

Tesis Doctoral

**LA DETECCIÓN DE SIMULACIÓN DE DÉFICITS
COGNITIVOS EN TRAUMATISMOS
CRANEOENCEFÁLICOS LEVES**



**Departamento de Personalidad, Evaluación y Tratamiento
Psicológico
Universidad de Granada**

**Doctoranda: Raquel Vilar López
Directores: Dr. Miguel Pérez García
Dr. Manuel Gómez Río**

Agradecimientos:

A Miguel Pérez, por sus incontables horas de dedicación.

A María Sánchez e Isabel Peralta.

A Manuel Gómez Río, Jose Manuel Llamas Elvira, y al resto de los profesionales del
Servicio de Medicina Nuclear del Hospital Virgen de las Nieves.

A mi familia, a Sergio Bastida y a Beatriz Sánchez, porque sin su constante apoyo y
confianza esta tesis tampoco hubiese sido posible.

ÍNDICE

Presentación	1
Capítulo 1: El traumatismo craneoencefálico leve y el síndrome post-conmocional	7
1. Introducción	9
2. El Traumatismo Craneoencefálico Leve y el Síndrome Post-Conmocional	9
3. Dificultades cognitivas en los TCE-L	14
3.1. Problemas de atención	15
3.2. Problemas de memoria	16
3.3. Problemas en la función ejecutiva	17
4. Evaluación de los TCE-L y del SPC mediante neuroimagen funcional	19
5. Factores que inciden en los pacientes con TCE-L para el desarrollo del SPC	21
Capítulo 2: Psicología Forense y Neuropsicología Forense	27
1. Introducción	29
2. Definición y ámbitos de aplicación de la Psicología Forense y la Neuropsicología Forense	29
3. Problemas de la Neuropsicología Forense	32
4. Situación actual de la Neuropsicología Forense	35
5. Cualificación del Neuropsicólogo Forense	38
6. Problemas éticos de la Neuropsicología Forense	40
7. La Neuropsicología Forense en España	42
Capítulo 3: La simulación del daño neuropsicológico	45
1. Introducción	46
2. Conceptualización de la simulación y terminología	46
2.1. ¿De qué debemos diferenciarla?	48
2.2. Más términos por aclarar...	51
3. Los modelos de simulación	52
3.1. El modelo patogénico	52
3.2. El modelo criminológico	52
3.3. El modelo adaptativo	53
4. Criterios diagnósticos de la simulación	54
5. La investigación de la simulación	58
6. Problemas metodológicos	63
7. Problemas éticos de la detección de simulación	68
Capítulo 4: Métodos para la detección de la simulación	73
1. Introducción	75
2. Métodos de detección de la simulación	75
2.1. El estudio de las inconsistencias	75
2.2. Métodos derivados de las pruebas específicamente diseñadas para la evaluación de la simulación	77
2.3. Métodos derivados de la ejecución en pruebas neuropsicológicas estándar	78
2.4. Análisis de los síntomas referidos por el paciente	80
2.5. Otros	80

3. La evaluación de los síntomas referidos por el paciente	81
4. El estudio psicofisiológico del engaño	85
4.1. El polígrafo	85
4.2. La detección psicofisiológica de la simulación del estrés post-traumático	90
4.3. La detección de simulación mediante el empleo de potenciales evocados y neuroimagen	91
5. Variables que influyen en la detección de simulación	94
5.1. El entrenamiento (“ <i>coaching</i> ”)	94
5.2. Otras variables	102
Capítulo 5: Instrumentos empleados en la detección de simulación	107
1. Introducción	109
2. Los test específicos	109
2.1. Tests de efecto suelo	109
2.2. Medidas de habilidades sobreaprendidas	112
2.3. Tests de elección forzosa	118
2.3.1. Estímulos Alternativos	126
2.3.2. Administración por ordenador	134
2.3.3. Otras variaciones	139
2.4. ¿Cuáles son los mejores tests específicos para detectar la simulación?	148
2.5. El empleo conjunto de diversos test específicos de simulación	150
3. Los tests neuropsicológicos no específicos de simulación	151
3.1. Baterías neuropsicológicas	162
3.2. El empleo de diversas medidas no específicas en la detección de la simulación	170
4. Conclusiones sobre las pruebas específicas y no específicas de simulación	171
Capítulo 6: Objetivos Generales	177
Capítulo 7; Estudio 1: Adaptación y estudio preliminar de las pruebas específicas de simulación	185
1. Introducción	187
2. Objetivos e hipótesis	188
3. Metodología	189
3.1. Participantes	189
3.2. Instrumentos	189
3.3. Las instrucciones de los instrumentos y su doble traducción cruzada	192
3.4. Procedimiento	193
3.5. Variables	194
3.6. Análisis estadísticos	195
4. Resultados	196
4.1. El Efecto de las instrucciones (Hipótesis 1)	196
4.2. Propiedades psicométricas de las pruebas (Hipót. 2 y 3)	196
4.3. Porcentaje de análogos clasificados correctamente (Hipótesis 4)	198
4.4. Comparación de los resultados de nuestra muestra con los de las muestras originales (Hipótesis 5)	199
5. Discusión	201

Capítulo 8; Estudio 2: Utilidad de las pruebas específicas de Simulación	211
1. Introducción	213
2. Objetivos e hipótesis	214
3. Metodología	216
3.1. Participantes	216
3.2. Instrumentos	217
3.3. Procedimiento	218
3.4. Variables	219
3.5. Análisis estadísticos	219
4. Resultados	220
4.1. Influencia de las variables demográficas y del tiempo transcurrido desde el traumatismo sobre los tests de simulación (Hipótesis 1)	220
4.2. Análisis de los “grupos de prevalencia diferencial” (Hipótesis 2)	221
4.3. Clasificación de los “grupos de prevalencia diferencial” (Hipótesis 3)	226
4.4. Análisis de “grupos conocidos” (Hipótesis 4)	230
4.5. Clasificación de los “grupos conocidos” (Hipótesis 5)	239
4.6. Sensibilidad, especificidad y puntos de corte (Hipót. 6)	243
4.7. Estabilidad de los hallazgos con las curvas ROC	247
4.8. Comparación con los resultados publicados en la literatura (Hipótesis 7)	248
5. Discusión	250
Capítulo 9; Estudio 3: Estudio de otros procedimientos en la detección de simulación	269
1. Introducción	271
2. Objetivos e hipótesis	272
3. Metodología	274
3.1. Participantes	274
3.2. Instrumentación	274
3.3. Procedimiento	277
3.4. Variables	279
3.5. Análisis estadísticos	283
4. Resultados	284
4.1. Análisis de los “grupos de prevalencia diferencial” en las pruebas neuropsicológicas no específicas de simulación (Hipótesis 1)	274
4.2. Análisis de “grupos conocidos” en las pruebas no específicas de simulación (Hipótesis 2)	290
4.3. Sensibilidad, especificidad y puntos de corte en las pruebas no específicas de simulación (Hipótesis 3)	298
4.4. Estabilidad de los hallazgos con las curvas ROC	301
4.5. La SPECT-PCR en la detección de simulación (Hipótesis 4)	302
4.6. Las medidas psicofisiológicas asociadas a la ejecución de una prueba específica de simulación (Hipótesis 5)	304

5. Discusión	312
Capítulo 10: Discusión General, Conclusiones y Perspectivas Futuras	327
1. Discusión General	329
2. Conclusiones	334
3. Perspectivas Futuras	335
Referencias	339
Apéndice 1: La doble traducción cruzada	361
Apéndice 2: Las instrucciones proporcionadas a los análogos	373
Apéndice 3: El informe posterior a la evaluación	379
Anexo 1: Puntos de corte, sensibilidades y especificidades de las pruebas específicas de simulación para diferentes probabilidades de base	383
Anexo 2: Puntos de corte, sensibilidades y especificidades de las pruebas neuropsicológicas no específicas de simulación para diferentes probabilidades de base	415

PRESENTACIÓN

Dentro de la Neuropsicología Clínica, el área que más rápidamente se ha extendido en los últimos años es el de la Neuropsicología Forense. Esta expansión puede apreciarse por el incremento de los artículos publicados sobre este tema (Bigler, 2006; Sweet, King, Manila, Bergman y Simons, 2002), así como por la proliferación de presentaciones en congresos y de libros de texto concernientes a esta materia. Dentro de la disciplina, el tópico de la simulación se ha convertido en el más estudiado (Sweet y cols., 2002). Este auge se debe a que cada vez más frecuentemente se demanda de los neuropsicólogos que presten testimonio experto en los casos de litigio por daño personal, particularmente en casos de daño cerebral leve o moderado.

En la actualidad, una gran proporción de las personas que se someten a evaluaciones neuropsicológicas en los EEUU están implicadas en procesos judiciales de los que pueden obtener grandes beneficios si se demuestra un daño cognitivo, sea este real o no. En muchos casos, los datos de los test neuropsicológicos son la única fuente de evidencia objetiva de déficits, especialmente en los casos de daño cerebral leve, donde la neuroimagen es generalmente negativa y los signos neurológicos están generalmente ausentes.

Por esto, el tema de la simulación de déficits cognitivos y sus posibles métodos de detección ha sido investigado y estudiado muy extensamente desde la década de los 80. Este minucioso estudio se debe, sin duda, a que los costes tanto personales como sociales derivados del resultado de estas evaluaciones (falsos positivos y falsos negativos) es altísimo. Para ilustrar estos costes basta con citar el ejemplo del estado de Louisiana, en el cual la suma a la que asciende el problema de la simulación teniendo en cuenta sentencias y veredictos, dinero empleado en tratamientos médicos innecesarios y costes de la defensa en contra de los simuladores asciende a un total de 107,2 millones de dólares al año (Gouvier, Less-Haley y Hammer, 2003). Si tomamos esto como una

media de lo que ocurre en todos los estados de EEUU, el coste derivado de la simulación asciende en este país a 5.36 billones de dólares al año, y esto sin tener en cuenta los días de trabajo perdidos ni las suplencias causadas por los simuladores. Con el dinero ahorrado de la detección de un simulador se podrían realizar 5000 evaluaciones neuropsicológicas, por lo que el resultado del coste/beneficio al emplear neuropsicólogos en el área forense parece sobradamente justificado. Si este es el coste de los falsos negativos, el de los falsos positivos quizá se aleje del económico, siendo más fuerte desde nuestro punto de vista el coste social o personal (pensemos en un sujeto al que se deja sin tratamiento por ser considerado un simulador cuando en realidad no lo es, o en las connotaciones sociales negativas que esta etiqueta lleva asociadas).

Por todas estas razones este trabajo pretende presentar el estado actual de la investigación en simulación, centrándonos en la población con mayor probabilidad de simulación y mayor dificultad en la detección como es el traumatismo craneo-encefálico leve (TCE-L). Para ello, la tesis consta de cinco capítulos teóricos dedicados a revisar: (i) el traumatismo craneoencefálico leve y la cronificación de sus síntomas en el síndrome post-conmocional (Capítulo 1), (ii) el estado actual de la Neuropsicología Forense, exponiendo sus principales problemas y cuestiones éticas, así como la cualificación necesaria para ejercer como neuropsicólogo forense (Capítulo 2), (iii) la definición y terminología relativa a la simulación, así como sus modelos conceptuales, criterios diagnósticos y la investigación de esta área, planteando los principales problemas metodológicos y éticos relacionados con la detección de la simulación (Capítulo 3), (iv) los métodos empleados para la detección de simulación, repasando aquí los métodos relativos al análisis de los síntomas referidos por el paciente y a la psicofisiología, para finalizar exponiendo algunas de las variables que influyen en la

detección de la simulación (Capítulo 4), y (v) los principales instrumentos empleados en la detección de síntomas cognitivos, centrándonos en las pruebas específicas de simulación y en las pruebas neuropsicológicas no específicas de simulación (Capítulo 5).

A continuación, presentamos tres estudios empíricos dirigidos a investigar: (i) la adecuación preliminar de las pruebas para su utilización en nuestra población (Capítulo 7), (ii) la utilidad de las pruebas específicas de simulación (Capítulo 8), (iii) la utilidad de otros procedimientos (pruebas neuropsicológicas no específicas, medidas psicofisiológicas y SPECT cerebral) en la detección de simulación (Capítulo 9).

Por último, discutiremos nuestros hallazgos, presentaremos las principales conclusiones que se extraen de nuestro trabajo y expondremos algunas perspectivas futuras de investigación derivadas de lo obtenido en esta tesis (Capítulo 10).

CAPÍTULO 1

EL TRAUMATISMO CRANEOENCEFÁLICO LEVE

Y EL SÍNDROME POST-CONMOCIONAL

1. Introducción:

El Traumatismo Craneoencefálico Leve (TCE-L) constituye el trastorno más frecuentemente evaluado entre los neuropsicólogos que se dedican al área de daños y perjuicios personales en el ámbito forense (Larrabee, 2005^a). Además, el TCE-L es la situación clínica con mayor probabilidad de simulación (Holmquist y Wanlass, 2002; Pelegrín-Valero, Fernández-Guinea, Tirapu-Ustarroz y Muñoz Céspedes, 2001), lo que ha llevado a que el estudio de la simulación se haya focalizado fuertemente en los casos de TCE-L (Bianchini, Greve y Love, 2003). Por ello, el tipo de pacientes elegidos para esta investigación han sido aquellos con quejas persistentes vinculadas a un TCE-L, que cumplieran los criterios del Síndrome Post-Conmocional (SPC). Como nos parece fundamental conocer el tipo de muestra con la que vamos a trabajar, comenzaremos este estudio con un breve repaso de los aspectos fundamentales de los TCE-L.

No es la finalidad de este capítulo realizar una revisión exhaustiva, sino centrarnos en algunos aspectos importantes por su vinculación con el tópico de la simulación. En concreto, realizaremos una breve aproximación al concepto de TCE-L y al SPC, haciendo especial hincapié en las dificultades cognitivas asociadas al mismo, a los principales hallazgos en los estudios de neuroimagen y a los factores que influyen sobre la incidencia, causa y persistencia de los déficits en este trastorno.

