio alto (IH)	B1 y alcanzaría el B2	B1+ -> B2
Avanzado bajo (AL)	B2	B2+
Avanzado medio (AM)	B2 y alcanzaría el C1	B2+ -> C1
Avanzado alto (AH)	C1 y alcanzaría el C2	C1+ -> C2
Superior	C2	C2+

Los solapamientos se dan entre todos los niveles. Son extensiones de un nivel a varios niveles: el A1 abarca inicial medio e inicial alto; el A2 abarca inicial alto, intermedio bajo e intermedio medio; el B1 abarca intermedio medio e intermedio alto; el B2 abarca intermedio alto, avanzado bajo y avanzado medio; el C1 abarca avanzado medio y alto; y el C2 abarca avanzado alto y superior. Quedan libres el inicial bajo: “menos que A1”, el intermedio bajo: A2, el avanzado bajo: B2; y el superior: C2.

Si quisiéramos solucionar estos solapamientos deberíamos subdividir los niveles del MCER en otros nuevos A1 y A1+, A2, A2+, A2++, B1 y B1+, B2 y B2+, C1 y C1+ y C2 y C2+, como hemos marcado en la columna de la derecha del cuadro. Pero aún así, seguirían sin ser niveles definidos, ya que el A1+ sería como el A2, el A2++ como el B1, el B1+ como el B2, el B2++ como el C1 y el C1+ como el C2.

El MCER hizo un primer escalonamiento con nueve niveles que luego desechó por poco operativos en los sistemas prácticos de calificar tradicionales. Estos tenían A2+, B1+ y C1+ (Consejo de Europa, 2003: 119). Los subniveles que nosotros necesitaríamos para evitar estos solapamientos serían otros, diferentes a los determinados al diseñar las escalas del MCER (A1+, A2++ y C2+) y convertirían a la nuevas escalas en escalas de 14 niveles sin operatividad.

La progresión que hemos mostrado en los cuadros 4.5.2 d y 4.5.2 e no sirve para dar un certificado de dominio, ya que en este tipo de certificados se hace constar solamente el nivel que se posee efectivamente, aunque el discurso presente características de otro nivel o subnivel. Nosotros hemos calificado las entrevistas de una manera absoluta respecto a los niveles, como se dice en §4.2.1- b, certificando un nivel solamente cuando se cumplen y se dominan la mayoría de las características del nivel descritas en las escalas del MCER. Esto es, haciendo una evaluación en la que se otorga una certificación de dominio²⁴¹.

Pero el MCER se puede utilizar también para hacer otros tipos de evaluación (sumativa, pedagógica, de diagnosis) además de usos muy útiles en el aula de lenguas, como el portfolio. Y, con estos otros objetivos, los niveles entre ambas escalas se podrían determinar de otra manera; por ejemplo, estableciendo el nivel al que debe acudir un estudiante o el nivel de lo que debe practicar y aprender, en lugar del nivel de

²⁴¹ Los resultados de esta evaluación para certificación se presentan en el apartado siguiente, §4.5.3, (cuadros 4.5.3 a. y 4.5.3 b.).

lo que ya efectivamente domina. En este otro tipo de evaluación son útiles las subdivisiones que el MCER propone de A2.1, A2.2, B1.1, B1.2, etc. En este caso las equiparaciones serían otras²⁴². Para averiguarlas debemos volver a los cuadros 4.5.2 a. y 4.5.2 b.

Fijémonos que en el cuadro numérico 4.5.2 b al candidato de A2 e IM le corresponderían 2.81 puntos. Esto es decir que este candidato lo calificamos como A2 porque presenta todas las características del discurso de A2 y es una persona que, aunque todavía no domina suficientemente las características del discurso del B1, estaría aprendiendo y manejándolas incipientemente. Desde este otro punto de vista se encontraría en B1.1 o B1.2, y en este caso concreto, más cerca del B1.2 por tener una nota de 2.81.

Si utilizamos una evaluación para certificación, solamente le otorgamos el nivel A2. Si utilizamos una evaluación de diagnóstico, para clasificación, o pedagógica, podemos decir que se encuentra en el B1.2.

Esto es lo que vamos a mostrar en los dos apartados siguientes.

4.5.3 Conclusiones de la equiparación para una evaluación para certificación de dominio general

Tomando las fórmulas señaladas en el apartado anterior y siguiendo los criterios de corrección y certificación expuestos, los cuadros de equiparaciones son los siguientes:

²⁴² Los resultados de este otro modelo de informar respecto a los niveles se presentan en §4.5.4, (cuadros 4.5.4- a y 4.5.4- b).

Cuadro 4.5.3 a. Resultados de la equiparación entre los niveles de ACTFL y el MCER, conociendo los valores del MCER

MCER	Equivalencia en ACTFL	
	Superior \	/
C2 es:	Avanzado alto \	C2 /
C1 es:	Avanzado medio \	C1 /
	Avanzado bajo \	B2 /
B2 es:	Intermedio alto \	B2 /
B1 es:	Interm. medio \	B1 /
A2 es:	Interm. bajo \	A2 /
	Inicial alto \	A1 /
A1 es:	Inicial medio \	A1 /

(Ejemplo de interpretación: si un candidato tiene un certificado del MCER que le da la calificación de B2, se equipara a un intermedio alto de ACTFL).

Cuadro 4.5.3 b. Resultados de la equiparación entre el test del MCER y de ACTFL, conociendo los niveles de ACTFL. Para certificación²⁴³ del MCER

ACTFL	Equivalencia del MCER
Superior es:	C2
Avanzado alto (AH) es:	C1
Avanzado medio (AM) es:	B2
Avanzado bajo (AL) es:	B2
Intermedio alto (IH) es:	B1
Intermedio medio (IM) es:	A2
Intermedio bajo (IL) es:	A2
Inicial alto (NH) es:	A1
Inicial medio (NM) es:	“- A1”
Inicial bajo (NL) es:	“- A1”

(Ejemplo de interpretación: si un candidato tiene un nivel intermedio alto de ACTFL, equivaldría a un B1 del MCER²⁴⁴).

²⁴³ Para certificación se toman valores enteros: es decir, si eres avanzado alto tienes 9 puntos y con ellos te corresponde ser C1. Claro que puedes ser un avanzado alto más alto y más próximo al superior que otros candidatos, pero las calificaciones de ACTFL no dan más grados. Un avanzado alto al que se le pudieran dar en lugar de 9 puntos, 9.90 (véase la tabla 4.5.2 b.) sería ya C2. Y así con todos.

Este mismo cuadro trasladado a un cuadro de seis niveles, tipo MCER, sería:

Superior	es:	C2
Avanzado alto	es:	C1

Avanzado bajo y medio	es:	B2
Intermedio alto	es:	B1

Intermedio bajo e intermedio medio	es:	A2
Inicial alto	es:	A1

4.5.4 Conclusiones de la equiparación para una evaluación para usos pedagógicos

Los cuadros que siguen muestran las escalas del MCER subdivididas en niveles como C1.1, C1.2, etc. Esto se ha hecho para señalar el nivel en el que se encuentra un candidato o estudiante en un curso de lengua, o en el que debe ser incluido para estudiar. Ello no supone una certificación automática. Esto es, por el hecho de haber estudiado o practicado las actividades o habilidades de ese nivel, no se logran dominar, y puede ser que al terminar no logre cumplir los objetivos.

Al cuadro 4.5.2 c, con la progresión del aprendizaje, añadimos ahora los puntos donde se empezaría a estudiar un curso con las características y actividades necesarias para alcanzar el dominio de un nivel de la escala del MCER (véase cuadro 4.5.4 a). Una vez dominado un nivel, por ejemplo el A1 y, mientras no se obtiene un certificado del siguiente nivel, A2, se deben estudiar o practicar las habilidades de A2.1 y A2.2 que llevan a lograr el certificado.

²⁴⁴ Repetimos que hay una diferencia entre el resultado que se alcanza si se parte de la nota de ACTFL o de la del MCER, ya que en una escala se alcanza un nivel antes que en la otra. Esto es por lo que los dos ejemplos de interpretación de los cuadros 4.5.3 a y 4.5.3 b. son estudiantes intermedio alto, pero en un caso es B1 y en otro B2. Si eres B1 eres solo intermedio medio, y para ser intermedio alto hay que ser B2. Pero si eres intermedio alto sólo eres B1.

Cuadro 4.5.4 a. Progresión de niveles desde un punto de vista pedagógico, con subdivisiones en los niveles del MCER

NL	NM	A1	NH	A2/IL	IM	B1	IH	B2	AL	AM	C1	AH	C2	S
		A2.1	A2.2	B1.1	B1.2	B2.1	B2.2	C1.1	C1.2	C2				

Para entender los siguientes dos cuadros podemos plantearnos estas preguntas:

(1). ¿Qué nivel de contenidos o actividades tiene el curso que está estudiando un candidato, según el MCER? Cuando domine esos contenidos tendrá las habilidades para ser un hablante de ese nivel.

(2). ¿Qué nivel de ACTFL aspira a obtener el candidato tras estudiar el curso?

Cuadro 4.5.4 aa. Resultados de la equiparación entre los niveles del test de ACTFL y el del MCER, conociendo los valores del nivel del MCER. Para usos pedagógicos

(1) Curso que se estudia (MCER).	(2) Dominio de ACTFL al que aspira.	Nivel de certificado del MCER que posee mientras realiza el curso.
C2 Superior	\	C1 /
C2 Avanzado alto	\	C1 /
C1.2: Avanzado medio	\	B2 /
B2.2/C1.1: Avanzado bajo	\	B1-B2 /
B1.2/B2.1: Intermedio alto	\	A2-B1 /
B1.1: Interm. medio	\	A2 /
A2.2: Interm. bajo	\	A1 /
A2.1: Inicial alto	\	A1 /
A1: Inicial medio	\	/
A1: Inicial bajo	\	/

(Ejemplo de interpretación: si un candidato se encuentra estudiando o practicando las actividades y las habilidades correspondientes al nivel B1.2 del MCER significa que está estudiando para lograr un certificado de B1, porque ya tiene un certificado de A2, y que estaría practicando las habilidades que le posibilitarían llegar a obtener un nivel intermedio alto de ACTFL).

Cuadro 4.5.4 b Resultados de la equiparación entre los niveles obtenidos en el test del MCER y de ACTFL, conociendo los niveles de ACTFL. Para usos pedagógicos²⁴⁵

(1) Curso que se estudia (MCER)	(2) Dominio de ACTFL al que aspira.	Nivel de certificado del MCER que ya posee.
C2+ ²⁴⁶	Superior	C1
C2	Avanzado alto	C1
C1.2	Avanzado medio	B2
C1.1	Avanzado bajo	B2
B2.2	Avanzado bajo	B1
B2.1	Intermedio alto	B1
B1.2	Intermedio alto	A2
B1.1	Intermedio medio	A2
A2.2	Intermedio bajo	A1
A2.1	Inicial alto	A1
A1	Inicial bajo e inicial medio	

(Ejemplo de interpretación: si un candidato se encuentra estudiando o practicando las actividades y las habilidades correspondientes al nivel B1.2 del MCER significa que está estudiando para lograr un certificado de B1 y que estaría practicando las habilidades que le posibilitarían llegar a obtener un nivel intermedio alto de ACTFL).

4.5.5 Comparación de estos resultados con otras equiparaciones existentes.

En §2.8 se ha comentado que existen algunas equiparaciones realizadas entre los niveles del MCER y ACTFL; la de Tschirner (2005) y Buitrago (2006)²⁴⁷. También

²⁴⁶ Debemos especificar este C2 + ya que el C2 normal equivale, según los cálculos expuestos en el cuadro 4.5.2 a, al nivel avanzado alto de ACTFL solamente. Por lo que si un candidato desea obtener el nivel superior de ACTFL deberá seguir estudiando, aprendiendo o practicando las nuevas destrezas y habilidades, que no se consideran en las escalas del MCER.

²⁴⁷ Como decimos en §2.8 no hemos tenido acceso por el momento al trabajo de Mosher.

nosotros hicimos nuestro propio análisis no empírico sobre las posibles equiparaciones entre las escalas, explicado en §3.2

Voy a comparar los resultados actuales, obtenidos tras el trabajo empírico y análisis estadístico, con los suyos y con los que dábamos en §3.2.

4.5.5- A) Comparación con la equiparación de Tschirner

Como hemos señalado en §2,8, Tschirner (2005) hace una comparación entre los niveles de ACTFL y del MCER y concluye con la siguiente equiparación entre las escalas del MCER y ACTFL:

MCER			A1		A2	B1		B2	C1	C2
ACTFL	NL	NM	NH	IL	IM	IH	AL	AM	AH	S

(Tschirner, 2005).

Si partimos del nivel de ACTFL²⁴⁸ para establecer el del MCER, coinciden todos los niveles, excepto que se alcanzaría el nivel A2 y B2 antes que en los resultados de Tschirner, ya que en nuestros resultados un intermedio bajo ya es A2 y un avanzado bajo ya es B2; y en los de Tschirner se debe llegar a intermedio medio para ser A2 y a avanzado medio para ser B2.

Lo que se muestra en el siguiente cuadro, marcado en negrita:

ACTFL	NL	NM	NH	IL	IM	IH	AL	AM	AH	S
MCER Tschirner			A1		A2	B1		B2	C1	C2
MCER Esta tesis			A1	A2	A2	B1	B2	B2	C1	C2

²⁴⁸ Utilizamos nuestro cuadro 4.5.3 b, de resultados de ACTFL conocidos y para certificación.

Ya vimos en §2.8 que Tschirner no hace un análisis empírico para llegar a estas conclusiones, sino que son sus propias conclusiones. Sin embargo sus resultados no se alejan demasiado de los que hemos logrado nosotros.

Distinta suerte corren los que intenta establecer Buitrago, que exponemos en el siguiente apartado.

4.5.5- B) Comparación con la equiparación de Buitrago

Como hemos señalado en §2.8, Buitrago (2006), establece las equiparaciones entre las escalas del MCER y la de ACTFL así:

MCER	A1 Acceso	A2 Plataforma		B1 Umbral		B2 Avanzado		C1 Dominio operativo eficaz		C2 Maestría
ACTFL	1. NL	2. NM	3. NH	4. IL	5. IM	6. IH	7. AL	8. AH	9. S	

Pero nuestros resultados han dado esta otra disposición de los niveles:

MCER	A1 Acceso	A2 Plataforma		B1 Umbral		B2 Avanzado		C1 Dominio operativo eficaz		C2 Maestría
ACTFL	1. NL	2. N M	3. NH	4. IL	5. IM	6. IH	7. AL	(8.) AM	8. AH	9. S

Para Buitrago equivaldría:

- El nivel A1 con el inicial bajo (y algo de inicial medio).
- El nivel A2 con el inicial medio (y algo de inicial alto).

- El nivel B1 (algo de inicial alto) con el intermedio bajo (y algo del intermedio medio).
- El nivel B2 con el intermedio medio y alto (y algo del avanzado).
- El nivel C1 con el avanzado y avanzado alto.
- El nivel C2 con el superior (y con algo del avanzado alto).

Y nuestros resultados muestran que las equivalencias son:

- El nivel A1 con el inicial bajo y medio (y algo del alto).
- El nivel A2 con el inicial alto e intermedio bajo.
- El nivel B1 con el intermedio medio (y algo del intermedio alto).
- El nivel B2 con el intermedio alto y el avanzado bajo.
- El nivel C1 con el avanzado medio.
- El nivel C2 con el avanzado alto.
- Un C2 más alto podría equivaler al superior.

Lo que se presenta en este cuadro de forma más visual:

Buitrago	Nuestra
A1 es inicial bajo (e inicial medio)	A1 es inicial bajo y medio (y alto)
A2 es inicial medio (e inicial alto)	A2 es inicial alto e intermedio bajo
B1 es (inicial alto) inter. bajo (e inter. medio)	B1 es intermedio medio (e intermedio alto)
B2 es intermedio medio y alto (y avanzado)	B2 es intermedio alto y el avanzado bajo
C1 es avanzado y avanzado alto	C1 es avanzado medio
C2 es (avanzado alto) superior	C2 es avanzado alto

4.5.5- C) Comparación con nuestra propia opinión cualitativa

En §3.2 ya presentamos nuestras opiniones respecto a lo que podían ser los niveles de ACTFL partiendo los del MCER y viceversa, en dos cuadros que traemos de nuevo aquí (3.2.5 a y 3.2.5 b) para comparar con los resultados actuales, dados en el cuadro 4.5.2 c.

1. Comparemos el cuadro 3.2.5 a con los resultados actuales del cuadro 4.5.2 c:

(Cuadro 3.2.5 a)

Lo que en el MCER es:	En ACTFL puede ser ²⁴⁹ :		
	Tabla 5	Tabla 4	Tabla 6
C2	AM-AH	AM	AM
C1	AM	AM	AM
B2	AL-AM	AL	IH
B1	IM- (IH)	IL-IM	IM
A2	IM	NM -IL	IL
A1	NM-NH	NM-NH	NH- IL

(Cuadro 4.5.2 c)

NL__NM__A1__NH__A2/IL__IM__B1__IH__B2__AL__AM__C1__AH__C2__S__
--

En el cuadro 3.2.5 a hemos visto que si en el MCER se definían unas habilidades de A1, estas ya eran las que se definían en ACTFL en el nivel NH (en las tablas 4 y 5) y hasta IL (en la tabla 6). Ahora vemos que si un candidato tiene el nivel A1 (o si todavía no logra la calificación de A2) puede llegar en ACTFL al NH (e incluso al IL, ya que se llega a IL antes que a A2, según el cuadro 4.5.2 a²⁵⁰).

Si en el MCER se definían unas habilidades de A2, estas ya eran las que se definían en ACTFL en el nivel IL (en las tablas 4 y 6) y hasta IM (en la tabla 5). Ahora vemos que si un candidato tiene el nivel A2 (o si todavía no logra la calificación de B1) puede ser en ACTFL IL y llegar al IM.

Si en el MCER se definían unas habilidades de B1, estas ya eran las que se definían en ACTFL en el nivel IL (en la tabla 4), e IM (tablas 4, 5 y 6) y hasta IH (en la tabla 5). Ahora vemos que si un candidato tiene el nivel B1 (o si todavía no logra la calificación de B2) puede ser en ACTFL IH.

²⁴⁹ Recordamos que estas tablas 4, 5 y 6 se refieren a las analizadas en el capítulo 3, (§3.1.6), escala de *Niveles cualitativos de referencia* (§3.2.1), escala de *Niveles comunes de referencia: escala global* y §3.2.3, escala de *Autoevaluación: interacción oral*. Todas ellas mostradas en el cuadro 3.2.5 a.

²⁵⁰ A2 lograría 3,92 puntos. Todavía no los 4,00 necesarios para ser IL.

Si en el MCER se definían unas habilidades de B2, estas ya eran las que se definían en ACTFL en el nivel IH (en la tabla 6), AL (tablas 4 y 5) y hasta AM (en la tabla 5). Ahora vemos que si un candidato tiene el nivel B2 (o si todavía no logra la calificación de C1) puede ser en ACTFL AL o AM.

Si en el MCER se definían unas habilidades de C1, estas ya eran las que se definían en ACTFL en el nivel AM (en las tablas 4, 5 y 6). Ahora vemos que si un candidato tiene el nivel C1 (o si todavía no logra la calificación de C2) puede ser en ACTFL AH.

Y si en el MCER se definían unas habilidades de C2, estas ya eran las que se definían en ACTFL en el nivel AM (en las tablas 4, 5 y 6) y hasta AH (en la tabla 5). Ahora vemos que si un candidato tiene el nivel C2 todavía puede no ser el superior de ACTFL.

Estos datos coinciden en todos los casos excepto el de C1, puesto que habíamos pronosticado que sería AM y en los resultados empíricos ha salido que puede ser más, hasta AH.

2. Comparemos el cuadro 3.2.5 b con los resultados actuales del cuadro 4.5.2 c:

(Cuadro 3.2.5 a)

Lo que en ACTFL es:	En el MCER puede ser ²⁵¹ :		
	Tabla 5	Tabla 4	Tabla 6
SUPERIOR	No hay	No hay	No hay
ADV. HIGH	C2	No hay	No hay
ADV. MID	C1 C2	C1 -C2	C2
ADV. LOW	B2	B2	C1
INTERM. HIGH	B1	B2	B2
INTERM. MID	A2 B1	B1	B1
INTERM. LOW	A2	A2 B1	A1 A2
NOVICE HIGH	A1	A1 A2	A1
NOVICE MID	A1	A1 A2	No hay
NOVICE LOW	No hay	No hay	No hay

²⁵¹ Recordamos que estas tablas 4, 5 y 6 se refieren a las analizadas en el capítulo 3, (§3.1.6), escala de *Niveles cualitativos de referencia*, §3.2.1, escala de *Niveles comunes de referencia: escala global* y §3.2.3, escala de *Autoevaluación: interacción oral*. Todas ellas mostradas en el cuadro 3.2.5- b.

(Cuadro 4.5.2 c)

<i>NL</i> ___ <i>NM</i> ___ A1 ___ <i>NH</i> ___ A2/IL ___ <i>IM</i> ___ B1 ___ <i>IH</i> ___ B2 ___ <i>AL</i> ___ <i>AM</i> ___ C1 ___ <i>AH</i> ___ C2 ___ <i>S</i> ___

En el cuadro 3.2.5 b hemos visto que si en ACTFL se definían unas habilidades de NM, estas ya eran las que se definían en el MCER en el nivel A1 (en las tablas 4 y 5) y hasta A2 (en la tabla 4). Ahora vemos que si un candidato tiene el nivel NM (o si todavía no logra la calificación de NH) puede ser en el MCER A1.

Si en ACTFL se definían unas habilidades de NH, estas ya eran las que se definían en el MCER en el nivel A1 (en las tablas 4, 5 y 6) y hasta A2 (en la tabla 4). Ahora vemos que si un candidato tiene el nivel NH (o si todavía no logra la calificación de IL) puede ser en el MCER A1 (e incluso A2, ya que se llega a IL antes que a A2, según el cuadro 4.5.2 b²⁵²).

Si en ACTFL se definían unas habilidades de IL, estas ya eran las que se definían en el MCER en el nivel A2 (en las tablas 4, 5 y 6) y hasta B1 (en la tabla 4). Ahora vemos que si un candidato tiene el nivel IL (o si todavía no logra la calificación de IM) sólo puede ser en el MCER A2.

Si en ACTFL se definían unas habilidades de IM, estas ya eran las que se definían en el MCER en el nivel B1 (en las tablas 4, 5 y 6). Ahora vemos que si un candidato tiene el nivel IM (o si todavía no logra la calificación de IH) puede ser en el MCER B1.

Si en ACTFL se definían unas habilidades de IH, estas ya eran las que se definían en el MCER en el nivel B1 (en la tabla 5) o B2 (en la tabla 6). Ahora vemos que si un candidato tiene el nivel IH (o si todavía no logra la calificación de AL) puede ser en el MCER B2.

Si en ACTFL se definían unas habilidades de AL, estas ya eran las que se definían en el MCER en el nivel B2 (en las tablas 4 y 5) y hasta C1 (en la tabla 6). Ahora vemos

²⁵² Se llega al A2 con 3.75 puntos, antes de llegar al IL, que son 4.00 puntos, según el cuadro 4.5.2 b.

que si un candidato tiene el nivel AL (o si todavía no logra la calificación de AM) sólo puede ser en el MCER B2.

Si en ACTFL se definían unas habilidades de AM, estas ya eran las que se definían en el MCER en el nivel C1 (tablas 4 y 5) o hasta C2 (en las tablas 4, 5 y 6). Ahora vemos que si un candidato tiene el nivel AM (o si todavía no logra la calificación de AH) sólo puede ser en el MCER C1.

Si en ACTFL se definían unas habilidades de AH, estas ya eran las que se definían en el MCER en el nivel C2 (solamente con la tabla 5). Ahora vemos que si un candidato tiene el nivel AH (o si todavía no logra la calificación de S) puede ser en el MCER C2.

Y si en ACTFL se definían unas habilidades de S, no había definición en el MCER que abarcara tanto nivel. Y ahora vemos que si un candidato tiene el nivel S tiene más nivel que el C2. Para lograr el S hay que obtener más de los 6 puntos que es C2, 6.06.

Estos datos coinciden en todos los casos excepto el de NM, AL y AM, puesto que habíamos pronosticado que NM sería A2 con la tabla 6 y en los resultados empíricos ha salido que puede ser solamente A1; igualmente, habíamos pronosticado que AL sería hasta C1 en la tabla 6 y en los resultados empíricos ha salido que puede ser solamente B2; y habíamos pronosticado que AM sería C2 (en las tres tablas) y es solamente C1.

No obstante, esos datos eran solamente una aproximación que debíamos comprobar con datos empíricos. Y las diferencias, además de ser pocas, pueden deberse a que en la investigación llevada a cabo se han utilizado otras escalas para realizar las calificaciones y entre ellas se han utilizado escalas de *Dominio de vocabulario* y de *Corrección gramatical*, que tiene una gradación más estricta.

Por ejemplo, si un candidato llega a AM en ACTFL pero tiene errores de gramática, aunque el dominio en otras habilidades de las escalas del MCER (las de las tablas 4, 5 y 6) pueda ser el mismo que el MCER define para C2, las calificaciones dadas por los profesores correctores de las entrevistas, utilizando escalas de gramática, penalizan el resultado y lo dejan solamente en C1.

En §3.2.5, en el cuadro 3.2.5 c, se hablaba de las incongruencias de intentar equiparar las calificaciones de un candidato en una escala con calificaciones automáticas en otra, pues podíamos encontrar a dos candidatos que tuvieran el mismo nivel según un sistema y, en cambio, dos niveles diferentes según el otro sistema. Esto continúa ocurriendo ahora, como se extrae de las conclusiones 1, 2 y 3, que se presentan en el siguiente capítulo.

CAPITULO 5

CONCLUSIONES

Este capítulo expone las conclusiones de la investigación (§5.1), las interpreta (§5.2), y explica sus limitaciones así como posibles trabajos para el futuro (§5.3).

5.1 Conclusiones

De los análisis expuestos en el capítulo anterior y especialmente en §4.5.1²⁵³, extraemos las siguientes conclusiones:

1. Posibilidades de equiparación entre los resultados obtenidos con nuestro test del MCER y el de ACTFL

Este trabajo ha intentado hacer una equiparación entre la OPI y un test del MCER para averiguar si sus escalas y sus formas de evaluar se pueden equiparar y con ello determinar si un B2 del MCER, por ejemplo, equivale a un avanzado de ACTFL; y así con el resto de niveles de cada sistema y escala.

El resultado ha sido que entre las escalas del MCER y la de ACTFL es posible realizar una equiparación matemática cuya fórmula se ha expuesto en §4.5.2 y repetimos ahora en los cuadros 5.1 a y 5.1 b. La equiparación entre ambas escalas es

²⁵³ En el capítulo anterior hemos visto que los tests tienen una alta correlación entre sí (*Coefficiente de correlación de Spearman* de .953), sirven para evaluar el mismo constructo, y la labor de los calificadoros es adecuada y consistente. Son, por lo tanto, efectivos y pueden prestarse a comparación.

posible, pero no es una correspondencia unívoca ni conmutativa. Salvo en los niveles de ACTFL que se dan en la conclusión número 4, son equiparaciones aproximadas.

Que no es conmutativa significa que si un candidato es de un nivel en una escala y le corresponde un nivel en la otra (por ejemplo: B2 en el MCER es intermedio alto en ACTFL) eso no significa que un candidato que sea de ese nivel en la segunda escala (en este caso intermedio alto) sea del mismo nivel en la primera (siguiendo el ejemplo anterior: intermedio alto en ACTFL no es B2 en el MCER sino B1) (véanse cuadros 4.5.2 a y 4.5.2 b).

Que no es unívoca significa que a cada nivel de una escala no le corresponde solamente un nivel en la otra. A veces le corresponden dos o incluso tres, como se verá en las conclusiones expuestas en los puntos siguientes. En realidad, para saber qué nivel exacto tiene un candidato en el otro sistema, debe realizar su examen correspondiente.

Las fórmulas para pasar de un sistema a otro son:

Cuadro 5.1 a. Fórmula para hallar el nivel que se tendría en ACTFL

MCER	ACTFL	Fórmula
A1, A2, B1, B2, C1 o C2	¿?	$ACTFL(x) = 1.317 + 1.335 \times MCER$

Cuadro 5.1 b. Fórmula para hallar el valor que se tendría en el MCER

ACTFL	MCER	Fórmula
NL, NM, NH, IL, IM, IH, AL, AM, AH o S	¿?	$MCER(x) = -0.448 + 0.655 \times ACTFL$

Los datos para cada valor se encuentran expuestos y comentados en §4.5.2 y suponen una equiparación que no es conmutativa ni unívoca, sino progresiva, como se ve en el cuadro 4.5.2 c, que repetimos aquí. Los niveles se alcanzan a diferente ritmo y en diferente momento y no son exactamente equivalentes.

Cuadro 4.5.2 c. Progresión de niveles del MCER y ACTFL

NL__NM__A1__NH__A2=IL__IM__B1__IH__B2 / AL__AM__C1__AH__C2__S_

2. Equiparaciones entre los niveles del MCER y ACTFL

No hay ningún nivel del MCER que se corresponda solamente con uno de ACTFL. A los niveles del MCER les corresponde una horquilla de dos o tres niveles de ACTFL, como se deduce del cuadro 4.5.2 a (véase §4.5.2), que aclaramos aquí (cuadro 5.1 c):

a- A los niveles del MCER A2 y C1 les corresponden tres niveles de ACTFL:
A2 puede ser NH, IL, IM; C1 puede ser AL, AM, AH.

b- A los niveles A1, B1, B2 y C2 les corresponden dos niveles de ACTFL:
A1 puede ser NM o NH; B1 puede ser IM o IH; B2 puede ser IH o AL; C2 puede ser AH o S.

