

Formación del profesorado de ELE y evaluación. Construcción de una guía para autenticar pruebas de español lengua extranjera

PAULA REYES ÁLVAREZ BERNÁRDEZ
Universidad de Vigo

CARLES MONEREO FONT
Universidad Autònoma de Barcelona

Recibido: 12 octubre 2014 / Aceptado: 8 junio 2015
ISSN: 1697-7467

RESUMEN: Este artículo estudia el impacto de un curso de formación en evaluación auténtica de competencias dirigido a profesores de Español Lengua Extranjera (ELE). Se partió de un diseño de investigación colaborativa en el que los participantes se involucraron en el análisis de sus propias pruebas de evaluación y en la construcción de una rúbrica o guía para mejorarlas. Partiendo del concepto de evaluación auténtica de competencias y de las dimensiones que la caracterizan – realismo, relevancia, socialización, coherencia, competencia y orientación-, la investigación demuestra que la formación resultó eficaz para la mejora de la mayoría de esas dimensiones en la mayor parte de las pruebas de evaluación en ELE.

Palabras clave: evaluación auténtica, competencias, formación del profesorado, español como lengua extranjera.

ELE teacher training and assessment. Construction of a guide to authenticate tests of Spanish as a Foreign Language

ABSTRACT: This article studies the impact of a training course in authentic assessment of competences aimed to teachers of Spanish as a Foreign Language (ELE). We proposed a collaborative research design where the participants are engaged in analyzing their own assessment tests and in the construction of a rubric or guide for improvement. Based on the concept of authentic assessment of competencies and the dimensions that characterize it - realism, relevance, socialization, coherency, competence and orientation-, this research shows that the training was effective in improving most of these dimensions in most assessment tests in ELE.

Keywords: Authentic Assessment, competences, Teacher Training, Spanish as a foreign language.

1. INTRODUCCIÓN

La formación inicial de los evaluadores de ELE para poner en marcha un dispositivo de evaluación de calidad, entendida esta en toda su amplitud, es decir, viable, sostenible y competencial, se centra en la mayor parte de los casos en cuestiones relativas a la redacción de especificaciones, elaboración de tareas, análisis y validación de ítems y documentos, etc.

Este proceso, suele incluir distintas fases e involucrar a varios profesionales en roles diversos que dificulta, a menudo, la planificación de la logística necesaria para desarrollar una evaluación suficientemente sensible para valorar las competencias reales que los alumnos han desarrollado.

Por otra parte, es de esperar que una buena prueba de evaluación cumpla con las condiciones básicas de validez, viabilidad y fiabilidad, características que el *Marco Común Europeo de Referencia para las lenguas*¹ (Consejo de Europa, 2002) ha señalado como propias de una evaluación efectiva y coherente. Además, se espera que las pruebas tengan un impacto positivo sobre evaluados y evaluadores. No obstante, señalar estos requisitos como referencia exclusiva y obligada para todos los sectores involucrados en la evaluación de la lengua española como lengua extranjera no garantiza el éxito completo en cuanto a aumentar la efectividad de las propias pruebas se refiere (Parrondo, 2004). Dificultad a la que sumamos los pocos esfuerzos que se han dedicado a proyectar un sistema con el que poder prever, describir y planificar los procesos de evaluación y las operaciones sociales y cognitivas que comporta la elaboración y resolución exitosa de una prueba de evaluación.

Si tomamos como referencia el modelo dominante actualmente en Educación, el enfoque constructivista de la enseñanza y el aprendizaje, comprobamos que una competencia no puede definirse sino en función de situaciones y por consiguiente evaluar si un alumno es competente supone, necesariamente, enfrentarlo a una situación-problema, en alguna medida, nueva o desconocida para comprobar si es capaz de movilizar, regular e integrar sus conocimientos y recursos para resolverla (Díaz, 2011; Monereo, 2009a). Lo que debe caracterizar a ese tipo de situaciones-problema es su grado de autenticidad, es decir, su grado de similitud con las situaciones que habitualmente se producen en la vida cotidiana, en nuestro caso en las situaciones comunicativas más frecuentes. A pesar de no haber sido incluida entre los criterios que garantizan la idoneidad de las pruebas de evaluación de ELE, la autenticidad constituye uno de los puntos esenciales en la evaluación de las competencias y, a su vez, es una de las máximas preocupaciones de los expertos (Bordón, 2004; Kohonen, 2000; Prati, 2007).

No obstante, en una línea paralela han aparecido otras iniciativas muy próximas a las de la evaluación auténtica entre las que destacamos el *Portfolio Europeo de las Lenguas* (PEL) y más recientemente el *Portfolio Electrónico Europeo de las Lenguas* (ePEL), iniciativas con las que se intenta que al alumno lleve a cabo un registro de sus tareas más representativas y reflexione sobre ellas. Sin embargo, el principal problema de estas propuestas es que la simple toma de conciencia, sin otros referentes que promuevan el cambio, puede conducir a un ejercicio de trivialización y a aprendizajes superficiales) que no lleguen a producir cambios reales en las prácticas, ya que estos requieren de aprendizajes más sustanciales y estratégicos (Burner, 2014; Jurisevic, Enever y Pizorn, 2014).