2. El Traumatismo Craneoencefálico Leve y el Síndrome Post-Conmocional

El primer problema al que nos enfrentamos en el estudio del TCE-L es que diversos términos se emplean como sinónimos, por lo que en los diferentes trabajos nos podemos encontrar con el uso indistinto de: Closed Head Injury, Concussion, Mild Head Injury, Mild Closed Head Injury, Minor Closed Head Injury, Head Injury o Postconcussive síndrome (Blostein, y Jones, 2003). Además de esta variabilidad

terminológica, tampoco existe consenso a la hora de definir qué constituye un TCE-L. Probablemente la definición que goza de una mayor aceptación es la propuesta por el American Congress of Rehabilitation Medicine (Kay y cols., 1993), según la cual un paciente que ha padecido un Traumatismo Craneoencefálico Leve es una persona que ha sufrido traumáticamente una disrupción de las funciones fisiológicas de su cerebro que se manifiestan *al menos* por uno de los siguientes:

1. Cualquier periodo de pérdida de conciencia.
2. Cualquier pérdida de memoria de los acontecimientos inmediatamente anteriores o posteriores al accidente.
3. Cualquier alteración en el estado mental en el momento del accidente (confusión, desorientación, aturdimiento...)
4. Déficits Neurológicos Focales que pueden o no ser transitorios pero en los que la severidad de la lesión no excedan:
 - .. Pérdida de conciencia < 30 minutos.
 - .. Nivel de Glasgow de 13 o mayor (13-15) sin deterioro posterior.
 - .. Amnesia postraumática menor de 24 h.

En ausencia de fracturas de cráneo con depresión de fragmentos, hematomas craneales u otra patología neuroquirúrgica.

Esta definición, como cualquier otra, no está exenta de problemas. Así, por ejemplo, alguien que se encuentre bajo los efectos agudos del estrés post-traumático, podría experimentar uno o varios de los síntomas del TCE-L (dificultad para recordar el evento, aturdimiento, confusión...) y ser erróneamente diagnosticado, especialmente si ha sufrido un accidente de tráfico. Por otro lado, personas que sufran un daño cerebral más severo, como una contusión cerebral, podrían ser clasificados incorrectamente si no

han sufrido una pérdida de conciencia o amnesia. Por tanto, estos criterios pueden resultar en el diagnóstico de TCE-L de personas que no lo han padecido realmente o de personas que sufren un daño cerebral más severo (Sbordone y Saul, 2000). La exactitud en la clasificación de los TCE-L no carece de importancia si tenemos en cuenta que constituyen del 80 al 90% del total de los traumatismos.

A pesar de que, como hemos mencionado previamente, en numerosas ocasiones el término TCE-L se intercambia con el de conmoción cerebral, los criterios diagnósticos propuestos por el DSM-IV (APA, 1994) para el Síndrome Post-Conmocional (SPC) son considerablemente más estrictos que los expuestos para el TCE-L, y consisten en:

1. Una pérdida de conciencia superior a 5 minutos
2. Un período de amnesia post-traumática anterógrada superior a 12 horas
3. Los síntomas cognitivos o conductuales han de persistir al menos 3 meses y han de ponerse de manifiesto mediante evaluación neuropsicológica.

Por tanto, no todos los sujetos que sufren un TCE-L pueden ser diagnosticados de SPC.

Otra definición muy extendida del SPC es la propuesta por La *International Statistical Classification of Diseases and Related Health Problems*, 10ª Edición (ICD-10, 1992), que incluye como parámetros clínicos los siguientes criterios:

- . Historia de traumatismo craneoencefálico con pérdida de conciencia, que precede a la aparición de los síntomas en un máximo de 4 semanas.
- . Tres o más categorías sintomatológicas entre las siguientes:
 - A. Cefalea, fatiga, intolerancia al ruido, vértigos-mareos, malestar.
 - B. Depresión, ansiedad, irritabilidad, labilidad emocional.

- C. Disminución de la capacidad de concentración, memoria o capacidad intelectual sin evidencia neuropsicológica de un deterioro marcado.
- D. Insomnio.
- E. Disminución de la tolerancia al alcohol.
- F. Preocupación por cualquiera de los síntomas anteriores y miedo a padecer un daño cerebral con componente hipocondríaco y adopción del papel de enfermo.

Uno de los principales problemas del SPC es que sus síntomas no parecen ser lo suficientemente sensibles ni específicos ya que, por un lado, su incidencia es muy alta entre la población normal (Youngjohn, Burrows y Erdal, 1995; Wang, Chan y Deng, 2006), y por otro lado las quejas del SPC son comunes a una gran variedad de condiciones médicas (como dolor crónico) y psicológicas (por ejemplo el estrés post-traumático), y por tanto no diferencian con precisión el SPC (Fox, Lees-Haley, Earnest y Dolezal-Wood, 1995). Además, aproximadamente el 50% de las personas que están en litigio, sin haber sufrido un traumatismo, refieren síntomas englobados dentro del SPC, que parecen ser causa del estrés vinculado al proceso legal (Lees-Haley y Brown, 1993). Los síntomas más comunes del SPC y su origen probable aparecen mencionados en la Tabla 1.

La historia natural de un paciente típico tras un TCE-L es su recuperación aproximadamente en 6-12 semanas. La mayoría de los síntomas suelen desaparecer o estabilizarse a los 6 meses del evento desencadenante (Larrabee, 2005^a), pero un pequeño porcentaje de pacientes (entre el 10 y el 25%) presentará síntomas persistentes.

Tabla 1. Componentes habituales del s. post-conmocional y su origen probable

Somáticos		
	Cefalea	Insuficiencia de la Art. Vertebral Insuficiencia de la Art. Basilar (lesión en latigazo)
	.. Migraña Post-Traumática	
	.. Cefalea tensional.	
	.. Aumento de una cefalea previa	
	Vértigos .. Periférico .. Central	Señales de entrada aberrantes de los propioceptores posicionales de columna vertebral Sobre-estimulación de terminales cervicales simpáticas Contusión Troncocerebral Compromiso Vascular Vertebral
	Perdida de Audición	Daño mecánico del conducto auditivo Daño mecánico de terminales nerviosas Disfunción electrofisiológica de Cóclea, VIII par o SNC
	.. De conducción	
	.. Neural	
	Tinnitus	
	Pérdida de Olfato y Gusto	Cizallamiento de tracto olfatorio con la lámina cribiforme. Daño mecánico de estructuras óseas nasales Contusiones frontales (/occipitales por contragolpe) o temporales.
	Trastornos del sueño	Miedo crónico
	Rigidez Cervical	
Neuro-psicológicos	Irritabilidad	Personalidad premorbida Estatus psicológico Lesiones orgánicas de corteza frontal, temporal, área amigdalina, orbito-frontal, conexiones del sistema límbico
	Labilidad	
	Ansiedad	
	Depresión	
Cognitivos	Concentración	Mayor probabilidad a mayor número de lesiones estructurales precoces. Lesiones Frontales: Funcionamiento Ejecutivo Lesiones Temporales: Memoria
	Atención	
	Retención (memoria).	
	Palabras adecuadas	
	Procesamiento de la información	
	Funcionamiento ejecutivo	

Suele considerarse al Síndrome Post-Conmocional como crónico cuando un año después del TCE-L el paciente presenta una sintomatología continuada que puede ser un síntoma aislado o un grupo de síntomas, e incluso puede apreciarse el empeoramiento del complejo sintomático inicial. Es curioso que, a pesar de que está comprobado que un seguimiento mínimo de estos pacientes por parte de los especialistas es efectivo a la hora de reducir las consecuencias sociales y la severidad de los síntomas tras un TCE-L (Wade, King, Wenden, Crawford y Caldwell, 1998), en un estudio en centros estadounidenses se ha puesto de manifiesto que sólo el 34% de los centros suelen remitir a los pacientes con TCE-L a algún especialista para su revisión después de abandonar el servicio de urgencias.

3. Dificultades cognitivas en los TCE-L

Después de un TCE-L son comunes las quejas referidas a las áreas de concentración y memoria. En apoyo a estas quejas, la investigación neuropsicológica ha documentado daños en la velocidad del procesamiento de la información, atención y concentración en la fase aguda de los TCE-L. Muchos pacientes también refieren problemas a la hora de iniciar y organizar actividades complejas. Los tests clínicos empleados en la evaluación neurológica no revelan estas dificultades, por eso cuando los síntomas persisten después de 6 meses es recomendable realizar una evaluación neuropsicológica, útil para determinar si los síntomas referidos por los pacientes con SPC tienen una base orgánica (Mateer y D'Arcy, 2000; Pelegrín-Valero, Fernández-Guinea, Tirapu-Ustarroz y Muñoz-Céspedes, 2001; Vos y cols., 2002). Otro apoyo al empleo de medidas neuropsicológicas en el SPC se deriva de que, si bien los resultados neuropsicológicos no correlacionan con los resultados de la neuroimagen en la fase aguda, sí lo hacen con los resultados de la resonancia magnética realizada a los 6 meses

del golpe (Willson, Wiedmann, Hadley, Condon, Teasdale, y Brooks, 1988). A continuación exponemos algunos de los principales hallazgos de los déficits cognitivos objetivos encontrados en los estudios neuropsicológicos de estos pacientes.

3.1. Problemas de atención

El enlentecimiento en el procesamiento de la información ha sido uno de los primeros déficits estudiados y probablemente el mejor documentado en los pacientes con TCE-L (Gronwall, 1987, 1991). Sin embargo, ya en los primeros trabajos se refiere que el resultado habitual es que se produzca una mejoría en el procesamiento de la información en las semanas y meses después del traumatismo (Gentilini, Nichelli y Schoenhuber, 1989), con un retorno a los niveles normales o cercanos a la normalidad a los 6 meses del daño (McFlynn, Montgomery, Fenton y Rutherford, 1984).

En cuanto a los problemas de distractibilidad de estos pacientes, aunque en la clínica son frecuentes las quejas referidas a la dificultad de realizar dos tareas a la vez o de trabajar en condiciones ruidosas, los resultados de la investigación son mixtos (Mateer, 2000). Cicerone (1996), por ejemplo, objetiviza dichas quejas al poner de manifiesto que los pacientes con TCE-L se diferencian de los sujetos control a la hora de realizar tareas del paradigma dual en condiciones ruidosas.

Otro aspecto fundamental de la atención se refiere a su focalización. En este sentido, Bohnen, Twijnstra y Jolles (1993) han puesto de manifiesto que los pacientes con SPC se diferencian de los controles en el test STROOP, evidenciando, por tanto, problemas en la atención selectiva con control de las interferencias en estos pacientes.

Por último, también se han encontrado déficits persistentes en la atención dividida de estos pacientes: Stablum, Mogentale y Umilta (1996) encontraron que la diferencia a la hora de realizar una tarea simple o doble era mayor en los pacientes con TCE-L (si eran mayores de 30 años o habían sufrido pérdida de conciencia) que en los

controles, y dicho déficit se mantenía dos años más tarde, por lo que parece ser persistente.

En resumen, parece que todos los problemas de atención en los pacientes con SPC están relacionados con la función ejecutiva. Los déficits atencionales de estos pacientes son sutiles, y se ponen de manifiesto sólo en condiciones que requieren esfuerzo o un control del procesamiento cognitivo. Puesto que las tareas empleadas en la evaluación clínica no suelen evaluar estos aspectos, es probable que la ausencia de hallazgos cognitivos en pacientes con TCE-L frecuentemente referida en la literatura (por ejemplo, Binder, 1997) sea debida a este motivo.

Por otro lado, Keller, Hiltbrunner, Dill y Kesselring (2000) han encontrado que los resultados en tareas de atención dividida eran similares en los pacientes con SPC o con daño cerebral severo. Sin embargo, los resultados de dichos pacientes con SPC pasaban a ser comparables a los de un grupo control después de una adecuada motivación verbal. Si bien es puramente especulativo, estos hallazgos podrían estar en consonancia con la teoría de la disfunción ejecutiva, ya que parece que la apatía es otro síntoma fundamental de estos pacientes, y que podría estar influyendo en los resultados de los tests neuropsicológicos.

3.2. Problemas de memoria

Algunos estudios han puesto de manifiesto dificultades en los sistemas de memoria anteriores en los pacientes con TCE-L, en especial en la memoria de trabajo, que se basa en la habilidad para mantener cierta cantidad de información en mente, internalizarla y emplearla para guiar la conducta para la consecución de metas externas. Los pacientes con TCE-L aparecen dañados en pruebas de atención que requieren memoria de trabajo (Stuss y Gow, 1992).

Sin embargo, no se aprecian dificultades a largo plazo en los test tradicionales de

memoria en estos pacientes (Ruff y cols., 1989; Gentilini, Nichelli, y Schoenhuber 1989; Newcombe, Rabbitt y Briggs, 1994) a pesar de sus quejas subjetivas. Esto puede explicarse por los problemas atencionales y de concentración en los pacientes con TCE-L a largo plazo, que los pacientes pueden interpretar como fallos de memoria.

3.3. Problemas en la función ejecutiva

El término función ejecutiva se refiere a funciones diversas mediadas por los sistemas neurales del córtex prefrontal, y su fallo implica problemas a la hora de conseguir metas, realizar tareas múltiples o secuenciar los pasos de una tarea. La corteza prefrontal es especialmente vulnerable a los efectos mecánicos de los TCE-L, y es, por su naturaleza, la región más susceptible a la disfunción causada por el daño axonal difuso. Los pacientes con traumatismos leves han mostrado déficits en las funciones ejecutivas vinculadas a las tareas de fluidez verbal, tanto semántica como fonética (Raskin y Rearick, 1996).

En los últimos años parece que se está produciendo un cambio en la evaluación de las funciones ejecutivas, y cada vez parece existir más consenso a la hora de reconocer las dificultades que plantean las pruebas neuropsicológicas clásicas de flexibilidad cognitiva (como el Wisconsin Card Sorting Test) a la hora de evaluar las funciones frontales. Esta postura crítica se basa en que los déficits en las funciones ejecutivas pueden no mostrarse en tareas explícitas y bien estructuradas, sino en las situaciones de incertidumbre típicas en nuestra vida cotidiana. Por eso, en la actualidad se están desarrollando pruebas de tareas múltiples (como la prueba R-SAT) que intentan superar estas limitaciones, si bien todavía no se ha realizado ningún estudio con este tipo de pruebas en esta población (Raskin, 2000).