Cuadro 5.1 c. Horquilla de niveles para ACTFL

MCER	ACTFL
A1 / .	NM Inicial medio
\ .	NH Inicial alto
A2 /→	IL Intermedio bajo
\ .	IM intermedio medio
B1 / .	IH Intermedio alto
\ .	IH Intermedio alto
B2 / .	AL Avanzado bajo
\ .	AM Avanzado medio
C1 /→	AM Avanzado medio
\ .	AH Avanzado alto
C2 / .	AH Avanzado alto
\ .	S Superior

Esto hace que los niveles A2, B2 y C2 del MCER se correspondan -en el nivel más bajo - con los *high* (“alto”) de ACTFL, NH, IH y AH, respectivamente. Y en sus niveles más altos la correspondencia es A1, B1 y C1 con NH, IH y AH, respectivamente.

3. Equiparaciones entre los niveles de ACTFL y del MCER

Los niveles de ACTFL se corresponden con un nivel del MCER o les corresponde una horquilla de dos niveles, como se deduce del cuadro 4.5.2 b (expuesto en §4.5.2) que aclaramos aquí (cuadro 5.1 d):

a- Los niveles de ACTFL NL, IL, AL y S se corresponden con uno solo del MCER: NL es “menos que A1”, IL es sólo A2, AL es sólo B2 y S es sólo C2.

b- A los niveles de ACTFL NM, NH, IM, IH, AM y AH les corresponden dos niveles del MCER: NM puede ser -A1 o A1; NH puede ser A1 o A2; IM puede ser A2 o B1; IH puede ser B1 o B2; AM puede ser B2 o C1; y AH puede ser C1 o C2.

c- No hay ningún nivel de ACTFL al que le correspondan tres niveles del MCER.

Cuadro 5.1 d. Horquilla de niveles para el MCER

ACTFL	MCER
Inicial bajo NL →	-A1
Inicial medio NM /	
	\ A1
Inicial alto NH /	
Intermedio bajo IL →\	A2
Intermedio medio IM /	
	\ B1
Intermedio alto IH /	
Avanzado bajo AL →\	B2
Avanzado medio AM /	
	\ C1
Avanzado alto AH /	
	\ C2
Superior S →	C2

4. Niveles mínimos alcanzados en las correspondencias

Según las conclusiones anteriores (de los cuadros 5.1 c. y 5.1 d.) podemos establecer los mínimos que se conseguirían en las equiparaciones (cuadro 5.1 e).

Cuadro 5.1 e. Comparaciones de los mínimos niveles alcanzados en cada escala

Del MCER >	> a ACTFL	De ACTFL >	> al MCER
MCER	mínimo de ACTFL	ACTFL	mín. del MCER
C2 >	Avanzado alto (AH)	Superior >	C2
C1 >	Avanzado bajo (AL)	Avanzado alto (AH) >	C1
B2 >	Intermedio alto (IH)	Avanzado bajo y medio (AL AM) >	B2
B1 >	Intermedio medio (IM)	Intermedio alto (IH) >	B1
A2 >	Inicial alto (NH)	Intermedio bajo y medio (IL IM) >	A2
A1 >	Inicial medio (NM)	Inicial alto (NH) >	A1

Según este cuadro, los niveles “2” del MCER alcanzan, como mínimo, los niveles alto (‘high’) de ACTFL (se muestra gráficamente en el cuadro 5.1 f): A2 es, como mínimo, NH; B2 es, como mínimo, IH; C2 es, como mínimo, AH.

Y los niveles “1” del MCER son: A1 es, como mínimo, NM; B1 es, como mínimo, IM y C1 es, como mínimo, AL.

Cuadro 5.1 f. Equiparación de mínimos para ACTFL

Dado un nivel del MCER...	...se obtendría un mínimo en ACTFL de:
C2 >	> AH
C1 >	> AL
B2 >	> IH
B1 >	> IM
A2 >	> NH
A1 >	> NM

Y en la otra dirección, de ACTFL al MCER, los niveles bajo y medio ('low' y 'mid') de ACTFL alcanzan, como mínimo, los niveles "2" (A2, B2 y C2) del MCER (se muestra gráficamente en el cuadro 5.1 g): IL e IM es, como mínimo, A2; AL y AM es, como mínimo, B2; S es, como mínimo, C2.

Y los niveles altos ('high') de ACTFL alcanzan, como mínimo, los niveles "1" del MCER: NH es, como mínimo, A1; IH es, como mínimo, B1; AH es, como mínimo, C1.

Cuadro 5.1 g. Equiparación de mínimos para el MCER

Dado un nivel de ACTFL... ...se obtendría un mínimo en el MCER de:		
Superior \	C2	/
Avanzado alto \	C1	/
Avanzado medio \	B2	/
Avanzado bajo \	B2	/
Intermedio alto \	B1	/
Interm. medio \	A2	/
Interm. bajo \	A2	/
Inicial alto \	A1	/
Inicial medio \		/
Inicial bajo \		/

5. El MCER abarca un dominio de lengua más limitado que ACTFL, tanto en los niveles bajos como en los altos

El nivel C2 del MCER no es el superior de ACTFL, es simplemente avanzado alto. Aunque el nivel superior de ACTFL sí es el C2 del MCER.

Esto quiere decir que se alcanzaría el nivel avanzado alto siendo C1 y si el hablante continuara mejorando su dominio de la lengua y alcanzaría el nivel C2, pero seguiría siendo todavía avanzado alto. Así, aún siendo de nivel C2 en el MCER, no se alcanzaría el nivel superior de ACTFL salvo que el hablante adquiriera unas habilidades extras que se describen en la escala de ACTFL pero no en las del MCER.

Por ello, el límite de la escala del MCER es avanzado alto. No obstante, un hablante de nivel C2 puede llegar a ser de nivel superior de ACTFL, si adquiere unas habilidades mayores de lo que dice la escala del MCER²⁵⁴.

Igualmente, en los niveles bajos ACTFL describe habilidades, aunque mínimas, para inicial bajo y medio; el MCER, en cambio, comienza con una descripción del nivel A1 en la que ya se poseen habilidades notablemente más altas que las de los niveles inicial bajo y medio de ACTFL. Esto hace que un hablante que posea el nivel inicial bajo o inicial medio de ACTFL no encuentre todavía descriptores en el MCER. Los niveles inicial bajo y medio de ACTFL son todavía “menos que el A1”. Una vez que se domina el inicial medio y se continúan aumentando las capacidades y habilidades, llegados a un nivel de 2.64 puntos, casi el inicial alto (que serían 3.00 puntos), se encuentra en el A1.

Esto hace que todas las habilidades descritas en el MCER quepan en la escala de ACTFL sobrando sitio por abajo, pues el inicial bajo no aparece descrito en las escalas del MCER, y por arriba, pues el superior tampoco aparece. Como se veía en el cuadro 4.5.3 a, en el que los niveles del MCER están completamente dentro del triángulo invertido de la escala de ACTFL.

5.2 Discusión de las conclusiones

Las principales consecuencias de este trabajo tienen que ver con las posibilidades de uso del MCER y con el análisis de las diferencias entre este sistema y el de ACTFL.

²⁵⁴ Para ser superior en las *guidelines* de ACTFL se necesita habilidad para realizar una serie de funciones con lenguaje abstracto en forma de discurso extendido; y en el MCER no se pide eso.

Con este trabajo hemos visto que se puede definir un constructo de lengua mediante el MCER y que este sirve para evaluar la lengua utilizando sus escalas. Igualmente, que el constructo puede parecerse a otros constructos que se evalúan mediante otros sistemas, en nuestro caso la OPI. Así como que se puede proceder a formar a un conjunto de evaluadores que utilizando las escalas del MCER actúen consistentemente para calificar ese test diseñado. Y finalmente, hemos visto unas posibles equiparaciones entre ambos tests, el nuestro del MCER y la OPI, con sus limitaciones y sus horquillas de niveles.

Pero el MCER no define contenidos para ninguna actividad ni nivel y esta labor es una labor difícil que, no obstante, hemos podido realizar. Con ello sabemos que hemos participado en la polémica sobre si el MCER sirve para evaluar. Efectivamente lo hemos conseguido pero, pensamos como Fulcher, el MCER no es una herramienta que por sí sola lo permita. No es tanto un marco para trabajar sino un modelo. Y para lograr evaluar mediante él hay que salvar la distancia entre su abstracción y los contenidos, actividades y tareas concretas que se necesitan. Y esto limita mucho sus posibilidades.

Nuestros resultados, además, han mostrado que no se puede establecer una equiparación clara y unívoca, ni menos aún conmutativa, entre los niveles del MCER y de ACTFL, ni aun eligiendo un test que evalúe su mismo constructo y sea paralelo. Ambas escalas evalúan de forma diferente el mismo constructo, lo compartimentan de forma diferente y ofrecen resultados distintos²⁵⁵.

Por ello queremos avisar de la precaución que se debe tener en todo momento al hacer las comparaciones. No se pueden equiparar los dos sistemas sin explicar cómo se está procediendo: hay que explicar qué sistema de informar de las calificaciones y resultados se sigue, y se debe tener claro cómo evalúa cada sistema, sus objetivos y limitaciones.

²⁵⁵ Esto sigue la lógica de que cada test solo sirve para evaluar aquello que él mismo define y por ello es difícil equiparar resultados obtenidos mediante herramientas diferentes, como también dice Fulcher (2004 b).

5.2.1 Las escalas de habilidad oral del MCER desde el punto de vista de la funcionalidad descrita en las *guidelines* orales de ACTFL

A pesar de las horquillas de niveles señaladas, hay una forma en que se simplifican las diferencias y se pueden equiparar los contenidos y descriptores de los niveles entre ambos sistemas y escalas: atendiendo a su funcionalidad. Si los analizamos desde un punto de vista funcional, usando las descripciones funcionales de la escala de ACTFL, los niveles que se agrupan parece que lo hacen porque todos ellos suponen unas habilidades y unas características similares.

Así, por ejemplo, el nivel A1 sería equivalente a un casi intermedio, con sus características funcionales sin dominar completamente: hasta el inicial alto.

El nivel A2, que puede ser tanto inicial alto como intermedio bajo o intermedio medio, quedaría como un nivel en el que las características funcionales básicas serían las del intermedio, ya que ser inicial alto supone poseerlas, aunque no al 100%, ser intermedio bajo supone poseerlas al 100%, y ser intermedio medio supone poseerlas con muestras de calidad y cantidad. En los tres niveles sería el mismo elemento el que actuaría; el A2 los conjuga en uno solo.

Igualmente el nivel B1, que puede ser tanto intermedio medio como intermedio alto, quedaría como un nivel “no avanzado” pero con sus características funcionales. Estas se pueden dar desde que se superan las del intermedio, tanto en intermedio medio, donde pueden apuntar, aunque mínimamente, como en intermedio alto, donde se dan pero no al 100%. Pero no se llega a alcanzar el dominio del avanzado hasta que no se dominan en su totalidad.

El nivel B2 continúa con las características funcionales del avanzado, aunque estas ya han comenzado en el B1. Este B2 alcanza desde que se poseen en una alta medida, más del 50% del tiempo, siendo intermedio alto (que también puede ser un B1), hasta que se dominan completamente, en el avanzado bajo.

El nivel C1 sigue en esta funcionalidad de avanzado, pero con calidad y cantidad, desde el mínimo exigible, avanzado bajo, hasta el avanzado medio.

Y el nivel C2 entra en la funcionalidad del superior, desde que esta aparece en el avanzado alto hasta que se logra completamente en el superior.

Lo hemos apuntado en §4.5.2 y en el cuadro 4.5.2 d.

Esta equiparación no sería exacta, pero supone un acercamiento que equipara los 6 niveles del MCER a los 10 de ACTFL, reduciendo estos a solamente 7, como en el siguiente cuadro.

Cuadro 5.2.1. Correspondencia funcional de niveles

MCER	Correspondencia funcional con ACTFL
- A1	Inicial: NL, NM
A1	No intermedio: NM, NH
A2	Intermedio: NH, IL, IM
B1	No avanzado: IM, IH
B2	Avanzado: IH, AL
C1	Avanzado con gran calidad: AL, AM, AH
C2	No superior: AH, S

Si analizamos estas correspondencias desde ACTFL al MCER se forma una escalera que superpone dos niveles del MCER en uno de ACTFL. Superpone el A1 y A2 en NH; el A2 y B1 en IM; el B1 y B2 en IH; el B2 y C1 en AM; y el C1 y C2 en AH. Pero esta escalera tiene cuatro puntos de descanso donde los niveles están claros. Curiosamente, estos se corresponden con cada uno de los niveles bajos de ACTFL, el inicial bajo, intermedio bajo, avanzado bajo y superior, que se corresponden enteramente con “menos que A1”, A2, B2 y C2 respectivamente.

5.2.2 Equiparaciones entre nuestra evaluación oral según el MCER y las *guidelines* de ACTFL, desde el punto de vista pedagógico

Otra forma de buscar que las equiparaciones tengan más sentido unívoco y a cada nivel le corresponda solamente uno en la otra escala, es diferenciar los niveles de las escalas del MCER en subniveles, tal y como propone el MCER y North explica en su estudio de escalonamiento de niveles apuntado en §2.5.3.

Si consideramos que existen los niveles A2.1 y A2.2, B1.1 y B1.2, B2.1 y B2.2, C1.1 y C1.2, entonces tenemos una escala de 10 niveles que es más fácilmente equiparable con la de ACTFL, de también 10 niveles.

Este sistema es el que hemos denominado en los análisis de las entrevistas y sus resultados expuestos en §4.5.4 y su cuadro 4.5.4 a. como pedagógico. Porque sirve para programar cursos y hacer evaluaciones progresivas o de diagnóstico y rendimiento.

Pero la equiparación resultante tampoco equivaldría exactamente a los niveles de ACTFL, como muestra el cuadro 4.5.4 b, que repetimos aquí, aunque soluciona varios de sus solapamientos. Pero junta en un solo nivel a los niveles B1.2 y B2.1, por un lado, y B2.2 y C1.1, por otro, como muestra el cuadro (en el que lo hemos señalado con negrita y los símbolos \ /):

(Cuadro 4.5.4 b. Equiparación con subdivisión de niveles del MCER)

Curso que se estudia (MCER)	Dominio de ACTFL al que aspira.	Nivel de certificado del MCER que posee mientras realiza el curso.
C2+ .	Superior	C1
C2 .	Avanzado alto	C1
C1.2 .	Avanzado medio	B2
C1.1 \ /		B2
B2.2 / \	Avanzado bajo	B1
B2.1 \ /		B1
B1.2 / \	Intermedio alto	A2
B1.1 .	Intermedio medio	A2
A2.2 .	Intermedio bajo	A1
A2.1 .	Inicial alto	A1
A1 .	Inicial medio	

Si elaboramos este cuadro con los resultados del MCER a ACTFL y de ACTFL al MCER, quedaría como sigue (cuadro 5.2.2).

Entender este cuadro supone responder a estas dos preguntas:

(1). ¿Qué nivel de contenidos o actividades tiene el curso que está estudiando un candidato, según el MCER? Cuando domine esos contenidos tendrá las habilidades para ser un hablante de ese nivel.

(2). ¿Qué nivel de ACTFL aspira a obtener el candidato tras estudiar el curso?

Cuadro 5.2.2. Equiparaciones pedagógicas para MCER / ACTFL

Del MCER >	> a ACTFL	De ACTFL >	> al MCER
(1) Curso que estudia con el MCER	(2) ACTFL (certificado al que aspira)	(2) ACTFL (certificado al que aspira)	(1) Curso que estudia
C2+ >	Superior (S)	Superior >	C2+
C2 >	Avanzado alto (AH)	Avanzado alto >	C2
C1.2 >	Avanzado medio (AM)	Avanzado medio >	C1.2
C1.1 >	Avanzado bajo (AL)	Avanzado bajo >	C1.1
B2 2 >	Intermedio alto (IH)	Avanzado bajo >	B2.2
B2.1 >	Intermedio alto (IH)	Intermedio alto >	B2.1
B1.2 >	Intermedio medio (IM)	Intermedio medio >	B1.2
B1.1 >	Intermedio medio (IM)	Intermedio medio >	B1.1
A2.2 >	Intermedio bajo (IL)	Intermedio bajo >	A2.2
A2.1 >	Inicial alto (NH)	Inicial alto >	A2.1
A1 >	Inicial medio (NM)	Inicial bajo e inicial medio >	A1

Según esto, evaluando a un mismo candidato con ambas escalas se conseguiría una calificación mucho más rica. Si con el MCER se da una calificación, las *guidelines* de ACTFL la ampliarían incluyendo más detalles. Si la primera calificación está hecha con la OPI, las escalas del MCER aumentarían las subdivisiones dentro de los niveles ya de por sí detallados en las *guidelines* de ACTFL.

Ambas son diferentes maneras de evaluar, como se viene señalando. Complementarias si se quiere, pero diferentes.

5.2.3 Posibilidades del MCER para desarrollos académicos

En cuanto a sus posibilidades como marco de referencia para usos en el aula, estas son grandes y entre ellas está su uso como medio de establecer niveles, hacer programaciones y calificar de forma horizontal el dominio lingüístico de los estudiantes (véase §2.7.1) una vez que se le dote de los contenidos y las implementaciones teóricas necesarias (§2.7.4). Es evidente que el uso de sus casi 60 escalas ofrece una riquísima descripción de la lengua y eso hace que la enseñanza se enriquezca no solamente con materiales dispersos en nuevos libros o con los gustos de cada autor o profesor.

Pero también en este sentido hay un aspecto especialmente significativo, ya que tiene consecuencias negativas: es el hecho de que en los niveles bajos no tenga suficientes descriptores y que hayamos encontrado que los estudiantes de menor nivel, después de un mes de estudios, no logran adquirir el nivel A1. No hay descriptores para decir cuál es su nivel, como hemos comprobado en dos ocasiones en nuestra investigación, citadas y explicadas en §4.3.1.

En cuanto a su uso como material para evaluación no académica y para certificación general, el MCER en sí mismo es, antes que nada, un compendio de escalas. No hay en él información para realizar una evaluación utilizándolas y tampoco en los documentos que intentan desarrollar la evaluación se dan las técnicas que lo posibiliten (Consejo de Europa, 2002, 2005, 2007). Una vez más estamos de acuerdo con Weir, para quien no se puede evaluar según el MCER sin ser especialista. Es, por lo tanto, un documento político que intenta establecer una forma de medir, que puede ser aceptada o no según se vean interesantes las descripciones que establece en sus 6 niveles. Y según sean los intereses del evaluador.

A nosotros, como hemos visto en la investigación, nos parecen insuficientemente expresadas y con lagunas importantes a la hora de evaluar ciertos aspectos, que identifican el A2 y el B1, o el B2 con el C1 y C2. Por lo que nos parece ambiguo e insuficiente.

Sus escalas sirven para evaluar, pero no para equiparar entre sí resultados, ni entre diferentes lenguas, ni en distintos contextos. Dependerá de las variables contextuales, del constructo seleccionado y de las escalas de medición utilizadas. Por ello creemos que el MCER sirve más bien como referente para su uso en el aula o en el portfolio, pero no para evaluación oral general de dominio lingüístico. Dependiendo del constructo de test que se utilice y de las escalas que intervengan en su construcción y en su evaluación se obtendrán unos u otros resultados. Por ejemplo, si no hubiéramos utilizado las escalas de *Dominio de vocabulario* y de *Corrección gramatical* en nuestro test, se habrían equiparado los niveles de C1 y C2 e incluso el B2, ya que utilizando otras escalas hay poca información para diferenciar entre esos tres niveles.

Hemos comprobado sus limitaciones como medio de diseño de tests y para evaluación de certificación general. En este sentido, hemos visto cómo al ponderar las notas entre varias escalas para dar una evaluación general ha habido que llegar a un acuerdo sobre la forma de establecer el resultado (expuesto en §4.2.1) que no era

sencillo pues necesitaba distinguir entre, por ejemplo, qué significaba decir que un estudiante era B2.2 y qué significaba que fuera B2. La información para usos pedagógicos puede señalar que un estudiante es B2.2, pero esto equivaldría, en un sistema de certificar dominio, que es todavía B1 y no B2; y esta aparente contradicción permanecía en la mente de nuestros evaluadores y también a la hora de interpretar los resultados al establecer las equiparaciones. Es por ello por lo que queremos, para no llevar a errores, remarcar la importancia de definir claramente, desde el comienzo del proceso de diseño del test y en sus especificaciones de contenido y administración, el sistema de informar de los resultados a las partes y usuarios e informar de esta ambivalente forma de entender cada resultado.

El MCER depende de su contexto de uso y de sus constructos y sistemas de evaluar. Su uso es flexible y está abierto a todo tipo de sistemas educativos, contextos de aprendizaje y necesidades del usuario, pero esta grandeza le resta uniformidad y con ello posibilidades de equiparación y homogeneidad; su abstracción lo convierte en un modelo en vez de en un marco. Como instrumento académico es todavía limitado, mientras permanezca carente de especificaciones de contenido; y como instrumento de evaluación es todavía más limitado, pues no solo dependerá de lo que en él se defina sino también de las maneras de definirlo y ejecutarlo. Si para diseñar un test se debe ser experto en la materia y demostrar su validez y fiabilidad, una vez hecho esto, utilizar el MCER como referente, o cualquier otro de los sistemas existentes u otros nuevos, dependerá de los intereses y necesidades de cada evaluador. Y será un sistema más entre muchos otros existentes.

El uso del MCER, otras escalas o un conjunto de ellas, dependerá de las condiciones para su administración y de la naturaleza de los candidatos y sus objetivos. Usarlo como guía en los centros de enseñanza de lenguas puede ser útil y no demasiado costoso, salvando la necesidad de hacer un curso estándar de formación que ayude a entenderlo e interpretarlo, solucionando sus limitaciones y evitando crear las falsas asunciones de que el MCER supone una explicación del desarrollo de la adquisición o el aprendizaje de las lenguas. Pero usarlo para evaluación de dominio general resulta muy complicado, además de caro, lo que lo hace impracticable para los objetivos que normalmente tiene este tipo de evaluación. Antes debe crearse una agencia lingüística que sea a la vez banco de ítems y tareas a la que puedan adscribirse los diferentes operadores educativos, investigadores y lingüistas.

También es tema de discusión el hecho fundamental, traído por Alderson y ya citado en §2.7 sobre si “mi B2 es igual que tu B2”. Nuestro trabajo ha encontrado serias limitaciones a esta afirmación. Es difícil demostrarlo: ello dependerá de los objetivos que defina cada test, y de su constructo.

Sería bueno que el Consejo de Europa, mediante una agencia europea que se encargara de ello en cada lengua, siguiera estableciendo los sistemas necesarios para solucionar estas y otras ambigüedades y problemas, entre ellas la necesaria formación e interpretación homogénea de todos los usuarios, para conseguir que finalmente se pueda decir con seguridad que mi B2 es igual al tuyo. Sería bueno también terminar de definir las escalas que vimos que carecen de descriptores, haciendo un esfuerzo más por llenarlas de contenido, lo cual a pesar de lo que se dice en el MCER puede ser posible.

5.2.4 Los niveles efectivos del MCER

Todavía nos queda otro aspecto que señalábamos en el capítulo 2 respecto a la práctica de la evaluación mediante el MCER. Veíamos entonces que en ocasiones solamente se diferencian tres niveles y dudábamos de que pudieran encontrarse más al calificar discursos reales. Ya hemos visto en el capítulo 4 que esto se ha podido solucionar a costa de utilizar escalas de *Dominio de vocabulario* y *Corrección gramatical*; lo que ha llevado a deslizar el constructo a evaluar en un sentido formal y a dar más peso a la corrección del vocabulario y gramática que a otros conceptos, a la hora de calificar y distinguir un C1 de un C2. Por lo tanto, sigue en pie que el MCER define o diferencia realmente menos de los 6 niveles que pretende, y ello dependerá de las escalas que se utilicen.

Con los resultados que hemos obtenido podemos elaborar una escala que diferenciaría cuatro niveles:

1. Un superior con errores o C1 y otro sin errores o C2, ambos en la funcionalidad del superior de ACTFL.
2. Un avanzado claro o B2, con toda la funcionalidad del avanzado de ACTFL, abarcando desde el intermedio medio hasta el avanzado medio.

3. Un intermedio, bajo o medio, con suficiente calidad, o calidad y cantidad, que serían A2 y B1 respectivamente.

4. Y, lógicamente, un principiante, que apenas sabe utilizar fórmulas memorizadas.

5.2.5 Diferencias entre el MCER y la OPI

Para terminar, queremos señalar las principales diferencias que hemos encontrado entre los resultados obtenidos mediante nuestro test, con el sistema de calificar y evaluar diseñado utilizando algunas escalas del MCER, y los resultados obtenidos mediante la OPI y las *guidelines* de ACTFL.

La primera de todas, señalada en la conclusión 5, el hecho de que el dominio definido en el MCER es menos que el definido en ACTFL. Con C2 del MCER no se alcanza el superior sino el AH. Y el A1 es ya el inicial alto de ACTFL.

En segundo lugar, el hecho de que el MCER es más rico y flexible y por ello útil para usos en el aula, aunque esto tiene la contrapartida de que es de difícil uso para otros contextos no académicos.

En tercer lugar, hemos visto que el nivel alto (C) se determinaba fundamentalmente por cuestiones de corrección de gramática y vocabulario, allí donde debía verse si un candidato era C1 o C2, o incluso apelando al siempre peligroso concepto del hablante nativo como referente de comparación. El MCER marca el nivel alto describiendo elementos formales que se aproximan a lo que un nativo sabe hacer. Se exige tener un dominio alto del léxico y coloquialismos o expresiones idiomáticas, el uso de énfasis para evitar ambigüedades, niveles connotativos del significado, etc. Pero para su diferenciación definitiva entre un C1 y un C2 se ha debido recurrir a las escalas más formales y ha surgido espontáneamente el criterio siempre discutible de comparación con un nativo. Las *guidelines* de ACTFL, en cambio, no dan tanto peso a los elementos gramaticales o formales. Un candidato puede expresarse sin cometer ningún error y quedarse solamente en el nivel avanzado medio o alto. Ser de nivel superior exige otras habilidades, que pueden llegar a ser algo que no todos los nativos sean capaces de dominar, como hacer discursos abstractos, defender opiniones y plantearse situaciones nuevas sobre las que elaborar hipótesis. Por ello, un candidato puede ser C2 y solamente avanzado alto. Igualmente hemos visto que un candidato puede ser calificado como de nivel superior con las *guidelines* de ACTFL, aún si hace ciertos errores no sistemáticos,

y quedar solamente como C1 según algunas escalas del MCER, teniendo en cuenta que esos errores le impiden llegar al nivel C2.

Y en cuarto lugar, estaría la forma de ponderar las calificaciones para dar el resultado final. En ACTFL los niveles están interrelacionados para dar una evaluación como un todo en una escala de dominio bajo, medio o alto, junto al cumplimiento de funciones de mayor complejidad y del tipo de discurso empleado. En el MCER depende en todo momento de las escalas utilizadas, como venimos señalando respecto a las de vocabulario y gramática.

5.3 Limitaciones de nuestra investigación y propuestas de trabajos futuros

Además de los resultados que hemos desgranado a lo largo de este capítulo, y además de la natural revisión o escrutinio a que puede y debe ser sometido todo trabajo científico, queremos proponer una serie de investigaciones e indagaciones que se pueden hacer continuando en la línea de buscar equivalencias entre las posibilidades de evaluar utilizando el MCER y ACTFL y que no hemos podido llevar a cabo debido a las limitaciones de metodología, contenido y trabajo de campo realizado.

En cuanto a nuestras limitaciones de método, nos hemos propuesto evaluar dos tests mediante escalas diferentes y hemos tenido que diseñar uno de ellos de forma paralela al otro. Esto lo hemos justificado porque creemos imposible evaluar el test de ACTFL mediante las escalas del MCER sin que surjan penalizaciones debidas a la especial estructura y objetivos de la OPI. Aunque nosotros hemos diseñado un test que ha obtenido una altísima correlación con el de ACTFL, por lo que lo consideramos como un test paralelo, nuestra forma de trabajo no deja de ser una forma indirecta de acceder a los resultados.

Un trabajo futuro podría ser diseñar y llevar a cabo una investigación en la que se evalúen la entrevista de ACTFL con las escalas del MCER, diseñando el trabajo de forma que pueda evitarse la posible interpretación del estrés producido en la OPI al llevar a los hablantes hasta más allá de sus posibilidades de expresión; y al contrario, intentando evaluar entrevistas del MCER con la escala de ACTFL, si es que se puede llegar a hacerlo sin que el resultado sea *non ratable* por incumplir los protocolos de ACTFL. Esta evaluación cruzada sería interesante para comprobar los resultados de forma directa.

En cuanto a las limitaciones del contenido de nuestro trabajo, hemos diseñado un test que se pareciera lo más posible al de la OPI tanto en funciones y concepto de lengua como en el sistema de elicitación y de calificación. Para ello, se eligieron escalas, tareas, contenidos y protocolos similares que acercaran los resultados. Y se ha conseguido, como se ve en los resultados estadísticos y de análisis factorial. Pero debería intentar averiguarse en qué medida otros tests con otros contenidos, otras tareas y otras escalas de calificación del MCER pueden lograr la equiparación o no con la OPI. Pues en gran medida eso dependerá del constructo a evaluar y de su sistema de realización.

Creemos que, por ejemplo, si se hubieran utilizado otras escalas en las que no se señalan diferencias entre el B2, el C1 y el C2, como es el caso de muchas escalas del MCER, o si no se hubieran utilizado escalas de *Dominio de vocabulario* y *Riqueza de vocabulario* o de *Corrección gramatical*, no se hubieran obtenido diferencias entre los niveles altos, con lo que la equiparación resultante sería diferente.