Por todo ello, nuestro propósito en este artículo es facilitar el proceso que los evaluadores llevan a cabo cuando construyen sus pruebas pensando en evaluar competencias. Para ello se diseña y desarrolla una propuesta de formación, con base en los principios y dimensiones que definen una evaluación auténtica, que permita a los profesores participantes planificar, analizar y elaborar pruebas de ELE hábiles para valorar las competencias de sus estudiantes (Kohonen, 2000; Monereo, 2009a; Prati, 2007).

¹ En adelante MCER.

Siendo más explícitos, presentamos a continuación las razones que fundamentan la elección de la evaluación auténtica como un sistema de calidad con el que poder conseguir cambios en la evaluación de ELE (Darling-Hammond, Anness y Falk, 1995; Monereo, Castelló, Durán y Gómez, 2009):

En primer lugar hemos elegido la evaluación auténtica porque representa la posibilidad de evaluar las competencias y la naturaleza estratégica que las sustenta (Consejo de Europa, 2002; Monereo, 2009a; Zabala y Arnau, 2008). Como veremos más adelante, las dimensiones: realismo, relevancia y socialización, sirven para evaluar aquello que el alumno es capaz de hacer en el idioma aprendido (rentabilidad del aprendizaje), y no solo lo que el alumno conoce de manera superficial (rendimiento alcanzado). Definir el contexto de la evaluación con base en estos principios conduce a consolidar y reorientar las prácticas académicas hacia el logro de aprendizajes significativos y a la concreción de un currículo centrado en el alumno.

En segundo lugar, tal y como apuntan estudios recientes (Coll, Rochera, Mayordomo y Naranjo, 2007; Monereo, 2009b), los efectos retroactivos que tiene la evaluación hacen que cuanto más auténtica sea la propia evaluación, más auténtica sea la enseñanza y el aprendizaje. Así, por ejemplo, si la prueba de evaluación está basada en una concepción comunicativa de la lengua, será más probable que las prácticas docentes se orienten hacia un tipo de constructos similares. Dicho de otro modo, si se pretendiera fomentar la implantación de metodologías procedimentales basadas en un método comunicativo o en un enfoque por tareas, lo lógico y lo coherente sería que las pruebas correspondientes estuvieran diseñadas siguiendo esos planteamientos.

Por último, el hecho de que las actividades de evaluación presenten altos índices de autenticidad parece promover una mayor motivación de los alumnos hacia el aprendizaje, una mejora en la comprensión de lo aprendido y, por ende, la permanencia de lo interiorizado (Antibi, 2005; Monereo, 2009b).

Existe acuerdo entre los estudiosos sobre este tema en considerar que para que una prueba de evaluación sea auténtica debe involucrar las siguientes dimensiones (Kohonen, 2000; Monereo, 2009a; Prati, 2007):

1. Realismo: Una evaluación es realista cuando las tareas que la componen guardan un alto grado de similitud con los principales problemas o situaciones que pueden tener lugar en la vida real y cotidiana del aprendiente de lenguas. Para ello, la representación de diferentes situaciones en las que el alumno pueda desenvolverse y actuar de manera espontánea, tal y como lo haría fuera del aula, son excelentes oportunidades con las que poder calibrar sus conocimientos de la lengua extranjera y visionar cuáles son sus necesidades y sus limitaciones.
2. Relevancia: Los alumnos priorizan en su recuerdo aquello que se ha comprendido y que resulta útil en el momento en que se enseña. Por eso, una evaluación será relevante cuando sea útil y consiga que el alumno rentabilice de manera bastante inmediata, en su entorno más próximo, todo lo que ha podido aprender en el aula.
3. Socialización: Este último aspecto tiene que ver con la posición que mantendrá el actante dentro del discurso. Es en el espacio creado durante la interacción, donde el estudiante de lenguas extranjeras debe definir su propia identidad e identificarse como miembro del grupo. Para ello, es indispensable que el alumno tenga acceso a todos los recursos que forman parte de la realidad en la que se encuentra inmer-

so. Disponer de diferentes herramientas, sean estas la ayuda del profesor, de otros compañeros, el uso de diccionarios, de apuntes, etc. repercute de manera positiva no solo en la ejecución de la tarea de evaluación, sino también en la creación de su identidad y, por supuesto, en la mejora de su proceso de aprendizaje.

En último lugar, recogemos e incluimos los ítems que señala el MCER, en el capítulo 9, como propios de una evaluación coherente de lenguas extranjeras; estos son: *validez*, *viabilidad* y *fiabilidad*. Para ser más exactos y respetar la denominación otorgada por el documento base, incluimos los tres ítems bajo la dimensión *Coherencia*. De esta manera entendemos la coherencia de una prueba auténtica de evaluación como aquella que se corresponde y respeta las programaciones curriculares y el proceso de instrucción (*validez*), es decir, si lo que se evalúa se corresponde con lo que está previsto en el currículo y en las adaptaciones de aula; mide la naturaleza real de lo que pretende medir (*fiabilidad*), valora las competencias que realmente fueron enseñadas durante las clases; y es factible, en la medida que puede realizarse sin problemas, ni obstáculos que dificulten su desarrollo (*viabilidad*).