En cuanto a las pruebas tradicionales, se ha encontrado que los pacientes con TCE-L suelen mostrar déficits en la flexibilidad cognitiva (con el Wisconsin Card

Sorting Test: Stuss y Gow, 1992), en las pruebas de razonamiento abstracto (Leninger, Gramling, Farrel y Kreuzer, 1990) o en las pruebas que analizan la conducta intencional (como el Tinkertoy Test: Martzke, Swan y Varney, 1991).

En resumen, la literatura indica que en la fase aguda de los TCE-L son frecuentes los problemas de atención, memoria y función ejecutiva, aunque dichos problemas no suelen ser persistentes. Hay que puntualizar, sin embargo, que no encontrar diferencias *estadísticamente* significativas entre un grupo con TCE-L y un grupo control no implica que algunos pacientes continúen mostrando déficits cognitivos. De hecho, la mayoría de los investigadores consideran actualmente que una proporción de pacientes con TCE-L tienen déficits persistentes que no pueden ser explicados por las reacciones emocionales a sus accidentes (Richardson, 2000).

También es necesario considerar que los resultados de las evaluaciones neuropsicológicas pueden estar sesgados por los siguientes motivos (Pelegrín-Valero y cols., 2001):

- Los antecedentes premórbidos (personalidad, baja capacidad intelectual, alteraciones del aprendizaje, alcoholismo...)
- La influencia que pueden tener en las pruebas psicométricas las secuelas motoras y sensoriales
- Los déficits cognitivos secundarios a trastornos psiquiátricos relacionados con el TCE, especialmente la depresión
- La edad avanzada
- Los déficits cognitivos asociados a factores físicos, como la fatiga o el dolor
- La posibilidad de simulación o falta de cooperación en caso de litigio o búsqueda de compensación económica.

4. Evaluación de los TCE-L y del SPC mediante neuroimagen funcional

Tras una revisión de los trabajos publicados entre 1966 y 2000, Davalos y Bennett (2002) concluyen que la SPECT (*Single-Photon Emisión Computerized Tomography*) es una herramienta útil para la detección del daño en los TCE-L. De hecho, esta técnica se ha mostrado superior a la Tomografía Axial Computadorizada (TAC) y a la Resonancia Magnética (RM) a la hora de detectar anomalías en los pacientes con TCE-L. Por ejemplo, Gowda y cols. (2006) detectaron anomalías en la SPECT del 63% de los pacientes estudiados con TCE-L, encontrando una TAC anormal sólo en el 34% de los casos. Del mismo modo, Audenaert y cols. (2003) encontraron anomalías en la SPECT de 6 de los 8 pacientes con TCE-L que evaluaron, mientras ninguno de ellos presentó hallazgos en la TAC. También Abdel-Dayem y cols. (1998) encontraron una SPECT anormal en 28 de los 41 TCE-L que estudiaron con una TAC normal. En cuanto a la RM, Hofman y cols. (2001) encontraron anomalías en la SPECT del 61% de los TCE-L y en la RM de sólo el 57%. Umile, Plotkin y Sandel (1998) también encuentran una superioridad de la SPECT con respecto a la TAC y a la RM. En definitiva, podemos concluir que los diferentes estudios muestran una mayor eficiencia de los estudios funcionales que de los estructurales a la hora de detectar anomalías en los pacientes con TCE-L, siendo el resultado más común una hipoperfusión frontal (Gowda y cols, 2006), si bien otros estudios refieren también hipoperfusión temporal (Audenaert y cols, 2003) o en áreas basales y tálamo, parietal, occipital e ínsula (Abdel-Dayem, y cols., 1998).

Estos hallazgos en la evaluación temprana han mostrado relación con el resultado del seguimiento a los 6 meses (Hofman y cols., 2001), y la SPECT también se ha mostrado útil a la hora de ofrecer una evidencia objetiva de la sintomatología post-conmocional persistente (Jacobs, Put, Ingels y Bossuyt, 1994; Kant, Smith-Seemiller,

Isaac y Duffy 1997).

Diversos estudios han intentado correlacionar la localización de los defectos en la perfusión con los resultados de los test neuropsicológicos, pero los resultados han sido bastante contradictorios. Masdeu y cols., (1994) encontraron que los pacientes con TCE-L y hallazgos en la SPECT también obtenían una puntuación anormal en una tarea de recuerdo selectivo. Ichise y cols. (1994) encontraron una correlación significativa entre el ratio Anterior:Posterior calculado para las imágenes de SPECT y RM y la ejecución en tests de atención, memoria y función ejecutiva en pacientes con TCE. Sin embargo, Varney y cols. (1995) encontraron que los pacientes con TCE-L y hallazgos en la SPECT realizaban correctamente la mayoría de los test neuropsicológicos, excepto las tareas de escucha dicótica y de fluidez de diseños. Tampoco Hofman y cols. (2001) encontraron diferencias entre los pacientes con RM con o sin hallazgos y la ejecución neuropsicológica a los 6 meses después del traumatismo, pero sí en la fase aguda y a los dos meses. Por su parte, Umile y cols. (1998) encontraron que la ejecución en los test neuropsicológicos predice los hallazgos en la SPECT, pero no a la inversa, es decir, los hallazgos en la SPECT no predicen la ejecución neuropsicológica.

Desde nuestro punto de vista, la variabilidad encontrada en los trabajos que estudian la relación entre resultados de la SPECT y de la neuropsicología podría ser inconsistente por varios motivos. En primer lugar, la concepción de daño neuropsicológico es diferente en los diversos estudios, siendo en algunos quizá demasiado estricto (dos desviaciones típicas por debajo de la media); además, los dominios neuropsicológicos medidos no son los mismos en los diversos trabajos, ni están medidos con las mismas pruebas. Por último, en muchos trabajos se considera un índice general de deterioro neuropsicológico que tiene en cuenta los resultados de todas las tareas aplicadas para hallar una sola puntuación indicativa o no de daño. Esta

estrategia puede no ser demasiado acertada, teniendo en cuenta la heterogeneidad que puede encontrarse en la afectación de los dominios cognitivos en los pacientes con traumatismos.

A modo de conclusión, podemos decir que el síndrome post-conmocional es un trastorno que aparece en una minoría de los pacientes que han sufrido un TCE-L y en el que influyen diversos factores tanto orgánicos como psicológicos. Los estudios no son concluyentes, pero parece que existen anormalidades tanto neuropsicológicas como en la neuroimagen funcional de estos pacientes, aunque dichas anormalidades no siempre correlacionan.

5. Factores que inciden en los pacientes con TCE-L para el desarrollo del SPC

¿Cuáles son, por tanto, las causas subyacentes al SPC? La controversia permanece vigente. Algunos autores hablan de causas orgánicas, otros de causas psicológicas (negando incluso la existencia del síndrome), algunos de factores anteriores y otros posteriores al traumatismo, si bien la mayoría aceptan que influyen muy diversos factores que tendrán diferente peso dependiendo del caso. Síntomas similares, pero con procesos etiológicos diferentes se darán en los distintos casos y posiblemente en el mismo individuo en diferentes momentos. El SPC parece, por tanto, el paradigma del modelo biopsicosocial, y debemos intentar valorar, mediante el juicio clínico, el peso que cada uno de los factores pueda tener en cada paciente (Muñoz-Céspedes, Pelegrín-Valero, Tirapu-Ustarroz y Fernández-Guinea, 1998; Pelegrín-Valero y cols., 2001)

King (2003) realiza un exhaustivo repaso de los estudios que apoya tanto los factores orgánicos del SPC como los psicológicos, citando como evidencia del primer tipo de factores la siguiente:

- Los estudios post-mortem en humanos y animales demuestran la existencia de daño axonal difuso consecuencia de TCE-L.
- En las primeras semanas después del daño entre un 8 y un 10% de los pacientes presentan lesiones cerebrales macroscópicas mediante resonancia magnética, aunque estas lesiones no correlacionan bien con los síntomas del SPC.
- Existen hallazgos de daño celular y anormalidades metabólicas en la sustancia blanca frontal en las primeras semanas del daño.
- Se producen anormalidades intraparenquimales en el flujo regional cerebrales las primeras semanas tras el TCE-L
- Se encuentran anormalidades electroencefalográficas y en medidas de respuestas evocadas en el tronco encefálico en las 48 horas posteriores al daño (no relacionadas con los síntomas post-conmocionales ni con los resultados neuropsicológicos).
- Pueden aparecer anormalidades en el flujo sanguíneo cerebral regional y reducción en el metabolismo de la glucosa en áreas temporales en pacientes con SPC incluso 5 años después del daño.

Otro tipo de evidencia a favor de los factores orgánicos se refiere a la peor recuperación ante la existencia de TCE-L previos, a que se produce una recuperación más lenta vinculada a la edad (punto de corte alrededor de los 40 años), y también en los pacientes con antecedentes de abuso de alcohol u otras sustancias.

En cuanto a la evidencia relacionada con los procesos psicológicos, King destaca:

- Las consecuencias del traumatismo son peores cuando existe un trastorno psicopatológico previo.
- Los mejores predictores tempranos de los síntomas post-conmocionales son los

factores psicológicos.

- Existe una elevada comorbilidad entre los síntomas ansiosos y depresivos y los síntomas del SPC, siendo muchos de estos síntomas idénticos.
- La prevalencia del SPC es mayor en mujeres, igual que la prevalencia de manifestaciones psicopatológicas.
- Existe una asociación entre severidad de los síntomas post-conmocionales o del tiempo necesario para incorporarse al trabajo tras un TCE-L y la búsqueda de compensación.

Este último punto es el que quizás ha suscitado más polémica en un síndrome ya de por sí controvertido. La idea de que los síntomas del SPC pueden ser debidos al litigio y no al daño cerebral es apoyada por diferentes autores (Youngjohn y cols., 1995, Fox y cols., 1995) aunque su más ferviente defensor es Binder. Este autor descubrió que el rendimiento neuropsicológico de los pacientes con TCE-L que buscan compensación es significativamente inferior al de los pacientes con un daño cerebral bien documentado sin incentivos financieros. De hecho, Binder y Willis (1991) descubrieron que la ejecución obtenida por tres grupos en búsqueda de compensación (TCE-L, pacientes con daño cerebral documentado y análogos –sujetos a los que se pide que simulen un daño) era inferior a la de tres grupos sin incentivos económicos (pacientes con daño cerebral, pacientes con trastornos afectivos y controles), concluyendo que la búsqueda de compensación explica una proporción del déficit neuropsicológico de los pacientes con TCE-L. Es más, parece existir una relación inversa entre severidad del daño y resultados neuropsicológicos en los pacientes con incentivos económicos, ya que el grupo con daño cerebral documentado obtiene una ejecución superior a la de los pacientes con traumatismos leves (Binder, 1993). Además, la compensación económica está relacionada con la severidad de los síntomas post-conmocionales y el tiempo

transcurrido hasta la reincorporación laboral, siendo más probable la búsqueda de compensación en aquellos pacientes con daños menos severos según los datos neurológicos (Binder y Rohling, 1996). Sin embargo, parece que los síntomas se mantienen después de la resolución de la compensación económica (Binder, 1986; Fee y Rutherford, 1988), si bien para Binder esto no le quita peso a la evidencia de que el proceso de litigio es un factor que influye en la persistencia del SPC.

Este tipo de investigación entronca con los estudios de los años 60, cuando se popularizó el término “neurosis de compensación”, que se guía por el principio de que no hay daño cerebral si no hay lesiones objetivas. Esta es la situación en la cual es más probable que se plantee la posibilidad de simulación, y por ello la investigación del esfuerzo insuficiente o la baja motivación se ha centrado en los pacientes con TCE-L, puesto que en este trastorno es típica la discrepancia entre los hallazgos subjetivos y objetivos y es frecuente que terminen en el ámbito médico-legal (Osimani, Alon, Berger y Abarbanel, 1997).

Una vez revisado el TCE-L y el SPC, así como su vinculación con el concepto de simulación, en el próximo capítulo vamos a hablar de la Psicología Forense, y especialmente de la Neuropsicología Forense, campo en el que la detección de simulación es uno de los tópicos fundamentales.

CAPÍTULO 2
PSICOLOGÍA FORENSE Y
NEUROPSICOLOGÍA FORENSE

1. Introducción

Una vez descrita la población con la que vamos a trabajar, en este capítulo revisaremos la disciplina en la que se engloba el estudio de la simulación, que es la Neuropsicología Forense. Para ello, en primer lugar aclararemos los conceptos de Psicología Forense y de Neuropsicología Forense, centrándonos en sus principales áreas de aplicación. A continuación explicaremos los problemas fundamentales a los que se ha enfrentando y se enfrenta esta disciplina, así como su situación actual. También describiremos el perfil que, a nuestro juicio, debería tener un profesional para ser neuropsicólogo forense, y algunos de los problemas éticos a los que frecuentemente debe enfrentarse. Por último, finalizaremos este capítulo realizando una breve descripción del estado en el que se encuentra la Neuropsicología Forense en nuestro país.

2. Definición y ámbitos de aplicación de la Psicología Forense y la Neuropsicología Forense

Podemos tomar como definición de *Psicología Forense* la siguiente: “toda psicología, bien experimental o clínica, orientada a la producción de investigaciones psicológicas y a la comunicación de sus resultados, así como a la realización de evaluaciones y valoraciones psicológicas para su aplicación en el contexto legal” (Garzón, 1990).

Los marcos de actuación más relevantes dentro de la psicología forense engloban, principalmente, los siguientes:

- Prevención del delito
- Tratamiento del delincuente/educación del menor
- Prevención de la victimización

- Búsqueda y tratamiento de la víctima
- Evaluación de los permisos de vigilancia penitenciaria
- Evaluación forense: penal, civil, trabajo, canónico.
- Asistencia a abogados
- Formación y selección de policías/jueces
- Mediación en la Sala de Justicia (procesos de separación)

Quizá la actuación más solicitada al psicólogo forense es la realización de informes periciales, entre los que podemos destacar:

Dentro del derecho civil:

- Internamiento en centros
- Incapacidades
- Evaluación de las secuelas de accidentes
- Mediación en la resolución de conflictos

En el derecho de familia:

- Separaciones
- Divorcios
- Evaluación psicológica de las causas de las nulidades civiles
- Acogimientos y adopciones
- Guardia y custodia
- Régimen de visitas
- Mediación familiar
- Informes para los tribunales eclesiásticos

En el derecho penal:

- Imputabilidad
- Evaluación de las eximentes incompletas

- Secuelas en víctimas
- Evaluación de la credibilidad de testimonios
- Realización de interrogatorios especiales (niños, violaciones...)
- Evaluación y/o elaboración de las ruedas de identificación

En el derecho del trabajo:

- Secuelas de accidentes
- Informes de incapacidad

El trabajo de Ackerman (2006) constituye una buena guía para la redacción de informes forenses. Según este autor, estos informes deben ser lo suficientemente extensos para cubrir toda la información necesaria, y todas las conclusiones han de estar basadas en la literatura científica sobre el tema en cuestión, datos objetivos e información colateral cuando sea necesaria. La cuestión última por la que fue encargado el informe debe quedar claramente señalada en la conclusión del informe, que deberá estar escrito a un nivel adecuado al de los lectores a los que va dirigido.