Igualmente, si se hubiera utilizado otro sistema de ponderación de resultados o se hubiera llegado a otro acuerdo de interpretación de las escalas y del nivel de los discursos de las entrevistas, también se hubieran obtenido otros resultados.

El MCER tiene enormes posibilidades de uso en diferentes campos de enseñanza, programación, pedagogía y evaluación. Eso es un punto fuerte que ayuda en determinados momentos a entender de forma unívoca los conceptos, pero hace que sean muchas las diferencias que se pueden marcar entre diferentes tests, con lo que se complica la posibilidad de realizar equiparaciones y comparaciones verdaderamente significativas. Estamos con Weir, quien criticaba (véase §2.7.4) que el MCER no es suficientemente claro para construir un constructo de lengua para evaluar, lo que impide hacer comparaciones entre tests, “dentro y a través de diferentes lenguas y niveles (...) Lo que provoca falsas asunciones de equivalencia al situar en el mismo punto de la escala tests contruidos para diferentes propósitos y audiencias, con diferentes constructos y de diferentes manera” (Weir, 2005: 281-282).

Un trabajo también interesante sería utilizar técnicas de análisis del discurso, para ver qué tipo de lenguaje se produce en el MCER con diferentes tareas y escalas de calificación, frente al que surge en la OPI. O incluso realizar ese mismo análisis y comparación con las entrevistas que hemos realizado para nuestra investigación.

También sería interesante estudiar en qué medida un test u otro interpreta y juzga de forma diferente el estrés que pueda aparecer ante situaciones de dificultad y si ello conlleva diferencias en las calificaciones finales. En nuestras entrevistas el elicitador-

entrevistador fue el mismo en todas las entrevistas y realizó la evaluación con el mismo método; por ello pudimos eliminar esta variable como factor causante de falta de homogeneidad en los resultados, pero esto no es sencillo cuando hay un grupo grande de evaluadores, salvo que hayan sido intensamente entrenados, como hace ACTFL.

Y por último, en cuanto a las limitaciones de nuestro trabajo de campo, se pueden y deben realizar investigaciones con otros tipos de candidatos, de diferente edad, educación, objetivos de certificación, nacionalidad o cultura, lengua materna, etc. Y con un grupo más numeroso de candidatos. Aunque nuestros análisis estadísticos ofrecen resultados con índices de fiabilidad y consistencia muy altos en todos sus sentidos, es una continuación lógica para todo trabajo científico que aspire a presentarse como definitivo. ACTFL ha sido sometida durante años a críticas por parte de diferentes investigadores y teóricos de la evaluación que han advertido de posibles lagunas en los diferentes campos de su evaluación y en su validación o fiabilidad. El MCER, nuevo todavía en su desarrollo y uso, debe someterse a este mismo escrutinio. Nosotros hemos conseguido un test con una alta correlación con el de ACTFL y unos calificadores con una consistencia alta. Habría que ver en qué medida es posible repetirlo en otras condiciones, dadas las lagunas que el MCER tiene para realizar, diseñar y validar tests, ya que si el diseñador es lego, el MCER por sí solo no le ofrece las herramientas con que hacer un test adecuado, válido y fiable.

Queremos terminar recordando unas ideas que ya se han señalado en esta tesis, sobre las limitaciones de la evaluación en general. Alderson / Banerjee plantean la necesidad de debatir e investigar más sobre los procesos de evaluación de la lengua hablada o los formatos de examen pues constantemente “se dibujan nuevas perspectivas en áreas como la lingüística aplicada, el análisis del discurso y la adquisición de segundas lenguas y evidentemente hay todavía campo para más investigaciones interdisciplinares” (Alderson / Banerjee, 2002: 95) así como sobre otras perspectivas alternativas o críticas (Alderson / Banerjee, 2001: 228) que tengan en cuenta la evaluación democrática con conciencia de que la evaluación es una práctica social.

La posibilidad de estos debates son un síntoma claro de que ahora hay más madurez y autocrítica, mayor confianza en las prácticas y sus efectos, teniendo en cuenta su impacto social y nuestras responsabilidades. Corolario de ello es que en el campo de la evaluación de lenguas el siglo XXI hay una conciencia de que los instrumentos de que se dispone son limitados e insuficientes para la labor que exige la sociedad como

consecuencia de los cambios producidos en la consideración de las lenguas en los últimos 25 años y la fuerza de los conceptos comunicativos en todos los campos. Esto ha llevado a que todas las instituciones se hayan planteado seriamente códigos, protocolos y criterios que tengan en cuenta la justicia y equidad en sus sistemas de evaluación. TOEFL ha variado su sistema de evaluar y sus escalas transformando el tradicional test en uno completamente nuevo, TOEFL iBT (Test of English as a Foreign Language internet-based test)²⁵⁶, UCLES y ELTS de Cambridge también.

La evaluación comienza a ampliar sus perspectivas con un gran sentimiento de responsabilidad y conocimiento de sus limitaciones, consciente de que es un asunto con complicadas restricciones. Es deber de quien trate con ella conocerlas para no establecer conclusiones falsas ni realizar inferencias o extraer resultados que el test no permita. Aquí las hemos establecido para las relaciones entre la OPI de ACTFL y nuestro test oral del MCER.

²⁵⁶ Para una explicación del significado de estos cambios y ejemplos de los nuevos ítems y test ver la reseña de Fulcher en *Guardian Weekly* el 18 de noviembre de 2005, o las páginas web de TOEFL: www.ets.org/Media/Tests/TOEFL/pdf/TOEFL_Tips.pdf o www.dxrgroup.com/DxR_TOEFL.

EPÍLOGO

Por la boca muere el pez

(Acervo popular)

Hablar es un goce para muchos. A través de ello hacen y controlan el mundo, o su mundo, o lo deshacen. Con el lenguaje, que todo lo domina y atrapa, y al que se supone creador de la propia conciencia, que nos hace y deshace, que hace lo que somos. Incluso hombres y mujeres que no tienen otra forma política o social de acción gozan con él. También políticos y planificadores, abogados, vendedores... El que habla, si está sano mentalmente, goza de su hablar y de la imagen que con ello crea de sí en los demás, su fantasma, su espejo, como nos enseña Lacan, y en sí mismo. Pero si el que habla goza, ¿qué decir del que logra hacer hablar a otros, o del que hace a otros hablar? Analistas, confesores, periodistas, jueces, profesores, evaluadores de lengua... ¿Y de los que hacen que otros digan/nos digan eso que queríamos oír? Aquí se encuentran los evaluadores de lengua, junto a amantes, encantadores, manipuladores...

No hace falta volver a citar los muchos trabajos que han estudiado y analizado las formas y maneras de evaluar oralmente la lengua de extranjeros y parafrasearlos para admitir que lo único que estamos dispuestos a evaluar es aquello que previamente habíamos planificado evaluar, y que queremos oír de manos de los candidatos (digo de manos para que quede clara la función práctica del uso del lenguaje, su acción, su hacer algo con él) lo que ya teníamos pensado que dijeran.

Evaluar oralmente la lengua nos acerca al papel de quien hace decir lo que otro quiere oír o lo que él quiere, sin importarle la realidad de los hechos. ¿Es constituyente

de una realidad lo obtenido de esta manera? No hace falta ir hasta Lewis Carroll para decir que eso depende de quién mande. En la ideología medieval había “verdades” que sólo se podían obtener mediante la confesión obtenida con la tortura: se trataba de hacer hablar a un yo interior que estaba preso y no podía expresarse de otra manera; y en esa creencia de verdad se basaba el criterio de autenticidad y de aceptación del mensaje así extraído. Lo que en términos pragmáticos de análisis de las condiciones del discurso griceanas vendría a ser aceptar que el incumplimiento de la máxima de cualidad o de cooperación viene satisfecho por el de relevancia. Por suerte Martínez de la Rosa terminó con la Inquisición tras el primer fallido intento de las Cortes de Cádiz. Pero seguimos viviendo en una sociedad donde muchos se ocupan de hacer decir a otros lo que estos no querían, no sabían o no se atrevían a decir; donde el valor de decir lo oculto e imprevisto tiene la catarsis de la limpieza y la autoayuda: desde la radio a la televisión, los padres, los jueces o los médicos -esos otros jueces-, sin citar otros medios más esotéricos.

Los dispositivos y técnicas para que otros hablen, -y peor aún, acepten ese hablar como suyo propio y verdadero de forma que como tal les comprometa o configure- sean de sofisticación sicologista o simplemente coercitivos o físicos, han abarcado en la historia desde la confesión cristiana (bien documentada e interpretada por Foucault, 1984) hasta los espías modernos; pasando por diferentes etapas de control del saber y del no saber abiertas, indirectas y sutiles. Sus protocolos abarcan por tanto campos tan dispares como el militar, el religioso, el jurídico o el médico y llegan a la entrevista radiofónica o periodística, y a la evaluación de lenguas.

Nosotros hemos tratado de este último paso del saber y del hacer hablar, con sus diferentes técnicas, definiendo y explicando sus modos y analizando lo que se usa hoy en día al respecto y sus límites (la influencia del método empleado o de la tarea utilizada, las condiciones del discurso para ser tomado en serio, los comportamientos de los participantes y sus influencias mutuas en la producción de diferentes modelos de discurso, etc.). En todos ellos hay una función implícita: la de que el lenguaje se puede usar (es un acto) para conseguir otras cosas, ya sea informando, ya sea delatando, ya ganando un status, etc.

Hemos tratado esta vertiente de analizar una de las técnicas -y una en concreto muy particular- de hacer hablar a otra persona para juzgar al que habla (no lo que dice), no por lo que dice sino por lo que es capaz de decir y por cómo lo dice. Parece retorcido y es, cuando menos, complicado e incompleto. Se han descrito diversos sistemas y se ha

intentado ver de qué manera y con qué limitaciones es posible; estas últimas se deben al comportamiento de los participantes y del evaluador responsable de crear y manejar la interacción, al sistema de realización de preguntas y elicitación del discurso, al manejo de las tareas y al propio formato y estructura del test. Así mismo, hemos visto que en las relaciones entre las diferentes escalas que hemos analizado hay lagunas, puntos de fricción y puntos de conexión. Por eso es de vital importancia analizar y definir la forma de las entrevistas, las escalas de valoración, los modelos de interacción utilizados, las técnicas de obtención de discurso, las tareas empleadas, el comportamiento de los participantes, los objetivos perseguidos, el conocimiento mutuo de estos objetivos, y en fin, la fiabilidad y homogeneidad del proceso. De lo contrario no solo no seríamos justos con los candidatos y la sociedad sino que nuestra deshonestidad -por acción u omisión, consciente o inconsciente- nos haría cómplices (estúpidos) de una práctica injustificada y de resultados falseados que se pretenderían hacer pasar por buenos.

¿Cuál es entonces el acto -o incluso la perversión de ese acto- que se produce/ocasiona en una entrevista oral de lengua extranjera para evaluación, en la que una persona (evaluador) hace a otra hablar solamente para juzgar la manera en que es capaz de usar la lengua, la habilidad en abstracto (no el contenido ni siquiera su formato ni la cultura ni la adecuación a la verdad), mientras la otra persona (el candidato) habla sin objeto aparente o sin que entre ambos actores quede claro el objeto salvo el de conseguir demostrar que sabe hablar y que el otro lo acepte (sin que sea fácil o necesario establecer un contexto adecuado de implicaturas y colaboración -por eso hemos hablado de perversión del acto-), para, indirectamente, lograr un certificado o un nuevo status? La pregunta es retórica ya que lleva implícitas sus respuestas: lo único que pretende ese acto es probar que el candidato sabe hablar y conseguir aumentar su estatus mediante un certificado. Lo que pretende el evaluador es algo que depende de escalas, protocolos y objetivos, a veces ocultos o ambiguos y no siempre gramaticales o morfológicos -algo que el candidato erróneamente suele intuir y perseguir-.

En nuestra sociedad, ¿a qué modelo de discurso se parece lo resultante (el discurso del candidato) obtenido mediante ese tipo de técnicas para elicitarlo? Es decir, en el mundo real de fuera del aula de lenguas, ¿tiene un referente realista con el que se pueda comparar?

Ya hemos discutido el valor del discurso en la interacción y su comparación con una conversación natural entre nativos o no nativos. Ahora nos queda plantearnos algo más allá, más filosófico, o más acá, más realista: ¿de qué manera y en qué medida, o

bajo qué condiciones, tenemos derecho a hacer “eso”, a juzgar “eso” y de “esas” maneras, a exigir “eso” a un candidato a veces desorientado respecto a lo que esperamos de él (aparte de que él se deje hacer con el objetivo de conseguir su ascenso social mediante el certificado)? Aquí dejamos la pregunta, las preguntas. La sociedad irá desgranando poco a poco sus respuestas. Por suerte, los últimos protocolos y las críticas y acercamientos a una evaluación democrática y social van por un camino que respeta a los candidatos y se plantea sus necesidades, sean estos estudiantes o inmigrantes.

El MCER, ACTFL y los protocolos de los códigos de ALTE, ILTA o EALTA entre otros, tratan de hacer de la evaluación una acción clara y un trabajo justo, positivo y razonado. ¿Hasta dónde es esto posible? Creo que hemos expuesto algunas de las limitaciones y reconversiones necesarias para que sea así. Pero creo que también está claro que queda mucho trabajo por delante. Interesante, importante, responsable, apasionante. Sin duda.

Al comienzo de este trabajo citaba una escena de la película *Blade Runner*. Ahora, para terminar, y siguiendo con esa puerta siempre entreabierta que es la ciencia ficción, quisiera plantear otra cuestión relacionada, sea de forma tangencial, con las posibilidades de la evaluación oral de lenguas. Una cuestión con un planteamiento diferente y abierto. Me serviré del eminente físico Roger Penrose y su libro *La nueva mente del emperador* (Penrose, 2007) y cito sus palabras:

En un famoso artículo de Alan Turing, titulado “Computing Machinery and Intelligence”, aparecido en 1950 en la revista filosófica Mind, se describió por primera vez la idea ahora conocida como test de Turing. Este pretendía ser una prueba acerca de si se puede decir razonablemente que una máquina piensa. De acuerdo con el test de Turing, el computador y algún voluntario humano se ocultan a la vista de un (perspicaz) interrogador. El interrogador tiene que tratar de decidir cuál de los dos es el computador y cuál el ser humano, mediante el simple procedimiento de plantear preguntas de prueba a cada uno de ellos. Estas preguntas, y lo que es más importante, las respuestas recibidas, se transmiten de forma impersonal; por ejemplo, pulsadas en un teclado o mostradas en una pantalla. Al interrogador no se le permite más información sobre cualquiera de las partes que la que obtiene en esta sesión de preguntas y respuestas. El sujeto humano responde a las preguntas sinceramente y trata de persuadirle de que él es realmente el ser humano y que el otro sujeto es el computador; pero el computador está programado para “mentir” y, por consiguiente, tratar de convencer al entrevistador de que él es, y no el otro, el ser humano. Si en el transcurso de una serie de tests semejantes, el interrogador es incapaz de identificar de una forma definitiva al sujeto humano real, se considera que el computador (o el programa del computador, o el programador, o el diseño, etc.) ha superado la prueba (Penrose, 2007: 34).

Hasta aquí la cita. Ahora el planteamiento.

En este trabajo se han visto algunos de los límites de la evaluación oral y la necesidad de ser conscientes de ellos, y señalamos que la prueba del nueve de la evaluación oral es evaluar a un nativo y que su calificación se extraiga mediante los criterios estrictos del test y no porque sepamos que es un nativo. Nos gustaría plantearlo dando un paso más allá, pasando del nativo a la máquina de que hablaba Turing: con las concepciones que actualmente tenemos sobre la lengua y los mecanismos de que disponemos para evaluarla oralmente, ¿seríamos capaces de evaluar a una máquina de esas características para decidir su nivel de lengua?, ¿servirían nuestros tests para diferenciar a la máquina del ser humano?, ¿obtendríamos datos acertados? En definitiva, ¿soportarían nuestros tests el test de Turing?

No queremos decir que nuestros actuales tests deban satisfacer los términos del test de Turing. Cada época necesita solucionar aquellos problemas que la responsabilidad social le exija. La evaluación, los que las instituciones y sus candidatos requieran. Pero sí nos parece interesante recalcar una vez más la importancia de la conciencia de nuestros límites y de los de la evaluación. Y de sus alcances y consecuencias.

APÉNDICES

APÉNDICE 1 CARACTERÍSTICAS DEL TEST UTILIZADO EN ESTA INVESTIGACIÓN

(Citado en §2.2 -cuadro 2.2 B- y §3.7).

El análisis de las tareas presentadas en el capítulo 3 lo acompañamos ahora de una ficha resumen que contiene una descripción de los elementos citados por Bachman / Palmer que permite delimitarlas y compararlas efectivamente. Estos investigadores, recordemos, consideran 5 elementos: ambiente (condiciones físicas y tiempo de realización), formato (organización e instrucciones generales), tareas (instrucciones, formato, lengua y naturaleza de las instrucciones usadas en las actividades), respuesta esperada (formato, lengua y naturaleza de la respuesta esperada) e interacción entre las instrucciones y la respuesta.

Ficha de las características de las tareas, realizada según el cuadro de Bachman / Palmer (1996: 108).

Características del test.	
Instrucciones	Las instrucciones se dan en español (y en inglés a los candidatos de poco nivel) leyendo y explicando cada actividad.
Lengua usada	En español.
Canal	Oralmente.
Explicación de procedimientos	Se explica cada tarea asegurándose de que se entiende.
Estructura	Una sucesión de diferentes tareas según los niveles de los candidatos.
Núm. secuenciación y diferenciación de las partes	<ol style="list-style-type: none"> 1. Presentación. 2. Prueba de OPI. 3. Prueba de tareas del MCER. La secuenciación es: <ol style="list-style-type: none"> a. Tareas de intercambio de información. b. Tareas de dominio. c. Tareas de argumentación. d. Cómic. <p>Cada tarea se termina diciendo “Bueno ya está, ahora vamos a hacer otra cosa”.</p> <ol style="list-style-type: none"> 4. Despedida.
Nivel	<ol style="list-style-type: none"> 1. A1 y A2 principalmente. 2. Desde B2. 3. B1 o más. 4. Desde B2. 5. Desde B1. 6.7.8. Todos los niveles desde A2 en adelante.
Importancia de esta parte	Todas son igualmente importantes y cada una mide unas características determinadas, salvo las de dominio que miden toda la gama de dominios. Únicamente no son importantes las tareas que, aunque se pida

	realizar a los candidatos, no les correspondan realmente por ser de nivel superior al que tras realizar la evaluación se otorgue a cada candidato. Estas tareas extras se utilizan para averiguar cómo y en qué medida se colapsa la lengua en estos candidatos.
Duración (limitada o libre)	Libre.
Sistema de corrección o puntuación	Grado de cumplimiento y de adecuación a las escalas correspondientes de habilidad y de dominio para la tarea.
Características del input Canal Forma Lengua usada Extensión Tipo Tiempo de procesamiento Fuente (en vivo, grabado) Características de la lengua	<p>Oral. Lengua. L2 excepto para A1 y A2 que se puede explicar la tarea de hacer de canguro también en inglés.</p> <p>Breve introducción del asunto. Explicación.</p> <p>Automático (excepto para la tarea de cómic, pues se da tiempo para comprender la historieta). Para la tarea 3, y en general para cuando sea necesario provocar el discurso del candidato, se ayuda mediante unas primeras preguntas de calentamiento, a seleccionar ese momento o actividad cultural interesante, mediante un tanteo hasta lograr que el candidato esté dispuesto a hablar de algún tema adecuado.</p> <p>En vivo.</p> <p>Frases simples explicadas de forma que cada candidato entienda la tarea. No es posible que la no realización se deba a las características del input ya que este se adaptará si es necesario al nivel de los candidatos.</p>
Características de la respuesta esperada Canal Forma Lengua usada Extensión Tipo Tiempo de respuesta Fuente (en vivo, grabado) Características de la lengua Tipo de contenido	<p>Oral. Lengua. Lengua 2, español. Variable y dependiendo de la cantidad de información que el candidato desee dar y la tarea provoque. Abierta. Desde pocos minutos a extensa al dar en opiniones, explicaciones y argumentaciones. En vivo y grabado en casete pequeño digital.</p> <p>Desde frases cortas y preguntas breves hasta argumentaciones complejas con estructuras variadas. Viene definido en las escalas de habilidad y de dominio citadas en su apartado correspondiente. Esto se analiza en §3.5., apartado dedicado al lenguaje usado. El relacionado con el team de la tarea.</p>
Relación input/respuesta Reacción	Recíproca y adaptativa en todas excepto las 7 y 8. Estas apenas provocan intervenciones del evaluador ya que tratan de

Dirección	provocar un monólogo argumentativo y mostrar cómo explica una historia de un cómic. No obstante si el candidato no produce suficiente información el evaluador puede intervenir. La información del input de las instrucciones no ayuda a cumplir la tarea.
-----------	--

APÉNDICE 2

GUIDELINES DE HABILIDAD ORAL DE ACTFL

(Citado en §2.6.2- 2, cuadro 2.6.2-e.).

Esta escala de 10 niveles procede de ACTFL, 1999. (Breiner-Sanders, 2000: 14-17, o www.actfl.org).

SUPERIOR

Speakers at the Superior level are able to communicate in the language with accuracy and fluency in order to participate fully and effectively in conversations on a variety of topics in formal and informal settings from both concrete and abstract perspectives. They discuss their interests and special fields of competence, explain complex matters in detail, and provide lengthy and coherent narrations, all with ease, fluency, and accuracy. They explain their opinions on a number of topics of importance to them, such as social and political issues, and provide structured argument to support their opinions. They are able to construct and develop hypotheses to explore alternative possibilities. When appropriate, they use extended discourse without unnaturally lengthy hesitation to make their point, even when engaged in abstract elaborations. Such discourse, while coherent, may still be influenced by the Superior speakers own language patterns, rather than those of the target language. Superior speakers command a variety of interactive and discourse strategies, such as turn-taking and separating main ideas from supporting information through the use of syntactic and lexical devices, as well as intonational features such as pitch, stress and tone. They demonstrate virtually no pattern of error in the use of basic structures. However, they may make sporadic errors, particularly in low-frequency structures and in some complex high-frequency structures more common to formal speech and writing. Such errors, if they do occur, do not distract the native interlocutor or interfere with communication.

ADVANCED HIGH

Speakers at the Advanced-High level perform all Advanced-level tasks with linguistic ease, confidence and competence. They are able to consistently explain in detail and narrate fully and accurately in all time frames. In addition, Advanced-High speakers handle the tasks pertaining to the Superior level but cannot sustain performance at that level across a variety of topics. They can provide a structured argument to support their opinions, and they may construct hypotheses, but patterns of error appear. They can discuss some topics abstractly, especially those relating to their particular interests and special fields of expertise, but in general, they are more comfortable discussing a variety of topics concretely. Advanced-High speakers may demonstrate a well-developed ability to compensate for an imperfect grasp of some forms or for limitations in vocabulary by the confident use of communicative strategies, such as paraphrasing, circumlocution, and illustration. They use precise vocabulary and intonation to express meaning and often show great fluency and ease of speech. However, when called on to perform the complex tasks associated with the Superior level over a variety of topics, their language will at times break down or prove inadequate, or they may avoid the task altogether, for example, by resorting to simplification through the use of description or narration in place of argument or hypothesis.

ADVANCED MID

Speakers at the Advanced-Mid level are able to handle with ease and confidence a large number of communicative tasks. They participate actively in most informal and some formal exchanges

on a variety of concrete topics relating to work, school, home, and leisure activities, as well as to events of current, public, and personal interest or individual relevance. Advanced-Mid speakers demonstrate the ability to narrate and describe in all major time frames (past, present, and future) by providing a full account, with good control of aspect, as they adapt flexibly to the demands of the conversation. Narration and description tend to be combined and interwoven to relate relevant and supporting facts in connected, paragraph-length discourse. Advanced-Mid speakers can handle successfully and with relative ease the linguistic challenges presented by a complication or unexpected turn of events that occurs within the context of a routine situation or communicative task with which they are otherwise familiar. Communicative strategies such as circumlocution or rephrasing are often employed for this purpose. The speech of Advanced-Mid speakers performing Advanced-level tasks is marked by substantial flow. Their vocabulary is fairly extensive although primarily generic in nature, except in the case of a particular area of specialization or interest. Dominant language discourse structures tend to recede, although discourse may still reflect the oral paragraph structure of their own language rather than that of the target language. Advanced-Mid speakers contribute to conversations on a variety of familiar topics, dealt with concretely, with much accuracy, clarity and precision, and they convey their intended message without misrepresentation or confusion. They are readily understood by native speakers unaccustomed to dealing with non-natives. When called on to perform functions or handle topics associated with the Superior level, the quality and/or quantity of their speech will generally decline. Advanced-Mid speakers are often able to state an opinion or cite conditions; however, they lack the ability to consistently provide a structured argument in extended discourse. Advanced-Mid speakers may use a number of delaying strategies, resort to narration, description, explanation or anecdote, or simply attempt to avoid the linguistic demands of Superior-level tasks.

ADVANCED LOW

Speakers at the Advanced-Low level are able to handle a variety of communicative tasks, although somewhat haltingly at times. They participate actively in most informal and a limited number of formal conversations on activities related to school, home, and leisure activities and, to a lesser degree, those related to events of work, current, public, and personal interest or individual relevance. Advanced-Low speakers demonstrate the ability to narrate and describe in all major time frames (past, present and future) in paragraph length discourse, but control of aspect may be lacking at times. They can handle appropriately the linguistic challenges presented by a complication or unexpected turn of events that occurs within the context of a routine situation or communicative task with which they are otherwise familiar, though at times their discourse may be minimal for the level and strained. Communicative strategies such as rephrasing and circumlocution may be employed in such instances. In their narrations and descriptions, they combine and link sentences into connected discourse of paragraph length. When pressed for a fuller account, they tend to grope and rely on minimal discourse. Their utterances are typically not longer than a single paragraph. Structure of the dominant language is still evident in the use of false cognates, literal translations, or the oral paragraph structure of the speaker's own language rather than that of the target language. While the language of Advanced-Low speakers may be marked by substantial, albeit irregular flow, it is typically somewhat strained and tentative, with noticeable self-correction and a certain grammatical roughness. The vocabulary of Advanced-Low speakers is primarily generic in nature. Advanced-Low speakers contribute to the conversation with sufficient accuracy, clarity, and precision to convey their intended message without misrepresentation or confusion, and it can be understood by native speakers unaccustomed to dealing with non natives, even though this may be achieved through repetition and restatement. When attempting to perform functions or handle topics associated with the Superior level, the linguistic quality and quantity of their speech will deteriorate significantly.

INTERMEDIATE HIGH

Intermediate-High speakers are able to converse with ease and confidence when dealing with most routine tasks and social situations of the Intermediate level. They are able to handle successfully many uncomplicated tasks and social situations requiring an exchange of basic information related to work, school, recreation, particular interests and areas of competence, though hesitation and errors may be evident. Intermediate-High speakers handle the tasks pertaining to the Advanced level, but they are unable to sustain performance at that level over a variety of topics. With some consistency, speakers at the Intermediate High level narrate and describe in major time frames using connected discourse of paragraph length. However, their performance of these Advanced-level tasks will exhibit one or more features of breakdown, such as the failure to maintain the narration or description semantically or syntactically in the appropriate major time frame, the disintegration of connected discourse, the misuse of cohesive devices, a reduction in breadth and appropriateness of vocabulary, the failure to successfully circumlocute, or a significant amount of hesitation. Intermediate-High speakers can generally be understood by native speakers unaccustomed to dealing with non-natives, although the dominant language is still evident (e.g. use of code-switching, false cognates, literal translations, etc.), and gaps in communication may occur.

INTERMEDIATE MID

Speakers at the Intermediate-Mid level are able to handle successfully a variety of uncomplicated communicative tasks in straightforward social situations. Conversation is generally limited to those predictable and concrete exchanges necessary for survival in the target culture; these include personal information covering self, family, home, daily activities, interests and personal preferences, as well as physical and social needs, such as food, shopping, travel and lodging. Intermediate-Mid speakers tend to function reactively, for example, by responding to direct questions or requests for information. However, they are capable of asking a variety of questions when necessary to obtain simple information to satisfy basic needs, such as directions, prices and services. When called on to perform functions or handle topics at the Advanced level, they provide some information but have difficulty linking ideas, manipulating time and aspect, and using communicative strategies, such as circumlocution. Intermediate-Mid speakers are able to express personal meaning by creating with the language, in part by combining and recombining known elements and conversational input to make utterances of sentence length and some strings of sentences. Their speech may contain pauses, reformulations and self-corrections as they search for adequate vocabulary and appropriate language forms to express themselves. Because of inaccuracies in their vocabulary and/or pronunciation and/or grammar and/or syntax, misunderstandings can occur, but Intermediate-Mid speakers are generally understood by sympathetic interlocutors accustomed to dealing with non-natives.

INTERMEDIATE LOW

Speakers at the Intermediate-Low level are able to handle successfully a limited number of uncomplicated communicative tasks by creating with the language in straightforward social situations. Conversation is restricted to some of the concrete exchanges and predictable topics necessary for survival in the target language culture. These topics relate to basic personal information covering, for example, self and family, some daily activities and personal preferences, as well as to some immediate needs, such as ordering food and making simple purchases. At the Intermediate-Low level, speakers are primarily reactive and struggle to answer direct questions or requests for information, but they are also able to ask a few appropriate questions. Intermediate-Low speakers express personal meaning by combining and recombining into short statements what they know and what they hear from their interlocutors. Their utterances are often filled with hesitancy and inaccuracies as they search for appropriate linguistic forms and vocabulary while attempting to give form to the message. Their speech is characterized by frequent pauses, ineffective reformulations and self-corrections. Their

pronunciation, vocabulary and syntax are strongly influenced by their first language but, in spite of frequent misunderstandings that require repetition or rephrasing, Intermediate-Low speakers can generally be understood by sympathetic interlocutors, particularly by those accustomed to dealing with non-natives.