Además, en nuestro afán de dotar al español como lengua extranjera de una evaluación auténtica de competencias propia, incluimos junto a la dimensión de autenticidad y los ítems que la integran, dos dimensiones que concretan y cierran nuestra propuesta. Estas son: *competencia* (enfatisa el abordaje de situaciones-problema específicos que permiten comprobar la pertinencia del aprendizaje de los alumnos y, por ende, su aptitud en los estadios propios de competencia lingüística, comunicativa y orientada a la acción) y *orientación* (se enfoca a desarrollar procesos de autorregulación en el alumno que se extienden a niveles diferentes que afectan no solo al desarrollo académico, sino también personal y social).

2. UNA INVESTIGACIÓN SOBRE LA FORMACIÓN EN EVALUACIÓN AUTÉNTICA DE COMPETENCIAS DEL PROFESORADO DE ELE

2.1. Objetivo de la investigación

La investigación que presentamos pretende comprobar si un sistema de formación colaborativo, fundamentado en los principios de la evaluación auténtica, puede modificar las prácticas evaluadoras de español como lengua extranjera, sobre las que el asesoramiento incidirá directamente, así como las creencias que los profesores participantes poseen al respecto. Partiendo de estos propósitos, nuestro trabajo tratará de averiguar qué pruebas de evaluación de ELE son más resistentes al cambio y qué aspectos relativos a la evaluación auténtica de competencias suponen más dificultad para los profesores a la hora de integrarlas en sus propias pruebas. Tal y como han demostrado los estudios más recientes realizados en Psicología de la Educación (Coll, Rochera, Mayordomo y Naranjo, 2007; Monereo, Castelló, Durán y Gómez, 2009) para que un cambio en la evaluación se pueda materializar en la práctica y mantenerse en el tiempo será necesario que ese cambio no sea superficial sino que influya sobre el modo en que los profesores participantes conceptualizan la evaluación.

2.2. Participantes

Para realizar este estudio contamos con la colaboración de nueve profesores de español lengua extranjera que formaban parte del cuerpo docente de un centro de enseñanza de lenguas de Madrid. El centro recibe a estudiantes de todas las edades (siendo la edad mínima para matricularse los 5 años) y sus cursos abarcan todos los niveles de conocimiento de la lengua (A1, A2, B1, B2, C1 y C2). A pesar de ser un centro abierto a todo el público, su alumnado más asiduo pertenece al sector empresarial. En relación a la formación del profesorado, siete de los colaboradores poseían formación específica en ELE, mientras que uno de ellos contaba con formación relacionada con la enseñanza de lenguas extranjeras (no específica en ELE). Por último, cabe señalar que dos de los participantes declararon que habían recibido formación previa en evaluación de lenguas extranjeras.

2.3. Método

El trabajo que se reporta se fundamenta metodológicamente en la denominada Investigación Basada en el Diseño -IBD- (Design-Based Research Collective, 2003) que hace referencia a los estudios de campo en los que un investigador, o equipo de investigación, interviene en un contexto académico particular para alcanzar, mediante una propuesta formativa determinada, una meta pedagógica explícitamente definida. Por esta razón, en toda IBD el diseño de la formación es clave para mejorar el aprendizaje, crear conocimiento útil y avanzar en la construcción de teorías sobre la enseñanza, el aprendizaje y la evaluación. Seguidamente referimos las dos fases en las que se estructuró el procedimiento:

La primera fase, orientada a la construcción y validación de las dimensiones de evaluación auténtica de competencias de ELE, fue realizada a partir de una sesión de reflexión grupal (*focus group*) en la que se identificaron y establecieron, mediante acuerdo con los nueve profesores mencionados, las tres dimensiones de partida: competencia, autenticidad y orientación. La sesión fue íntegramente descrita en un informe elaborado por el coordinador del grupo. Del mismo modo, también fueron transcritos los argumentos y exposiciones que justificaban el establecimiento de cada una de las dimensiones. Con el fin de medir la fiabilidad y validez de la propuesta, cuatro investigadores independientes de ELE, que no habían participado en la sesión de *focus group*, analizaron cada una de los enunciados transcritos y los adscribieron a las dimensiones establecidas. El grado de fiabilidad de este análisis, según el índice de concordancia Kappa, fue del 75 %.

La segunda fase consistió en la organización de un seminario de *Evaluación Auténtica de Competencias en ELE*, con el compromiso de mantener esta actividad a lo largo de 30 horas en la modalidad de formación semipresencial. Durante las jornadas se revisaron algunos conceptos clave de la autenticidad y se elaboró una “Guía para autenticar pruebas de evaluación de ELE (GAPELE)”, con el objetivo de que fuese aplicada a sus pruebas de evaluación. Las prácticas e intervenciones realizadas nos permitieron documentar los cambios y resistencias que se producen en la evaluación de ELE: redefinición de las concepciones dominantes de los profesores sobre evaluación, exposición y conocimiento de los dilemas y temores, así como la posibilidad de experimentar procedimientos de evaluación auténtica en su propio proceso formativo, al actuar como alumnos.