Podemos definir la *Neuropsicología Forense* como una subespecialidad de la neuropsicología clínica que aplica directamente los principios y prácticas neuropsicológicas a cuestiones referidas a la toma de decisiones legales (Hom, 2003) o simplemente como la aplicación de la neuropsicología a los procedimientos legales civiles y criminales (Larrabee 2000, 2005^b). Por tanto, la neuropsicología forense se considera un híbrido que combina la ciencia y práctica neuropsicológica con la ciencia y práctica forense (Heilbroner, 2004). Según Larrabee (2000) los principales ámbitos de aplicación de la neuropsicología forense serían los siguientes:

En el ámbito forense:

- Daños y perjuicios personales:

Traumatismos craneoencefálicos

Daños neurotóxicos

Secuelas neurológicas causadas por corriente eléctrica

- Competencias civiles:

Tutela

Competencia para la aceptación de tratamientos médicos

Capacidad testamentaria

En el ámbito criminal:

- Capacidad para participar en el juicio
- Determinación de la responsabilidad criminal.

El tema de esta tesis, es decir, la evaluación de la simulación, sería fundamental en todos estos ámbitos de aplicación, ya que la única diferencia entre el dominio civil y criminal reside en la motivación para simular: en el primero de los casos estaría relacionada con una compensación económica y en el segundo con evitar una pena o castigo. De todos modos, los neuropsicólogos forenses han focalizado sus trabajos en los litigios civiles, y concretamente en aquellos referidos a traumatismos craneoencefálicos, por lo que existe muy poco material escrito sobre el neuropsicólogo en el ámbito criminal (Denney y Wynkoop, 2000).

3. Problemas de la neuropsicología forense

Existen diversos artículos que ponen de manifiesto las limitaciones de la neuropsicología forense, y que incluso cuestionan su existencia. Entre estos artículos escépticos destaca la revisión crítica sobre los principales problemas de la neuropsicología forense realizada por Faust (1991), que destaca como principales los siguientes:

- la neuropsicología no es necesaria cuando se trata con casos de daño cerebral severo,

puesto que otras disciplinas ofrecen una evidencia definitiva, mientras que cuando se enfrenta con daños leves la neuropsicología se ve reducida a la especulación y las suposiciones;

- falta de datos científicos que respalden la descripción y predicción del funcionamiento de la vida diaria;
- el desconocimiento de los porcentajes de base para los distintos trastornos dentro del contexto forense;
- la ausencia de prácticas estandarizadas;
- la falta de capacidad para detectar simulación de déficits;
- los problemas a la hora de determinar el funcionamiento intelectual, neuropsicológico y emocional premórbido;
- dificultades a la hora de integrar datos complejos;
- la experiencia y la confianza de los expertos en su testimonio no parece estar relacionada con la exactitud de los diagnósticos.

En cuanto a este último punto es interesante citar los estudios pioneros de Heaton, Smith, Lehman y Vogt (1978) y de Faust, Hart y Guilmette (1988). En estos primeros estudios se ponía de manifiesto la capacidad de los sujetos (incluso niños) para ser entrenados en la simulación de un trastorno concreto que los profesionales diagnosticaban sin ser conscientes de dicha simulación. Estas investigaciones estimularon el desarrollo de los estudios sobre simulación, y una de las explicaciones ofrecidas es que los neuropsicólogos no detectaban la simulación simplemente porque no la tenían en cuenta es su diagnóstico diferencial. De hecho, Trueblood y Binder (1995) realizaron un estudio en el que advertían a los psicólogos sobre la posibilidad de simulación, y estos obtuvieron una exactitud muy elevada a la hora de distinguir sujetos simuladores de pacientes con traumatismo craneoencefálico.

Las contestaciones a los artículos que cuestionan la utilidad de la neuropsicología forense fueron numerosas. Por ejemplo, Barth, Ryan y Hawk (1992) declaran que, si bien es necesaria una cualificación adecuada y el empleo de métodos apropiados, los neuropsicólogos pueden ofrecer un valioso testimonio experto dentro de los límites impuestos por el sistema legal. Además, ponen de manifiesto numerosos problemas de los estudios en los que se basan los autores escépticos para llegar a la conclusión de que el testimonio experto de los neuropsicólogos no debería ser aceptado en los juicios, concluyendo que, si bien es cierto que los profesionales de la neuropsicología debemos seguir trabajando para remediar los déficits de nuestra disciplina, podemos aportar con nuestra evaluación una amplia variedad de datos muy útil para diversos tipos de procesos judiciales.

Uno de los principales problemas que sigue vigente en el ámbito de la neuropsicología forense es el de la estimación del funcionamiento premórbido. No se trata, ni mucho menos, de una dificultad exclusiva de esta disciplina, sino que ha sido y sigue siendo una preocupación constante en todos los ámbitos de la neuropsicología. La cuestión central es que no podemos determinar las consecuencias de un daño en un área si no conocemos el funcionamiento del sujeto en dicho área antes del daño. Aunque ninguna está exenta de problemas, disponemos de distintas estrategias de estimación del funcionamiento premórbido (Verdejo-García, Alcázar, Gómez-Jarabo y Pérez García, 2004):

- El método del mejor rendimiento, que consiste en tomar las puntuaciones más elevadas obtenidas por el paciente como indicativas de su funcionamiento general antes del daño.
- La evaluación de habilidades cognitivas resistentes a daños neurológicos y psicopatológicos, como por ejemplo la lectura.

- Métodos basados en la relación existente entre ciertas variables sociodemográficas y las puntuaciones obtenidas en los test de inteligencia.
- Métodos de regresión mixtos que incluyen variables sociodemográficas y habilidades resistentes al daño.
- Si bien todos los métodos anteriores se centran en la estimación del CI premórbido, en la actualidad los esfuerzos se están dirigiendo a la estimación de otras habilidades más interesantes para el ámbito forense, como por ejemplo la memoria, aunque todavía se necesita mucha investigación al respecto.

4. Situación actual de la neuropsicología forense

Hay que destacar que los artículos escépticos fueron escritos, en su mayoría, a principios de los 90, y que el panorama de la disciplina ha cambiado ampliamente. De hecho, la neuropsicología forense es el área que más rápidamente se ha extendido en los últimos años dentro de la neuropsicología clínica, existiendo una revista desde 1999 dedicada exclusivamente a este área (*The Journal of Forensic Neuropsychology*). Otra revista, *The Clinical Neuropsychologist*, le dedica una sección (*Courting the Clinician*), y *The Journal of Head Trauma Rehabilitation* le ha dedicado dos ediciones especiales (Heilbroner, 2004). Cada vez se publican más artículos sobre el tópico de neuropsicología forense (Bigler, 2006). Para documentar dicho progreso, Sweet, King y cols. (2002) han analizado los artículos publicados en *Archives of Clinical Neuropsychology*, *Journal of Clinical and Experimental Neuropsychology* y *The Clinical Neuropsychology* (elegidas por ser las que sustentan un mayor número de suscripciones), llegando a la conclusión de que los artículos sobre neuropsicología forense han pasado de un 4% en 1990 a un 14% en el 2000. Del mismo modo, las presentaciones forenses han pasado del 3,9 al 11,3% en los congresos de la NAN y del

2,3 al 11,7% en los congresos de la División 40 de la APA, llegando a constituirse en uno de los campos que más está aportando al conjunto de la neuropsicología.

Además de esta expansión científica, la práctica forense es, en la actualidad, un dominio muy importante dentro de la neuropsicología clínica (Sweet, Peck, Abramowitz y Etzweiler, 2002). El tiempo que los neuropsicólogos le dedican la práctica forense ha sufrido un notable incremento. De hecho, los abogados constituyen en la actualidad la tercera fuente de referencia de pacientes (superados sólo por psiquiatras y neurólogos) (Sweet, Moberg y Suchy, 2000), y la primera fuente de referencia si consideramos sólo el contexto privado (Hom, 2003).

Demostrada la expansión de la disciplina, el panorama actual tiene poco que ver con el que existía en la época en la que Faust escribió sus críticas. El crecimiento de la neuropsicología forense es un resultado directo del crecimiento de la neuropsicología clínica, que ha establecido principios de la relación cerebro-conducta así como una metodología válida y fiable para medir dichas relaciones. En la actualidad, dado que los neuropsicólogos clínicos están ampliamente familiarizados con el escrutinio de su disciplina, con el procedimiento clínico con énfasis en la toma de decisiones basada en datos objetivos y con una aproximación de comprobación de hipótesis para dar respuesta a las cuestiones que se le plantean, parece razonable pensar que los neuropsicólogos clínicos cumplen los requisitos necesarios para cubrir el papel cada vez más amplio que se les demanda desde el área forense (Heilbrunner, 2004).

A pesar de que, como hemos mencionado anteriormente, la neuropsicología forense está considerada como un área dentro de la neuropsicología clínica, existen ciertas diferencias entre las evaluaciones clínicas y las forenses que las hacen sustancialmente diferentes (Bush, NAN Policy & Planning Committee, 2005). Denney y Wynkoop (2000) establecen las siguientes:

1) La meta de la evaluación clínica es generalmente la de poder establecer un plan de intervención que ayude a la persona a aliviar su sufrimiento, mientras la evaluación forense se pone al servicio del sistema penal. Así, en la clínica, el clima es de colaboración y confianza, mientras que una evaluación forense probablemente sea considerada como adversa.

2) El rol del clínico es el de “ayudar al paciente”, y el del forense el de “buscador de la verdad”, por lo que mientras el clínico establece una alianza con el paciente, la alianza del forense es con la verdad y no con el sujeto, puesto que la evaluación ha de ser neutra y objetiva.

3) Estas diferencias se extienden también a la metodología, ya que la evaluación forense requiere más fuentes de información que la clínica (médica, educativa...) y por tanto más tiempo y minuciosidad [de hecho las evaluaciones forenses son, con diferencia, a las que más tiempo se les dedica dentro de la neuropsicología clínica (Sweet, Peck y cols., 2002)], así como adjudicar un mayor peso a las pruebas objetivas que a las quejas subjetivas y conceder mayor importancia a la simulación .

Lo que se demanda en última instancia al neuropsicólogo forense es que arroje luz sobre la “cuestión forense”, que no se refiere a la existencia o no de déficits en un individuo, sino que debe aclarar si dicha disfunción resulta del evento bajo consideración legal (Larrabee, 2000; Heilbroner, 2004). Por ejemplo, en un juicio por accidente de tráfico en el que el demandante reclama una compensación económica a causa de las secuelas neuropsicológicas sufridas a raíz del accidente, la cuestión forense crítica no es si el paciente tiene un daño cognitivo, sino si los déficits son el resultado del accidente. El daño neuropsicológico por sí solo no responde la cuestión forense (como en el caso de un individuo acusado de un crimen y que a pesar de tener déficits neuropsicológicos está capacitado para ofrecer su testimonio en el juicio y para hacerse

cargo de las consecuencias de sus actos).

5. Cualificación del Neuropsicólogo Forense

Teniendo todo lo anterior en cuenta, ¿quién está cualificado para practicar la neuropsicología forense? En la actualidad no existen criterios establecidos para ello ni siquiera en los EEUU, donde la disciplina está mucho más extendida y avanzada, por lo que cualquier profesional puede autodenominarse neuropsicólogo forense en 49 de los 50 estados (con excepción de Louisiana) si considera que reúne los conocimientos necesarios para serlo. No existe certificación, credenciales específicas o programas de licenciatura para la práctica de la Neuropsicología forense. Indudablemente, esto puede acarrear serios problemas entre los que podemos citar el desprestigio del área si los profesionales que la ejercen no lo hacen con la debida responsabilidad. Nuestro punto de vista coincide con el de Larrabee (2000) y Heilbroner (2004), al señalar que un profesional adecuado sería aquel con una probada formación en neuropsicología clínica (en concreto en el área pertinente para el caso particular: TCE, demencia...) mediante la certificación pertinente (*The American Board of Clinical Neuropsychology* o *The American Board of Professional Neuropsychology*) y en psicología forense (debe estar familiarizado con la *American Academy of Forensic Psychology* y la *American-Law Society's Specialty Guidelines for Forensic Psychologists –Comitee on Ethical Guidelines for Forensic Psychologists, 1991-*), así como un adecuado conocimiento de los aspectos jurídicos que sean relevantes para el caso. Hay que tener en cuenta que un profesional especializado en litigios civiles puede no estar preparado para enfrentarse a casos criminales o viceversa, es decir, un neuropsicólogo debe poseer el conocimiento psicológico, neuropsicológico y forense necesario para dar respuesta al caso concreto para el que sus servicios son demandados. Al tener un campo de aplicación potencial

tan amplio, es prácticamente imposible que un profesional sea poseedor del conocimiento necesario para afrontar casos de todos los dominios posibles, por lo que es realmente importante la relación con otros profesionales de nuestro área y del resto de las disciplinas forenses (abogados, jueces, médicos forenses...) para ejercer la profesión de modo responsable (para un análisis de las relaciones entre neuropsicólogos y abogados a través de las distintas fases de un proceso judicial, recurrir al trabajo de Greiffenstein y Cohen, 2005).

Estrechamente vinculada a la capacitación de un profesional está la aceptación de su testimonio. En EEUU, durante 70 años el "standard" para la admisibilidad del testimonio experto ha estado basada en el caso Frye vs. United States (1923), quedando establecido que la evidencia científica admisible es aquella que posee reconocimiento y aceptación general en el campo específico al que pertenece. No se hace mención, por tanto, a la *validez* de la técnica empleada. Esta situación cambia en 1993 (caso *Dauvert vs. Merrel Dow Pharmaceuticals*), fecha en la que la admisibilidad de la evidencia pasa a estar basada en inferencias y metodología derivadas del método científico, y ser correctamente aplicadas al hecho en cuestión. Los criterios concretos para la decisión de aceptabilidad se basan en los siguientes factores:

- 1) si la teoría o técnica puede ser (y ha sido) probada,
- 2) si ha estado sujeta a publicaciones y revisiones de otros expertos,
- 3) si se conoce el error de base potencial de los métodos usados y
- 4) si posee aceptación general dentro de la comunidad científica.