NOVICE HIGH

Speakers at the Novice-High level are able to handle a variety of tasks pertaining to the Intermediate level, but are unable to sustain performance at that level. They are able to manage successfully a number of uncomplicated communicative tasks in straightforward social situations. Conversation is restricted to a few of the predictable topics necessary for survival in the target language culture, such as basic personal information, basic objects and a limited number of activities, preferences and immediate needs. Novice-High speakers respond to simple, direct questions or requests for information; they are able to ask only a very few formulaic questions when asked to do so. Novice-High speakers are able to express personal meaning by relying heavily on learned phrases or recombinations of these and what they hear from their interlocutor. Their utterances, which consist mostly of short and sometimes incomplete sentences in the present, may be hesitant or inaccurate. On the other hand, since these utterances are frequently only expansions of learned material and stock phrases, they may sometimes appear surprisingly fluent and accurate. These speakers first language may strongly influence their pronunciation, as well as their vocabulary and syntax when they attempt to personalize their utterances. Frequent misunderstandings may arise but, with repetition or rephrasing, Novice-High speakers can generally be understood by sympathetic interlocutors used to non-natives. When called on to handle simply a variety of topics and perform functions pertaining to the Intermediate level, a Novice- High speaker can sometimes respond in intelligible sentences, but will not be able to sustain sentence level discourse.

NOVICE MID

Speakers at the Novice-Mid level communicate minimally and with difficulty by using a number of isolated words and memorized phrases limited by the particular context in which the language has been learned. When responding to direct questions, they may utter only two or three words at a time or an occasional stock answer. They pause frequently as they search for simple vocabulary or attempt to recycle their own and their interlocutor words. Because of hesitations, lack of vocabulary, inaccuracy, or failure to respond appropriately, Novice-Mid speakers may be understood with great difficulty even by sympathetic interlocutors accustomed to dealing with non-natives. When called on to handle topics by performing functions associated with the Intermediate level, they frequently resort to repetition, words from their native language, or silence.

NOVICE LOW

Speakers at the Novice-Low level have no real functional ability and, because of their pronunciation, they may be unintelligible. Given adequate time and familiar cues, they may be able to exchange greetings, give their identity, and name a number of familiar objects from their immediate environment. They are unable to perform functions or handle topics pertaining to the Intermediate level, and cannot therefore participate in a true conversational exchange.

APÉNDICE 3

EJEMPLO DE ENTREVISTA ORAL DE ACTFL (OPI)

(Citado en §2.6.2- 3).

ENTREVISTA DE MARCIE

Veamos una entrevista completa y la explicación de sus fases para asegurar que el nivel de esta chica en la OPI es intermedio medio. Esta entrevista no es parte de las que se han realizado para esta investigación sino que procede de Hernández, (2006).

Fase 1 (calentamiento):

- *¿Estás bien?*
- Sí.
- *¿Sí? Cuéntame, ¿qué haces estos días? ¿Ya te vas a Estados Unidos?*
- Ah... sí, sí.
- *¿Cuándo?*
- Veintinueve.
- *El veintinueve y ¿qué vas a hacer estos días?*
- Ah... alquilar, alquilar un coche am... porque... más rápido, creo que autobús...sí... y am... la último ah... semana... am mi con... eh... am... cinc or cuatr, cuatro chicas, amigas, ah... ir a... or vamos ar... Gibraltar para cuatro días.
- *¿A la playa?*
- Sí.
- *¿Por qué vas a Gibraltar?*
- No sé, porque... un poco aburrido aquí en Granada.
- *¿No te gusta Granada?*
- No, me gusta mucho pero... para... no em... tengo muchas cosas aquí para...
- *Y ¿vais a alquilar un coche?, ¿tú sabes conducir los coches europeos?*
- Ah...creo que sí.
- *Pero ¿tienen marchas?*
- ¡Sí, sí, yo sé!
- *¡Ah vale! entonces no pasa nada*
- (risas).

Fase 2 (prueba de nivel):

- *Y ¿qué tal tu familia aquí en Granada?*
- Muy, muy bien. Me gusta mucho.
- *¿Sí?, ¿cómo son? Cuéntame.*
- Ah... muy tranquila, ah...
- *¿Es una chica, una señora o una familia?*
- Una familia de dos niñas, am.. el padre se llama Jaime y la madre se llama, ah... María del Mar.
- *Es una familia normal o ¿cómo son?, ¿trabajan?, ¿estudian?*
- Sí... padre, el padre trabaja en el campo todos los dos semanas.
- *¿En el campo?*

- Y... las chicas no trabajan y no... em.. ah... (silencio) no estudio, estudian, solo viven en la casa.
- *A ver explícame, ¿Cuántos años tienen?*
- Veintecuatro, veintuno.
- *¿Y no hacen nada?*
- Sí, no sé por qué...
- *¿Qué hacen?, ¿cómo es su vida?, ¿qué sabes tú de ellas?*
- Am.. (pausa larga) de...
- *De las chicas.*
- Am... van a calle, por la calle, las noches, a veces, no siempre. Y ah... am... ayudan su madre en, en la cocina y... am... para limpiar la casa y... un ... ah... chica... ah... va a escuela para conducir y... am... creo que esta es el último semana para clases de conducir. Ah...
- *¿Son simpáticas? ¿tienen novio?*
- Oh, sí, una, se llama Mar, Mar, Maria del Mar pero la familia llam, (pausa larga) llame, llame ella Mar, tiene un novia, novio.
- *¿Guapo?*
- No sé.
- *¡Ah! ¿No lo conoces?*
- No.
- *Pero ¿esa es una familia normal?, ¿tú que crees?*
- Sí.
- *Sí, ¿por qué?*
- Ah... porque la madres, mm... eh... (silencio largo) no sé... comid... am... preparar comida y... am... el padre trabajo en el campo y much... creo que muchas am... gentes en Granada, am, trabajan en el campos también sí, y ah...
- *Vivir aquí con una familia española como ésta, ¿no? que tú dices que es normal pero las hijas no estudian, no trabajan, no hacen nada...*
- Es normal para España.
- *Bueno, no sé. (risas).*
- *¿Sí?*
- *Creo que tienen que hacer algo más, no sé ¿Por qué no estudian?*
- No.
- *¿No hablas con ellas?*
- Un poco. Mar quiere, quiere niños pero... no partir muy lejos y... (pausa) quier, quieren vivir aquí en Granada todos... su vida.

Hasta aquí, la chica se limita a describir su entorno familiar, en este caso su familia española. Podemos observar que el entrevistador le pide su opinión intentando que ella intente un tipo de discurso de un nivel más alto, (¿pero esa es una familia normal?, ¿tú que crees? Sí, ¿por qué?), pero ella se limita a hablar de los padres brevemente.

Fase 3 (búsqueda de limitaciones o prueba de nuevo nivel más difícil):

Con estas dos preguntas, el entrevistador ya sube a la candidata a una tarea de nivel avanzado para que cumpla con la función de compara y la lengua se le rompe.

- *Por eso te pregunto. Vivir con una familia española como ésta que tú has conocido es muy diferente que vivir en Estados Unidos. ¿Son familias y conceptos de la vida diferentes?*
- *Sí porque...*
- *¿Me puedes explicar las diferencias de la familia de Estados Unidos y aquí en España?*
- *En Estados Unidos, ah... muy... ah... la familia de mí en Estados Unidos, am... trabaja mucho, mucho, am... nunca en la casa nunca porque hermanos están en escuela y padres están en ah... trabajo y ... am... cuando no tra, tienen trabajar or... am... escuela em... mi familia ir a... partidos de fútbol, fútbol americano or.. partidos de... am... mi hermanos cuando... ah... tienen partidos de béisbol, fútbol o... no sé. Ah... pero.... aquí, mi familia aquí siempre en la casa, siempre pero mi familia de Estados nunca en la casa y para... (silencio) co... comida... cuando, com... comer, mi familia Estados es muy raro y solamente... tres días de la semana ah... comen ah... la cena con juntos pero aquí siempre, todos los días, ah... com...*

El evaluador sigue probándola en nivel avanzado, pidiéndole su opinión. Le da una segunda oportunidad. Si bien es cierto que su respuesta se entiende, no llega a comunicarse con fluidez y expresar su opinión, que es lo que debe poder hacer un avanzado.

- *¿Y qué te parece mejor una familia como ésta que están juntos siempre, y comen juntos y están siempre juntos o una familia americana? que son más independientes las familias americanas también. ¿qué te parece mejor?*
- *Me encanta comer juntos pero es muy aburrido cuando siempre están en la casa, sí. Y no me gusta. Y la madre sólo cocinar y limpiar la casa y no me gusta.*

Fase 2 nuevamente (vuelta al nivel de comodidad, nivel más bajo):

En esta secuencia el entrevistador ha bajado de nuevo el nivel a intermedio medio. La chica al principio sufre lo que llamamos en este trabajo estrés comunicativo. Al final de la secuencia, vuelve a poder comunicarse a su nivel, como al principio de la entrevista.

- *¿Y tu madre qué hace? ¿Tu madre hace otras cosas?*
- *Sí, ah... ah... cocinar pero, ah... no todos los días, no. Ah... muchas veces tiene... ah, comida de otro... am (pausa, piensa) de almuerzo y cenar de otro día en la frig, frigo, ¿frigo? y... am... y ah... porque es en el frigo sólo es necesito, necesita para ella cocinar tres días.*
- *Y tu madre trabaja. ¿Cómo es su vida?*
- *Sí, sí, sí, sí.*
- *Háblame un poco de tu madre.*
- *Ok, ah... mi madre, am... te quiero, te quiera, ¿como se dice?, se quiero, te quiero, oh, ¡no sé! Ok, ella trabaja en un iglesia de Estados Unidos con jóvenes ah... y... ah... es y ella ayudar el (no se entiende) es su... (silencio) no sé pero trabaja en la iglesia todos los días.*
- *¿Y gana dinero con eso?*
- *Mmmmm...*
- *¿O es voluntario?*

- No mucho. Oh, no, no, gana dinero pero no mucho.
- *¿Y tú también quieres trabajar en la iglesia? ¿Sois una familia muy religiosa?*
- Sí, sí.
- *¿O quieres hacer otras cosas?*
- Esta verano quiero trabajar en la iglesia con mi... am... con mi madre con jóvenes también. Pero mi... toda mi vida no quiero trabajar, trabajar en la iglesia. Sí.

Fase 3 nuevamente (una prueba de mayor nivel):

Nuevamente se sube la dificultad a un nivel avanzado pidiéndole otra opinión. Y de nuevo se desestructura el discurso. Tiene ideas y sabe que decir pero no lo transmite hasta el punto de que el entrevistador intenta traducir o clarificar su discurso.

- *Tú llevas aquí ya unos cuantos meses, cinco meses, ¿no? Has visto que la religión el concepto de la religión en España es diferente que en Estados Unidos o no es diferente, ¿qué te parece?*
- Amm, sí... ah.. porque muchísimos católicos y... pero am.. creo que la religión aquí es más, am... sólo para televisión y... no para la vida, ¿sí? am... tam... es lo mismo en Estados Unidos pero aquí es más. Porque todos los gentes, no todos pero muchísimos gentes, am... son católicos pero... sólo dicen “yo soy católicos”, pero, “católico” pero nunca van a la iglesia, por ejemplo mi familia... aquí en Granada, am... y... ah... en Estados más cristian... (risas) y... no sé... pero muchísimos católicos.
- *Ya entiendo. Que en Estados Unidos es todo más real, la religión se vive más y aquí es todo más tradiciones y más fiestas sí, ¿pero eso que te parece a ti?, ¿es bueno o es malo? Los españoles son católicos pero no van a la Iglesia y no hacen nada.*
- No, creo que no necesitan para persona... ir a iglesia todos los días o ah... domingos pero... ah... ah... (silencio) creo que es mejor cuando ah... vivi, vive, vive la... religión ah... y... pien, piensan, am... sobre religión, am... no sé (silencio).

Fase 4 (vuelta a su nivel de comodidad o salida):

Este es el cierre de la entrevista. El entrevistador vuelve a su nivel para que se vuelva a sentir cómoda. Aunque le cuesta mucho y no es capaz de dar todo lo que sí podía al principio de la entrevista.

- *Bueno, vamos a volver a ti, a tu vida. Tú en Estados Unidos. ¿Tú que haces? Tú estudias, ¿no?, ¿tú trabajas también?*
- Sí, cuéntame.
- Yo tuve, ah... tuve or... (silencio) am cuatro trabajadores.
- *¡Cuatro trabajos!, ¡Uh! ¿Por qué cuatro trabajos?*
- Porque necesito para ganar dinero.
- *Y ¿cuál es el más bonito? ¿el más interesante?*
- Ah... trabaja en el... am... de mi trabajadores
- Sí.

- (silencio largo) ¡ah! Ah, ok, am... (pausas) trabajadora de estudiantes en la “admisión” donde am, personas, gentes... Cuando gentes am... que no ir a mi escuela quiere... quieren ir.
- *¿Y ese trabajo es bonito?*
- No bonito pero... me gusta mucho porque muchas cosas para... para mi... hacer y los gentes quien yo... trabajo con, am... muy simpáticos.
- *¿Y ahora cuando vuelvas a Estados Unidos vas a trabajar otra vez?*
- (asiente) Ah... mismo... am... lugar, admisiones. Pero sólo admisiones, no, no tengo cuatro trabajadores. Sí...
- *Ah bueno, vale.*
- Y...

Fase 5 (role play):

Se le pide a la candidata que haga preguntas ella misma, para mostrar que sabe tomar la iniciativa y llevar una conversación por sí misma.

- *Bueno Marcie, yo te he hecho muchas preguntas. Tú a mi, hazme dos o tres preguntas, cualquier cosa.*
- *¿Cuándo tú, am... (pausa) va a Estados Unidos y... ¿cómo se dice? Tú estás en España, cuando llegar aquí.*
- *Llegar.*
- *¿Qué echar de menos, or echas de menos en Estados Unidos?*
- *¿De estados Unidos? El trabajo bien organizado que hay en Estados Unidos, sí que está muy bien, aquí no.*
- *Y muchos... cuando gentes en...*
- *No, no eso a mi no me importa. Pero el trabajo de Estados Unidos, me gusta mucho la organización. La gente también es muy simpática y muy humana en Estados Unidos, más directa, más cercana y aquí no. Aquí los españoles parece que son muy simpáticos pero después no son tan simpáticos, tan familiares como los americanos.*
- *Y que hechas de menos cuando está en...*
- *Bueno, pues la vida de la calle. España tiene vida de la calle, las tapas. Todo eso, la noche, la calle...*
- *Sí... (piensa).*
- *Cualquier pregunta.*
- (larga pausa) Ah... (silencio largo) *¿Por qué no esta en clase?*
- *Ahora tengo vacaciones. Porque junio, julio, agosto y septiembre voy a trabajar y no tengo vacaciones este verano. Tengo este mes, mayo tengo vacaciones.*
- *Y... am... has ir, ido, am... una vez.*
- *No, estoy aquí en Granada, son unas vacaciones que no son reales.*
- *¿Y qué...?*
- *Descansar. Nada especial.*
- *¿Qué haces?*
- *Nada, nada especial.*
- *¿Nunca?*
- *Este mes. Normalmente sí, normalmente voy a la playa o a algún país, hago algún viaje. Pero este año no. Este año es especial.*
- *Bueno, Marcie, ya está.*

APÉNDICE 4 FORMULARIOS PARA RELACIONAR EL TEST NUESTRO CON EL MCER

(Citado en §2.7.5).

Este apéndice presenta algunas de las 30 fichas que el Consejo de Europa ha editado para demostrar que el trabajo de diseño, validación y calificación de un test se ha hecho de acuerdo a unas normas generales y se puede relacionar con el MCER, con sus escalas y conceptos. Las fichas proceden de *Relating Language Examinations to the CEF, Manual Preliminary Pilot Version. Introduccion and Feedback to the pilot phase*, (Consejo de Europa, 2007). (Versión actual en la página web: www.coe.int/t/dg4/linguistic/Manuel1_EN.asp).

Hemos rellenado estas fichas como si fueran oficialmente para entregar ante el Consejo de Europa demostrando que nuestro test se relaciona adecuadamente con el MCER. Pero no hemos sido exhaustivos ya que se remite a los diferentes capítulos de la tesis para recabar la información concreta y completa que se pide en las fichas.

En el capítulo 3 de la tesis se han explicado pormenorizadamente los aspectos referentes al test diseñado según el MCER. El capítulo 4 en lo referente a las calificaciones.

Form A1: General Examination Description

GENERAL EXAMINATION DESCRIPTION	
General Information	
Name of examination	EXAMEN ORAL BASADO EN EL MCER
Language tested	Español
Examining institution	CLM (Universidad de Granada).
Date of this version	<input type="checkbox"/> International <input type="checkbox"/> National <input type="checkbox"/> Regional <input type="checkbox"/> Institutional
Type of examination	Examen oral
Purpose	<input type="checkbox"/> Lower Sec <input type="checkbox"/> Upper Sec <input type="checkbox"/> Uni/College Students
Target population	<input type="checkbox"/> Adult
No. of test-takers per year	Estudiantes de todos los niveles desde inicial absoluto hasta reciclaje de profesores. ---
What is the overall aim?	
Averiguar su nivel según las escalas y niveles del MCER.	
What are the more specific objectives? If available describe the needs of the intended users on which this examination is based.	
Demostrar habilidades orales de interacción, fluidez, argumentación, descripción, etc. Están descritas en el capítulo 3 de la tesis.	

Form A2: Test Development

Test development	Short description and/or references
What organisation decided that the examination was required?	<input type="checkbox"/> Own organisation/school <input type="checkbox"/> A cultural institute <input type="checkbox"/> Ministry of Education <input type="checkbox"/> Ministry of Justice <input type="checkbox"/> Other: specify: _____
If an external organisation is involved, what influence do they have on design and development?	<input type="checkbox"/> Determine the overall aims <input type="checkbox"/> Determine level of language proficiency <input type="checkbox"/> Determine examination domain or content <input type="checkbox"/> Determine exam format and type of test tasks <input type="checkbox"/> Other: specify: _____
If no external organisation was involved, what other factors determined design and development of examination? Compararlo con el test de ACTFL, OPI.	<input checked="" type="checkbox"/> A needs analysis <input type="checkbox"/> Internal description of examination aims <input checked="" type="checkbox"/> Internal description of language level <input type="checkbox"/> A syllabus or curriculum <input type="checkbox"/> Profile of candidates
In producing test tasks are specific features of candidates taken into account? Se han tenido en cuenta sus hábitos al respecto y la posibilidad de que las tareas sean naturales y las puedan realizar en su lengua. Se explica en el capítulo 3 de la tesis.	<input checked="" type="checkbox"/> Linguistic background (L1) <input checked="" type="checkbox"/> Language learning background <input type="checkbox"/> Age <input checked="" type="checkbox"/> Educational level <input type="checkbox"/> Socio-economic background <input checked="" type="checkbox"/> Social-cultural factors <input type="checkbox"/> Ethnic background <input type="checkbox"/> Gender
Who writes the items or develops the test tasks?	Nosotros, con ayuda y opiniones de 10 profesores.
Have test writers guidance to ensure quality? El trabajo de campo con estudiantes voluntarios y profesores.	<input checked="" type="checkbox"/> Training <input checked="" type="checkbox"/> Guidelines <input checked="" type="checkbox"/> Checklists <input type="checkbox"/> Examples of valid, reliable, appropriate tasks: <input checked="" type="checkbox"/> Calibrated to CEF level description <input type="checkbox"/> Calibrated to other level description: _____
Is training for test writers provided?	<input type="checkbox"/> Yes <input type="checkbox"/> No
Are test tasks discussed before use? Sí, entre candidatos voluntarios y un equipo de profesores.	<input checked="" type="checkbox"/> Yes <input type="checkbox"/> No
If yes, by whom?	<input type="checkbox"/> Individual colleagues <input checked="" type="checkbox"/> Internal group discussion <input type="checkbox"/> External examination committee <input checked="" type="checkbox"/> Internal stakeholders <input type="checkbox"/> External stakeholders
Are tests tasks pre-tested?	<input checked="" type="checkbox"/> Yes <input type="checkbox"/> No
If yes, how?	Con candidatos voluntarios de diferentes niveles en el CLM.
If no, why not?	
Is the reliability of the test estimated?	<input checked="" type="checkbox"/> Yes

	<input type="checkbox"/> No
If yes, how? Coefficiente de Spearman.	<input checked="" type="checkbox"/> Data collection and psychometric procedures <input type="checkbox"/> Other: specify: _____
Are different aspects of validity estimated?	<input checked="" type="checkbox"/> Face validity <input checked="" type="checkbox"/> Content validity <input checked="" type="checkbox"/> Concurrent validity <input checked="" type="checkbox"/> Predictive validity <input checked="" type="checkbox"/> Construct validity
If yes, describe how?	Se explica en el capítulo 3 de la tesis.

Form A3: Marking

Marking: Subtest	<i>Complete a copy of this form for each Subtest. Short description and/or reference</i> Cada tarea se explica detenidamente en el capítulo 3 de la tesis. Vamos a rellenar esta hoja como ejemplo de una de ellas.
How are the test tasks marked?	For receptive test tasks: <input type="checkbox"/> Optical mark reader <input type="checkbox"/> Clerical marking For productive or integrated test tasks: <input checked="" type="checkbox"/> Trained examiners <input checked="" type="checkbox"/> Teachers
Where are the test tasks marked? El sistema de corrección se explica en el capítulo 4. Tres fuentes de datos: un equipo de evaluadores entrenados, trabajando en equipo de dos o tres personas para llegar a un acuerdo sobre la calificación final; y dos profesores entrenados, que dan la calificación dos y tres de forma independiente.	<input type="checkbox"/> Centrally <input checked="" type="checkbox"/> Locally: <input checked="" type="checkbox"/> By local teams <input checked="" type="checkbox"/> By individual examiners
What criteria are used to select markers?	Compañeros de trabajo con experiencia. Y estudiantes del III Máster de enseñanza de español para extranjeros de la Universidad de Granada.
How is accuracy of marking promoted?	<input type="checkbox"/> Regular checks by co-ordinator <input checked="" type="checkbox"/> Training of markers/raters <input checked="" type="checkbox"/> Moderating sessions to standardise judgements <input checked="" type="checkbox"/> Using standardised examples of test tasks: <input type="checkbox"/> Calibrated to CEF <input type="checkbox"/> Calibrated to another level description <input type="checkbox"/> Not calibrated to CEF or other description
Describe the specifications of the rating criteria of productive and/or integrative test tasks. Cada tarea se evaluaba según una escala del MCER con la que se había diseñado	<input checked="" type="checkbox"/> one holistic score for each task <input checked="" type="checkbox"/> marks for different aspects for each task <input checked="" type="checkbox"/> rating scale for overall performance in test <input checked="" type="checkbox"/> rating grid for aspects of test performance <input checked="" type="checkbox"/> rating scale for each task <input type="checkbox"/> rating grid for aspects for each task

<p>y relacionado, y con un conjunto de otras escalas del MCER hasta un total de 13. Pero finalmente se da una nota global para cada tarea y una nota final para el test completo.</p>	<input type="checkbox"/> rating scale bands are defined, but not to CEF <input checked="" type="checkbox"/> rating scale bands are defined in relation to CEF
<p>Are productive or integrated test tasks single or double rated?</p>	<input type="checkbox"/> Single rater <input type="checkbox"/> Two simultaneous raters <input type="checkbox"/> Double marking of scripts / recordings <input type="checkbox"/> Other: specify: 3 calificaciones para cada entrevista
<p>If double rated, what procedures are used when differences between raters occur? De las tres calificaciones se ha tomado la nota final como la de la mayoría. Pero se han tenido en cuenta todas las notas para los análisis de fiabilidad y otras calibraciones estadísticas.</p>	<input type="checkbox"/> Use of third rater and that score holds <input checked="" type="checkbox"/> Use of third marker and two closest marks used <input type="checkbox"/> Average of two marks <input type="checkbox"/> Two markers discuss and reach agreement <input type="checkbox"/> Other: specify: _____
<p>Is inter-rater agreement calculated?</p>	<input checked="" type="checkbox"/> Yes <input type="checkbox"/> No

Form A4: Grading

Grading: _____	Subtest	<i>Complete a copy of this form for each Subtest.</i> Short description and/or reference
Are pass marks and/or grades given?		<input type="checkbox"/> Pass marks <input checked="" type="checkbox"/> Grades
Describe the procedures used to establish pass marks and/or grades and cut scores		Leyendo las escalas del MCER se decide en qué nivel se engloba el tipo de discurso desplegado en cada tarea y se califica con esa nota, de forma general y utilizando diversas escalas, por ejemplo <i>Argumentación, Fluidez y Dominio de vocabulario</i>, etc.
If grades are given, how are the grade boundaries decided?		Leyendo las escalas del MCER.
If only pass / fail is reported, How are the cut-off scores for pass / fail set?		
How is consistency in these standards maintained?		La consistencia entre evaluadores e ha calculado y es muy alta. Véase capítulo 5 de la tesis.

Form A5: Reporting Results

Results	Short description and/or reference
<p>What results are reported to candidates? Se podrían dar todo tipo de explicaciones. Para la tesis se ha utilizado la nota global y las de cada tarea.</p>	<input checked="" type="checkbox"/> Global grade or pass / fail <input type="checkbox"/> Grade or pass / fail per subtest <input type="checkbox"/> Global grade plus profile across subtests <input type="checkbox"/> Profile of aspects of performance per subtest

In what form are results reported?	<input type="checkbox"/> Raw scores <input checked="" type="checkbox"/> Undefined grades (e.g. "C") <input type="checkbox"/> Level on a defined scale <input type="checkbox"/> Diagnostic profiles
On what document are results reported?	<input type="checkbox"/> Letter or email <input type="checkbox"/> Report card <input type="checkbox"/> Certificate / Diploma
Is information provided to help candidates to interpret results? Give details.	
Do candidates have the right to see the corrected and scored examination papers?	
Do candidates have the right to ask for remarking?	

Form A6: Data Analysis

Test Analysis and Post-examination Review	Short description and/or reference
Is feedback gathered on the examinations? Por ahora no hay previsto usarlo más veces. Si se va a utilizar se harán ajustes debido a los resultados obtenidos hasta ahora.	<input type="checkbox"/> Yes <input type="checkbox"/> No
If yes, by whom?	<input type="checkbox"/> Internal experts (colleagues) <input type="checkbox"/> External experts <input type="checkbox"/> Local examination institutes <input type="checkbox"/> Test administrators <input type="checkbox"/> Teachers <input type="checkbox"/> Candidates
Is the feedback incorporated in revised versions of the examinations? Se hizo en el periodo de pruebas piloto.	<input checked="" type="checkbox"/> Yes <input type="checkbox"/> No
Is data collected to do analysis on the tests?	<input checked="" type="checkbox"/> On all tests <input type="checkbox"/> On a sample of test-takers: How large: _____. How often: _____ <input type="checkbox"/> No
If yes, indicate how data are collected?	<input checked="" type="checkbox"/> During pre-testing <input type="checkbox"/> During live examinations <input checked="" type="checkbox"/> After live examinations
For which features is analysis on the data gathered carried out? Ver el capítulo 3 de la tesis.	<input checked="" type="checkbox"/> Difficulty <input checked="" type="checkbox"/> Discrimination <input checked="" type="checkbox"/> Reliability <input checked="" type="checkbox"/> Validity
State which analytic methods have been used (e.g. in terms of psychometric procedures).	Spearman, fiabilidad entre evaluadores, consistencia del test y de los ítems.
Are performances of candidates from different groups analysed? If so, describe how.	En las pruebas piloto se tuvieron en cuenta todas las actuaciones por parte de diferentes voluntarios de diferentes niveles y culturas: americanos, orientales,

	 europeos y norteafricanos.
Describe the procedures to protect the confidentiality of data.	Se pidió a cada candidato cambiar su nombre durante las pruebas y entrevistas. Estas no se van a utilizar en otra cosa que en los análisis de los resultados del test, para la comparación con otro test similar hecho mediante la OPI.
Are relevant measurement concepts explained for test users? If so, describe how.	No. Salvo el hecho de explicarles que era un test oral para evaluar la competencia general y no exclusivamente la corrección gramatical o el vocabulario. Pidiéndoles que utilicen estrategias de comunicación ante cualquier problema que les surgiera.

Form A7: Rationale for Decisions

Rationale for making decisions	Short description and/or reference
Give the rationale for the decisions that have been made in relation to the examination or the test tasks in question.	La comparación con los descriptores de las escalas del MCER. En el capítulo 4 se explican las formas de hacer los juicios, con ejemplos de los diferentes.

Form A8: Impression of Overall Examination Level

<i>Initial Impression of Overall CEF Level</i>		
<input type="checkbox"/> A1	<input type="checkbox"/> B1	<input type="checkbox"/> C1
<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
<input type="checkbox"/> A2	<input type="checkbox"/> B2	<input type="checkbox"/> C2
<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Short rationale, reference to documentation		
El test se ha hecho para evaluar candidatos de todos los niveles.		

Form A11: Spoken Interaction

<i>Spoken Interaction</i>	<i>Short description and/or reference</i>
Which situations, content categories, domains are the test takers expected to show ability in? ➤ Table 5 in CEF 4.1 might be of help as a reference.	Se explica en el capítulo 3 de la tesis.
Which communication themes are the test takers expected to be able to handle? ➤ The lists in CEF 4.2 might be of help as a reference.	Se explica en el capítulo 3 de la tesis.
Which communicative tasks are the test takers	Se explica en el capítulo 3 de la tesis.