3. RESULTADOS

A continuación se muestran los resultados hallados siguiendo el orden en que se desarrolló el estudio: construcción de GAPELE, cambios en los procedimientos (variaciones en las dimensiones que componen la evaluación auténtica) y cambios en las creencias sobre la evaluación por parte de los profesores participantes.

3.1. Construcción de GAPELE

La guía para la autenticación de pruebas de evaluación de ELE (GAPELE) surge como resultado de la formación, con el fin de que pudiese ser aplicada a las pruebas de evaluación de los profesores participantes y que, a su vez, estos pudiesen reelaborarlas siguiendo los ítems incluidos (la guía puede consultarse en el Anexo 1).

Para su construcción se tomó como modelo la “Guía para el análisis de pruebas de evaluación desde la perspectiva PISA” y se revisaron los procesos de autenticación seguidos en diversas pruebas de evaluación de otras áreas. La guía mencionada planteaba la posibilidad de emplear los principios psicoeducativos que subyacen en la elaboración del proyecto PISA² para suscitar cambios en las concepciones y prácticas educativas del profesorado de secundaria hacia una mayor autenticidad y complejidad cognitiva de las actividades, y una mayor promoción de la autonomía de aprendizaje de los alumnos (Monereo, Castelló, Durán y Gómez, 2009).

En su forma final, GAPELE, fruto de un proceso conjunto de análisis y reflexión, está compuesta por 15 ítems distribuidos entre las dimensiones: autenticidad, competencia y orientación:

- Nivel de autenticidad: Esta dimensión contaba con 9 ítems referidos al grado de realismo, relevancia, socialización y coherencia (validez, viabilidad y fiabilidad) de las actividades de evaluación que componen las pruebas.
- Nivel de competencia: Se distribuye en 3 ítems referidos al grado del aporte lingüístico (competencia lingüística), comunicativo (competencia comunicativa) y orientado a la acción (competencia orientada a la acción) que los contenidos incluidos en las pruebas poseen, así como el grado de complejidad cognitiva asociado a cada una de ellas (recuperación simple de la información, interpretación y reflexión o resolución de situaciones-problema).
- Nivel de Orientación: Este nivel, formado por 3 ítems, trata de mostrar el aporte que la prueba promueve en relación al desarrollo académico (ayuda al alumno a conocer qué ha aprendido y qué aspectos debe mejorar), personal (permite al evaluado ser consciente de su proceso de aprendizaje, es decir, cómo ha aprendido) y social (enseña a actuar y tomar decisiones ante diferentes problemas y situaciones comunicativos).

² El *Programme for Informational Student Assessment* –PISA- es un programa internacional impulsado por la OCDE que tiene como finalidad la evaluación del progreso de los sistemas educativos de los países que la integran a partir de la evaluación de las competencias básicas de los estudiantes que finalizan la etapa de enseñanza secundaria obligatoria, fundamentalmente en las áreas de lectura, ciencias y matemáticas

Con el fin de obtener datos comparables se creó un instrumento que, respetando todas las dimensiones e ítems que la guía incluía, permitiese valorar el nivel de evaluación auténtica de ELE en las pruebas iniciales y reconstruidas. El instrumento, denominado “Rúbrica GAPELE”, valoró sobre 15 puntos cada una de las dimensiones comprendidas y fue aplicado tanto en el análisis de las pruebas iniciales (PAU-1) como en el de las pruebas reconstruidas y autenticadas (PAU-2). En este punto, resulta necesario destacar que mientras la revisión de las PAU-1 fue realizada por el propio profesor-autor de la prueba; la revisión de las pruebas reconstruidas (PAU-2) fue realizada por dos profesores participantes que, de manera independiente, ejercieron de evaluadores de las pruebas del resto de participantes. No obstante, en aquellos casos en los que las valoraciones de los revisores no se aproximaron (diferencia de puntuación total superior a 5 puntos), se introdujo la revisión de un tercer evaluador.

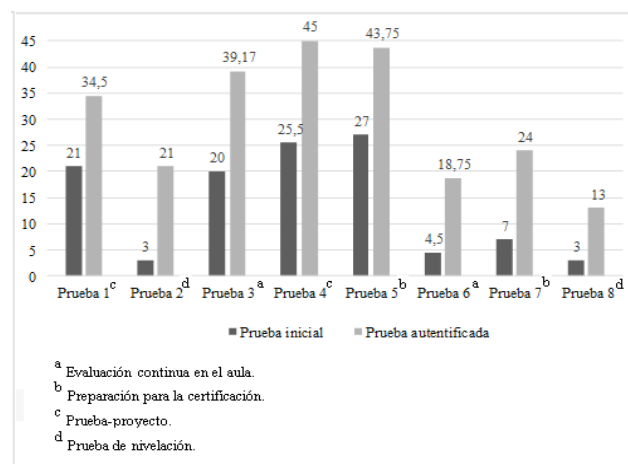
Por último, los resultados obtenidos permitieron establecer una clasificación de las pruebas según su nivel de proximidad a la evaluación auténtica de competencias de ELE, siguiendo los criterios que se exponen a continuación:

- Evaluación auténtica de competencias con índice alto: > 40 (89% - 100%)
- Evaluación auténtica de competencias con índice medio: 30-39’9 (67% - 88%)
- Evaluación auténtica de competencias con índice bajo: 23-29’9 (51% - 66%)
- Evaluación no auténtica: < 22’9 (50%)

3.2. Cambios en los procedimientos evaluadores

En la Figura 1 se recogen los resultados generales obtenidos tras el cómputo de las valoraciones otorgadas a cada dimensión (competencia, autenticidad y orientación), así como la tipología de las pruebas analizadas. De igual manera se muestra la evolución en las pruebas de los docentes, entre la prueba inicial y la prueba autenticada, y su aproximación al nivel de evaluación auténtica de competencias de ELE.

Figura 1. Aproximación al nivel de evaluación auténtica de competencias de ELE en las pruebas inicial y autenticada.



En general, puede establecerse que tras la reconstrucción realizada por los participantes, las pruebas se aproximan a una evaluación auténtica. Sin embargo, no siempre consiguen afianzarse como pruebas de evaluación auténtica de competencias de ELE, es decir, no logran superar el nivel medio (51%) que implica el alcance mínimo de 23 puntos en su estimación final.

Como vemos, solo en su valoración inicial dos de las pruebas analizadas obtienen una estimación positiva en su nivel de autenticidad (pruebas 4 y 5). Las pruebas 1 y 3, a pesar de estar cerca, no lo logran, mientras que las pruebas restantes (pruebas 2, 6, 7, y 8) son claramente pruebas no auténticas de evaluación.

Tras la reconstrucción, las pruebas 4 y 5 son los casos que consiguen afianzarse con el índice de autenticidad más alto ya que ambas superan los 40 puntos en su valoración final.

La prueba 4 se fundamenta en un proyecto de grupo, distribuido en varias actividades que los alumnos evaluados deben realizar en el aula a lo largo de una semana (la evaluación no se realiza en un mismo día). Además, entre las características que rodean a esta prueba se encuentra la heterogeneidad en relación al nivel de conocimiento de la lengua (cada grupo está formado por cuatro alumnos cuyo nivel de conocimiento de ELE es diferente al de sus compañeros), lo que favorece el aprendizaje entre los propios miembros. Por último, la introducción de algunas modificaciones permitió incrementar todos los niveles que componen la evaluación auténtica de ELE: competencia, autenticidad y orientación por lo que alcanzó la puntuación máxima en su valoración final: 45 puntos.

Por su parte, en el caso de la prueba 5, aun tratándose de una prueba de preparación para la certificación de ELE en la que, por ende, se respetan los cánones que estas pruebas imponen, vemos como la prueba consigue alcanzar un índice del 97 % en cuanto al nivel de evaluación auténtica se refiere. Un análisis más profundo de la prueba permitió atestiguar que, aunque en todas las dimensiones se aprecia un incremento de nivel, es notorio el incremento experimentado en la dimensión “orientación” en la que consigue, tras el proceso de reconstrucción, el total de 15 puntos estimados para esa dimensión.

Dentro del nivel “Evaluación auténtica de competencias con índice medio”, encontramos los casos de las pruebas 1 (21 puntos) y 3 (20 puntos) las cuales, tras el proceso de reconstrucción, incrementan respectivamente en 34'5 y 39'17 su puntuación.

La prueba 7 es el único caso que se clasifica como “evaluación auténtica de competencias con índice bajo”. Se trata de una prueba de preparación para la certificación cuya valoración inicial era de 7 puntos sobre 45 y que finalmente obtiene un resultado total de 24 puntos. A pesar de haber notado un incremento en los índices de las dimensiones “competencia” y “autenticidad”, su análisis nos permitió percibir que sigue necesitando refuerzo en el nivel de “orientación”, donde obtiene una puntuación mínima de 5 puntos sobre los 15 totales que se establecen para esta dimensión.

Las pruebas 2, 6 y 8, tras el proceso de autenticación y aun mostrando una evolución positiva en sus dimensiones, no consiguieron afianzarse como pruebas auténticas de ELE. Resulta pertinente destacar que las pruebas 2 y 8 eran pruebas de nivelación de lenguas cuyo formato de origen se limitaba al examen tipo test y cuya valoración inicial fue inferior a 5 puntos (11%) en ambos casos. En cuanto a la prueba 6, aun tratándose de una prueba de evaluación continua en el aula, su análisis posterior denotó que, a pesar de la introducción de algunos cambios substanciales, la prueba seguía presentando una fuerte carga del componente lingüístico por encima del comunicativo y orientado a la acción, al mismo tiempo que se echaban en falta claves para que pudiesen orientar el aprendizaje.