La contribución de la neuropsicología forense al campo legal depende, por tanto, de la aplicación apropiada de una metodología científicamente válida para dar respuesta a la cuestión forense propuesta. En la actualidad disponemos de un cuerpo de investigación y conocimiento que lo permite, y es responsabilidad del neuropsicólogo

forense estar al día en estos hallazgos para desempeñar su función de modo competente.

6. Problemas éticos de la neuropsicología forense

Debido a las diferencias mencionadas con anterioridad entre la neuropsicología clínica y la forense no se considera que exista una verdadera relación neuropsicólogo-paciente en el contexto forense. Sin embargo, el neuropsicólogo debe actuar conforme al Código Ético de la *American Psychological Association* (2002). Este código, a diferencia de su predecesor redactado en 1992, no incluye una sección específica dedicada a la neuropsicología forense. Los cambios del Código que pueden tener una mayor repercusión en el desempeño de la neuropsicología forense se refieren a la cesión de los datos del paciente, la necesidad de obtener un consentimiento informado para la evaluación, y la utilización de intérpretes para las evaluaciones. Recientemente ha sido publicado un informe oficial de la NAN (Bush, NAN Policy & Planning Committee, 2005) con directrices dirigidas a la consecución de una evaluación forense independiente y objetiva. Entre los principales obstáculos para la práctica ética de la disciplina podemos citar los siguientes (adaptado de Grote, 2005 y Larrabee, 2000):

- *Competencia.* Los psicólogos deben aceptar sólo aquellos casos para los que su educación, entrenamiento y conocimiento es suficiente para ofrecer una opinión experta. Debe, por tanto, poseer conocimiento sobre la literatura relevante y experiencia clínica con el tipo de paciente y problema referido.
- *Método del anticipo.* Los abogados están acostumbrados a cobrar sus honorarios una vez la sentencia ha sido dictada, y pueden proponer lo mismo al profesional del que demandan servicios. Parece difícil, sin embargo, mantener la objetividad cuando el cobro de los honorarios depende de la decisión judicial, por lo que esta práctica debe evitarse.

- *Revisión de datos y examen del paciente.* El psicólogo no debe conformarse con examinar la información que un abogado o compañía de seguros le envía para un determinado caso, sino que debe escrutar de forma rutinaria toda la información que considere necesaria para llegar a una conclusión. En la mayoría de los casos el psicólogo hace su propio examen del paciente, pero esto no siempre es posible. Todos los obstáculos encontrados en el proceso y su posible efecto sobre la formación de opiniones deben incorporarse en el informe.
- *Presencia de terceros.* En muchos casos, el abogado solicita estar presente durante la evaluación de su defendido. La presencia de terceros o la grabación durante la evaluación es inadmisibles por dos razones fundamentales (Larrabee, 2000): a) puede afectar la validez de la ejecución y b) se puede comprometer la confidencialidad de los tests.
- *Extensión y pormenores de la entrevista.* Es frecuente que los pacientes interpreten que el psicólogo estará a su favor (si es referido por su abogado) o en su contra (si es referido por la compañía de seguros). El evaluador debe esforzarse por convencer al evaluado de su objetividad, y en ningún caso la actitud del paciente ha de determinar el número, extensión o tipo de preguntas que se le realizan.
- *Relación con litigantes y demandantes.* El neuropsicólogo debe explicar las circunstancias de la evaluación, los límites de la confidencialidad y que se realizará un informe escrito de los resultados. Ha de hacerse un consentimiento firmado, cuya realización puede basarse en el propuesto por la NAN (Bush, NAN Policy & Planning Committee, 2005).
- *Realización de una evaluación competente.* Para ello el neuropsicólogo debe mantenerse actualizado en los avances de su disciplina.

- *Informe completo de los resultados.* No debe omitirse información relevante para el caso, por mucho que esta pueda resultar embarazosa para el paciente.
- *Exactitud del testimonio.* Debe darse una opinión completa y exacta de todos los hallazgos.
- *Sesgos.* Los abogados pueden intentar, de forma más o menos sutil, que sesguemos nuestros resultados. El paciente y el propio neuropsicólogo son también fuentes de sesgo potencial. Uno de los principales sesgos es de confirmación (tendencia a buscar evidencia que confirme nuestra hipótesis e ignorar la evidencia que la cuestione), las correlaciones ilusorias... Conocer y evitar dichos sesgos es responsabilidad del neuropsicólogo.

De modo más general, Bigler (2006) señala un sesgo potencial que podría estar afectando al conjunto de la disciplina neuropsicológica forense. Dada la naturaleza adversa y hostil del sistema judicial, en la que dos partes se ven enfrentadas, a qué parte esté vinculado el neuropsicólogo podría hacer que los resultados de lo que se publica estén sesgados. Dependiendo de la tendencia en sus publicaciones, así serán llamados por una u otra de las partes. Sin embargo, no se puede determinar si existe este sesgo vinculado al conflicto de intereses, ya que sólo el 10% de los artículos sobre tópicos forenses especifican la información necesaria para dilucidar este asunto. Especificar dicha información sería otra de las cuestiones éticas que atañen a los neuropsicólogos.

7. La neuropsicología forense en España

Bajo nuestro punto de vista no es exagerada la afirmación de que la neuropsicología forense española es, prácticamente, inexistente. Esto no es de extrañar, puesto que la neuropsicología clínica está todavía emergiendo en nuestro país. Estamos seguros de que esta situación cambiará con el paso de los años, ya que disponemos de

profesionales que están poniendo todo su esfuerzo para el desarrollo de esta disciplina.

Verdejo y cols., (2004) señalan como principales retos para el desarrollo de la neuropsicología forense en nuestro país los siguientes:

- El problema de la estimación del funcionamiento premórbido
- La predicción del funcionamiento diario y la validez ecológica de los instrumentos
- El problema de los porcentajes de base
- El problema de la detección de simuladores

En gran medida estas dificultades son dolencias generales de la neuropsicología clínica, que se hacen más patentes cuando nos centramos en el ámbito forense. Todos los profesionales deberíamos hacer un esfuerzo investigador para intentar reducirlas. De hecho, esta es la finalidad de este trabajo con respecto a los dos últimos problemas mencionados.

Una vez descrita la disciplina en la que se engloba esta tesis, en el siguiente capítulo nos centraremos en uno de sus tópicos fundamentales, que constituye el foco de este trabajo: la simulación del daño neuropsicológico.

CAPÍTULO 3
LA SIMULACIÓN DEL DAÑO
NEUROPSICOLÓGICO

1. Introducción

La finalidad de este capítulo es realizar una descripción de los conceptos básicos para el estudio de la simulación. Para ello, en primer lugar definiremos la simulación y algunos términos relacionados con ella, estableciendo su diagnóstico diferencial. En segundo lugar, explicaremos los modelos teóricos básicos que permiten explicar la simulación, pasando, a continuación, a la explicación de sus criterios diagnósticos. En los últimos apartados del capítulo, describiremos los modelos de los que disponemos en la actualidad para estudiar la simulación, exponiendo los problemas que plantean, para finalizar con algunas de las cuestiones éticas asociadas a la detección de la simulación.

2. Conceptualización de la simulación y terminología

Aunque el engaño ha acompañado al hombre a lo largo de su historia y lo hace también a lo largo de su vida, ha empezado a estudiarse dentro de un marco científico muy recientemente. A pesar de esta demora, parece haberse producido un boom en las últimas décadas, de modo que la simulación se ha convertido en el tópico más estudiado dentro de la neuropsicología forense (apareciendo en un 86% de los artículos forenses) (Sweet, King y cols., 2002), por lo que es común encontrarse en las revistas vinculadas con la neuropsicología al menos un artículo basado en este tema (Lynch, 2004).

En la actualidad, la definición más extendida y aceptada de simulación es la empleada por el DSM-IV (American Psychiatric Association, 1994), que la entiende como la producción intencional de síntomas físicos o psíquicos falsos o extraordinariamente exagerados, con el objetivo de conseguir una recompensa externa identificable como obtener compensaciones económicas o drogas, evadir el trabajo o responsabilidades criminales...

Uno de los problemas que se derivan de esta conceptualización de la simulación es

su visión dicotómica o visión de todo o nada, según la cual un sujeto simula o no simula; su ejecución en los test es válida o inválida. Muchos autores, aunque aceptan la definición del DSM-IV, son más partidarios de ver la simulación como un contínuum en el que se sitúa al sujeto (Haines y Norris, 1995; Gouvier y cols., 2003). De acuerdo con esta perspectiva, Rogers analiza distintos estilos de respuesta que pueden guiar la conducta del simulador (adaptado de Rogers, 1997^a y Rogers y Bender, 2003):

- Simulación (*malingering*): Siguiendo al DSM-IV, Rogers la define como la fabricación o extraordinaria exageración consciente de síntomas físicos o psicológicos buscando un objetivo externo.

- Minimización (*defensiveness*): es el polo opuesto de la simulación. Se refiere a la negación o extraordinaria minimización consciente de síntomas con el fin de obtener una recompensa externa.

- Respuesta irrelevante: se refiere a un estilo de respuesta en el cual el individuo no llega a implicarse en el proceso de evaluación y las respuestas dadas no están necesariamente relacionadas con el contenido de la petición clínica (por ejemplo, no leer los ítems de un test o no esforzarse por responder a lo que se le pregunta).

- Respuesta al azar: es un subconjunto de las respuestas irrelevantes en las que se puede identificar un patrón al azar. Este estilo de respuesta es observado frecuentemente en medidas con un formato de elección forzosa.

- Respuesta fingida: es aquella dirigida a la fabricación o exageración de síntomas, pero de la que no podemos asumir su finalidad.

- Esfuerzo insuficiente (también denominado incompleto o subóptimo): es un término descriptor que indica que no se realizó una ejecución máxima por parte del sujeto. Puede deberse a estados internos (fatiga, frustración) o comórbidos (por ejemplo, depresión debida a un daño cerebral). Sólo debe considerarse que el sujeto finge cuando

el esfuerzo insuficiente es extremo, aunque también en este caso han de determinarse el estado interno y la comorbilidad. De todos modos, estamos de acuerdo con Boone, Lu y Herzberg (2002) que opinan que este término no es demasiado acertado, ya que muchas veces los pacientes realmente ejercen un esfuerzo considerable para producir síntomas falsos, por lo que consideran más adecuado el término esfuerzo sospechoso.

- Respuesta honesta: se refiere al patrón de respuesta reflejado por el intento sincero del paciente de ser exacto en sus respuestas. Un enorme problema para los clínicos es cómo establecer la respuesta honesta ya que la ausencia de simulación, aunque crítica, no es suficiente.

- Respuesta híbrida: Se refiere a cualquier combinación de los estilos de respuesta previos. Aunque clínicamente observada, la incidencia de los estilos de respuesta híbridos permanece completamente desconocida. (Un ejemplo podría ser el del drogadicto que es honesto en lo referido a su sintomatología pero altamente minimizador con respecto a su consumo de sustancias).

2.1. ¿De qué debemos diferenciar la simulación?

El DSM-IV recoge la necesidad de realizar un diagnóstico diferencial entre la simulación y los trastornos somatoforme y facticio. Para ello es necesario tener en cuenta que en la simulación la conducta está bajo el control volitivo de la persona y el tipo de incentivo buscado es externo (recompensas económicas, evitación de sanciones...). En el trastorno facticio la conducta es controlada por el sujeto, pero el tipo de incentivo buscado es psicológico (adoptar el rol de enfermo, recibir atención...). Por último, en el trastorno somatoforme la conducta del sujeto no está bajo control volitivo, y los incentivos perseguidos son de tipo psicológico (por ejemplo controlar el estrés o algún conflicto interno) (ver Tabla 2).

Tabla 2. Diagnóstico diferencial de la simulación

	Control de la conducta	Incentivo externo
Simulación	SÍ	SÍ
Trastorno facticio	SÍ	NO
Trastorno somatoforme	NO	NO

Además de tener en cuenta estos trastornos, cuando nos enfrentamos a la posibilidad de simulación debemos considerar que la magnificación de síntomas o un bajo esfuerzo en los test neuropsicológicos puede tener muchas causas subyacentes. De este modo, que un paciente indique síntomas no neurológicos o un esfuerzo atípico dentro del contexto de evaluación neuropsicológica no indica automáticamente que el individuo esté simulando. La presión y estrés asociados a los procesos de litigio, una necesidad psicológica de mantener el rol de enfermo, la hipersugestionabilidad, distorsiones depresivas, somatizaciones, hipocondrías y reacciones de conversión deberían ser consideradas en la evaluación y el proceso diagnóstico diferencial de la simulación (Iverson y Binder, 2000). Otros factores que podemos citar como causas de una baja ejecución son la fatiga o privación de sueño, la medicación o abuso de sustancias, problemas emocionales o psiquiátricos, condiciones de evaluación inadecuadas, falta de rapport, funcionamiento premórbido dañado, daño periférico motor o sensorial, dolor agudo, estresores psicosociales, problemas culturales o lingüísticos ante las pruebas...Otras fuentes externas al paciente que conducen a errores en el diagnóstico son los problemas metodológicos (por ejemplo el problema de los porcentajes de base o la validez ecológica de los instrumentos) o los sesgos propios del examinador (por ejemplo el sesgo confirmatorio).