<p>expected to be able to handle?</p> <ul style="list-style-type: none"> ➤ The lists in CEF 4.3 might be of help as a reference. 	
<p>What kind of communicative activities and interaction strategies are the test takers expected to be able to handle?</p> <ul style="list-style-type: none"> ➤ The lists in CEF 4.4.3.1 and 4.4.3.5 might be of help as a reference. 	Se explica en el capítulo 3 de la tesis.
<p>What kind of texts and text-types are the test takers expected to be able to handle?</p> <ul style="list-style-type: none"> ➤ The lists in CEF 4.6.2 and 4.6.3 might be of help as a reference. 	Se explica en el capítulo 3 de la tesis.
<p>What kind of tasks are the test takers expected to be able to handle?</p> <ul style="list-style-type: none"> ➤ The description in CEF 7.1, 7.2 and 7.3 might be of help as a reference. 	Se explica en el capítulo 3 de la tesis.
<p>After reading the scale for Overall Spoken Interaction, given below, indicate and justify at which level(s) of the scale the subtest should be situated.</p> <ul style="list-style-type: none"> ➤ The sub-scales for spoken interaction in CEF 4.4.3.1 listed after the scale might be of help as a reference. 	Level
	<p>Las tareas engloban diferentes niveles y el test completo sirve para evaluar todos los niveles.</p> <p><i>Justification (incl. reference to documentation). Ver el capítulo 3.</i></p>

Form A13: Spoken Production

Spoken Production	Short description and/or reference
<p>Which situations, content categories, domains are the test takers expected to show ability in? Table 5 in CEF 4.1 might be of help as a reference</p>	Ver el capítulo 3 de la tesis.
<p>Which communication themes are the test takers expected to be able to handle?</p> <ul style="list-style-type: none"> ➤ The lists in CEF 4.2 might be of help as a reference. 	Ver el capítulo 3 de la tesis.
<p>Which communicative tasks are the test takers expected to be able to handle?</p> <ul style="list-style-type: none"> ➤ The lists in CEF 4.3 might be of help as a reference. 	Ver el capítulo 3 de la tesis.
<p>What kind of communicative activities and strategies are the test takers expected to be able to handle?</p> <ul style="list-style-type: none"> ➤ The lists in CEF 4.4.1.1 might be of help as a reference. 	Ver el capítulo 3 de la tesis.
<p>What kind of texts and text-types are the test takers expected to be able to handle?</p> <ul style="list-style-type: none"> ➤ The lists in CEF 4.6.2 and 4.6.3 might be of help as a reference. 	Ver el capítulo 3 de la tesis.
<p>What kind of tasks are the test takers expected to</p>	Ver el capítulo 3 de la tesis.

be able to handle? ➤ The description in CEF 7.1, 7.2 and 7.3 might be of help as a reference.	
After reading the scale for Overall Spoken Production, given below, indicate and justify at which level(s) of the scale the subtest should be situated. ➤ The sub-scales for spoken production in CEF 4.4.1.1 listed after the scale might be of help as a reference.	Level Las tareas engloban diferentes niveles y el test completo sirve para evaluar todos los niveles.
	<i>Justification (incl. reference to documentation)</i> Ver el capítulo 3 de la tesis.

Form A15: Integrated Skills Combinations

Integrated Skills Combinations	Subtest it occurs in
Listening and Note-taking <input type="checkbox"/>	
Listening and Spoken Production <input checked="" type="checkbox"/>	En todas las tareas hay interacción entre el evaluador y el evaluado: este debe oír, reaccionar y contestar a las preguntas del entrevistador.
Listening and Written Production <input type="checkbox"/>	
Reading and Note-taking <input type="checkbox"/>	
Reading and Spoken Production <input type="checkbox"/>	
Reading and Written Production <input type="checkbox"/>	
Listening and Reading, plus Note-taking <input type="checkbox"/>	
Listening and Reading, plus Spoken Production <input type="checkbox"/>	
Listening and Reading, plus Written Production <input type="checkbox"/>	

Form A16: Integrated Skills

	<i>Complete for each combination noted above</i>
Integrated Skills	Short description and/or reference
Which skills combinations occur? ➤ Refer to your entry in Form A15.	La tarea 8 consiste en ver y describir un cómic.
Which text-to-text activities occur? ➤ Table 6 in CEF 4.6.4 might be of help as a reference.	Oír, reaccionar y contestar a las preguntas del entrevistador.
Which situation, content categories, domains are the test takers expected to show ability in? ➤ Table 5 in CEF 4.1 might be of help as a reference.	La descripción de cada tarea en el capítulo 3 lo explica detenidamente.
Which communication themes are the test takers expected to be able to handle? ➤ The lists in CEF 4.2 might be of help as a reference.	La descripción de cada tarea en el capítulo 3 lo explica detenidamente.
Which communicative tasks are the test takers expected to be able to handle? ➤ The lists in CEF 4.3 might be of help as a	La descripción de cada tarea en el capítulo 3 lo explica detenidamente.

reference.	
What kind of texts and text-types are the test takers expected to be able to handle? ➤ The lists in CEF 4.6.2 and 4.6.3 might be of help as a reference.	La descripción de cada tarea en el capítulo 3 lo explica detenidamente.
After reading the scales for Processing Text, given below, plus Listening/Reading Comprehension and Written Production given earlier, indicate and justify at which level(s) of the scale the subtest should be situated. ➤ The sub-scale for Note-taking in CEF 4.6.3 might also be of help as a reference.	Level Las tareas engloban diferentes niveles y el test completo sirve para evaluar todos los niveles. Justification (incl. reference to documentation). Ver capítulo 3.

Form A19: Aspects of Language Competence in Reception

Linguistic Competence	Short description and/or reference
What is the range of lexical and grammatical competence that the test takers are expected to be able to handle? ➤ The lists in CEF 5.2.1.1 and 5.2.1.2 might be of help as a reference.	Se explica en el capítulo 3 de la tesis, al analizar el tipo de lenguaje que se espera produzca cada tarea.
After reading the scale for Linguistic Competence in Table 4.3, indicate and justify at which level(s) of the scale the examination should be situated.	Level El test tiene tareas para diferentes niveles: véase capítulo 3. En general compone un test para evaluar todos los niveles. Justification (incl. reference to documentation) Capítulo 3 de la tesis.
Socio-linguistic Competence	Short description and/or reference
What are the socio-linguistic competences that the test takers are expected to be able to handle: linguistic markers politeness conventions, register, adequacy, dialect/accent, etc? ➤ The lists in CEF 5.2.2 might be of help as a reference.	Lo especificado en las escalas del MCER utilizadas para evaluar el discurso de los candidatos, a diferentes niveles. Ver lista con esas escalas en el Apéndice 6.
After reading the scale for Socio-linguistic Competence in Table 4.3, indicate and justify at which level(s) of the scale the examination should be situated.	Level Ver el capítulo 3 de la tesis. Justification (incl. reference to documentation) Ver el capítulo 3 de la tesis.
Pragmatic Competence	Short description and/or reference
What are the pragmatic competences that the test takers are expected to be able to handle: discourse competences, functional competences? ➤ The lists in CEF 5.2.3 might be of help as a reference.	Ver el capítulo 3 de la tesis.

After reading the scale for Pragmatic Competence in Table 4.3, indicate and justify at which level(s) of the scale the examination should be situated.	Level Ver el capítulo 3 de la tesis.
	Justification (incl. reference to documentation). Ver el capítulo 3
Strategic Competence	Short description and/or reference
What are the strategic competences that the test takers are expected to be able to handle? ➤ The discussion in CEF 4.4.2.4. might be of help as a reference.	Ver el capítulo 3 de la tesis.
After reading the scale for Strategic Competence in Table 4.3, indicate and justify at which level(s) of the scale the examination should be situated.	Level Ver el capítulo 3 de la tesis.
	Justification (incl. reference to documentation) Ver el capítulo 3 de la tesis.

Form A20: Aspects of Language Competence in Interaction

<i>Linguistic Competence</i>	<i>Short description and/or reference</i>
What is the range of lexical and grammatical competence that the test takers are expected to be able to handle? The lists in CEF 5.2.1.1 and 5.2.1.2 might be of help as a reference.	Ver el capítulo 3 de la tesis.
What is the range of phonological and orthographic competence that the test takers are expected to be able to handle? The lists in CEF 5.2.1.4 and 5.2.1.5 might be of help as a reference.	Ídem.
After reading the scales for Range and Accuracy in Table 4.4, indicate and justify at which level(s) of the scale the examination should be situated. The scales for Phonological Control in CEF 5.2.1.4 and for Orthographic Control in 5.2.1.5 might also be of help as a reference.	Level Misma respuesta que antes. Justification (incl. reference to documentation) Ídem.
Socio-linguistic Competence	<i>Short description and/or reference</i>
What are the socio-linguistic competences that the test takers are expected to be able to handle: linguistic markers politeness conventions, register, adequacy, dialect/accent, etc.? The lists in CEF 5.2.2 might be of help as a reference).	Ídem.
After reading the scale for Socio-linguistic Competence in Table 4.4, indicate and justify at which level(s) of the scale the examination should be situated.	Level Ídem. Justification (incl. reference to documentation) Ídem.
Pragmatic Competence	<i>Short description and/or reference</i>

What are the pragmatic competences that the test takers are expected to be able to handle: discourse competences, functional competences? ➤ The lists in CEF 5.2.3 might be of help as a reference.	Ídem.
After reading the scale for Fluency in Table 4.4, indicate and justify at which level(s) of the scale the examination should be situated.	Level Ídem.
	Justification (incl. reference to documentation) Ídem.
Strategic Competence	Short description and/or reference
What are the interaction strategies that the test takers are expected to be able to handle? ➤ The discussion in CEF 4.4.3.5 might be of help as a reference.	Ídem.
After reading the scale for Interaction in Table 4.4, indicate and justify at which level(s) of the scale the examination should be situated.	Level Ídem.
	Justification (incl. reference to documentation) Ídem.

Form A21: Aspects of Language Competence in Production

Linguistic Competence	Short description and/or reference
What is the range of lexical and grammatical competence that the test takers are expected to be able to handle? ➤ The lists in CEF 5.2.1.1 and 5.2.1.2 might be of help as a reference.	Ver el capítulo 3 de la tesis.
What is the range of phonological and orthographic competence that the test takers are expected to be able to handle? ➤ The lists in CEF 5.2.1.4 and 5.2.1.5 might be of help as a reference.	
After reading the scales for Range and Accuracy in Table 4.4, indicate and justify at which level(s) of the scale the examination should be situated. The scales for Phonological Control in CEF 5.2.1.4 and for Orthographic Control in 5.2.1.5 might also be of help as a reference.	Level Ídem.
	Justification (incl. reference to documentation) Ídem.
Socio-linguistic Competence	Short description and/or reference
What are the socio-linguistic competences that the test takers are expected to be able to handle: linguistic markers politeness conventions, register, adequacy, dialect/accent, etc? The lists in CEF 5.2.2 might be of help as a reference.	Ídem.
After reading the scale for Socio-linguistic Competence in Table 4.5, indicate and justify at which level(s) of the scale the examination should be situated.	Level Ídem.
	Justification (incl. reference to documentation) Ídem.

Pragmatic Competence	<i>Short description and/or reference</i>
What are the pragmatic competences that the test takers are expected to be able to handle: discourse competences, functional competences? ➤ The lists in CEF 5.2.3 might be of help as a reference.	Ídem.
After reading the scale for Pragmatic Competence in Table 4.4, indicate and justify at which level(s) of the scale the examination should be situated.	Level Ídem. <i>Justification (incl. reference to documentation) Ídem.</i>
Strategic Competence	<i>Short description and/or reference</i>
What are the production strategies that the test takers are expected to be able to handle? The discussion in CEF 4.4.1.3 might be of help as a reference).	Ídem.
After reading the scale for Strategic Competence in Table 4.4, indicate and justify at which level(s) of the scale the examination should be situated.	Level Ídem. <i>Justification (incl. reference to documentation) Ídem.</i>

Form A23: Graphic Profile of the Relationship of the Examination to CEF Levels (Example)

C2								
C1								
B2.2								
B2								
B1.2								
B1								
A2.2								
A2								
A1								
Overall	Liste- ning	Reading	Social Conver- sation	Informat- ion Exchange	Notes, Messages and Forms	Sociolin- guistic	Pragmatic	Linguis- tic

En el capítulo 4, en el cuadro 4.5.a, damos la lista completa de calificaciones.

Form B1 Standardisation Record Form

Standardisation Record Form			
Location		Date:	
Coordinator	Name:	Institution/Project	
Stage	Familiarisation Training Benchmarking Performances Standard-setting	<input type="checkbox"/> x <input type="checkbox"/> x <input type="checkbox"/> <input type="checkbox"/>	
Area(s)	Assessing Spoken Samples Assessing Written Samples Test Tasks/ Item: ▪ Listening ▪ Reading ▪ Language work Other: _____	<input type="checkbox"/> x <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>	
Participants	Number:	Functions:	
Activities completed	Familiarisation Illustration with standardised exemplars Controlled/free practice with calibrated exemplars Benchmarking local performance samples Follow-up meeting of judges Training with calibrated items Judging item difficulty Feedback of actual item difficulty Other _____	<input checked="" type="checkbox"/> <input checked="" type="checkbox"/>	
Materials Used	CEF Standardised performance samples <input checked="" type="checkbox"/> CEF Rating instruments (Tables 5.4, 5.5 ,5.8) <input checked="" type="checkbox"/> Local performance samples Adapted rating instruments (to be appended) CEF Calibrated test tasks and items Local test tasks and items Other: Ejemplos de entrevistas y escalas del MCER.	<input type="checkbox"/> <input checked="" type="checkbox"/> <input checked="" type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>	
Information on Tasks and Items	De cada tarea se informó respecto a su nivel requerido para cumplirla, el tipo de lenguaje que se esperaba que produjera, y las escalas de las que se había diseñado y con las que se debía calificar.		
Additional Comment			
Dissemination procedures planned			

Form B2: Analytic Rating Form

Eurocentres (North 1991/1992)/Swiss Project (Schneider and North 2000)

Disponemos de información de las calificaciones de cada evaluador para cada tarea con varias escalas. En el capítulo 4 de la tesis se dan ejemplos significativos con una nota global y algunos comentarios sobre aspectos de habilidades independientes.

LEARNER'S NAME _____

Ihr Name/Votre nom: _____

Niveaus/Niveaux: **B, W, W+, T, T+, V, V+, E, M**

1. Initial Impression

Einstufung mit der
Globalskala
Classement - échelle globale

2. Detailed Analysis with Grid / Beurteilung mit Raster / Estimation – grille

RANGE	ACCURACY	FLUENCY	INTERACTIO N	COHERENC E
Spektrum Étendue	Korrektheit Correction	Flüssigkeit Aisance	Interaktion Interaction	Kohärenz Cohérence

3. Considered Judgement

Abschliessende Einstufung
Classement final

Form B3: Holistic Rating Form (DIALANG)

Los resultados de cada candidato y su calificación se han dado en el capítulo 4.

Skill: _____	Level assigned	Comments
Sample/Task 1		
Sample/Task 2		
Sample/Task 3		
Sample/Task 4		

This is an example of a simple rating sheet which requires the participant to give one global judgement about the level of each sample or task. This rating sheet can be used to rate either performances or test items.

Form B4: Collation Global Rating Form (DIALANG)

En el capítulo 4 de la tesis se dan ejemplos de las calificaciones obtenidas por cada candidato, tarea a tarea.

	<i>Mickey Mouse</i>	<i>Donald Duck</i>	<i>Groover</i>	<i>Fred</i>	<i>Henry VII</i>	<i>Susi Q</i>	<i>Other names</i>	<i>Other names</i>	<i>Other names</i>
<i>Item 1</i>									
<i>Item 2</i>									
<i>Item 3</i>									
<i>Item 4</i>									

Form B5: Item Rating Form (DIALANG)

En el capítulo 4 de la tesis se dan las calificaciones obtenidas por cada candidato.

Skill _____	Descriptor Operationalised <i>(List subscale and level)</i>	CEF level assigned	Comments <i>(Include references to Form A10)</i>
Item 1			
Item 2			
Item 3			
Item 4			

APÉNDICE 5

FICHA USADA CON LOS EXPERTOS PARA EL EXPERIMENTO DE INTERPRETACIÓN DE NIVELES DE LA ESCALA DE *EXPRESIÓN ORAL EN GENERAL* DEL MCER

(Citado en §3.1.7).

Te presentamos 7 descriptores de habilidad lingüística de *Expresión oral en general* del Marco común europeo de referencia (MCER). Te pedimos dos cosas:

A. Están en desorden y debes numerarlos de menor nivel a mayor nivel, del 1 al 7.

B. Después debes decidir a qué niveles corresponden del MCER: A1, A2, B1, B2, C1 o C2.

Pero tienes un descriptor más que niveles. Hay 6 niveles y 7 descriptores.

Esto es porque un nivel -por poner un ejemplo ficticio, el A1- está subdividido en dos subniveles y cada uno lleva su descriptor correspondiente que llamaríamos por ejemplo A1.a y A1.b (siendo a de menor nivel y b de mayor nivel).

Escribe el nivel al que crees que corresponde cada uno especificando cuál es el que tiene esos dos subniveles.

Número

! Nivel

! !

		a. Realiza descripciones y presentaciones claras y detalladas sobre una serie de asuntos relacionados con su especialidad, ampliando y defendiendo sus ideas con aspectos complementarios y ejemplos relevantes.
		b. Realiza descripciones y presentaciones claras y detalladas sobre temas complejos, integrando otros temas, desarrollando ideas concretas y terminando con una conclusión adecuada.
		c. Sabe hacer una descripción o presentación sencilla de personas, condiciones de vida o trabajo, actividades diarias, cosas que le gustan o no le gustan, en una breve lista de frases y oraciones sencillas.
		d. Realiza descripciones y presentaciones claras y sistemáticamente desarrolladas, resaltando adecuadamente los aspectos significativos y los detalles relevantes que sirvan de apoyo.
		e. Puede llevar a cabo, con razonable fluidez, una descripción sencilla de una variedad de temas que sean de su interés, presentándolos como una secuencia lineal de elementos.
		f. Puede expresarse con frases sencillas y aisladas relativas a personas y lugares.
		g. Produce discursos claros, fluidos y bien estructurados cuya estructura lógica resulta eficaz, ayuda al oyente a fijarse en elementos significativos y a recordarlo.

APÉNDICE 6

ESCALAS DEL MCER CON LAS QUE SE HAN CORREGIDO LAS TAREAS DE NUESTRO NUEVO TEST

(Citado en §4.1).

A continuación se presentan las escalas utilizadas en la calificación de las entrevistas realizadas con el test del MCER:

- 1- *Intercambiar información.*
- 2- *Desarrollo de Descripciones y narraciones.*
- 3- *Monólogo sostenido: argumentación.*
- 4- *Fluidez.*
- 5- *Competencia lingüística general.*
- 6- *Riqueza de vocabulario.*
- 7- *Dominio de vocabulario.*
- 8- *Compensación.*
- 9- *Control y corrección.*
- 10- *Corrección gramatical.*
- 11- *Flexibilidad.*
- 12- *Expresión oral en general.*
- 13- *Autoevaluación.*

1. INTERCAMBIAR INFORMACIÓN (Consejo de Europa, 2002: 80).

INTERCAMBIAR INFORMACIÓN	
C2	Como en B2
C1	Como en B2
B2.2	- Comprende e intercambia información compleja y consejos sobre todos los asuntos relacionados con su profesión. (1)
B2.1	- Comunica con total certeza información detallada. (2) - Realiza descripciones claras y detalladas acerca de la forma de llevar a cabo un procedimiento. (3) - Sintetiza y comunica información y argumentos procedentes de varias fuentes. (4)
B1.2	- Intercambia, comprueba y confirma con cierta confianza información concreta sobre asuntos, cotidianos o no, dentro de su especialidad. (5) - Describe la forma de hacer algo dando instrucciones detalladas. (6) - Resume y da su opinión sobre relatos, artículos, charlas, discusiones, entrevistas o documentales breves y responde a preguntas complementarias que requieren detalles. (7)
B1.1	- Sabe como encontrar y comunicar información concreta y sencilla. (8) - Pide y comprende indicaciones detalladas para ir a un lugar. (9) - Obtiene información más detallada. (10)
A2.2	- Comprende lo suficiente como para desenvolverse sin esfuerzo en intercambios sencillos y habituales. (11) - Se desenvuelve en situaciones prácticas de la vida diaria: encuentra y comunica información concreta y sencilla. (12) - Plantea y contesta preguntas sobre costumbres y acciones de la vida cotidiana. (13) - Plantea y contesta preguntas sobre pasatiempos y actividades pasadas. (14) - Da y comprende indicaciones e instrucciones sencillas; por ejemplo, explica cómo ir a un lugar. (15)

A2.1	<ul style="list-style-type: none"> - Se comunica en charlas sencillas y cotidianas que requieren un intercambio de información sencillo y directo. (16) - Intercambia información limitada sobre gestiones cotidianas y habituales. (17) - Plantea y sabe cómo responder a preguntas relativas a lo que se hace en el trabajo y en el tiempo libre. (18) - Pide y da indicaciones que se refieren a un mapa o a un plano. (19) - Pide y ofrece información personal. (20)
A1	<ul style="list-style-type: none"> - Comprende preguntas e instrucciones que se le formulan despacio y con cuidado, y comprende indicaciones breves y sencillas sobre cómo ir a un lugar. (21) - Plantea y contesta preguntas sencillas, realiza afirmaciones sencillas sobre temas de necesidad inmediata o muy cotidiana y responde a este tipo de afirmaciones. (22) - Plantea y contesta preguntas sobre sí mismo y sobre otras personas: dónde vive, personas a las que conoce, cosas que tiene. (23) - Es capaz de hacer indicaciones temporales mediante frases como, por ejemplo, “la semana que viene, el pasado viernes, en noviembre, a las tres”. (24)

2. DESARROLLO DE DESCRIPCIONES Y NARRACIONES (Consejo de Europa, 2002: 122).

DESARROLLO DE DESCRIPCIONES Y NARRACIONES.	
C2	Como en C1
C1	Realiza descripciones y narraciones detalladas integrando varios temas, desarrollando aspectos concretos y terminando con una conclusión apropiada.
B2	Realiza descripciones y narraciones claras ampliando y apoyando sus puntos de vista sobre los aspectos principales, con detalles y ejemplos adecuados.
B1	Realiza, con razonable fluidez, narraciones y descripciones sencillas, mediante una secuencia lineal de elementos.
A2	Cuenta historias o describe algo con la ayuda de una lista sencilla de elementos.
A1	No hay descriptor posible.

3. MONÓLOGO SOSTENIDO: ARGUMENTACIÓN (Consejo de Europa, 2002: 63)

MONÓLOGO SOSTENIDO: ARGUMENTACIÓN.	
C2	- No hay descriptor disponible para C2.
C1	- No hay descriptor disponible para C1.
B2	<ol style="list-style-type: none"> 1. Desarrolla argumentos sistemáticamente, dando un énfasis apropiado a los aspectos importantes y apoyándose en detalles adecuados. 2. Desarrolla argumentos con claridad, ampliando con cierta extensión y defendiendo sus puntos de vista con ideas complementarias y ejemplos adecuados. 3. Construye cadenas argumentales razonadas. 4. Explica puntos de vista sobre un tema, proponiendo las ventajas y desventajas de ciertas opciones.
B1	<ol style="list-style-type: none"> 5. Desarrolla argumentos lo bastante bien como para que se puedan comprender sin dificultad la mayor parte del tiempo. 6. Ofrece breves razonamientos y explicaciones de opiniones, planes y acciones.
A1 A2	- No hay descriptor disponible para A1 y A2

4. FLUIDEZ ORAL (Consejo de Europa, 2002: 126)

FLUIDEZ ORAL.	
C2	Se explica mediante un discurso amplio y seguido con una fluidez natural, sin esforzarse ni dudar. Se detiene sólo para considerar las palabras más adecuadas con las que expresar sus pensamientos o para buscar un ejemplo o explicación adecuada.

C1	Se expresa con fluidez y espontaneidad, casi sin esfuerzo. Sólo un tema conceptualmente difícil puede obstaculizar la fluidez natural del discurso.
B2.2	Se comunica espontáneamente, a menudo mostrando una fluidez y una facilidad de expresión notables incluso en períodos más largos y complejos.
B2.1	Produce discursos con un ritmo bastante regular, aunque puede dudar mientras busca estructuras y expresiones; provoca pocas pausas largas. Participa en la conversación con un grado de fluidez y espontaneidad que hace posible la interacción habitual con hablantes nativos sin producir tensión en ninguno de los interlocutores.
B1.2	Se expresa con relativa facilidad. A pesar de algunos problemas al formular su discurso, que dan como resultado pausas y “callejones sin salida”, es capaz de seguir adelante con eficacia y sin ayuda.
B1.1	Es capaz de mantener su discurso, aunque sean muy evidentes las pausas para ordenar la gramática y el léxico, y de corregirse; sobre todo, en períodos largos de producción libre.
A2.2	Se hace entender en intervenciones breves, aunque resulten muy evidentes las pausas, las dudas iniciales y la reformulación.
A2.1	Construye frases sobre temas cotidianos con la facilidad suficiente como para desenvolverse en breves intercambios, a pesar de tener dudas muy evidentes y tropiezos en el comienzo.
A1	Se desenvuelve con enunciados muy breves, aislados y preparados, y con muchas pausas para buscar expresiones, articular palabras menos habituales y salvar la comunicación.

5. COMPETENCIA LINGÜÍSTICA GENERAL (Consejo de Europa, 2002: 107)

COMPETENCIA LINGÜÍSTICA GENERAL	
C2	Saca provecho de un dominio amplio y fiable de un completo repertorio de elementos lingüísticos para formular pensamientos con precisión, poner énfasis, diferenciar y eliminar la ambigüedad. No manifiesta ninguna limitación de lo que quiere decir.
C1	Elige una formulación apropiada de una amplia serie de elementos lingüísticos para expresarse con claridad y sin tener que limitar lo que quiere decir.
B2	Se expresa con claridad y sin manifestar ostensiblemente que tenga que limitar lo que quiere decir. Dispone de suficientes elementos lingüísticos como para hacer descripciones claras, expresar puntos de vista y desarrollar argumentos, utilizando para ello algunas oraciones complejas y sin que se le note mucho que está buscando las palabras que necesita.
B1	Dispone de suficientes elementos lingüísticos como para describir situaciones impredecibles, para explicar los puntos principales de una idea o un problema con razonable precisión y para expresar pensamientos sobre temas abstractos o culturales tales como la música y las películas. Dispone de suficientes elementos lingüísticos como para desenvolverse y suficiente vocabulario para expresarse con algunas dudas y circunloquios sobre temas como la familia, aficiones e intereses, trabajo, viajes y hechos de actualidad, pero las limitaciones léxicas provocan repeticiones e incluso, a veces, dificultades en la formulación.
A2	Tiene un repertorio de elementos lingüísticos básicos que le permite abordar situaciones cotidianas de contenido predecible, aunque generalmente tiene que adaptar el mensaje y buscar palabras. Produce expresiones breves y habituales con el fin de satisfacer necesidades sencillas y concretas: datos personales, acciones habituales, carencias y necesidades, demandas de información. Utiliza estructuras sintácticas básicas y se comunica mediante frases, grupos de palabras y fórmulas memorizadas al referirse a sí mismo y a otras personas, a lo que hace, a los lugares, a las posesiones, etc. Tiene un repertorio limitado de frases cortas memorizadas que incluye situaciones predecibles de supervivencia. Puede incurrir en malentendidos e interrupciones si se trata de una situación de comunicación poco frecuentes.
A1	Tiene un repertorio muy básico de expresiones sencillas relativas a datos personales y a necesidades de tipo concreto.

6. RIQUEZA DE VOCABULARIO (Consejo de Europa, 2002: 109)

RIQUEZA DE VOCABULARIO	
C2	Tiene un buen dominio de un repertorio léxico muy amplio que incluye expresiones idiomáticas y coloquiales; muestra que es capaz de apreciar los niveles connotativos del significado.
C1	Tiene un buen dominio de un amplio repertorio léxico que le permite superar con soltura sus deficiencias mediante circunloquios; apenas se le nota que busca expresiones o que utiliza estrategias de evitación. Buen dominio de expresiones idiomáticas y coloquiales.
B2	Dispone de un amplio vocabulario sobre asuntos relativos a su especialidad y temas más generales. Varía la formulación para evitar la frecuente repetición, pero las deficiencias léxicas todavía pueden provocar vacilación y circunloquios.
B1	Tiene suficiente vocabulario para expresarse con algún circunloquio sobre la mayoría de los temas pertinentes para su vida diaria como, por ejemplo, familia, aficiones e intereses, trabajo, viajes y hechos de actualidad.
A2	Tiene suficiente vocabulario para desenvolverse en actividades habituales y en transacciones cotidianas que comprenden situaciones y temas conocidos.
A1	Tiene suficiente vocabulario para expresar necesidades comunicativas básicas. Tiene suficiente vocabulario para satisfacer necesidades sencillas de supervivencia.