Con la finalidad de mostrar más claramente el índice de mejoría que alcanzan las pruebas en cada una de las dimensiones analizadas, detallamos a continuación los resultados generales obtenidos en relación a los niveles de autenticidad, competencia y orientación. Se toma como criterio para la cuantificación de cada ítem, el baremo ofrecido en la “Rúbrica GAPELE”.

3.2.1. Nivel de autenticidad

Con relación al nivel de autenticidad estamos en disposición de avalar que el aspecto más resistente al cambio continúa siendo la *socialización* del alumno durante la prueba. A pesar de que ahora se introducen recursos propios y comunes a las situaciones que se plantean y que permiten conectar al alumno con el entorno real y próximo, siguen echándose en falta la introducción de actividades que promuevan la interacción, la cooperación y el contacto con los otros. La *relevancia*, es decir, la utilidad de los aprendizajes que las pruebas evalúan, junto con la *coherencia*, alcanzan los niveles más elevados tras la reconstrucción. Por último, se aprecia una evolución muy favorable en la dimensión *realismo*.

3.2.2. Nivel de competencia³

En su inicio, las pruebas mostraban un fuerte desequilibrio en relación con las demandas cognitivas que las actividades de evaluación requerían: mientras que había una fuerte carga de actividades destinadas a la simple recuperación de la información (8 pruebas) y a su interpretación (6 pruebas), una minoría de pruebas incluía actividades de resolución (4 pruebas) y reflexión (1 prueba). Tratando de buscar un equilibrio en cuanto a la tipología de demandas cognitivas se refiere, comprobamos que en el cómputo final las pruebas lograron incrementar sus niveles de exigencia en relación a las actividades que demandaban resolución de situaciones-problema (8 pruebas) y reflexión (6 pruebas) y, asimismo, disminuyó el nivel de recuperación de información (6 pruebas). No obstante, se mantuvo el número de pruebas que introducen actividades para la interpretación (6 pruebas).

La modificación de la complejidad cognitiva en las actividades de evaluación, permitió registrar cambios en la evaluación de las competencias lingüística, comunicativa y orientada a la acción. Tras la modificación de las pruebas se observa un cambio interesante: el incremento de la evaluación de las competencias de los alumnos en sus vertientes *comunicativa* y *orientada a la acción* por encima de la competencia lingüística.

3.2.3. Nivel de orientación

A pesar de la evolución favorable en este nivel, la *orientación* sigue siendo el aspecto que más complicaciones presenta a la hora de incluirlo en las pruebas. Las pruebas continúan evaluando el resultado final y se echan en falta actividades que permitan evaluar y enseñar a regular el proceso de toma de decisiones que ha seguido el alumno para afrontar la situación-problema (*desarrollo personal*). Del mismo modo, el *desarrollo social*, ligado inexcusablemente a la falta de procesos de socialización descritos anteriormente, es otro

³ En este caso el cómputo se realizó en base al total de pruebas analizadas, es decir, sobre ocho pruebas de evaluación de ELE.

de los aspectos que debe seguir mejorándose. No obstante, se percibe una evolución muy favorable con relación al *desarrollo académico* que las pruebas introducen.

3.3. Cambios en las concepciones

Al margen de los resultados obtenidos en las pruebas tras su reconstrucción, la investigación demostró que la elaboración de las actividades inspiradas en las dimensiones de la evaluación de competencias de ELE y consecuentemente en GAPELE, producía ciertos efectos colaterales muy interesantes como la revisión de determinadas concepciones y prácticas evaluadoras de los profesores participantes. En las citas siguientes se ofrece una muestra de los comentarios más significativos que consideramos de interés para valorar la funcionalidad de GAPELE y su implementación:

Antes de iniciar las sesiones veía la evaluación como parte del proceso educativo, tal y como debe ser. No obstante, no me gustaba considerarla como un fin en sí misma, sino como parte del proceso pero no veía con claridad cómo hacerlo. Ahora, después de estas sesiones y gracias a la guía, veo que el concepto de evaluación que yo tenía obedece a un esquema constatable que todos los formadores deberían tener en mente a la hora de evaluar. Creo que este modelo que ahora está racionalizado y justificado en mi concepción de evaluación debería ser el modelo a seguir para que el proceso educativo sea constructivo y continuo.

En este caso observamos un cambio de relativo a la posibilidad de tratar la evaluación como un elemento integrado en el proceso de enseñanza-aprendizaje y no como un aspecto aislado e independiente.

Enseñar al alumno a autoevaluarse así como hacerle participe y protagonista de su aprendizaje es algo que había descuidado como profesora de ELE. GAPELE enseña a evaluar. La evaluación auténtica enseña a aprender.

Aquí, el cambio consiste en considerar la evaluación como herramienta que permite al alumno auto-regular el propio aprendizaje, es decir, una evaluación centrada en el alumno y no solo en los contenidos.

Partía de la creencia de que una evaluación es un tipo de prueba encaminada a medir los conocimientos y habilidades de un candidato y, aunque sigo pensando lo mismo, he descubierto que las pruebas que realizamos siempre se pueden mejorar, especialmente en tres aspectos: motivación, significación y realismo.