Teniendo en cuenta todos estos factores parece claro que una ejecución en los test indicativa de esfuerzo insuficiente o baja motivación no implica directamente que el

sujeto esté simulando, sino que debemos considerar toda una serie de explicaciones alternativas. Paralelamente, que un sujeto realice bien los test de simulación no elimina directamente la posibilidad de que dicho sujeto esté simulando: si entendemos la ejecución de un individuo simulador como un *continuum* es fácil determinar que un sujeto puede simular consistentemente durante toda la evaluación o hacerlo de modo intermitente, del mismo modo que puede decidir simular de un modo general o hacerlo en ciertos síntomas (por ejemplo emocionales), pero no en otros (por ejemplo cognitivos). Por tanto, ni una ejecución indicativa de simulación en los tests implica simulación, ni una puntuación en los tests no indicativa de simulación elimina la posibilidad de que el sujeto esté simulando. Entonces, ¿en qué debemos basarnos para concluir si alguien es simulador? El diagnóstico final de simulación debe estar basado en un minucioso examen de la historia del individuo (toda la información que podamos conseguir de la vida del sujeto antes del daño –nacimiento y desarrollo, datos médicos, educativos, laborales, legales, referidos a salud mental, abuso de sustancias, datos militares...-, en el momento del daño –informes de testigos-, y en la actualidad –examen médico, neurológico, psicológico...-), proporcionada tanto por el paciente como por informadores colaterales (familiares, compañeros de trabajo, médicos) y de la evaluación que realicemos nosotros mismos al paciente (la entrevista, su ejecución en los test neuropsicológicos...). Por tanto, el diagnóstico de simulación recae, en último término, en el juicio clínico.

Si bien la simulación es mucho más probable en el contexto forense que en el clínico, no debemos infravalorarla en este último ámbito. Como recomienda la NAN (Bush, Ruff y cols., 2005), si bien el examen de la validez de la ejecución de los sujetos ha de realizarse de modo más exhaustivo y completo en el contexto forense, en algunos casos el peso potencial de la simulación en el contexto clínico puede ser igual de

importante, como en la determinación de la incapacidad de un sujeto con propósitos rehabilitadores, aunque en el momento de la evaluación esta no tenga relevancia forense.

2.2. Más términos por aclarar...

Un tema de especial relevancia en cualquier área emergente en una disciplina es el de la terminología. Así, siguiendo con la conceptualización de Rogers (1997^a), es necesario aclarar una serie de términos equívocos relacionados con la simulación:

Disimulación (*dissimulation*): término general empleado para describir un individuo que está distorsionando deliberadamente sus síntomas. Suele emplearse cuando son inaplicables términos más específicos como simulación o minimización.

Falta de fiabilidad (*unreliability*): término no específico usado para describir clínicamente las características de un individuo cuyo estilo de respuesta no es honesto, aunque no se puede clarificar más con respecto a sus intenciones.

Engaño (*deception*): término para describir cualquier intento de un individuo de distorsión o infra-representación de lo que informa. Por ejemplo, algunos pacientes no son honestos sobre sus conductas, actitudes o percepciones pasadas o actuales, lo cual puede estar completamente separado de la descripción del paciente de sus funciones psicológicas.

Ganancia secundaria: es un término clínico inespecífico que debería evitarse en las evaluaciones forenses. Se usa para describir la perpetuación o aumento de síntomas como respuesta a fuerzas internas (para los modelos psicodinámicos) o externas (para los modelos de la medicina conductual) no controladas por el sujeto.

Autoapertura (*self-disclosure*): se refiere a cuánto de sí mismo revela un individuo, es decir, al grado de apertura. No implica falta de honestidad, sino simplemente falta de voluntad para compartir información personal.

Deseabilidad social: es un constructo muy vinculado a la minimización, pero es más extenso. Se compone de dos elementos: la negación de características y la atribución de cualidades positivas.

Manipulación de las apariencias (*impression management*): es un constructo ampliamente investigado para explicar las conductas sociales que intentan crear una imagen positiva y evitar perturbaciones y emociones negativas.

La paradoja de la simulación: se refiere a los problemas de la investigación ejemplificados por los estudios de simulación en los cuales se les pide a los participantes que finjan problemas o síntomas para estudiar a pacientes que fingen realmente sus síntomas cuando se les pide que realicen lo mejor que puedan las evaluaciones médicas o psicológicas.

3. Los modelos de simulación:

Siguiendo a Rogers (1990, 1997^a), existen tres modelos o puntos de vista desde los cuales podemos entender la simulación:

3.1. El modelo patogénico

Este modelo postula que la fuerza subyacente en la simulación es el desorden mental. En un intento por controlar los síntomas emergentes, el paciente crea los síntomas y los enmarca como genuinos. El resultado es un empeoramiento del desorden mental y una aparición de verdaderos síntomas. El modelo patogénico se ha abandonado porque sus predicciones no han sido demostradas y porque la percepción de la simulación se ha modificado en las últimas décadas.

3.2. El modelo criminológico

Nace con la publicación del DSM-III. Los modelos DSM presuponen que es probable que la simulación ocurra cuando: la persona está diagnosticada de trastorno

antisocial de la personalidad, cuando las evaluaciones se realizan con propósitos forenses, cuando la persona no coopera con la evaluación y el tratamiento, y cuando las quejas de la persona son discrepantes con los hallazgos objetivos. Para Rogers este modelo carece de base empírica y conceptual. En primer lugar, la asociación entre trastorno antisocial de la personalidad y simulación es ilusorio. Cuando las personas se enfrentan a una evaluación forense, pueden adoptar un rango de estilos de respuesta (por ejemplo, un agresor sexual es improbable que exagere su conducta; más bien la negará y adaptará un estilo de respuesta defensivo). Por otro lado, la selección del criterio de no cooperación es contraintuitivo, ya que excluiría a pacientes verdaderos (como en los trastornos de la comida o en los de abuso de sustancias) y no identificaría a verdaderos simuladores que cooperan con el evaluador para conseguir los incentivos perseguidos. Por último, la discrepancia con hallazgos objetivos es, para Rogers, una falacia, ya que asume erróneamente la existencia de datos “objetivos” contra los que las inconsistencias pueden ser medidas. Si los hallazgos objetivos para los desórdenes mentales existiesen, la simulación no sería un problema.

3.3. El modelo adaptativo

Según este modelo, la simulación sería más probable cuando: a) el contexto de la evaluación es percibido como adverso; b) el riesgo personal es muy alto; c) no parecen viables otras alternativas. Además, el análisis coste-beneficio está probablemente influido por la estimación de las habilidades del paciente para triunfar, mientras los datos sugieren que los simuladores sobre-estiman ampliamente su capacidad para fingir sin ser detectados. Los datos descriptivos de este modelo se ven apoyados por datos de prevalencia más altos en situaciones adversas o cuando el riesgo personal es particularmente alto. El DSM-IV recoge este modelo al asumir que la simulación puede representar una conducta adaptativa, como cuando un prisionero de guerra finge alguna

enfermedad. De todos modos, debemos tener siempre presente que no podemos asumir que una persona con la motivación potencial para simular lo hará realmente.

4. Criterios diagnósticos de la simulación

Es sorprendente que no exista un acuerdo sobre los criterios diagnósticos aplicables para determinar la probabilidad de simulación, ya que además de facilitar la práctica clínica y la comunicación profesional, dichos criterios facilitarían enormemente la investigación sistemática de la simulación (Slick, Sherman e Iverson, 1999). De hecho, como veremos más adelante, la falta de criterios es uno de los puntos más débiles en la investigación de la simulación.

El DSM-IV sugiere que debemos considerarla en los siguientes casos:

- 1) Cuando el sujeto es referido por un abogado para su evaluación.
- 2) Cuando existe una discrepancia considerable entre las quejas referidas y los hallazgos objetivos de la evaluación.
- 3) Si se aprecia una falta de cooperación durante la evaluación y a la hora de cumplir los tratamientos prescritos.
- 4) Ante la presencia de trastorno antisocial de la personalidad.

Sin embargo, los criterios diagnósticos del DSM-IV adolecen de muchos problemas (recordemos las críticas de Rogers vistas anteriormente), y no son lo suficientemente específicos y formales, por lo que distintos autores han propuesto criterios diagnósticos alternativos. Así, Greiffenstein, Baker y Gola (1994) han propuesto un conjunto de cuatro criterios para el diagnóstico de simulación diseñado específicamente para los pacientes con síndrome postconmocional en procesos de litigio:

- 1) Una ejecución improbablemente baja (por ejemplo, 4 Sd) en dos o más medidas neuropsicológicas.

- 2) Impedimento para desarrollar un rol social importante (por ejemplo, un ama de casa sin amnesia post-traumática que alega que su marido e hijos tienen que hacer todas las tareas domésticas debido a sus problemas de confusión y secuenciación).
- 3) Contradicción entre fuentes colaterales y la historia de síntomas (por ejemplo, un paciente que afirma haber estado en coma 2 días, pero en la hoja de registro del hospital aparece que dicha pérdida de conciencia fue muy breve).
- 4) Pérdida de memoria remota (por ejemplo, un paciente que se queja de haber olvidado leer y enroscar una bombilla).

Otro ejemplo de criterios es el propuesto por Prigatano, Smason, Lamb y Bortz (1997), que proponen para el diagnóstico de simulación:

- 1) que las quejas subjetivas del daño neuropsicológico del paciente excedan de lejos lo esperado dada su historia médica
- 2) que el nivel y patrón de ejecución en los test neuropsicológicos sea muy improbable dada la historia del paciente.

Faust y Ackley (1998) analizan diversas formas de inexactitud intencional de la información originada por el paciente:

- Pobre esfuerzo en la realización de los test.
- Fabricación de síntomas
- Exageración de síntomas
- Encubrir intencionalmente o engañar sobre su causalidad
- Presentar una línea base falsa
- No reconocer los puntos fuertes, habilidades positivas o áreas positivas de funcionamiento.

Teichner y Wagner (2004) basan sus criterios para el establecimiento de la

simulación en los siguientes supuestos:

- 1) El paciente está en un contexto médico-legal (litigio, petición de incapacidad...).
- 2) Existen incentivos identificables para simular.
- 3) Los síntomas no tienen sentido médico, neuropsicológico o psicológico.
- 4) La incapacidad reclamada excede los hallazgos objetivos.
- 5) Falta de cooperación durante la evaluación.
- 6) Se observan inconsistencias entre quejas y conducta del paciente.
- 7) Aparecen contradicciones entre quejas subjetivas y datos médicos.

Otro ejemplo de criterios diagnósticos son los propuestos por Rogers (1990) referidos a la simulación de trastornos psiquiátricos.

Sin embargo, a Slick, Sherman e Iverson (1999), les debemos los criterios más sistemáticos y específicos publicados hasta la fecha sobre la simulación de déficits cognitivos. En los últimos años, sus criterios han sido con diferencia los más empleados por las distintas investigaciones, y además poseen una gran utilidad clínica, por eso los explicaremos con detenimiento. Estos autores parten de una definición de simulación de disfunciones neurocognitivas basada en la fabricación o exageración volitiva de disfunción cognitiva, por un lado, y ganancias sustanciales externas por otro. Las categorías diagnósticas que estos autores proponen para la simulación de disfunciones neurocognitivas son: definitiva, probable y posible.

Un diagnóstico definitivo de simulación conlleva una evidencia clara de exageración o fabricación volitiva de daño cognitivo, en ausencia de explicaciones alternativas plausibles. Sus criterios diagnósticos son:

1. Presencia de incentivos externos sustanciales
2. Definitivo sesgo de respuesta negativo, entendido como la ejecución por debajo

del azar ($p < .05$) en una o más medidas de función cognitiva de elección forzosa.

3. La conducta del sujeto no puede ser explicada por factores psiquiátricos, neurológicos o del desarrollo.

Para el diagnóstico de simulación probable se mantienen los criterios 1 y 3, pero el criterio 2 está basado en una evidencia menor. Así, los autores proponen que deben cumplirse dos o más de los siguientes tipos de evidencia:

- A) Probable sesgo de respuesta: la ejecución es consistente con simulación en uno o más de los test psicométricos o índices bien validados.
- B) Discrepancia entre los datos de los test y los patrones de funcionamiento cerebral.
- C) Discrepancia entre los datos de dos o más test neuropsicológicos de un dominio y la conducta observada.
- D) Discrepancia entre los datos de dos o más test neuropsicológicos de un dominio e informes colaterales fiables.
- E) Discrepancia entre los datos de dos o más test estandarizados de función cognitiva para un dominio y la historia antecedente documentada.

De no cumplirse dos de los tipos de evidencia anteriores, sino que se cumple sólo uno de ellos, es además necesario uno o más de los siguientes tipos de evidencia:

- a) La historia referida es discrepante con la historia documentada.
- b) Los síntomas referidos son discrepantes con los patrones conocidos de funcionamiento cerebral.
- c) Los síntomas referidos son discrepantes con las observaciones conductuales.
- d) Los síntomas referidos son discrepantes con la información obtenida de informadores colaterales.

- e) Evidencia de exageración o fabricación de disfunción psicológica. Por ejemplo, indicada por escalas o índices bien validados de medidas de ajuste psicológico (por ej. MMPI-2).

Para el diagnóstico de simulación posible, además del criterio 1 (presencia de incentivos externos sustanciales), para el criterio 2 han de cumplirse uno o más de los criterios a-e, además de cumplirse el tercer criterio (la conducta del sujeto no puede ser explicada por factores psiquiátricos, neurológicos o del desarrollo). Otra posibilidad de simulación posible se produce cuando se cumplen los criterios para la simulación definitiva o probable excepto el criterio 3 (la conducta del sujeto no puede ser explicada por factores psiquiátricos, neurológicos o del desarrollo). En este último caso, han de especificarse las etiologías alternativas que no pueden ser excluidas (ver Tabla 3).

5. La investigación de la simulación:

Quizá el principal problema de la investigación de la simulación es que es excepcionalmente poco frecuente que alguien admita que está simulando. Concretamente, los investigadores se han valido de dos aproximaciones para intentar solventar este obstáculo: el empleo de personas a las que les pedimos que simulen un daño (análogos), y el empleo de pacientes que presentan un elevado riesgo de exagerar sus déficits o que cumplen ciertos criterios de déficits inconsistentes o no creíbles (Lezack, Howieson y Loring, 2004).

De estas dos aproximaciones se derivan cuatro tipos de diseños de investigación en el campo de la simulación (Rogers, 1997^b): estudio de casos, estudios de prevalencia diferencial, diseños de grupos conocidos y estudio de análogos.