7. DOMINIO DEL VOCABULARIO (Consejo de Europa, 2002: 109).

DOMINIO DEL VOCABULARIO	
C2	Utiliza con consistencia un vocabulario correcto y apropiado.
C1	Pequeños y esporádicos deslices, pero sin errores importantes de vocabulario.
B2	Su precisión léxica es generalmente alta, aunque tenga alguna confusión o cometa alguna incorrección al seleccionar las palabras, sin que ello obstaculice la comunicación.
B1	Manifiesta un buen dominio del vocabulario elemental, pero todavía comete errores importantes cuando expresa pensamientos más complejos o cuando aborda temas y situaciones poco frecuentes.
A2	Domina un limitado repertorio relativo a necesidades concretas y cotidianas.
A1	No hay descriptor disponible

8. COMPENSACIÓN (Consejo de Europa, 2002: 67).

COMPENSACIÓN	
C2	Sustituye una palabra que no recuerda por un término equivalente de una forma tan sutil que apenas se nota.
C1	Como B2.
B2	Utiliza circunloquios y paráfrasis para suplir carencias de vocabulario y de estructura.
B1	Define las características de algo concreto cuando no recuerda la palabra exacta que lo designa. Es capaz de transmitir significado modificando una palabra que significa algo parecido (por ejemplo, un camión para personas = autobús). Utiliza una palabra sencilla que significa algo parecido al concepto que quiere transmitir y pide a su interlocutor que le corrija. Adapta una palabra de su lengua materna y pide confirmación.
A2	Sabe cómo aclarar lo que quiere decir mediante gestos si sólo dispone en su repertorio de palabras inadecuadas. Puede hacer ver o identificar aquello a lo que quiere referirse señalándolo o apuntando con el dedo (por ejemplo, «Quiero esto, por favor.»).
A1	No hay descriptor disponible.

9. CONTROL Y CORRECCIÓN (Consejo de Europa, 2002: 68).

	CONTROL Y CORRECCIÓN
C2	Sortea las dificultades con tanta discreción que el interlocutor apenas se da cuenta de ello.
C1	Es capaz de sortear una dificultad cuando se topa con ella y vuelve a formular lo que quiere decir sin interrumpir totalmente la fluidez del discurso.
B2	Corrige deslices y errores si se da cuenta de ellos o si dan lugar a malentendidos. Toma nota de los errores más frecuentes y controla conscientemente el discurso en esas ocasiones.
B1	Puede corregir confusiones de tiempos verbales o de expresiones que pueden dar lugar a malentendidos siempre que el interlocutor indique que hay un problema. Pide confirmación de que la forma utilizada es la correcta. Vuelve a comenzar utilizando una táctica diferente cuando se interrumpe la comunicación.
A2 A1	No hay descriptor disponible.

10. CORRECCIÓN GRAMATICAL (Consejo de Europa, 2002: 111).

	CORRECCIÓN GRAMATICAL
C2	Mantiene un consistente control gramatical sobre un repertorio lingüístico complejo incluso cuando su atención se centra en otras actividades (por ejemplo, en la planificación de lo que sigue, en el seguimiento de las reacciones de los demás).
C1	Mantiene un alto grado de corrección gramatical de modo consistente; los errores son escasos y apenas se notan.
B2	Buen control gramatical; todavía puede cometer «deslices» esporádicos, errores no sistemáticos y pequeños fallos en la estructura de la frase, pero son escasos y a menudo puede corregirlos retrospectivamente. Manifiesta un grado relativamente alto de control gramatical. No comete errores que produzcan malentendidos.
B1	Se comunica con razonable corrección en situaciones cotidianas: generalmente tiene un buen control gramatical, aunque con una influencia evidente de la lengua materna. Comete errores, pero queda claro lo que intenta expresar. Utiliza con razonable corrección un repertorio de «fórmulas» y estructuras habituales relacionadas con las situaciones más predecibles.
A2	Utiliza algunas estructuras sencillas correctamente, pero sigue cometiendo errores básicos sistemáticamente; por ejemplo, suele confundir tiempos verbales y olvida mantener la concordancia; sin embargo, suele quedar claro lo que intenta decir.
A1	Manifiesta un control limitado sobre unas pocas estructuras gramaticales y sintácticas sencillas dentro de un repertorio aprendido.

11. FLEXIBILIDAD (Consejo de Europa, 2002: 121).

	FLEXIBILIDAD
C2	Manifiesta gran flexibilidad al reformular ideas con distintos elementos lingüísticos para poner énfasis, diferenciar según la situación, el interlocutor, etc., y eliminar la ambigüedad.
C1	Como B2+.
B2	Adapta lo que dice y los medios de expresarlo a la situación y al receptor, y adopta un nivel de formalidad adecuado a las circunstancias. Se sabe ajustar a los cambios de dirección, estilo y énfasis que se producen normalmente en la conversación. Varía la formulación de lo que quiere decir.
B1	Adapta su expresión para abordar situaciones menos habituales e incluso difíciles. Utiliza una gran diversidad de elementos lingüísticos sencillos con flexibilidad para expresar gran parte de lo que quiere
A2	Adapta frases sencillas, bien ensayadas y memorizadas, a circunstancias particulares mediante una sustitución léxica limitada. Amplía frases aprendidas volviendo a combinar de un modo sencillo determinados elementos.
A1	No hay descriptor disponible.

12. EXPRESIÓN ORAL EN GENERAL (Consejo de Europa, 2002: 62).

EXPRESIÓN ORAL EN GENERAL	
C2	Produce discursos claros, fluidos y bien estructurados cuya estructura lógica resulta eficaz y ayuda al oyente a fijarse en elementos significativos y a recordarlos.
C1	Realiza descripciones y presentaciones claras y detalladas sobre temas complejos, integrando otros temas, desarrollando ideas concretas y terminando con una conclusión adecuada.
B2.2	Realiza descripciones y presentaciones claras y sistemáticamente desarrolladas, resaltando adecuadamente los aspectos significativos y los detalles relevantes que sirvan de apoyo.
B2.1	Realiza descripciones y presentaciones claras y detalladas sobre una amplia serie de asuntos relacionados con su especialidad, ampliando y defendiendo sus ideas con aspectos complementarios y ejemplos relevantes.
B1	Puede llevar a cabo con razonable fluidez una descripción sencilla de una variedad de temas que sean de su interés, presentándolos como una secuencia lineal de elementos.
A2	Sabe hacer una descripción o presentación sencilla de personas, condiciones de vida o trabajo, actividades diarias, cosas que le gustan o no le gustan, en una breve lista de frases y oraciones sencillas.
A1	Puede expresarse con frases sencillas y aisladas relativas a personas y lugares.

13. NIVELES COMUNES DE REFERENCIA: CUADRO DE AUTOEVALUACIÓN (Consejo de Europa, 2002: 30-31).

	NIVELES COMUNES DE REFERENCIA:	CUADRO DE AUTOEVALUACIÓN
HABLAR	A1	A2
Interacción oral	Puedo participar en una conversación de forma sencilla siempre que la otra persona esté dispuesta a repetir lo que ha dicho o a decirlo con otras palabras y a una velocidad más lenta y me ayude a formular lo que intento decir. Planteo y contesto preguntas sencillas sobre temas de necesidad inmediata o asuntos muy habituales.	Puedo comunicarme en tareas sencillas y habituales que requieren un intercambio simple y directo de información sobre actividades y asuntos cotidianos. Soy capaz de realizar intercambios sociales muy breves, aunque, por lo general, no puedo comprender lo suficiente como para mantener la conversación por mí mismo.
Expresión oral	Utilizo expresiones y frases sencillas para describir el lugar donde vivo y las personas que conozco.	Utilizo una serie de expresiones y frases para describir con términos sencillos a mi familia y otras personas, mis condiciones de vida, mi origen educativo y mi trabajo actual o el último que tuve.
	B1	B2
Interacción oral	Sé desenvolverme en casi todas las situaciones que se me presentan cuando viajo donde se habla esa lengua. Puedo participar espontáneamente en una conversación que trate temas cotidianos de interés personal o que sean pertinentes para la vida diaria (por ejemplo, familia, aficiones, trabajo, viajes y acontecimientos actuales).	Puedo participar en una conversación con cierta fluidez y espontaneidad, lo que posibilita la comunicación normal con hablantes nativos. Puedo tomar parte activa en debates desarrollados en situaciones cotidianas explicando y defendiendo mis puntos de vista.

Expresión oral	Sé enlazar frases de forma sencilla con el fin de describir experiencias y hechos, mis sueños, esperanzas y ambiciones. Puedo explicar y justificar brevemente mis opiniones y proyectos. Sé narrar una historia o relato, la trama de un libro o película y puedo describir mis reacciones.	Presento descripciones claras y detalladas de una amplia serie de temas relacionados con mi especialidad. Sé explicar un punto de vista sobre un tema exponiendo las ventajas y los inconvenientes de varias opciones.
	C1	C2
Interacción oral	Me expreso con fluidez y espontaneidad sin tener que buscar de forma muy evidente las expresiones adecuadas. Utilizo el lenguaje con flexibilidad y eficacia para fines sociales y profesionales. Formulo ideas y opiniones con precisión y relaciono mis intervenciones hábilmente con las de otros hablantes.	Tomo parte sin esfuerzo en cualquier conversación o debate y conozco bien modismos, frases hechas y expresiones coloquiales. Me expreso con fluidez y transmito matices sutiles de sentido con precisión. Si tengo un problema, sorteo la dificultad con tanta discreción que los demás apenas se dan cuenta.
Expresión oral	Presento descripciones claras y detalladas sobre temas complejos que incluyen otros temas, desarrollando ideas concretas y terminando con una conclusión apropiada.	Presento descripciones o argumentos de forma clara y fluida y con un estilo que es adecuado al contexto y con una estructura lógica y eficaz que ayuda al oyente a fijarse en las ideas importantes y a recordarlas.

APÉNDICE 7

FORMULARIO UTILIZADO PARA LA VALIDACIÓN DE LAS TAREAS Y SU TIPO DE LENGUAJE CON EXPERTOS

(Citado en §3.6.1).

Os pido ayuda para **ver si son válidas y útiles 9 tareas para evaluar oralmente a los estudiantes de español** que he diseñado con relación al Marco común europeo de referencia.

Tengo opiniones de los propios estudiantes (los usuarios últimos de las tareas) pero para analizar si son válidas hacen falta opiniones de expertos (vosotros) además de las de los estudiantes. Os lo agradecería enormemente (porque va a ser algo de trabajo y tendréis que pensar un poco) si me contestáis a las **6 preguntas** que os hago, en la hoja de respuestas.

No hay una respuesta buena y otras malas, ya que se trata de saber qué opináis y en relación a vuestras opiniones habrá que rediseñar, cambiar o eliminar las tareas según hayáis dicho que valen o no.

Se trata de que leáis las 9 tareas y luego vayáis contestando a las preguntas:

Tened en cuenta que son para evaluar la habilidad de hablar de estudiantes de español como lengua extranjera o sea su habilidad de expresión oral.

Vamos allá.

Esta es la lista y número de las 9 tareas (luego van con sus instrucciones y escalas del MCER con las que se van a evaluar y de las que se han diseñado).

1. Encuesta de trabajo para conseguir un trabajo de canguro.
2. Explicar una receta de comida.
3. Explicar el camino a tu casa (por ejemplo para una fiesta) en un plano.
4. Opinión sobre una actividad cultural
5. Explicar ciertos aspectos de tu trabajo para un colega que va a ayudarte o sustituirte durante una semana.
6. Elegir una:
 - Argumentar a favor o en contra de la guerra de Israel contra Líbano.
 - Argumentar a favor o en contra de la guerra de Estados Unidos en Irak.
7. Explicar el viaje o experiencia más enriquecedor o interesante de los realizados durante la estancia en España durante este mes (o en otro momento).
8. Explicar las ventajas de vivir y estudiar español en España en vez de en tú país.
9. Explicar un cómic (Luego os enseñe la fotocopia).

Y esta la primera pregunta:

Pregunta 1 (sobre el examen en su totalidad).

(Quizá prefieras dejarla para el final, cuando te hayas familiarizado con las tareas)

En general, con un test en el que figuren estas 9 tareas, ¿crees que se podría medir el nivel general de habilidad oral de un candidato?

¿Haría falta algo más? ¿Qué?

Pregunta 2 (sobre la autenticidad de las tareas).

A continuación tienes la descripción de cada una de las 9 tareas.

¿Crees que son naturales, realistas y aceptables para los candidatos? (No se trata de si son difíciles o fáciles sino si son algo existente en la vida real o más bien son cosas que solo aparecen en los exámenes).

Contesta de 0 a 5 (0 Son cosas solo de exámenes / 5 Son cosas propias de la vida real)

Pregunta 3 (sobre las habilidades que pretendemos medir con las tareas).

El tipo de lenguaje que van a usar los estudiantes para realizarlas ¿serviría para medir alguna de estas dos habilidades?: (Contesta sí o no y en su caso qué habilidad).

a. **INTERCAMBIAR INFORMACIÓN.**

b. **MONOLOGO SOSTENIDO: ARGUMENTACIÓN**

(Debajo tienes las escalas de habilidad para saber a qué se refiere el MCER en cada caso).

Con otras palabras: pretendo medir dos habilidades: intercambiar información y argumentación. ¿Sirven estas tareas? ¿Cuáles para cada caso?

Pregunta 4 (sobre el nivel para el que van a servir).

Las tareas -tal y como se definen- ¿tienen relación con alguno de los niveles descritos en las escalas? O con otras palabras: ¿Hay en las escalas alguna frase (cada una está numerada) que se pueda corresponder con el contenido de las tareas? Por favor, marca el número (del 1 al 20) o el nivel (del A1 al C2) al lado de cada tarea, lo que te sea más sencillo.

Pregunta 5 (sobre el sistema de evaluarlas o calificarlas/puntuarlas).

¿Crees que las escalas del MCER (de *Argumentación* y de *Intercambio de información*) servirían para evaluar a los candidatos sobre el grado de habilidad para realizar las tareas en español?

Pregunta 6 (sobre contenido lingüístico que generan las tareas).

Marca de 0 a 5 en qué medida crees que el lenguaje necesario para realizar las tareas va a:

- el lenguaje necesario va a ser lingüísticamente complejo
- el lenguaje necesario va a necesitar organización retórica
- el lenguaje necesario va a depender de saberes culturales concretos
- el lenguaje necesario va a producir un discurso con bastantes elementos pragmáticos

Vuelve a mirar la pregunta 1, porque es sobre el examen en su totalidad.

TAREAS

1. Trabajo de canguro (“baby sitter”)

Vas a trabajar en una casa ayudando a la familia con su hijo pequeño y enseñándole un poco de tu lengua (inglés, francés, etc.). El padre de la familia te va a hacer unas preguntas antes de empezar el trabajo. Después si debes hacerle preguntas sobre su hijo y el trabajo, su vida cotidiana y costumbres (del hijo y de la familia): horarios, comidas, etc.

2. Receta de comida

Explica a tu profesor cómo se hace una comida que te guste y sepas hacer. Y contesta a sus preguntas para aclararle las cosas que no entienda bien cómo hacer.

3. Dar indicaciones de un plano (usando el plano de Granada)

Vas a dar una fiesta este fin de semana en tu casa de Granada y has invitado a unos amigos y a tu profesor. Explícale al profesor el camino para llegar a tu casa. Si vives lejos, utiliza un lugar céntrico y conocido para empezar a explicar el camino.

4. Opinión sobre actividad cultural

Has estado un mes en Granada viviendo y estudiando español. Seguro que has ido al cine o a algún espectáculo, o has leído algo interesante sobre Granada o España y su cultura. ¿Puedes decir a tu profesor qué cosas te han gustado más y charlar con él sobre una de ellas expresando tus opiniones y contestando a sus preguntas para dar más información o detalle?

5. Explicar ciertos aspectos de su trabajo para un colega que va a ayudarlo o sustituirle durante unos días

Vas a irte de vacaciones una semana y tienes que explicarle a un colega cómo es el trabajo que debe hacer, que es el que tú realizas habitualmente. Y a responder a sus preguntas sobre lo que él no entienda.

6a. Argumentar a favor o en contra de la actual guerra de Israel contra Líbano

Como quizás sepas, estos días Israel está atacando a Líbano y ha matado a varios cientos de personas inocentes civiles, unas 300, y está destruyendo importantes servicios de uso civil, porque las milicias de Hezbolá secuestraron a dos soldados en la frontera norte de Israel. ¿Te parece que esto es justo, que Israel tiene el derecho a defenderse, como han dicho los estados unidos, o crees que en realidad y aunque tiene el derecho a defenderse, debe usar otros medios o tener más cuidado con sus acciones?

6b. Argumentar a favor o en contra de la guerra de Estados Unidos en Irak

¿Crees que la guerra de Irak ha sido justa y necesaria? Estás de acuerdo con las afirmaciones de EE.UU. de que ahora el mundo es más seguro y era necesario terminar con un dictador y asesino como Sadam Hussein.

7. Explicar el viaje o experiencia más enriquecedora o interesante de los realizados durante la estancia en España durante este mes

- Para los niveles A1, A2 y B1:

¿Puedes contarme algún viaje que has hecho durante este mes?

Usar las preguntas normales sobre viajes para que genere respuestas (dónde, con quién, cuándo, cómo, cuántos días, qué tal, experiencias interesantes, visitas, alojamiento, incidentes, vuelta a Granada...).

- Para los niveles B2, C1 y C2:

Explícame el viaje o experiencia más enriquecedora o interesante de los realizados durante la estancia en España durante este mes.

8. Explicar las ventajas de vivir y estudiar en el país

- Para los niveles A1, A2 y B1:

Explícame las ventajas de vivir en un país extranjero con una familia, frente a vivir en tu propio país, para aprender una lengua.

¿Qué crees que es lo mejor de vivir en un país extranjero para estudiar una lengua? ¿Y vivir con una familia esté bien? ¿Por qué?

- Para los niveles B2, C1 y C2:

Explícame las ventajas de vivir y estudiar en el país con una familia frente a vivir con otros estudiantes de tu misma nacionalidad.

9. Explicar el siguiente cómic²⁵⁷

Estas son las escalas de *Argumentación, Intercambiar información y Expresión oral en general*, necesarias para los juicios u opiniones:

MONOLOGO SOSTENIDO: ARGUMENTACIÓN	
	- No hay descriptor disponible para C1 y C2
B2	<u>1.</u> Desarrolla argumentos sistemáticamente, dando un énfasis apropiado a los aspectos importantes y apoyándose en detalles adecuados. <u>2.</u> Desarrolla argumentos con claridad, ampliando con cierta extensión y defendiendo sus puntos de vista con ideas complementarias y ejemplos adecuados. <u>3.</u> Construye cadenas argumentales razonadas. <u>4.</u> Explica puntos de vista sobre un tema, proponiendo las ventajas y desventajas de ciertas opciones.
B1	<u>5.</u> Desarrolla argumentos lo bastante bien como para que se puedan comprender sin dificultad la mayor parte del tiempo. <u>6.</u> Ofrece breves razonamientos y explicaciones de opiniones, planes y acciones.
	- No hay descriptor disponible para A1 y A2

²⁵⁷ El cómic se les dio en fotocopia. Se encuentra en §3.4.8.

INTERCAMBIAR INFORMACIÓN	
C2 C1 B2	<p><u>1.</u> Comprende e intercambia información compleja y consejos sobre todos los asuntos relacionados con su profesión.</p> <p><u>2.</u> Comunica con total certeza información detallada.</p> <p><u>3.</u> Realiza descripciones claras y detalladas acerca de la forma de llevar a cabo un procedimiento.</p> <p><u>4.</u> Sintetiza y comunica información y argumentos procedentes de varias fuentes.</p>
B1	<p><u>5.</u> Intercambia, confirma y comprueba con cierta confianza información concreta sobre asuntos, cotidianos o no, dentro de su especialidad.</p> <p><u>6.</u> Describe la forma de hacer algo dando instrucciones detalladas.</p> <p><u>7.</u> Resume y da su opinión sobre relatos, artículos, charlas, discusiones, entrevistas o documentales breves y responde a preguntas complementarias que requieren detalles.</p> <p><u>8.</u> Sabe como encontrar y comunicar información concreta sencilla.</p> <p><u>9.</u> Pide y comprende indicaciones detalladas para ir a un lugar.</p> <p><u>10.</u> Obtiene información más detallada.</p>
A2	<p><u>11.</u> Se desenvuelve en situaciones prácticas de la vida diaria: encuentra y comunica información concreta sencilla.</p> <p><u>12.</u> Plantea y contesta preguntas sobre costumbres y acciones de la vida cotidiana.</p> <p><u>13.</u> Plantea y contesta preguntas sobre pasatiempos y actividades pasadas.</p> <p><u>14.</u> Da y comprende indicaciones e instrucciones sencillas; por ejemplo, explica cómo ir a un lugar.</p> <p><u>15.</u> Se comunica en charlas sencillas y cotidianas que requieren un intercambio de información sencillo y directo.</p> <p><u>16.</u> Intercambia información limitada sobre gestiones cotidianas y habituales.</p> <p><u>17.</u> Plantea y sabe cómo responder a preguntas relativas a lo que se hace en el trabajo y en el tiempo libre.</p> <p><u>18.</u> Pide y da indicaciones que se refieren a un mapa o a un plano.</p> <p><u>19.</u> Pide y ofrece información personal.</p>
A1	<p><u>20.</u> Plantea y contesta preguntas sencillas, realiza afirmaciones sencillas sobre temas de necesidad inmediata o muy cotidianos o responde a ese tipo de afirmaciones.</p> <p><u>21.</u> Plantea y contesta preguntas sobre sí mismo y sobre otras personas: dónde vive, personas a las que conoce, cosas que tiene.</p> <p><u>22.</u> Es capaz de hacer indicaciones temporales mediante frases como, por ejemplo, “la semana que viene, el pasado viernes, en noviembre, a las tres”.</p>

EXPRESIÓN ORAL EN GENERAL	
C2	Produce discursos claros, fluidos y bien estructurados cuya estructura lógica resulta eficaz, ayuda al oyente a fijarse en elementos significativos y a recordarlos.
C1	Realiza descripciones y presentaciones claras y detalladas sobre temas complejos, integrando otros temas, desarrollando ideas concretas y terminando con una conclusión adecuada.
B2.2	Realiza descripciones y presentaciones claras y sistemáticamente desarrolladas, resaltando adecuadamente los aspectos significativos y los detalles relevantes que sirvan de apoyo.
B2:1	Realiza descripciones y presentaciones claras y detalladas sobre una serie de asuntos relacionados con su especialidad, ampliando y defendiendo sus ideas con aspectos complementarios y ejemplos relevantes.
B1	Puede llevar a cabo, con razonable fluidez, una descripción sencilla de una variedad de temas que sean de su interés, presentándolos como una secuencia lineal de elementos.
A2	Sabe hacer una descripción o presentación sencilla de personas, condiciones de vida o trabajo, actividades diarias, cosas que le gustan o no le gustan, en una breve lista de frases y oraciones sencillas.
A1	Puede expresarse con frases sencillas y aisladas relativas a personas y lugares.

HOJA DE RESPUESTA SOBRE LAS TAREAS

1. Trabajo de canguro

Vas a trabajar en una casa ayudando a la familia con su hijo pequeño y enseñándole un poco de tu lengua (inglés, francés, etc.). El padre de la familia te va a hacer unas preguntas antes de empezar el trabajo. Después si debes hacerle preguntas sobre su hijo y el trabajo, su vida cotidiana y costumbres (del hijo y de la familia): horarios, comidas, etc.

Pregunta 2

1	2	3	4	5

Pregunta 3

SI
NO
Argumentación
Intercambiar información

Pregunta 4

La tarea se corresponde con el descriptor número: o con el nivel A1 A2 B1 B2 C1 C2

Pregunta 5

Para evaluar o calificar a los candidatos serviría la escala de <i>Argumentación</i> .
Para evaluar a los candidatos serviría la escala de <i>Intercambiar información</i> .

Pregunta 6

el lenguaje necesario va a ser lingüísticamente complejo	1	2	3	4	5
el lenguaje necesario va a necesitar organización retórica					
el lenguaje necesario va a depender de saberes culturales concretos					
el lenguaje necesario va a producir un discurso con bastantes elementos pragmáticos					

2. Receta de comida

Explica a tu profesor cómo se hace una comida que te guste y sepas hacer. Y contesta a sus preguntas para aclararle las cosas que no entienda bien cómo hacer.

Pregunta 2

1	2	3	4	5

Pregunta 3

SI
NO
<i>Argumentación</i>
<i>Intercambiar información</i>

Pregunta 4

La tarea se corresponde con el descriptor número: o con el nivel A1 A2 B1 B2 C1 C2

Pregunta 5

Para evaluar o calificar a los candidatos serviría la escala de <i>Argumentación</i> .
Para evaluar a los candidatos serviría la escala de <i>Intercambiar información</i> .

Pregunta 6

el lenguaje necesario va a ser lingüísticamente complejo	1	2	3	4	5
el lenguaje necesario va a necesitar organización retórica					
el lenguaje necesario va a depender de saberes culturales concretos					
el lenguaje necesario va a producir un discurso con bastantes elementos pragmáticos					

3. Dar indicaciones de un plano (usando el plano de Granada)

Vas a dar una fiesta este fin de semana en tu casa de Granada y has invitado a unos amigos

y a tu profesor. Explícale al profesor el camino para llegar a tu casa.

Si vives lejos, utiliza un lugar céntrico y conocido para empezar a explicar el camino.

Pregunta 2

1	2	3	4	5

Pregunta 3

SI
NO
Argumentación
Intercambiar información

Pregunta 4

La tarea se corresponde con el descriptor número: o con el nivel A1 A2 B1 B2 C1 C2

Pregunta 5

Para evaluar o calificar a los candidatos serviría la escala de <i>Argumentación</i> .
Para evaluar a los candidatos serviría la escala de <i>Intercambiar información</i> .

Pregunta 6

el lenguaje necesario va a ser lingüísticamente complejo	1	2	3	4	5
el lenguaje necesario va a necesitar organización retórica					
el lenguaje necesario va a depender de saberes culturales concretos					
el lenguaje necesario va a producir un discurso con bastantes elementos pragmáticos					

4. Opinión sobre actividad cultural

Has estado un mes en Granada viviendo y estudiando español. Seguro que has ido al cine o a algún espectáculo, o has leído algo interesante sobre Granada o España y su cultura. ¿Puedes decir a tu profesor qué cosas te han gustado más y charlar con él sobre una de ellas expresando tus opiniones y contestando a sus preguntas para dar más información o detalle?

Pregunta 2

1	2	3	4	5

Pregunta 3

SI
NO
Argumentación
Intercambiar información

Pregunta 4

La tarea se corresponde con el descriptor número: o con el nivel A1 A2 B1 B2 C1 C2

Pregunta 5

Para evaluar o calificar a los candidatos serviría la escala de <i>Argumentación</i> .
Para evaluar a los candidatos serviría la escala de <i>Intercambiar información</i> .

Pregunta 6

el lenguaje necesario va a ser lingüísticamente complejo	1	2	3	4	5
el lenguaje necesario va a necesitar organización retórica					
el lenguaje necesario va a depender de saberes culturales concretos					
el lenguaje necesario va a producir un discurso con bastantes elementos pragmáticos					

5. Explicar ciertos aspectos de su trabajo para un colega que va a ayudarlo o sustituirle durante unos días

Vas a irte de vacaciones una semana y tienes que explicarle a un colega cómo es el trabajo que debe hacer, que es el que tú realizas habitualmente. Y a responder a sus preguntas sobre lo que él no entienda.

Pregunta 2

1	2	3	4	5

Pregunta 3

SI
NO
<i>Argumentación</i>
<i>Intercambiar información</i>

Pregunta 4

La tarea se corresponde con el descriptor número: o con el nivel A1 A2 B1 B2 C1 C2

Pregunta 5

Para evaluar o calificar a los candidatos serviría la escala de <i>Argumentación</i> .
Para evaluar a los candidatos serviría la escala de <i>Intercambiar información</i> .

Pregunta 6

el lenguaje necesario va a ser lingüísticamente complejo	1	2	3	4	5
el lenguaje necesario va a necesitar organización retórica					
el lenguaje necesario va a depender de saberes culturales concretos					
el lenguaje necesario va a producir un discurso con bastantes elementos pragmáticos					

6a. Argumentar a favor o en contra de la actual guerra de Israel contra Líbano

Como quizás sepas, estos días Israel está atacando a Líbano y ha matado a varios cientos de personas inocentes civiles, unas 300, y está destruyendo importantes servicios de uso civil, porque las milicias de Hezbolá secuestraron a dos soldados en la frontera norte de Israel. ¿Te parece que esto es justo, que Israel tiene el derecho a defenderse, como han dicho los estados unidos, o crees que en realidad y aunque tiene el derecho a defenderse, debe usar otros medios o tener más cuidado con sus acciones?

6b. Argumentar a favor o en contra de la guerra de Estados Unidos en Irak

¿Crees que la guerra de Irak ha sido justa y necesaria? Estás de acuerdo con las afirmaciones de EE.UU. de que ahora el mundo es más seguro y era necesario terminar con un dictador y asesino como Sadam Hussein.

Pregunta 2

1	2	3	4	5

Pregunta 3

SI
NO
<i>Argumentación</i>
<i>Intercambiar información</i>

Pregunta 4

La tarea se corresponde con el descriptor número: o con el nivel A1 A2 B1 B2 C1 C2

Pregunta 5

Para evaluar o calificar a los candidatos serviría la escala de <i>Argumentación</i> .
Para evaluar a los candidatos serviría la escala de <i>Intercambiar información</i> .

Pregunta 6

el lenguaje necesario va a ser lingüísticamente complejo	1	2	3	4	5
el lenguaje necesario va a necesitar organización retórica					
el lenguaje necesario va a depender de saberes culturales concretos					
el lenguaje necesario va a producir un discurso con bastantes elementos pragmáticos					

7. Explicar el viaje o experiencia más enriquecedora o interesante de los realizados durante la estancia en España durante este mes

- Para los niveles bajos

¿Puedes contarme algún viaje que has hecho durante este mes?

Usar las preguntas normales sobre viajes para que genere respuestas (dónde, con quién, cuándo, cómo, cuántos días, qué tal, experiencias interesantes, visitas, alojamiento, incidentes, vuelta a Granada...).