Este comentario refleja una nueva percepción sobre la evaluación, no solo como instrumento estático que permite acreditar un nivel de conocimientos, sino como motor del aprendizaje.

GAPELE me ha enseñado a tener en cuenta en las pruebas aspectos “personalizadores” para cada alumno. Además, ahora pongo a disposición de mis alumnos

recursos que antes no ponía y que son básicos para que puedan desarrollar sus competencias de manera real.

El participante subraya en su comentario la necesidad de personalizar la evaluación, una idea contraria a una evaluación estandarizada y homogénea; por otra parte señala la necesidad de poner a disposición de los alumnos recursos de la vida cotidiana, similares a los que se encontrará cuando deba ejercer como hablante de español.

Cuando he incorporado los aspectos de la evaluación auténtica a la prueba he percibido que esta se vuelve menos estricta y tradicional, y pasa a ser más flexible y ajustada a los objetivos marcados.

Esta última aseveración enfatiza la idea de una evaluación adaptativa, adecuada a cada situación en la que se inscribe, capaz de cumplir con los objetivos de aula y dar respuesta a las necesidades de los estudiantes.

Consideramos que estos enunciados son indicios que manifiestan cambios significativos en las creencias de los participantes, propiciados por la formación y que avalan una futura permanencia de estos cambios

4. CONCLUSIONES

Como se ha podido comprobar, tras el proceso de reconstrucción con GAPELE, todas las pruebas evolucionan favorablemente en su aproximación al nivel de evaluación auténtica de competencias de ELE. Del total de las ocho pruebas analizadas, cinco consiguen afianzarse como pruebas de evaluación auténtica de ELE, llegando dos de ellas a alcanzar las puntuaciones más elevadas en su proceso de reconstrucción (pruebas 4 y 5). El incremento ha sido notable tanto en la dimensión de competencia como en la autenticidad. No ha sido así en la dimensión de orientación, hecho que puede deberse a que la formación incidió exclusivamente en la modificación de las pruebas que empleaban los profesores y las modificaciones que esos cambios suscitaban en sus concepciones y procedimientos. Es posible que una ayuda centrada en la enseñanza explícita de estrategias de auto-regulación (casos de pensamientos, incidentes críticos, etc.) hubiese resultado beneficiosa.

En relación con este último punto, observamos que no todas las pruebas presentan la misma disposición para integrar el nivel de orientación hacia el desarrollo de la competencia deseado. Nos referimos, concretamente, a las pruebas de nivelación de lenguas, pruebas que permiten clasificar a los alumnos en el estadio que les corresponde y en las que el examen tipo test continua siendo el instrumento por excelencia. Por esta razón nos preguntamos si estas pruebas son instrumentos válidos para evaluar competencias o deberían revisarse, tratando de optimizar las dimensiones de autenticidad defendidas en este artículo.

Por último destacamos el interés que puede tener el modelo de investigación colaborativa desarrollado como alternativa al modelo de formación tradicional. Involucrar a los profesores en su formación y en la investigación sobre la misma, a partir de un diálogo sobre sus propias prácticas, en nuestro caso sobre sus pruebas de evaluación, les convierte en protagonistas de los cambios que se produzcan y, por consiguiente, aumenta la posibilidad de que esos cambios se sostengan en el tiempo.