TABLA 3. Criterios diagnósticos para la simulación de daño neurocognitivo, adaptado de Slick, Sherman e Iverson (1999)

Simulación de disfunción neuropsicológica DEFINITIVA	Simulación de disfunción neuropsicológica PROBABLE	Simulación de disfunción neuropsicológica POSIBLE
<p>1. Presencia de incentivos externos sustanciales</p> <p>2. Definitivo sesgo de respuesta negativo, entendido como la ejecución por debajo del azar ($p < .05$) en una o más medidas de función cognitiva de elección forzada.</p> <p>3. La conducta del sujeto no puede ser explicada por factores psiquiátricos, neurológicos o del desarrollo.</p>	<p>1. Presencia de incentivos externos sustanciales</p> <p>2. Dos o más tipos de evidencia de los test neuropsicológicos (Criterio A)</p> <p>o</p> <p>un tipo de evidencia de los test neuropsicológicos y uno o más tipos de evidencia de lo referido por el sujeto (Criterio B).</p> <p>3. La conducta del sujeto no puede ser explicada por factores psiquiátricos, neurológicos o del desarrollo.</p>	<p>1. Presencia de incentivos externos sustanciales</p> <p>2. Evidencia de lo referido por el sujeto (Criterio B).</p> <p>3. La conducta del sujeto en el criterio B no puede ser explicada por factores psiquiátricos, neurológicos o del desarrollo</p> <p>o</p> <p>de no poder descartar alguna de estas etiologías, ha de ser especificada.</p>
Criterio A		
<p>A) Probable sesgo de respuesta: la ejecución es consistente con simulación en uno o más de los test psicométricos o índices bien validados.</p> <p>B) Discrepancia entre los datos de los test y los patrones de funcionamiento cerebral.</p> <p>C) Discrepancia entre los datos de dos o más test neuropsicológicos de un dominio y la conducta observada.</p> <p>Discrepancia entre los datos de dos o más test neuropsicológicos de un dominio e informes colaterales fiables.</p> <p>E) Discrepancia entre los datos de dos o más test estandarizados de función cognitiva para un dominio y la historia antecedente documentada.</p>		
Criterio B		
<p>a) La historia referida es discrepante con la historia documentada.</p> <p>b) Los síntomas referidos son discrepantes con los patrones conocidos de funcionamiento cerebral.</p> <p>c) Los síntomas referidos son discrepantes con las observaciones conductuales.</p> <p>d) Los síntomas referidos son discrepantes con la información obtenida de informadores colaterales.</p> <p>e) Evidencia de exageración o fabricación de disfunción psicológica. Por ejemplo, indicada por escalas o índices bien validados de medidas de ajuste psicológico (por ej. MMPI-2).</p>		

No nos detendremos en los estudios de casos debido a su bajo rigor experimental, pero hay que mencionar que han tenido su importancia para la demostración inicial de que es posible detectar la simulación (por ejemplo, Hiscock y Hiscock, 1989), pero en la actualidad carecen de importancia. En cuanto a los diseños de prevalencia diferencial, se basan en la comparación de un grupo en el que se conoce que existe una elevada probabilidad de simulación (típicamente sujetos buscadores de compensación) con un grupo en el que no se sospecha que haya simuladores. Siguiendo a Rogers, estos diseños son muy poco útiles, ya que no podemos saber quién está simulando ni cuántos sujetos están simulando en cada grupo, por lo que su validez (tanto interna como externa) es muy limitada.

Los diseños de grupos conocidos se componen de dos fases: el establecimiento de grupos atendiendo a criterios (simuladores y pacientes) y el análisis sistemático de las similitudes y diferencias entre ambos grupos. Lo más complicado de estos diseños es la primera fase, pero desde la publicación de los criterios de Slick y cols. (1999) estos estudios se han visto muy beneficiados y su empleo se ha extendido. Su validez externa es muy excelente, pero su validez interna es limitada (Rogers y Bender, 2003).

Por último, los estudios con análogos típicamente emplean personas sin daño a las que se les dan una serie de instrucciones para que intenten simular un daño de modo creíble en un contexto imaginario de litigio. En los diseños de análogos simuladores, este autor recomienda el empleo de cuatro grupos de sujetos:

- sujetos normales a los que se les pide que simulen (análogos);
- sujetos normales honestos;
- sujetos con el trastorno mental o daño cerebral que estamos estudiando a los que se pide que simulen;

- sujetos honestos con el trastorno mental o daño cerebral que estemos estudiando.

El empleo de grupos de comparación clínicos es esencial en estos estudios (a pesar de que raramente se emplea), puesto que de otro modo no podemos determinar si las diferencias son debidas a la simulación o son genuinas del desorden que estamos estudiando. Los estudios de análogos tienen una validez interna muy buena, pero su validez externa es limitada. Sin embargo, a pesar de esta limitación, los experimentos que emplean análogos han contribuido al desarrollo de nuevas medidas para la detección de simulación, han sugerido el empleo de determinados puntos de corte y perfiles cualitativos de ejecución de ciertos tests que han servido de gran ayuda a la hora de clasificar a los simuladores de la “vida real” (Erdal, 2004).

Además, Rogers ofrece una serie de pautas a seguir en la investigación con análogos para aumentar su generabilidad y rigor experimental:

En primer lugar, las *instrucciones* que se ofrecen a los sujetos en las condiciones experimentales deben cumplir una serie de requisitos:

- deben ser comprensibles para los participantes, quedándoles muy claro qué se espera que hagan;
- deben ser específicas;
- deben ofrecer escenarios con los que los participantes puedan identificarse;
- deben ser relevantes para los individuos (explicándoles los costes para la seguridad social, que repercuten en ellos, o desafiándoles para ver si son capaces de simular en un test);
- deben tener en cuenta la motivación para simular, considerando los incentivos externos tanto positivos como negativos;

- debe indicarse que la simulación ha de ser lo suficientemente creíble para evitar la detección.

En segundo lugar, debemos cuidar la *preparación y participación de los sujetos*, dándoles tiempo e información para preparar sus estrategias.

A la hora de pensar en los *incentivos a ofrecer*, debemos tener en cuenta varias cuestiones:

- la magnitud (el incentivo debe serlo realmente para las personas a las que va dirigido);
- el tipo de incentivo;
- la probabilidad del incentivo (no sabemos qué provoca una mayor motivación, una alta probabilidad de un incentivo pequeño o una probabilidad baja de un incentivo grande);
- los incentivos negativos (para que la condición experimental se aproxime más al mundo real).

Es también aconsejable un *informe posterior* a la participación del sujeto en la investigación que nos permita comprobar el recuerdo de las instrucciones, la comprensión de las mismas y la implicación de los participantes. También podemos comprobar aquí la comprensión y recuerdo de los materiales proporcionados, la motivación de los sujetos para simular, o los límites en la participación. Este informe es muy útil a la hora de eliminar de la investigación a los sujetos que no se han implicado lo suficiente en el estudio.

La mejor validación de las medidas de simulación sería una combinación de estudios de análogos y grupos conocidos (Rogers y Bender, 2003), ya que así se maximiza tanto la validez interna (grupos de análogos) como la externa (comparación de grupos conocidos).

6. Problemas metodológicos:

El primer problema que nos encontramos en el estudio de la simulación es el de establecer su **porcentaje de base** (*base rate*). Este término hace referencia a la frecuencia con la que algo ocurre. Así, si 15 de cada cien personas en procesos de litigio son simuladoras, el porcentaje de base de la simulación será del 15%. Para comprender la importancia de este concepto debemos aclarar primero 4 términos:

La *sensibilidad* hace referencia a la probabilidad de que un signo (por ejemplo, una puntuación indicativa de simulación en un test neuropsicológico) sea positivo cuando un trastorno (en este caso simulación) esté presente.

La *especificidad* hace referencia a la probabilidad de que un signo sea negativo cuando el trastorno no está presente.

Estos dos valores no se ven influidos por los porcentajes de base.

El *valor predictivo positivo* indica la probabilidad de que un sujeto tenga un trastorno cuando el signo es positivo (que sea simulador cuando el test lo indica).

El *valor predictivo negativo* es la probabilidad de que el sujeto no tenga el trastorno cuando el signo es negativo (que no sea simulador cuando el test indica que no lo es).

Lo que diferencia en última instancia al valor predictivo es que su valor se ve afectado directamente por el porcentaje de base. Para comprenderlo mejor, vamos a poner un ejemplo concreto: la probabilidad de que un sujeto esté simulando si obtiene una puntuación de 14 ítems difíciles correctos en el Victoria Symptom Validity Test (una prueba de la que hablaremos más tarde) es del 24% si asumimos un porcentaje de base del 5%, pero la probabilidad de que sea simulador asciende al 86% si el porcentaje de base es del 50%.

Hemos considerado pertinente la aclaración de estos términos porque es frecuente caer en el error de considerar que la sensibilidad y la especificidad son iguales al valor predictivo, cuando obviamente no es así.

Una vez aclarada la importancia del porcentaje de base, ¿cuál es la prevalencia de la simulación? Por desgracia, los datos de los que disponemos son de lo más dispares. Parece existir cierto consenso a la hora de establecer la prevalencia de la simulación entre un 33 y un 60% (Greiffenstein y cols., 1994; Haines y Norris, 1995), aunque Reynolds (1998) indica que entre los pacientes con daño cerebral en procesos de litigio dicha prevalencia se sitúa entre un 1 y un 50%, y Duncan (2005) dice que la simulación de daño cognitivo entre los pacientes forenses psiquiátricos está entre el 13 y el 17%. En los extremos podemos situar a Lezak (1995), que opina que es “muy raro” encontrarse con casos de simulación, y a Greiffenstein y cols., (1994) que sitúan la prevalencia de los pacientes con síndrome post-conmocional (que son los casos más comunes en los procesos de litigio) en un 66%. En medio nos encontramos con todo tipo de posicionamientos: Trueblood y Schmidt (1993) encontraron que el 15% de sus pacientes aparecían con una pobre motivación, Rogers, Sewell y Goldstein (1994) opinan que la simulación se da en el 17% de las evaluaciones forenses; Guilmette, Whelihan, Sparadeo y Buongiorno (1994) encontraron evidencia de una baja motivación en el 18% de los sujetos que esperaban obtener una minusvalía, Constantinou y cols. (2005) hablan de un 30% de simuladores entre los pacientes con TCE-L en litigio, y Binder (1993) la sitúa en el 33% en esta misma población. Gervais, Rohling, Green y Ford (2004) encontraron que un 43% de pacientes con diferentes diagnósticos en búsqueda de compensación eran simuladores, y Youngjohn, Burrows y Erdal (1995) encuentran que la prevalencia de simulación es del 48% en los pacientes con síndrome post-conmocional en procesos de litigio (ver Tabla 4).

Algunos factores que, desde nuestro punto de vista, contribuyen a explicar estas variaciones en la incidencia de la simulación son los siguientes: 1) qué test se han utilizado (test con diferentes sensibilidades darán lugar a líneas base diferentes), 2) qué criterios se siguen para diagnosticar la simulación (una ejecución indicativa de simulación en un test o en dos, necesidad de realizar los test significativamente por debajo del azar o no...), 3) el punto de corte seleccionado para los test (ya que normalmente aparecen varios en la literatura), 4) efectos de la posición de los test (un test de simulación no obtiene los mismos resultados si se administra al principio o al final de la evaluación), 5) las características del grupo evaluado (en procesos de litigio, población neuropsicológica general, población psiquiátrica), 6) el tipo de grupo (análogos simuladores o sospechosos de simulación), 7) el tipo de daño que presentan los pacientes evaluados (traumatismos, ACVs, demencias, esquizofrenia...) y la severidad del mismo, 8) factores dependientes del evaluador (astucia, escepticismo, sesgos...).

A pesar de todas estas diferencias, hoy en día disponemos de dos estudios que nos permiten arrojar un poco de luz al problema de la estimación de los porcentajes de base. El primero de ellos es el de Mittenberg, Patton, Canary y Condit (2002), en el que se analizan 33.531 casos de pacientes que buscan una compensación por los déficits alegados, y que han sido evaluados por los miembros de la American Board of Clinical Neuropsychology. Mediante este método, los autores establecen el porcentaje de base de la simulación para distintos trastornos del siguiente modo: 39% para los pacientes con TCE-L; 35% en casos de fatiga crónica o fibromialgia; 31% en pacientes con dolor crónico; 27% en casos de daño neurotóxico y 22% en los pacientes con daño debido a accidentes eléctricos.

El segundo de los estudios al que queremos hacer referencia es el de Larrabee (2003^a), que analiza 11 estudios relevantes para determinar el porcentaje de base de la simulación en contextos neuropsicológicos con ganancias secundarias potenciales. En el 40% de los casos (548 de un total de 1363) se encontró un déficit en la ejecución sugerente de simulación.

Por tanto, parece que existe evidencia suficiente para establecer que el porcentaje de base de simulación de los TCE-L, que es el trastorno más frecuentemente fingido (Haines y Norris, 1995), se sitúa entorno al 40%. En la Tabla 4 se presenta un resumen de las diferentes prevalencias de base encontradas en los distintos estudios.