Pregunta 2

1	2	3	4	5

Pregunta 3

Preguntas 4 y 5

SI	<i>Pregunta 4 ></i>	La tarea se corresponde con el descriptor número: o con el nivel A1 A2 B1 B2 C1 C2
NO		
<i>Argumentación</i>	<i>Pregunta 5 ></i>	Para evaluar serviría la escala de <i>Argumentación</i> .
<i>Intercambiar información</i>	<i>></i>	Para evaluar serviría la escala de <i>Intercambiar información</i> .

Pregunta 6

el lenguaje necesario va a ser lingüísticamente complejo	1	2	3	4	5
el lenguaje necesario va a necesitar organización retórica					
el lenguaje necesario va a depender de saberes culturales concretos					
el lenguaje necesario va a producir un discurso con bastantes elementos pragmáticos					

- Para los niveles altos: Explicame el viaje o experiencia más enriquecedora o interesante de los realizados durante la estancia en España durante este mes.

Pregunta 2

1	2	3	4	5

Pregunta 3

Preguntas 4 y 5

SI	<i>Pregunta 4 ></i>	La tarea se corresponde con el descriptor número: o con el nivel A1 A2 B1 B2 C1 C2
NO		
<i>Argumentación</i>	<i>Pregunta 5 ></i>	Para evaluar serviría la escala de <i>Argumentación</i> .
<i>Intercambiar información</i>	<i>></i>	Para evaluar serviría la escala de <i>Intercambiar información</i> .

Pregunta 6

el lenguaje necesario va a ser lingüísticamente complejo	1	2	3	4	5
el lenguaje necesario va a necesitar organización retórica					
el lenguaje necesario va a depender de saberes culturales concretos					
el lenguaje necesario va a producir un discurso con bastantes elementos pragmáticos					

8. Explicar las ventajas de vivir y estudiar en el país

- Para los niveles bajos:

Explícame las ventajas de vivir en un país extranjero con una familia, frente a vivir en tu propio país, para aprender una lengua.

¿Qué crees que es lo mejor de vivir en un país extranjero para estudiar una lengua?
¿Y vivir con una familia esté bien? ¿Por qué?

Pregunta 2

1	2	3	4	5

Pregunta 3

Preguntas 4 y 5

SI	<i>Pregunta 4 ></i>	La tarea se corresponde con el descriptor número: o con el nivel A1 A2 B1 B2 C1 C2
NO		
<i>Argumentación</i>	<i>Pregunta 5 ></i>	Para evaluar serviría la escala de <i>Argumentación</i> .
<i>Intercambiar información</i>	<i>></i>	Para evaluar serviría la escala de <i>Intercambiar información</i> .

Pregunta 6

el lenguaje necesario va a ser lingüísticamente complejo	1	2	3	4	5
el lenguaje necesario va a necesitar organización retórica					
el lenguaje necesario va a depender de saberes culturales concretos					
el lenguaje necesario va a producir un discurso con bastantes elementos pragmáticos					

- Para los niveles altos:

Explícame las ventajas de vivir y estudiar en el país con una familia frente a vivir con otros estudiantes de tu misma nacionalidad.

Pregunta 2

1	2	3	4	5

Pregunta 3

Preguntas 4 y 5

SI	<i>Pregunta 4 ></i>	La tarea se corresponde con el descriptor número: o con el nivel A1 A2 B1 B2 C1 C2
NO		
<i>Argumentación</i>	<i>Pregunta 5 ></i>	Para evaluar serviría la escala de <i>Argumentación</i> .
<i>Intercambiar información</i>	<i>></i>	Para evaluar serviría la escala de <i>Intercambiar información</i> .

Pregunta 6

el lenguaje necesario va a ser lingüísticamente complejo	1	2	3	4	5
el lenguaje necesario va a necesitar organización retórica					
el lenguaje necesario va a depender de saberes culturales concretos					
el lenguaje necesario va a producir un discurso con bastantes elementos pragmáticos					

9. *Explicar el siguiente cómic* (Ver la fotocopia). El cómic se encuentra en §3.4.8.

Pregunta 2

1	2	3	4	5

Pregunta 3

SI
NO
<i>Argumentación</i>
<i>Intercambiar información</i>

Pregunta 4

La tarea se corresponde con el descriptor número: o con el nivel A1 A2 B1 B2 C1 C2

Pregunta 5

Para evaluar o calificar a los candidatos serviría la escala de <i>Argumentación</i> .
Para evaluar a los candidatos serviría la escala de <i>Intercambiar información</i> .

Pregunta 6

el lenguaje necesario va a ser lingüísticamente complejo	1	2	3	4	5
el lenguaje necesario va a necesitar organización retórica					
el lenguaje necesario va a depender de saberes culturales concretos					
el lenguaje necesario va a producir un discurso con bastantes elementos pragmáticos					

Pregunta 1

Para el test en general, completo con las 9 tareas:

Con un test en el que figuren estas 9 tareas, ¿crees que se podría medir el nivel general de habilidad oral de un candidato?

SI
NO
¿Haría falta algo más? ¿Qué?

APÉNDICE 8

DESCRIPCIÓN DE LA TAREA 9

(Citado en §3.6.2 y §3.6.2-i).

TAREA 9

Planeada para A2

Dar indicaciones para explicar el camino en un plano.
--

<p><i>Vas a dar una fiesta este fin de semana en tu casa de Granada y has invitado a unos amigos y a tu profesor. Explícale al profesor el camino para llegar a tu casa. Si vives lejos, utiliza un lugar céntrico y conocido para empezar el camino.</i></p>

Explicación de la tarea 9

Esta tarea era una de las diseñadas para evaluar a candidatos de nivel bajo y trataba de que el candidato describiera la forma de ir a su casa a un grupo de amigos o a su profesor, con motivo de una fiesta, dando instrucciones. Se pretendía obtener un lenguaje básico cercano a su descriptor de la escala de *Intercambiar información*, “da y comprende indicaciones e instrucciones sencillas; por ejemplo, explica cómo ir a un lugar”. El tipo de lenguaje que buscábamos con ella era el de dar indicaciones explicando cómo ir a un lugar.

No obstante, como se explica en §3.6.2- i, tras su pilotaje con candidatos voluntarios y las opiniones con expertos, se desechó su uso en el nuevo test.

APÉNDICE 9

RESULTADOS DE LAS CALIFICACIONES EN LA OPI Y EN EL TEST DEL MCER

(Citado en el capítulo 4).

Este apéndice muestra los resultados de las calificaciones de las entrevistas:

1. Notas obtenidas en el test de ACTFL y del MCER.
2. Calificaciones en el test del MCER y calificaciones dadas por cada evaluador.
3. Calificaciones dadas por el evaluador número 1 en las 8 tareas del test del MCER.
4. Calificaciones dadas por el evaluador número 2 en las 8 tareas del test del MCER.
5. Calificaciones dadas por el evaluador número 3 en las 8 tareas del test del MCER.
6. Promedio de las calificaciones de cada tarea y cada candidato en el test del MCER.

9. 1. Notas obtenidas en el test de ACTFL y del MCER

Valores:

MCER: Menos que A1: 0, A1: 1, A2: 2, B1: 3, B2: 4, C1: 5, C2: 6.

ACTFL: Inicial bajo: 1, Inicial medio: 2, Inicial alto: 3, Intermedio bajo: 4, Intermedio medio: 5, Intermedio alto: 6, Avanzado bajo: 7, Avanzado medio: 8, Avanzado alto: 9, Superior: 10.

Número de candidato	Calificación ACTFL	Calificación MCER
1	2,00	0,00
2	2,00	0,00
3	3,00	1,00
4	3,00	1,00
5	2,00	1,00
6	4,00	2,00
7	3,00	1,00
8	4,00	2,00
9	4,00	3,00
10	5,00	3,00
11	5,00	3,00
12	4,00	3,00
13	5,00	4,00
14	5,00	4,00
15	6,00	4,00
16	6,00	4,00
17	6,00	4,00
18	7,00	4,00
19	7,00	4,00
20	7,00	4,00
21	8,00	4,00
22	9,00	6,00
23	10,00	6,00
24	10,00	5,00
25	10,00	6,00
26	7,00	4,00
27	8,00	4,00
28	9,00	6,00

9. 2. Calificaciones en el test del MCER y calificaciones dadas por cada evaluador

Valores:

MCER: Menos que A1: 0, A1: 1, A2: 2, B1: 3, B2: 4, C1: 5, C2: 6.

Número de candidato	Calificación en el MCER	Calificación en el MCER Juez 1	Calificación en el MCER Juez 2	Calificación en el MCER Juez 3
1	0,00	0,00	0,00	0,00
2	0,00	0,00	0,00	0,00
3	1,00	1,00	1,00	1,00
4	1,00	1,00	1,00	1,00
5	1,00	1,00	1,00	1,00
6	2,00	2,00	3,00	2,00
7	1,00	1,00	1,00	2,00
8	2,00	2,00	2,00	2,00
9	3,00	3,00	3,00	2,00
10	3,00	3,00	3,00	3,00
11	3,00	3,00	3,00	3,00
12	3,00	3,00	3,00	3,00
13	4,00	3,00	4,00	4,00
14	4,00	4,00	4,00	4,00
15	4,00	4,00	4,00	4,00
16	4,00	4,00	4,00	4,00
17	4,00	4,00	4,00	4,00
18	4,00	4,00	4,00	4,00
19	4,00	4,00	4,00	4,00
20	4,00	4,00	4,00	4,00
21	4,00	4,00	4,00	5,00
22	6,00	6,00	5,00	6,00
23	6,00	6,00	6,00	6,00
24	5,00	5,00	5,00	6,00
25	6,00	6,00	6,00	5,00
26	4,00	4,00	4,00	5,00
27	4,00	4,00	4,00	4,00
28	6,00	6,00	6,00	6,00

9. 3. Calificaciones dadas por el evaluador número 1 en cada una de las 8 tareas del test del MCER

Valores:

MCER: Menos que A1: 0, A1: 1, A2: 2, B1: 3, B2: 4, C1: 5, C2: 6.

Juez 1 tarea 1	Juez 1 tarea 2	Juez 1 tarea 3	Juez 1 tarea 4	Juez 1 tarea 5	Juez 1 tarea 6	Juez 1 tarea 7	Juez 1 tarea 8	Número de candidato
0,00	0,00				0,00	0,00	0,00	1
1,00	1,00				0,00	0,00	0,00	2
1,00					1,00	1,00	1,00	3
1,00	1,00	1,00			1,00	1,00	2,00	4
1,00					1,00	1,00	1,00	5
2,00		3,00		2,00	3,00	2,00	3,00	6
1,00	1,00	1,00		1,00	1,00	1,00	1,00	7
2,00	2,00				2,00	2,00	2,00	8
	2,00	3,00			2,00	3,00	3,00	9
2,00		3,00		3,00	3,00	3,00	3,00	10
2,00		3,00		3,00	3,00	3,00	3,00	11
	2,00	3,00		4,00	3,00	3,00	3,00	12
	3,00	3,00	3,00	3,00	3,00	3,00	3,00	13
	4,00	4,00		4,00	4,00	4,00	4,00	14
2,00	3,00	4,00		4,00	4,00	4,00	4,00	15
			4,00	4,00	4,00	4,00	3,00	16
			4,00	4,00	4,00	4,00	4,00	17
			4,00	4,00	5,00	4,00	5,00	18
	4,00	4,00	4,00	4,00	4,00	4,00	5,00	19
			4,00	4,00	4,00	4,00	4,00	20
			5,00	4,00	5,00	5,00	4,00	21
			6,00	6,00	6,00	6,00	6,00	22
			6,00	6,00	6,00	6,00	6,00	23
		5,00		5,00	5,00	5,00	5,00	24
	6,00		6,00	6,00	6,00	6,00	6,00	25
4,00		4,00		4,00		4,00	4,00	26
4,00		4,00		4,00		4,00	4,00	27
	6,00		6,00	6,00	6,00	6,00	6,00	28

9. 4. Calificaciones dadas por el evaluador número 2 en cada una de las 8 tareas del test del MCER

Valores:

MCER: Menos que A1: 0, A1: 1, A2: 2, B1: 3, B2: 4, C1: 5, C2: 6.

Juez 2 tarea 1	Juez 2 tarea 2	Juez 2 tarea 3	Juez 2 tarea 4	Juez 2 tarea 5	Juez 2 tarea 6	Juez 2 tarea 7	Juez 2 tarea 8	Número de candidato
0,00	0,00				0,00	0,00	0,00	1
1,00	0,00				0,00	1,00	0,00	2
1,00					1,00	1,00	1,00	3
1,00	1,00	1,00			1,00	1,00	2,00	4
1,00					1,00	1,00	1,00	5
3,00		3,00		3,00	3,00	3,00	3,00	6
1,00	1,00	1,00		1,00	1,00	1,00	1,00	7
2,00	2,00				2,00	2,00	2,00	8
	2,00	3,00			3,00	3,00	3,00	9
3,00		3,00		3,00	3,00	3,00	3,00	10
3,00		2,00		3,00	3,00	3,00	3,00	11
	3,00	3,00		4,00	3,00	3,00	3,00	12
	4,00	4,00	4,00	4,00	4,00	4,00	4,00	13
	4,00	4,00		4,00	4,00	4,00	4,00	14
	4,00	4,00		4,00	4,00	4,00	4,00	15
			4,00	4,00	4,00	4,00	4,00	16
			4,00	4,00	4,00	4,00	4,00	17
			4,00	4,00	4,00	4,00	4,00	18
	4,00	4,00	4,00	4,00	4,00	4,00	5,00	19
			4,00	4,00	4,00	4,00	4,00	20
			4,00	4,00	4,00	4,00	4,00	21
			6,00	5,00	6,00	5,00	5,00	22
			6,00	6,00	6,00	6,00	6,00	23
		5,00		5,00	5,00	5,00	5,00	24
	6,00		6,00	6,00	6,00	6,00	6,00	25
4,00		4,00		5,00		4,00	4,00	26
4,00		4,00		4,00		5,00	4,00	27
	6,00		6,00	6,00	6,00	6,00	6,00	28

9. 5. Calificaciones dadas por el evaluador número 3 en cada una de las 8 tareas del test del MCER

Valores:

MCER: Menos que A1: 0, A1: 1, A2: 2, B1: 3, B2: 4, C1: 5, C2: 6.

Juez 3 tarea 1	Juez 3 tarea 2	Juez 3 tarea 3	Juez 3 tarea 4	Juez 3 tarea 5	Juez 3 tarea 6	Juez 3 tarea 7	Juez 3 tarea 8	Número de candidato
0,00	0,00				0,00	0,00	0,00	1
0,00	0,00				0,00	1,00	0,00	2
1,00					1,00	1,00	1,00	3
1,00	0,00	1,00			1,00	1,00	1,00	4
1,00					1,00	1,00	1,00	5
1,00		2,00		2,00	2,00	2,00	1,00	6
2,00	2,00	2,00		0,00	2,00	2,00	2,00	7
2,00	2,00				1,00	2,00	2,00	8
	2,00	2,00			2,00	2,00	2,00	9
2,00		3,00		3,00	3,00	3,00	3,00	10
2,00		2,00		3,00	3,00	3,00	3,00	11
	2,00	3,00		4,00	3,00	3,00	3,00	12
	4,00	4,00	4,00	4,00	4,00	4,00	4,00	13
	4,00	4,00		4,00	4,00	4,00	4,00	14
	4,00	4,00		4,00	4,00	4,00	4,00	15
			4,00	4,00	5,00	5,00	5,00	16
			4,00	4,00	4,00	4,00	4,00	17
			4,00	4,00	4,00	4,00	4,00	18
			4,00	4,00	5,00	5,00	5,00	19
			4,00	4,00	5,00	5,00	4,00	20
			6,00	5,00	6,00	5,00	5,00	21
			6,00	6,00	6,00	6,00	6,00	22
			6,00	6,00	6,00	6,00	6,00	23
		6,00		5,00	6,00	6,00	6,00	24
	6,00		5,00	6,00	6,00	5,00	5,00	25
4,00		5,00		5,00		5,00	5,00	26
4,00		4,00		4,00		5,00	4,00	27
	6,00		6,00	6,00	6,00	6,00	6,00	28

9. 6. Promedio de las calificaciones de cada tarea y cada candidato en el test del MCER

Valores:

MCER: Menos que A1: 0, A1: 1, A2: 2, B1: 3, B2: 4, C1: 5, C2: 6.

Tarea 1	Tarea 2	Tarea 3	Tarea 4	Tarea 5	Tarea 6	Tarea 7	Tarea 8	Número de candidato
0,00	0,00				0,00	0,00	0,00	1
0,67	0,33				0,00	0,67	0,00	2
1,00					1,00	1,00	1,00	3
1,00	0,67	1,00			1,00	1,00	1,67	4
1,00					1,00	1,00	1,00	5
2,00		2,67		2,33	2,67	2,33	2,33	6
1,33	1,33	1,33		0,67	1,33	1,33	1,33	7
2,00	2,00				1,67	2,00	2,00	8
	2,00	2,67			2,33	2,67	2,67	9
2,33		3,00		3,00	3,00	3,00	3,00	10
2,33		2,33		3,00	3,00	3,00	3,00	11
	2,33	3,00		4,00	3,00	3,00	3,00	12
	3,67	3,67	3,67	3,67	3,67	3,67	3,67	13
	4,00	4,00		4,00	4,00	4,00	4,00	14
	3,67	4,00		4,00	4,00	4,00	4,00	15
			4,00	4,00	4,33	4,33	4,00	16
			4,00	4,00	4,00	4,00	4,00	17
			4,00	4,00	4,33	4,00	4,33	18
			4,00	4,00	4,33	4,33	5,00	19
			4,00	4,00	4,33	4,33	4,00	20
			5,00	4,33	5,00	4,67	4,33	21
			6,00	5,67	6,00	5,67	5,67	22
			6,00	6,00	6,00	6,00	6,00	23
		5,33		5,00	5,33	5,33	5,33	24
	6,00		5,67	6,00	6,00	5,67	5,67	25
4,00		4,33		4,67		4,33	4,33	26
4,00		4,00		4,00		4,67	4,00	27
	6,00		6,00	6,00	6,00	6,00	6,00	28

APÉNDICE 10 CUADROS DE ANÁLISIS ESTADÍSTICOS DE CORRELACIÓN Y FIABILIDAD

(Citado en el capítulo 5).

Este apéndice muestra 7 análisis realizados con el test y las tareas:

1. Correlaciones no paramétricas (Spearman), entre los resultados de la OPI y del test del MCER.
2. Regresión. Fórmula para averiguar el nivel en ACTFL predefiniendo el del MCER.
3. Regresión. Fórmula para averiguar el nivel en el MCER predefiniendo el de ACTFL.
4. Correlaciones entre evaluadores (Pearson).
5. Correlaciones no paramétricas entre evaluadores.
6. Análisis de fiabilidad de las tareas 5, 6, 7, y 8.
7. Análisis factorial de las tareas 4, 5, 6, 7, y 8.

10. 1. Correlaciones no paramétricas. Entre los resultados de la OPI y el test del MCER

			ACTFL	MCER
Tau_b de Kendall	ACTFL	Coeficiente de correlación	1,000	,886(**)
		Sig. (bilateral)	.	,000
		N	28	28
	MCER	Coeficiente de correlación	,886(**)	1,000
		Sig. (bilateral)	,000	.
		N	28	28
Rho de Spearman	ACTFL	Coeficiente de correlación	1,000	,953(**)
		Sig. (bilateral)	.	,000
		N	28	28
	MCER	Coeficiente de correlación	,953(**)	1,000
		Sig. (bilateral)	,000	.
		N	28	28

** La correlación es significativa al nivel 0,01 (bilateral).

10. 2. Regresión. Fórmula para averiguar el nivel en ACTFL predefiniendo el del MCER

Variables introducidas/eliminadas(b)

Modelo	Variables introducidas	Variables eliminadas	Método
1	MCER(a)	.	Introducir

a Todas las variables solicitadas introducidas

b Variable dependiente: ACTFL

Resumen del modelo

Modelo	R	R cuadrado	R cuadrado corregida	Error típ. de la estimación
1	,935(a)	,875	,870	,90809

a Variables predictoras: (Constante), MCER

ANOVA(b)

Modelo		Suma de cuadrados	gl	Media cuadrática	F	Sig.
1	Regresión	149,810	1	149,810	181,669	,000(a)
	Residual	21,440	26	,825		
	Total	171,250	27			

a Variables predictoras: (Constante), MCER

b Variable dependiente: ACTFL

Coefficientes(a)

Modelo		Coefficients no estandarizados		Coefficientes estandarizados	t	Sig.
		B	Error típ.	Beta	B	Error típ.
1	(Constante)	1,317	,371		3,551	,001
	MCER	1,335	,099	,935	13,478	,000

a Variable dependiente: ACTFL

$$\text{ACTFL} = 1.317 + 1.335 (\text{MCER})$$

Ejemplo: para un valor del MCER de 1 (A1), uno de ACTFL de 2.64 (inicial medio).

10. 3. Regresión. Fórmula para averiguar el nivel en el MCER predefiniendo el de ACTFL

Variables introducidas/eliminadas(b)

Modelo	Variables introducidas	Variables eliminadas	Método
1	ACTFL(a)	.	Introducir

a Todas las variables solicitadas introducidas

b Variable dependiente: MCER

Resumen del modelo

Modelo	R	R cuadrado	R cuadrado corregida	Error típ. de la estimación
1	,935(a)	,875	,870	,63640

a Variables predictoras: (Constante), ACTFL

ANOVA(b)

Modelo		Suma de cuadrados	gl	Media cuadrática	F	Sig.
1	Regresión	73,577	1	73,577	181,669	,000(a)
	Residual	10,530	26	,405		
	Total	84,107	27			

a Variables predictoras: (Constante), ACTFL

b Variable dependiente: MCER

Coefficientes(a)

Modelo		Coefficients no estandarizados		Coefficientes estandarizados	t	Sig.
		B	Error típ.	Beta	B	Error típ.
1	(Constante)	-,448	,304		-1,470	,153
	ACTFL	,655	,049	,935	13,478	,000

a Variable dependiente: MCER

$$\text{MCER} = -0.44 + 0.65 (\text{ACTFL})$$

Ejemplo: para un valor de ACTFL de 4 (intermedio bajo), uno del MCER de 2.16 (A2).

10. 4. Correlaciones entre evaluadores (Pearson)

Correlaciones

		juez1	juez2	juez3
juez1	Correlación de Pearson	1	,982(**)	,961(**)
	Sig. (bilateral)		,000	,000
	N	28	28	28
juez2	Correlación de Pearson	,982(**)	1	,954(**)
	Sig. (bilateral)	,000		,000
	N	28	28	28
juez3	Correlación de Pearson	,961(**)	,954(**)	1
	Sig. (bilateral)	,000	,000	
	N	28	28	28

** La correlación es significativa al nivel 0,01 (bilateral).

10. 5. Correlaciones no paramétricas entre evaluadores

Correlaciones

			juez1	juez2	juez3
Tau_b de Kendall	juez1	Coefficiente de correlación	1,000	,962(**)	,919(**)
		Sig. (bilateral)	.	,000	,000
		N	28	28	28
	juez2	Coefficiente de correlación	,962(**)	1,000	,923(**)
		Sig. (bilateral)	,000	.	,000
		N	28	28	28
	juez3	Coefficiente de correlación	,919(**)	,923(**)	1,000
		Sig. (bilateral)	,000	,000	.
		N	28	28	28
Rho de Spearman	juez1	Coefficiente de correlación	1,000	,980(**)	,961(**)
		Sig. (bilateral)	.	,000	,000
		N	28	28	28
	juez2	Coefficiente de correlación	,980(**)	1,000	,969(**)
		Sig. (bilateral)	,000	.	,000
		N	28	28	28
	juez3	Coefficiente de correlación	,961(**)	,969(**)	1,000
		Sig. (bilateral)	,000	,000	.
		N	28	28	28

** La correlación es significativa al nivel 0,01 (bilateral).

10. 6. Análisis de fiabilidad de las tareas 5, 6, 7, y 8

Escala: TODAS LAS VARIABLES

Resumen del procesamiento de los casos

		N	%
Casos	Válidos	19	67,9
	Excluidos(a)	9	32,1
	Total	28	100,0

a. Eliminación por lista basada en todas las variables del procedimiento.

Estadísticos de fiabilidad

Alfa de Cronbach	N de elementos
,993	4

Estadísticos de los elementos

	Media	Desviación típica	N
t5promedio	4,0877	1,32796	19
t6promedio	4,2281	1,29602	19
t7promedio	4,1404	1,26352	19
t8promedio	4,1404	1,27326	19

Estadísticos total-elemento

	Media de la escala si se elimina el elemento	Varianza de la escala si se elimina el elemento	Correlación elemento-total corregida	Alfa de Cronbach si se elimina el elemento
t5promedio	12,5088	14,548	,967	,995
t6promedio	12,3684	14,628	,988	,989
t7promedio	12,4561	14,842	,993	,988
t8promedio	12,4561	14,854	,981	,991

Estadísticos de la escala

Media	Varianza	Desviación típica	N de elementos
16,5965	26,106	5,10939	4

10. 7. Análisis factorial de las tareas 4, 5, 6, 7, y 8

Comunalidades

	Inicial	Extracción
t4promedio	1,000	,968
t5promedio	1,000	,978
t6promedio	1,000	,982
t7promedio	1,000	,986
t8promedio	1,000	,922

Método de extracción: Análisis de Componentes principales.

Varianza total explicada

Componente	Autovalores iniciales			Sumas de las saturaciones al cuadrado de la extracción		
	Total	% de la varianza	% acumulado	Total	% de la varianza	% acumulado
1	4,836	96,715	96,715	4,836	96,715	96,715
2	,105	2,095	98,810			
3	,030	,590	99,400			
4	,018	,359	99,759			
5	,012	,241	100,000			

Método de extracción: Análisis de Componentes principales.

Matriz de componentes(a)

	Componente
	1
t7promedio	,993
t6promedio	,991
t5promedio	,989
t4promedio	,984
t8promedio	,960

Método de extracción: Análisis de componentes principales.

a. 1 componentes extraídos

Matriz de componentes rotados(a)

a. Sólo se ha extraído un componente. La solución no puede ser rotada.

BIBLIOGRAFÍA

AVERMAET, P. VAN / KUIJPER, H. / SAVILLE N. (2005), *Setting and Monitoring Professional Standards: a QMS approach*, [pdf], (consulta: 21 de diciembre de 2007), <http://www.cambridgeesol.org/rs_notes/rs_nts22.pdf>.

AMERICAN COUNCIL ON THE TEACHING OF FOREIGN LANGUAGES (ACTFL) (1986), *ACTFL Proficiency Guidelines*, Yonkers, NY: ACTFL.

AMERICAN COUNCIL ON THE TEACHING OF FOREIGN LANGUAGES (ACTFL) (1999), *Standards for Foreign Language Learning in the 21st Century. National Standards in Foreign Language Education Project*, Yonkers, NY: ACTFL

ALDERSON J. Ch. / CLAPHAM, C. / WALL, D. (1998), *Exámenes de idiomas. Elaboración y evaluación*, Madrid: Cambridge University Press. [Título original: *Language Test Construction and Evaluation*, Cambridge: Cambridge University Press (1995)].

ALDERSON, J. Ch. (2000), *Assessing Reading*, Cambridge: Cambridge University Press.

ALDERSON, J. Ch. (ed.) (2002), *Common European Framework of Reference for Languages: Learning, Teaching, Assessment: Case Studies*, Strasbourg: Council of Europe.

ALDERSON, J. Ch. (2005), "Editorial", *Language Testing*, 22: 257-260.

ALDERSON, J. Ch. (2007), "The CEFR and the Need for More Research", *Modern Language Journal*, 91 (4): 659- 663.

ALDERSON, J. Ch. / BANERJEE, J. (2001), "Language Testing and Assessment (Part 1)", *Language Testing*, 34: 213-37.

ALDERSON, J. Ch. / BANERJEE, J. (2002), "Language Testing and Assessment (Part 2)", *Language Testing*, 35: 79-113.

ALDERSON, J. Ch. / FIGUERAS, N. / KUIJPER, H. / NOLD, G. / TAKALA, S. / TARDIEU, C. (2004), *The development of specifications for item development and classification within the European Framework of Reference for Languages: Learning, Teaching, Assessment, Reading and Listening. Final Report of The Dutch CEF Construct Project*, Lancaster: Lancaster University.

ALDERSON J. Ch. / HUHTA, A. (2005), "The development of a suite of computer-based diagnostic tests base don the Common European Framework", *Language Testing*, 22 (3): 301-20.

ASSOCIATION OF LANGUAGE TESTERS IN EUROPE (ALTE) (2001 a), *Principles of Good Practice for ALTE Examinations*, [pdf], (consulta: 21 de diciembre de 2007), <http://www.alte.org/quality_assurance/code/good_practice.pdf>.

ASSOCIATION OF LANGUAGE TESTERS IN EUROPE (ALTE), (2001 b), *The ALTE Code of Practice and Quality Management Checklists*, [en línea], (consulta: 21 de diciembre de 2007), <http://www.alte.org/quality_assurance/code/checklist.php>.

BACHMAN, L. F. / SAVIGNON, S. J. (1985), "Evaluation of Communicative Competence: The ACTFL Provisional Proficiency Guidelines", *The Modern Language Journal*, 69 (2): 129-34.

BACHMAN, L. F. / SAVIGNON, S. J. (1986), "The Evaluation of Communicative Language Proficiency: A Critique of the ACTFL Oral Interview", *The Modern Language Journal*, 70 (4): 380-390.

BACHMAN, L. F. (1988), "Problems in Examining the Validity of the ACTFL Oral Proficiency Interview", *Studies in Second Language Acquisition*, 10 (2): 149- 163.

BACHMAN, L. F. (1990), *Fundamental Considerations in Language Testing*, Oxford: Oxford University Press.

BACHMAN, L. (2004), *Statistical Analyses for Language Assessment*, Cambridge: Cambridge University Press.