5. REFERENCIAS BIBLIOGRÁFICAS

- Antibi, A. (2005). *La constante macabra o cómo se ha desmotivado a muchos estudiantes*. Madrid: El rompecabezas.
- Bordón, T. (2004). “La evaluación de la expresión oral y la comprensión auditiva”, en J. Sánchez y I. Santos (eds.), *Vademécum para la formación de profesores. Enseñar español como segunda lengua (L2)/Lengua extranjeras (LE)*. Madrid: SGEL, 983-1005.
- Burner, T. (2014). “The potential formative benefits of portfolio assessment in second and foreign language writing contexts: A review of the literature”, en *Studies in Educational Evaluation*, 43: 139-149.
- Coll, C., Rochera, M.J., Mayordomo, R.M. y Naranjo, M. (2007). “Evaluación continua y apoyo al aprendizaje. Análisis de una experiencia de innovación en educación superior”, en *Revista electrónica de investigación psicoeducativa*, 13, 5: 783-804.
- Consejo de Europa (2002). *Marco Común Europeo de Referencia para las lenguas: aprendizaje, enseñanza, evaluación*. Madrid: Secretaria General Técnica del MECD y Grupo Anaya.
- Darling-Hammond, L., Ancess, J. y Falk, B. (1995). *Authentic assessment in action*. New York: Teachers College Press.
- Design-Based Research Collective (2003). “Design-based research: An emerging paradigm for educational inquiry”, en *Educational Researcher*, 32 (1): 5-8.
- Díaz, A. (2011). “Competencias en educación. Corrientes de pensamiento e implicaciones para el currículo y el trabajo en el aula”, en *Revista Iberoamericana de Educación Superior (RIES)*, 2, 5: 1-13.
- Jurisevic, M., Enever, J. y Pizorn, K. (2014). “Triple Tool Effect: Professional Portfolios in Teaching Foreign Languages”, en *Porta Linguarum*, 21: 7-24.
- Kohonen, V. (2000). “La evaluación auténtica en la educación afectiva de lenguas extranjeras”, en J. Arnold (ed.), *La dimensión afectiva en el aprendizaje de idiomas*. Cambridge: Cambridge University Press, 295-309.
- Monereo, C. (ed.) (2009a). *La evaluación auténtica en Enseñanza Secundaria y Universitaria: investigación e innovación*. Barcelona: Edebé.
- Monereo, C. (coord.) (2009b). *Pisa como excusa*. Barcelona: Graó.
- Monereo, C., Castelló, M., Durán, D. y Gómez, I. (2009): “Las bases psicoeducativas del proyecto PISA como guía para el cambio en las concepciones y prácticas del profesorado de secundaria”, en *Infancia y Aprendizaje*, 32, 1: 13-19.
- Parrondo, J.R. (2004). “Modelos, tipos y escalas de evaluación”, en J. Sánchez y I. Santos (eds.), *Vademécum para la formación de profesores. Enseñar español como segunda lengua (L2)/ Lengua extranjeras (LE)*. Madrid: SGEL, 967-983.
- Prati, S. (2007). *Elaboración de exámenes. La evaluación en español Lengua Extranjera*, Argentina: Libros de la Araucaria.
- Zabala, A. y Arnau, L. (2008). *11 ideas clave: Cómo aprender y enseñar competencias*. Barcelona: Graó.

ANEXO 1

GUÍA PARA AUTENTIFICAR PRUEBAS DE EVALUACIÓN DE ELE (GAPELE)

AUTENTICIDAD*Realismo:*

1. Las actividades o tareas que se plantean, ¿poseen alguna de estas características?

Pregunta 1	Pregunta 2	Pregunta 3	Pregunta 4	Pregunta 5
Real				
Social				
Próxima				
Motivadora				

2. En cuanto al contenido, ¿los textos que se proponen son, en su mayoría...?

- Textos continuos (narrativos, expositivos, descriptivos, argumentativos).
 Textos discontinuos (imágenes, tablas, esquemas, mapas, cuestionarios, anuncios).
 Textos continuos y discontinuos.

Relevancia:

3. ¿Existe un hilo conductor (temático, argumental...) a lo largo de toda la prueba?

- SÍ NO

4. ¿Las tareas se relacionan con problemas o situaciones que el alumno podría encontrarse en su vida cotidiana?

- SÍ NO

Socialización:

5. Para realizar la tarea, ¿se permite el uso de los recursos y las ayudas que estarían presentes en esa situación?

- SÍ NO

En caso *afirmativo*, indica cuándo se comunica al alumno que dispone de esas ayudas y recursos:

- Al inicio de la actividad.
 Durante la ejecución de la tarea.
 Casi al finalizar la prueba.

6. ¿Las tareas ayudan a que el alumno se muestre de manera natural y espontánea (sin sobreactuaciones, simulaciones, imitaciones, etc.) como aprendiente de español?

- SÍ NO

Coherencia:

7. ¿Las tareas que se proponen evalúan las estrategias y competencias que se han enseñado en el aula?
 SÍ NO

8. ¿En todas las preguntas se explican con claridad y precisión los criterios de corrección?
 SÍ NO

9. ¿El tiempo para realizar la prueba es suficiente?
 SÍ NO

COMPETENCIA

10. ¿Qué competencias se desarrollan en cada pregunta?

PREGUNTA N°	COMPETENCIA		
	Lingüística	Comunicativa	Orientada a la acción

11. En la prueba, ¿se parte de una situación problemática, en forma de texto coherente (oral o escrito), a partir del cual se realizan diferentes actividades?
 SÍ NO

12 Según la destreza evaluada, ¿qué tipo de proceso es necesario llevar a cabo para realizar la tarea?

	Compresión escrita	Expresión escrita	Compresión oral	Expresión oral
Recuperación de la información (reproducir, copiar, decir de memoria)				
Interpretación de la información (explicar, comparar, diferenciar, resumir, esquematizar)				
Resolución a partir de la información (planificar una solución y ejecutarla)				
Reflexión a partir de la información (autoevaluar, corregir, ajustar la actuación)				

ORIENTACIÓN

13. ¿La tarea ayuda al alumno a conocer qué ha aprendido y qué aspectos debe mejorar?

SÍ NO

En caso *afirmativo*, indica en qué pregunta(s): _____

14. ¿Las actividades permiten al alumno ser consciente de su proceso de aprendizaje, es decir, cómo ha aprendido?

SÍ NO

En caso *afirmativo*, indica en qué pregunta(s): _____

15. Esta tarea, ¿enseña a actuar y tomar decisiones ante diferentes situaciones y problemas comunicativos?

SÍ NO

En caso *afirmativo*, indica en qué pregunta(s): _____