BACHMAN, L. F. (1991), "What does language testing have to offer?", *TESOL Quarterly*, 25 (4): 671-704.

BACHMAN, L. / PALMER, A. (1996), *Language Testing in Practice*, Oxford: Oxford University Press.

BAILEY, P. / ONWUEGBUZIE, A. / DALEY, C. (2000), "Correlates of anxiety at three stages of the foreign language learning process", *Journal of Language and Social Psychology*, 19: 474-490.

BALSEBRE, A. / VIDAL, D. / MATEU, M. (1998), *La entrevista en radio, prensa y televisión*, Madrid: Cátedra.

BARALO, M. (2005), "La competencia léxica en el «Marco común europeo de referencia»", *Carabela*, 58: 27- 48.

BARNWELL, D. (1989), "«Naive» native speakers and judgements of oral proficiency in Spanish", *Language Testing*, 6:152-163.

BONNET, G. (2007), "The CEF Rand Education Policies in Europe", *Modern Language Journal*, 91 (4): 669- 672.

BORDÓN, T. (2006), *La evaluación de la lengua en el marco de E/L2: Bases y procedimientos*, Madrid: Arco/ Libros, S. L.

BORDÓN, T. (2007), “La autoevaluación de adultos por medio de los cuadros propuestos en el PEL”, *Carabela*, 60: 73-98.

BREINER-SANDERS, K. E. / LOWE, P. / MILES J. / SWENDER, E. (2000), “ACTFL proficiency guidelines-speaking, revised”, *Foreign Language Annals*, 33 (1): 13-18.

BROWN, A. (2003), “Interviewer variation and the co-construction of speaking proficiency”, *Language Testing*, 20 (1): 1-25.

BUITRAGO, A. (2006), “La evaluación del español como lengua extranjera”, [pdf], Salamanca: Universidad de Salamanca. [Documento inédito, aportado en un taller de evaluación de la Universidad de Salamanca de mayo de 2006].

BYRNES, H. (2007), “Perspectives“, *Modern Language Journal*, 91 (4): 641-645.

CALSAMIGLIA, H. / TUSÓN A. (2002), *Las cosas del decir. Manual de Análisis del discurso*, Barcelona: Ariel Lingüística.

CANALE, M. / SWAIN, M. (1980), “Theoretical Bases of Communicative Approaches to Second Language Teaching and Testing“, *Applied Linguistics*, 1:1- 47.

CAPLAN, D. (1996), *El lenguaje. Estructura, procesamiento y trastornos*, Buenos Aires: Docencia.

CARROLL, B. J. (1983), *Testing Communicative Performance. An Interym Study*, Oxford: Pergamon Press.

CASADO, M. A. / DERESHIWSKY, M. I. (2001), “Foreign language anxiety of university students“, *College Students Journal*, 35: 539-551.

CASSANY, D. (2002), “El portafolio europeo de lenguas“, *Aula de Innovación educativa*, 117: 13-17.

CASSANY, D. (coord.) (2004), *Guía didáctica Portfolio europeo de lenguas para Enseñanza secundaria*, [pdf], (consulta: 21 de diciembre de 2007), <http://www.mec.es/programas-europeos/docs/pel_guia_didactica_secundaria.pdf>.

CASSANY, D. (2007), “Del portfolio al e-PEL“, *Carabela*, 60: 5- 21.

CHALHOUB-DEVILLE, M. (1997), “Theoretical models, assessment frameworks and test construction”, *Language Testing*, 14 (1): 3-22.

CHALHOUB-DEVILLE, M. / FULCHER, G. (2003), “The Oral Proficiency Interview: A Research Agenda”, *Foreign Language Annals*, 36 (4): 498- 506.

CHENG, Y. (2002), “Factors associated with foreign language writing anxiety”, *Foreign Language Annals*, 35 (6): 647-656.

CLARK, D. / CLIFFORD, R. T. (1988), “The FSI/ACTFL Proficiency Scales and testing techniques, Development, Current Status, and Needed research”, *Second Language Acquisition*, 10: 129-147.

CONSEJO DE EUROPA (2002), *Marco común europeo de referencia para las lenguas: aprendizaje, enseñanza, evaluación*, Madrid: Ministerio de Educación, Cultura y Deporte; Instituto Cervantes. [Título original: *Common European Framework of Reference for Languages Learning, teaching, assessment*, (2001), Strasbourg: Council of Europe].

CONSEJO DE EUROPA (2002 b), *Relating Language Examinations to the CEF, Preliminary Pilot Version*, [pdf], Strasbourg: Council of Europe, (consulta: 8 de agosto de 2005), <<http://www.coe.int/lang>>.

CONSEJO DE EUROPA (2005), *Relating Language Examinations to the CEF, Manual Preliminary Pilot Versión. Guidance for Feedback, Piloting and Case Studies, Version 3*, [pdf], Strasbourg: Council of Europe, (consulta: 16 de junio de 2007), <<http://www.coe.int/lang>>.

CONSEJO DE EUROPA (2007), *Relating Language Examinations to the CEF, Manual Preliminary Pilot Versión. Introduccion and Feedback to the pilot phase*, [pdf], Strasbourg: Council of Europe, (consulta: 2 de octubre de 2007), <<http://www.coe.int/lang>>.

DAVISON, F. / FULCHER, G. (2007), “The Common European Framework of Reference (CEFR) and the design of language tests: A matter of effect”, *Language Teaching*, 40 (3): 231- 241.

EUROPEAN ASSOCIATION FOR LANGUAGE TESTING AND ASESMENT (EALTA) (2004), *Directrices de EALTA para una buena práctica en el desarrollo de pruebas y la evaluación de lenguas*, [en línea], (consulta : 5 de diciembre de 2007), <<http://www.ealta.eu.org/documents/archive/guidelines/Spanish.pdf>>.

ELLIS, R. (1990), *Instructed Second Language Acquisition*, Oxford: Basil Blakwel.

ELLIS, R. (2001), “Some thoughts on testing grammar: An SLA perspective”, en ELDER, C. / BROWN, A. / GROVE, E. / HILL, K. / IWASHITA, N. / LUMLEY, T. / MCNAMARA, T. / O’LOUGHLIN, K. (eds), *Experimenting with uncertainty. Essays in honour of Alan Davies*, Cambridge: Cambridge University Press: 251- 263.

FIGUERAS, N. (2004), “Docencia, evaluación y contexto de uso real de la lengua oral”, *Carabela*, 55: 63-84.

FIGUERAS, N. (2005 a), “El Marco común europeo de referencia para las lenguas: de la teoría a la práctica”, *Carabela*, 57: 5-23.

FIGUERAS, N. (2005 b), “Relating Examinations to the Common European Framework: a manual”, *Language Testing*, 22 (3): 261-79.

FIGUERAS, N. (2007), “The CEFR, a Lever for the Improvement of Language Professionals in Europe”, *Modern Language Journal*, 91 (4): 673-675.

FOUCAULT, M. (1984), *La voluntad de saber*, Madrid: Siglo XXI.

FULCHER, G. (2003), *Testing Second Language Speaking*, London: Pearson/Longman.

FULCHER, G. (2004 a), “Are Europe's tests being built on an 'unsafe' framework?”, *Guardian Weekly TEFL Supplement*, [en línea], 18 de marzo de 2004, (consulta: 21 de diciembre de 2007), <<http://education.guardian.co.uk/tefl/story/0,,1170569,00.html>>.

FULCHER, G. (2004 b), “Deluded by Artifices? The Common European Framework and Harmonization”, *Language Assessment Quarterly*, 1 (4): 253- 266.

FULCHER, G. (2005), “Better communications test will silence critics”, en *Guardian Weekly*, [en línea], 18 de noviembre de 2005, (consulta: 21 de diciembre de 2007), <<http://education.guardian.co.uk/tefl/story/0,,1645011,00.html>>.

NEWFIELDS, T. (2007), “An Interview with Glenn Fulcher”, *JALT Testing & Evaluation SIG Newsletter*, [en línea], 12 de agosto de 2007, (consulta: 21 de diciembre de 2007), 12: 9 – 17, <<http://www.tnewfields.info/Articles/intFulcher.htm>>.

GARCÍA RAMOS, J. M. (1989), *Bases pedagógicas de la evaluación*, Madrid: Síntesis.

HALLECK, G. B. (1992), “The Oral Proficiency Interview: Discrete Point test or a Measure of Communicative Language Ability?”, *Foreign Language Annals*, 25 (3): 227-231.

HALLECK, G. B. (1995), “Assessing Oral Proficiency: A Comparison of Holistic and Objective Measures”, *The Modern Language Journal*, 79 (2): 223- 234.

HALLECK, G. B. (1996), “Interrater Reliability of the OPI: Using Academic Trainee Raters”, *Foreign Language Annals*, 29 (2): 223- 238.

HE, A. W. / YOUNG, R. (1998), “Language Proficiency Interviews: A Discourse Approach”, en YOUNG, R. / HE, A. W. (eds.) (1988), *Talking and Testing, Discourse Approaches to the Assessment of Oral Proficiency*, Amsterdam/Philadelphia: John Benjamin: 1-23.

HENNING, G. (1987), *A Guide to Language Testing. Development, Evaluation, Research*, Cambridge, Mass.: Newbury House.

HERNÁNDEZ, P. (2006), *El discurso hablado de extranjeros: cómo detectar las causas y los efectos de un mal planteamiento de preguntas y de la errónea interpretación de sus respuestas en la evaluación oral y en el aula de lenguas*. [Memoria inédita del I Máster de ELE del Departamento de Lengua Española de la Universidad de Granada, dirigida por MARTÍNEZ BAZTÁN, A].

HORWITZ, E. K. / HORWITZ, M. B. / COPE, J. (1986), “Foreign language classroom anxiety”, *Modern Language Journal*, 70: 125-132.

HOUTEN, J. B, VAN (2007), “Language Learning Policies in the United States”, Estrasburgo: Intergovernmental Policy Forum. [Documento de Power Point, presentado en el Intergovernmental Policy Forum organizado por el Consejo de Europa en Estrasburgo durante los días 6 - 8 de febrero de 2007].

HUHTA, A. / LUOMA, S. / OSCARSON, M. / SAJAVAARA, K. / TAKALA, S. / TEASDALE, A. (2002), “A diagnostic language assessment system for adult learners”, en ALDERSON, J. Ch. (ed.) (2002), *Common European Framework of Reference for Languages: Learning, Teaching, Assessment: Case Studies*, Strasbourg: Council of Europe: 130-46.

HULSTIJN, H. J. (2007), “The Shaky Ground Beneath the CERF: Quantitative and Qualitative Dimensions of Language Proficiency”, *Modern Language Journal*, 91 (4): 663- 666.

INTERNATIONAL LANGUAGE TESTERS ASSOCIATION (ILTA) (2000), *ILTA: Draft Code of Practice, Version 3*, [pdf], (consulta: 5 de diciembre de 2007), <http://www.iltaonline.com/CoP_3.1.htm>.

INTERNATIONAL LANGUAGE TESTERS ASSOCIATION (ILTA) (2000), *Code of Ethics for ILTA*, [en línea], (consulta: 5 de diciembre de 2007), <<http://www.iltaonline.com/code.pdf>>.

INSTITUTO CERVANTES (2006), *Plan curricular del Instituto Cervantes. Niveles de referencia para el español*, Madrid: Biblioteca Nueva.

JOHNSTON, M. (1987), "Understanding Learner Language", en NUMAN, D. (ed.), *Applying Second Language Acquisition Research*, Adelaida, Australia: AMEP: 5-44.

JONES, N. (2002), "Relating the ALTE framework to the Common European Framework of reference", en ALDERSON, J. Ch. (ed.), *Common European Framework of Reference for Languages: Learning, Teaching, Assessment: Case Studies*, Strasbourg: Council of Europe: 167-83.

JONES, N., (2005), *Seminar to calibrate examples of spoken performance*, [en línea], Cambridge: Cambridge ESOL / Council of Europe, (consulta: 21 de enero de 2008), <http://www.coe.int/T/DG4/Portfolio/?L=E&M=/main_pages/illustrationse.html>.

KRAMSCH, C. (1987), "The Proficiency Movement: second Language Acquisition Perspectives", *Studies in Second Language Acquisition*, 9 (3): 355- 62.

KRUMM, H. J. (2007), "Profiles Instead of Levels: The CEFR and its (Ab)Uses in the Context of Migration", *Modern Language Journal*, 91 (4): 667-669.

LANTOFF, J. P. / FRAWLEY, W. (1988), "Proficiency: Understanding the construct", *Studies in Second Language Acquisition*, 10: 181-195.

LAZARATON, A. (1992), "The Structural Organization of Language Interview: A Conversation Analytic Perspective", *System*, 20 (3): 373-386.

LAZARATON, A. (1996), "Interlocutor support in oral proficiency interviews: the case of CASE", *Language Testing*, 13 (2): 151-72.

LAZARATON, A. (2002), *A Qualitative Approach to the Validation of Oral language Tests*, Cambridge: UCLES, Cambridge University Press.

LEUNG, C. / LEWKOWICZ, J. (2000), "Expanding Horizons and Unresolved Conundrums: Language Testing and Assessment", *TESOL Quarterly*, 40: 211-240.

LISKIN-GASPARRO, J. (1987), "The ACTFL Proficiency Guidelines: A Historical Perspective", en HIGGS, T. V. (ed.) (1987), *Teaching for proficiency, the Organizing Principle*, Lincolwood, IL: National Textbook Co.: 11-42.

LISKIN-GASPARRO, J. (2003), "The ACTFL proficiency Guidelines and the Oral Proficiency Interview: A Brief History and Analysis of Their Survival", *Foreign Language Annals*, 36 (4): 483-491.

LITTLE, D. / LAZENBY SIMPSON, B. / O'CONNOR, F. (2002), "Meeting the English language needs of refugees in Ireland", en ALDERSON, J. Ch. (ed.) (2002), *Common European Framework of Reference for Languages: Learning, Teaching, Assessment: Case Studies*, Strasbourg: Council of Europe: 53-67.

LITTLE, D. / PERCLOVÁ, R. (2003), *El Portfolio Europeo de las Lenguas: guía para profesores y formadores de profesores*, [pdf], Madrid: Ministerio de Educación, Cultura y Deporte, (consulta: 21 de diciembre de 2007), <<http://www.mec.es/programas-europeos/docs/guiaprof.pdf>>.

LITTLE, D. (2005), "The Common European Framework and the European Language Portfolio: involving learners and their judgements in the assessment process", *Language Testing*, 22 (3): 321-36.

LITTLE, D. (2007), "The Common European Framework of Reference for Languages: Perspectives on the Making of Supranational Language Education Policy", *Modern Language Journal*, 91 (4): 645-655.

LLORIÁN, S. / RODRIGO, C. (2005), "Diseño y desarrollo del currículo de ELE a la luz del Marco común europeo de referencia para las lenguas", *Carabela*, 57: 49-80.

LONG, M. (1988), "Instructed Interlanguage Development", en BEEBE (ed.), *Issues in Second Language Acquisition: Multiple Perspectives*, Rowley, Mass: Newbury House.

LOWE, P. (1986), "Proficiency: Panacea, Framework, Process?: A Reply to Kramsch, Schulz and Particulary, to Bachman and Savignon", *The Modern Language Journal*, 70 (4): 391-397.

LOWE, P. (1998), "Keeping the Optic Constant: A Framework of Principles for Writing and Specifying the AEI Definitions of Language Abilities", *Foreign Language Annals*, 31 (3): 358-380.

LUOMA, S. (2004), *Assessing Speaking*, Cambridge: Cambridge University Press.

MACINTYRE, P. D. / GARDNER, R. (1989), "Anxiety and second-language learning. Toward a theoretical clarification", *Language Learning*, 39: 251-275.

MAGNAN, S. S. (1988), "Grammar and the ACTFL Oral Proficiency Interview: Discussion and Data", *The Modern Language Journal*, 72 (3): 266-276.

MAGNAN, S. / BACK, M. (2007), "Social Interaction and Linguistic Gain During Study Abroad", *Foreign Language Annals*, 40 (1): 43-78.

MARTÍNEZ BAZTÁN, A. (1995), "¿Les enseñamos, aprenden o aprenden a pesar de nosotros?", en *Aspectos de la enseñanza de lenguas extranjeras*, Granada: Método ediciones: 357-69.

MARTÍNEZ BAZTÁN, A. (2005), "Consideraciones sobre la evaluación del español como lengua de personas inmigradas adultas", *Glosas Didácticas*, [en línea], (consulta: 5 de diciembre de 2007), 15: 114-21, <<http://www.rediris.es/list/info/dilenguas.es.html>>.

MARTÍNEZ BAZTÁN, A. (2007), "Consideraciones sobre la evaluación de la lengua de personas inmigradas adultas y adolescentes", *Linred*, [pdf], (consulta de 5 de diciembre de 2007), 1-14, <http://www.linred.com/monograficos_pdf/LR_monografico1_articulo1.pdf>.

MCALPINE, D. / DHONAU, S. (2007), "Creating a Cultura for the Preparation of an ACTFL/NCATE Program Review?", *Foreign Language Annals*, 40 (2): 247-259.

MCNAMARA, T. F. (1995), "Modelling Performance: Opening Pandora's Box", *Applied Linguistics*, 16:159-179.

MCNAMARA, T. F. / LUMLEY, L. (1997), "The effect of interlocutor and assessment mode variables in overseas assessments of speaking skills in occupational settings", *Language Testing*, 14: 140-156.

MEREDITH, R. A. (1990), "The Oral proficiency Interview in real life: Sharpening the Scale", *The Modern Language Journal*, 74 (2): 288-296.

MIQUEL, L. / BAULEANS, N. S. (1985), *¿A que no sabes...?*, Madrid: Edi6.

MODER, C. L. / HALLECK, G. B. (1998), "Framing the Language Proficiency Interview as a Speech Event: Native and Non-Native Speaker's Questions", en YOUNG, R. / HE, A. W. (eds.) (1998), *Talking and Testing, Discourse Approaches to the Assessment of Oral Proficiency*, Amsterdam/Philadelphia: John Benjamin: 117-146.

MUÑOZ LICERAS, J. M. (1992), "Hacia un modelo de análisis de la interlengua" en MUÑOZ LICERAS, J. M (ed.), *La adquisición de las lenguas extranjeras*, Madrid: Visor: 11-27.

MUÑOZ LICERAS, J. (1996), *La adquisición de las segundas lenguas y la gramática universal*, Madrid: Síntesis.

NOIJONS, J. / KUIJPER, H. (2004), *Mapping the Dutch Foreign Language State Examinations onto the Common European Framework of reference*, Arnhem, The Netherlands: National Institute of Educational Measurement.

NORTH, B. (2002), "Developing descriptor scales of language proficiency for the CEF Common reference levels", en ALDERSON, J. Ch. (ed.) (2002), *Common European Framework of Reference for Languages: Learning, Teaching, Assessment: Case Studies*, Strasbourg: Council of Europe: 87-105.

NORTH, B. (2004), "Europe's framework promotes language discussion, not directives", *Guardian Weekly*, [en línea], 15 de abril de 2004, (consulta: 21 de diciembre de 2007), <<http://education.guardian.co.uk/tefl/story/0,,1191130,00.html>>.

NORTH, B. (2007), "The CEFR Illustrative Descriptor Scales", *Modern Language Journal*, 91 (4): 656-659.

NORTH, B. / LEPAGE, S. (2005), *Seminar to calibrate examples of spoken performances in line with the scales of CEFR*, [pdf], Strasbourg: Council of Europe, (consulta: 21 de enero de 2008), <<http://www.coe.int/T/DG4/Portfolio/documents/reportsevres.pdf>>.

NORTH, B. / LEPAGE, S. (2005 b), *Guide for the organization of a seminar to calibrate examples of spoken performances in line with the scales of CEFR*, [pdf], Strasbourg: Council of Europe, (consulta: 21 de enero de 2008), <<http://www.coe.int/T/DG4/Portfolio/documents/guideseminar.pdf>>.

NORTH, B. / MOYER, M. C. (2004), *Exemples de productions orales illustrant, pour le francais, les niveaux du Cadre européen commun de référence pour les langues*, [pdf], Strasbourg: Council of Europe, (consulta : 21 de enero de 2008), <www.coe.int/T/DG4/Portfolio/?L=E&M=/main_pages/illustrationse.html 8>

ORTEGA OLIVARES, J. (1996), "Evaluación en el aula de español/LE: algunas consideraciones generales", en RUEDA, M. / PRADO, E. / LEMEN, J. / GRANDE, F. J. (eds.), *Tendencias actuales en la enseñanza del español lengua extranjera II: actas del VI Congreso Internacional de ASELE*, León: Universidad de León: 19-33.

PARRONDO, J. R. (2004), "Modelos, tipos y escalas de evaluación", en SÁNCHEZ LOBATO, J. / SANTOS GARGALLO, I. (dirs.), *Vademécum para la formación de profesores*, Madrid: SGEL.

PEARSON, L. / B. FONSECA-GREBER / FOELL, K. (2006), "Advanced Proficiency for Foreign Language Teacher Candidates: What Can We Do to Help Them Achieve This Goal?", *Foreign Language Annals*, 39 (3): 507-519.

PENROSE, R. (2007), *La nueva mente del emperador*, Madrid: De Bolsillo. [Título original: *The Emperor's New Mind*, (1989), Oxford: Oxford University Press].

PHILLIPS, E. (1992), "The effects of language anxiety on students' oral test performance and attitudes", *Modern Language Journal*, 76: 14-26.

PINKER, S. (2001), *El instinto del lenguaje*, Madrid: Alianza Editorial.

PURPURA, J. E. (1999), *Learner strategy and performance on language tests: A structural equation modelling approach*, Cambridge: Cambridge University Press.

REAL ACADEMIA ESPAÑOLA (RAE) (2005), *Diccionario Panhispánico de dudas*, Madrid: Santillana.

RAFFALDINI, T. (1988), "The Use of Situation Test as Measures of Communicative Ability", *Studies on Second Language Acquisition*, 10: 197-216.

RINGS, L. (2006), "The Oral Interview and Cross-Cultural Understanding in the Foreign language Classroom", *Foreign Language Annals*, 30 (1): 43- 53.

ROSS, S. (1992), "Accommodative questions in oral proficiency interviews", *Language Testing*, 9: 173- 186.

ROSS, S. (1998), "Divergent Frame Interpretations in Language Proficiency Interview Interaction", en YOUNG, R. / HE, A. W. (eds.), *Talking and Testing, Discourse Approaches to the Assessment of Oral Proficiency*, Amsterdam/Philadelphia: John Benjamin: 333-353.

ROSS, S. / BERWICK, R. (1992), "The Discourse of Accommodation in Oral Proficiency Interviews", *Studies in Second Language Acquisition*, 14: 159-176.

SAITO, Y. / HORWITZ, E. K. / GARZA, T. J. (1999), "Foreign language reading anxiety", *Modern Language Journal*, 83: 202-218.

SALABERRY, R. (2000), "Revising the revised format of the ACTFL Oral Proficiency Interview", *Language Testing*, 17 (3): 287-310.

SAVIGNON, S. (1985), "Evaluation of Communicative Competence: The ACTFL Provisional Proficiency Guidelines", *The Modern Language Journal*, 69 (2): 129-134.

SAVILLE, N. (2005), "An Interview With John Trim at 80", *Language Assessment Quarterly*, 2 (4): 263-288.

SHOHAMY, E. (1988), "A proposed framework for testing the oral language of second foreign language learners", *Second language acquisition*, 10: 165-179.

SHULZ, R. (2002), "Changing Perspectives in Foreign Language Education: Where Do We Come From? Where Are we Going?", *Foreign Language Annals*, 35 (3): 285- 292.

SLAGTER, P. J. (1979), *Un Nivel Umbral*, Starsbourg: Council of Europe.

SOLER, E. A. (2000), "Desarrollo de la competencia discursiva oral en el aula de lenguas extranjeras: perspectivas metodológicas y de investigación", en MUÑOZ, C. (ed.), *Segundas Lenguas. Adquisición en el aula*, Barcelona: Ariel Lingüística.

SPOLSKY, B. (1995), *Measured Words, the Development of Objective Language Testing*, Oxford: Oxford University Press.

SPOLSKY, B. (2000), "Language Testing in «The Modern Language Journal»", *The Modern Language Journal*, 84: 536-552.

SURFACE, E. A. / DIERDORFT, E. C. (2003), "Reliability and the ACTFL Oral Proficiency Interview: Reporting Indices of Interrater Consistency and Agreement for 19 Languages", *Foreign Language Annals*, 36 (4): 507- 519.

SWENDER, E. (1989), *The Oral Proficiency Interview Tester Training Manual*, (Kathryn Buck ed.), Yonkers, NY: ACTFL.

SWENDER, E. (2003), "Oral Proficiency Testing in the Real World: Answers to Frequently Asked Questions", *Foreign Language Annals*, 36 (4): 520- 526.

TAGLIANTE, Ch. (2007), "Liens entre les certifications de français langue étrangère du MENESR et le CECR", CIEP. [Documento de Power Point, presentado en el *Policy Forum* organizado por el Consejo de Europa en Estrasburgo durante los días 6 - 8 de febrero de 2007].

THOMPSON, I. (1995), "A Study of Interrater Reliability of the ACTFL Oral Proficiency Interview in Five European Languages: Data from ESL, French, German, Russian, and Spanish", *Foreign Language Annals*, 28 (3): 407-22.

TRIM, J. L. / RICHTERICH, M. R. / EK, J. A., VAN / WILKINS, D. A. (1980), *Systems Development in Adult language Learning*, Oxford: Pergamon Press.

TRIM, J. L. M. (2001), *The work of the Council of Europe in the field of Modern Languages, 1957-2001*, [pdf], Graz: European Center for Modern Languages (ed.), (consulta: 21 de enero de 2008), <www.ecml.at/efsz/files/Trim.pdf>.

TRIM, J. L. M. (2007), *Modern Languages in the Council of Europe 1954-1997, International co-operation in support of lifelong language learning for effective communication, mutual cultural enrichment and democratic citizenship in Europe*, [en línea], Strasbourg: Council of Europe, (consulta: 18 de enero de 2008), <http://www.coe.int/t/dg4/linguistic/Source/TRIM_21janv2007_%20EN.doc>.

TSCHIRNER, E. (2005), "Das ACTFL OPI undder Europäische Referenzrahmen", *Babylonia*, 2: 50-55.

TURING, A. M. (1950), "Computing machinery and intelligence", *Mind* 59, 433-460.

VALDMAN, A. (1988), "Introduction", *Studies in Second Language Acquisition*, 10: 121- 128.

VANPATTEN, B. / CADIerno, T. (1993), "Input processing and second Language Acquisition: A role for Instruction", *The Modern Language Journal*, 77 (1): 45- 57.

VANPATTEN, B. / LEE, J. F. / BALLMAN, T. L. (2000), *¿Sabías que...?*, New York: McGraw-Hill Companies.

VANPATTEN, B. / WONG, W. (2003), "The evidence in IN: drills are OUT", *Foreign Language Annals*, 36 (3): 403-423.

VÁZQUEZ MACÍAS, C. (2006), *La evaluación de la expresión escrita según el "Marco común europeo de referencia para las lenguas*. [Memoria inédita del II Máster de ELE del Departamento de Lengua Española de la Universidad de Granada, dirigida por MARTÍNEZ BAZTÁN, A].

VERDÍA, E. (2002), "Comentarios al «Marco común europeo de referencia para las lenguas»", *Mosaico*, 9: 4-7.

WALTERS, F. S. (2007), "A conversation- analytic hermeneutic rating protocol to assess L´oral pragmatic competent", *Language Testing*, 2: 155-183.

WARREN, K. (2008), *La Evaluación Oral y el Estrés comunicativo: Identificar, localizar y explicar casos de estrés comunicativo en dos modelos de examen distintos*. [Memoria inédita del II Máster de ELE del Departamento de Lengua Española de la Universidad de Granada, dirigida por MARTÍNEZ BAZTÁN, A].

WEIR, J. W. (2005), "Limitations of the Common European Framework for developing comparable examinations and tests", *Language Testing*, 22: 281- 300.

WESTHOFF, G. (2007), "Challenges and Opportunities of the CEFR for Reimagining Foreign language Pedagogy", *Modern Language Journal*, 91 (4): 676- 679.

WIDOWSON, H. (2001), "Communicative language testing: The art of the possible",

en ELDER, C. / BROWN, A. / GROVE, E. / HILL, K. / IWASHITA, N. / LUMLEY, T. / MCNAMARA, T. / O'LOUGHLIN, K. (eds), *Experimenting with uncertainty. Essays in honour of Alan Davies*, Cambridge: Cambridge University Press: 12-21

WIDOWSON, H. (2003), *Defining Issues in English Language Teaching*. Oxford: Oxford University Press.

WILKINS, D. A. (1976), *Notional Syllabuses*, Oxford: Oxford University Press.

WINFORD, D. (2003), *An introduction to contact linguistics*, London: Blacwell.

YOFFE, L. (1997), "An overview of ACTFL proficiency interviews: a test of speaking ability", *Shiken: JALT Testing & Evaluation SIG Newsletter*, [en línea], (consulta: 21 de diciembre de 2007), 1 (2): 3 – 9, <www.jalt.org/test/yof_1.htm>.

YOUNG, R. / MILANOVIC, M. (1992), "Discourse Variation in Oral Proficiency Interviews", *Studies in Second Language Acquisition*, 14: 403-424.

YOUNG, R. (1995), "Conversation Styles in Language Proficiency Interviews", *Language Learning*, 45: 3-42.

YOUNG, R. / HALLECK, G. B. (1988), "«Let Them Eat Cake!» or How to Avoid Losing Your Head in Cross-Cultural Conversations", en YOUNG, R. / HE, A. W. (eds.), *Talking and Testing, Discourse Approaches to the Assessment of Oral Proficiency*, Amsterdam/Philadelphia: John Benjamin: 355-382.