Tabla 4: Estimación de la prevalencia de base de la simulación en diferentes estudios

Estudio	Población	Estimación de la prevalencia de simulación
Trueblood y Schmidt (1993)	Pacientes neuropsicológicos	15%
Duncan (2005)	Pacientes forenses psiquiátricos	13-17%
Rogers, Sewell y Goldstein (1994)	Pacientes forenses	17%
Guilmette, Whelihan, Sparadeo y Buongiorno (1994)	Pacientes en búsqueda de compensación	18%
Constantinou y cols. (2005)	Pacientes con TCE-L en litigio	30%
Binder (1993)	Pacientes con TCE-L en procesos de litigio	33%
Mittenberg, Patton, Canyock y Condit (2002)	Pacientes en búsqueda de compensación con TCE-L	39%
Larrabee (2003 ^a)	Pacientes en búsqueda de compensación	40%
Gervais, Rohling, Green y Ford (2004)	Pacientes en búsqueda de compensación	43%
Youngjohn, Burrows y Erdal (1995)	Pacientes con SPC en procesos de litigio	48%
Greiffenstein y cols., (1994)	Pacientes con SPC	66%

Nota: TCE-L= traumatismo craneoencefálico leve; SPC= síndrome post-conmocional

Otro problema directamente vinculado al del porcentaje de base es que en el estudio de la simulación se carece de un criterio externo o “**gold estándar**” a la hora de designar a una persona como simuladora. Durante mucho tiempo se ha tomado como gold estándar un criterio conservador, que requiere la existencia de una ejecución por debajo del azar para hablar de simulación. No nos detendremos en este punto porque será tratado debidamente en el siguiente capítulo, sin embargo, hemos de puntualizar que numerosos estudios han demostrado que este criterio no se ha mostrado lo suficientemente sensible (Bender y Rogers, 2004; Gervais y cols., 2004; Guilmette, Hart y Giuliano, 1993; Guilmette, Hart, Giuliano y Leininger, 1994; Greiffenstein y cols., 1994; Hiscock, Branham y Hiscock, 1994; Holmquist y Wanlass, 2002; Martin, Bolter, Todd, Gouvier y Nicolls, 1993; Martin, Hayes y Gouvier, 1996; Rose, Hall y Szalda-Petree, 1995; Slick, Hopp, Strauss, Hunter y Pinch, 1994). El problema que subyace al del establecimiento del gold estándar es que es excepcionalmente poco frecuente que un sujeto admita que está simulando, por lo que, como ya hemos explicado, se recurre al empleo de análogos simuladores. El problema obvio consiste en extraer conclusiones de una población que no es la objetivo. En este sentido, es probable que por mucho que intentemos igualar las situaciones su motivación por aparecer como dañados no sea la misma que la de un verdadero simulador (pues los beneficios secundarios que dependen de ello son bien distintos), el miedo a ser detectado también parece difícilmente comparable, así como tampoco serán iguales las consecuencias de la detección, especialmente las negativas (Rogers y Cruise, 1998). Por tanto, su estado emocional y motivacional ya es, de por sí, distinto. Otro aspecto todavía más difícil de igualar es la ocurrencia real de un evento traumático como, por ejemplo, un accidente de tráfico (Haines y Norris, 1995). A pesar de todas estas complicaciones, algunos estudios han puesto de manifiesto que la ejecución de los análogos y los simuladores reales es

comparable en las evaluaciones neuropsicológicas, apoyando así su utilización en la investigación de la simulación (Sweet, Wolfe y cols., 2000).

Otra de las limitaciones del estudio de la simulación es que se ha focalizado casi exclusivamente en un tipo de población: traumatismos craneoencefálicos leves de habla inglesa, por lo que no se sabe prácticamente nada de la simulación en trastornos más severos (una excepción es el trabajo de Bianchini y cols., 2003) con traumatismos moderados y severos) o en poblaciones distintas a la norteamericana (existe algún trabajo publicado con sujetos de habla germana, como el realizado por Merten, Green, Henry, Blaskewitz y Brockhaus (2005) y, hasta donde nuestro conocimiento alcanza, sólo dos en población española: Ramírez, Chirivella-Garrido, Caballero, Ferri-Campos, y Noé-Sebastián, 2004, y Vilar, Gómez-Río, Verdejo-García, Llamas-Elvira, y Pérez-García, In Press.

Otras limitaciones importantes de los estudios de simulación son los siguientes: no suelen someterse a validaciones cruzadas, no suelen evaluar la efectividad de los test de simulación dentro de una batería neuropsicológica y no comparan directamente la efectividad de test individuales con otros tests y su efectividad en combinación (Inman y Berry, 2002).

7. Problemas éticos de la detección de simulación.

Uno de los múltiples problemas éticos que plantea la detección de la simulación hace referencia a advertir o no al paciente sobre la utilización de índices de simulación dentro de la evaluación. Youngjohn, Lees-Haley y Binder (1999) no están de acuerdo con los autores que sugieren que proveer a los sujetos con dichas advertencias reduce las conductas simuladoras, ya que opinan que su resultado es el de hacer a los simuladores más sutiles y por tanto más creíbles y difíciles de detectar, disminuyendo

por tanto la validez de las evaluaciones neuropsicológicas forenses. Debido a esto, proponen como alternativa preguntar a los sujetos en las evaluaciones forenses si hay algo ajeno al daño que limite su capacidad para emplear todo su conocimiento y poner todo su esfuerzo a la hora de realizar los tests, preguntarles si están de acuerdo en poner todo su esfuerzo en la ejecución de las tareas y que los evaluadores ayuden a mantener una elevada motivación. Estas recomendaciones son consistentes con la práctica neuropsicológica aceptada y permiten mantener un buen nivel de esfuerzo sin reducir la seguridad de las medidas de simulación. Sin embargo, en un informe de la NAN publicado recientemente (Bush y NAN Policy & Planning Committee, 2005), en el ejemplo de consentimiento informado que proponen para que firmen los sujetos en las evaluaciones forenses se especifica que se evaluará la exactitud de las respuestas y el grado de esfuerzo de los sujetos durante la realización de las pruebas, por lo que esta parece la directriz ética a seguir en las evaluaciones neuropsicológicas. De todos modos, en el siguiente capítulo examinaremos este punto con más detenimiento, haciendo una revisión de la literatura que estudia los efectos de la advertencia sobre los resultados en los test de simulación (ver Capítulo 4, apartado 5.1: El entrenamiento).

Otra cuestión relacionada con la ética profesional es la de dar instrucciones engañosas, diciéndole a los evaluados, por ejemplo, que una prueba sencilla tiene un grado elevado de complejidad. Esto suele tenerse muy en cuenta en todos los test de simulación, empleando expresiones del tipo: "...podrías encontrar el próximo bloque más difícil...", evitando las afirmaciones falsas.

Un aspecto muy importante es el del papel del evaluador en contextos adversos. Si bien es aconsejable mantener el *rapport*, es necesario extremar las precauciones para no "modelar" el engaño del paciente, de modo que este aproxime sus resultados hacia el terreno de la simulación como consecuencia de las creencias del examinador. Como

puntualiza Rogers, sería irónico acusar a alguien de simular en respuesta a nuestro engaño (Rogers y Cavanaugh, 1983).

En definitiva, el papel del evaluador ha de ser objetivo. Para ello, han de evitarse asunciones moralistas, del tipo “nunca se debe mentir”. Además, el evaluador no ha de dejarse llevar por su propia visión general del hombre, situada en algún punto entre las dos visiones tradicionales filosóficas basadas o bien en la teoría de Thomas Hobbes en torno a que “el hombre es un lobo para el hombre” o en la opuesta de J.J. Rousseau de que “el hombre es bueno por naturaleza”. Por otro lado, como indican Millis y Putnam (1996) no debemos interpretar la simulación de un sujeto como algo personal (“está engañándose”). Para evitar este tipo de sesgos del evaluador podemos adoptar la perspectiva “adaptativa” de Rogers (1990, 1997^a) que ha sido explicada anteriormente, de modo que intentemos comprender la conducta simuladora del individuo como resultado de su interpretación de que esa es la mejor alternativa de las que dispone para solucionar la situación en la que se encuentra.

El término “simulador” se considera generalmente peyorativo, y su uso es controvertido. Algunos clínicos pueden ser muy reticentes al empleo de este término, a pesar de que posean una fuerte evidencia que lo apoye (Iverson y Binder, 2000). Esto es debido a las indudables consecuencias nefastas que puede tener una incorrecta clasificación de un individuo como simulador, tanto médicas como legales, económicas y sociales. Los profesionales parecen sentirse mucho más cómodos a la hora de emplear etiquetas como “esquizofrenia”, a pesar de que parecen igualmente peyorativas y sus consecuencias similarmente negativas. Si bien debemos extremar las precauciones de su uso, más en un contexto médico-legal que en uno investigador, en el informe ha de quedar bien claro el nivel de esfuerzo y la validez de los resultados de las pruebas. Es extremadamente difícil diferenciar si los objetivos que mueven la conducta de alguien

son exclusivamente internos o externos, por lo que coincidimos con la opinión de Millis y Putnam (1996) de que puede emplearse el término simulación en la investigación para referirse a la exageración de síntomas a pesar de que puedan estar motivados por un conjunto de factores tanto externos (condición indispensable) como internos. Por otro lado, la opción aplicada más plausible nos parece el empleo de una graduación probabilística como la ya expuesta en los criterios de Slick y cols. (1999). También nos parece útil el empleo de frases descriptivas, como propone Larrabee (2000), del tipo “una probabilidad de simulación del ___% basado en la investigación de ___”. Se opte por una u otra estrategia, siempre ha de evitarse una terminología vaga o imprecisa cuando queramos describir que la ejecución de un sujeto es inválida.

En este capítulo hemos expuesto los principios generales para la comprensión de la simulación del daño neuropsicológico, por lo que a continuación pasaremos a revisar los principales métodos de los que disponemos para su detección.

CAPÍTULO 4
MÉTODOS PARA LA DETECCIÓN
DE LA SIMULACIÓN

1. Introducción

En este capítulo comenzaremos con una breve descripción de los principales métodos de los que disponemos en la actualidad para detectar simulación: el estudio de inconsistencias, los métodos derivados de las pruebas específicas de simulación, los métodos derivados de la ejecución en pruebas neuropsicológicas estándar, el análisis de los síntomas referidos por el paciente y los métodos psicofisiológicos y de neuroimagen.

Además, en este capítulo profundizaremos en algunos de estos métodos, especialmente en la evaluación de los síntomas referidos por el paciente y en el estudio psicofisiológico del engaño, abordando los métodos derivados de las pruebas específicas de simulación y de la ejecución en pruebas neuropsicológicas en profundidad en el capítulo siguiente.

Para finalizar, repasaremos algunas de las variables que influyen en la capacidad para eludir la detección de los sujetos.

2. Métodos de detección de la simulación

Considerando el trabajo de diferentes autores (Bender y Rogers, 2004; Bush, Ruff y cols., 2005; Demakis, 1999; Haines y Norris, 1995; Hilsabeck y Gouvier, 2005; Iverson y Binder, 2000; Larrabee, 2000, 2005^b; Meyers y Volbretch, 2003; O'Bryant, Duff, Fisher y McCaffrey, 2004; Gutiérrez, & Gur, 1998; Slick y cols., 1999; Strauss y cols., 2000; Strauss y cols., 2002; Willison y Tombaugh, 2006), podemos considerar como los principales métodos de detección de simulación los siguientes:

2.1. El estudio de las inconsistencias:

Existe consenso a la hora de señalar la inconsistencia como el sello distintivo de la simulación. Larrabee (2000) lo expresa diciendo que la clave en la evaluación de la simulación es que todo debe tener “sentido neuropsicológico”, y Reynolds (1998) opina

que esta tiene tres componentes: congruencia, congruencia y congruencia. Dichas inconsistencias pueden darse:

- Entre la historia referida y la documentada
- Entre los síntomas, es decir, inconsistencia interna en la presentación de los síntomas referidos por el paciente, bien porque no tengan sentido al no corresponderse con ningún síndrome conocido como por ser altamente inusuales y no coincidir con los patrones conocidos de funcionamiento cerebral
- Entre los síntomas referidos y la conducta observada
- Entre los síntomas referidos y la información obtenida de informadores colaterales fiables
- Entre los síntomas y el nivel de ejecución en los tests psicométricos
- Entre las puntuaciones obtenidas y las esperadas por la severidad del daño documentado
- Entre las respuestas de los tests y la conducta observada o las actividades de la vida cotidiana
- Entre los resultados de los tests e informes colaterales fiables
- Entre las habilidades premórbidas y las medidas después del daño
- Entre dominios neuropsicológicos
- Entre evaluaciones repetidas

Este último punto es sobre el que existe más literatura (Reitan y Wolfson, 1995^a, 1997; Sewick y Sobota, 1996). Las diferencias en la consistencia intraindividual han sido empleadas con éxito para detectar grupos de simuladores con la batería Halstead-Reitan (Reitan y Wolfson, 1995^a, 1997) y con el Victoria Symptom Validity Test (VSVT; Strauss y cols., 2000). También Strauss y cols. (2002) encontraron las

inconsistencias a través de evaluaciones repetidas muy útiles en la detección de la simulación, especialmente con el VSVT, pero también en el Dot Counting y en el subtest dígitos del WAIS-III. Sin embargo, Demakis (1999) realizó un estudio empleando medidas de fluencia verbal y no verbal en el que advirtió que los simuladores obtenían un elevado nivel de consistencia inter e intra test, muy similar a la de los sujetos controles. A pesar de ello, al realizar evaluaciones cualitativas los simuladores presentaban una ejecución más inconsistente (un patrón de recuerdo más inconsistente que el de los controles a través de los cinco ensayos de aprendizaje del CVLT). Por otro lado, las inconsistencias parecen el patrón típico en los pacientes con daño cerebral leve (Bleiberg, Garmoe, Halpern, Reeves y Nadler, 1997). Así pues, a pesar de mostrarse como una estrategia prometedora, las inconsistencias inter-evaluación han de tomarse con precaución en determinadas poblaciones, y parecen no ser igual de efectivas con las diferentes pruebas, mostrando una superioridad las pruebas específicas para la detección de simulación.

Para terminar con este método de detección podemos concluir que, a pesar del consenso en cuanto a la importancia de las inconsistencias, existen realmente pocas investigaciones que las estudien, basándose la mayoría de los análisis en la opinión clínica del profesional.

2.2. Métodos derivados de las pruebas específicamente diseñadas para la evaluación de la simulación

Podemos diferenciar tres métodos fundamentales:

- Efecto suelo: Se refiere a fallar tareas que incluso individuos severamente dañados son capaces de realizar. La medida más extendida que sigue este método es el test de los 15-ítems de Rey. El problema de este método es su baja sensibilidad, aunque también plantea problemas su especificidad.

- Producción de respuestas inusuales: Es, quizá, el método menos estudiado dentro de los específicos. Las tareas se basan en habilidades muy sobreaprendidas (como reconocer letras o contar puntos) y por tanto preservadas en pacientes con daño real y que los simuladores fallan. Algunas investigaciones apoyan su utilidad en la detección de simulación (Boone y cols., 2000; Boone, Lu y Herzberg, 2002; Boone, Lu, Back y cols., 2002).
- El método de elección forzosa o *Symptom Validity Testing*: Es el más extensamente empleado y estudiado en la simulación de déficits cognitivos (Gervais y cols., 2004). En este paradigma es presentado un estímulo objetivo seguido por la presentación de dos estímulos entre los que el participante ha de seleccionar el objetivo. Son una excepción entre los métodos de detección de la simulación por su elevada sensibilidad (Slick y cols., 1999; Willison y Tombaugh, 2006), y su uso está aconsejado por la NAN siempre que sea necesario evaluar el esfuerzo del paciente (Bush, Ruff y cols., 2005) por lo que su contribución a las evaluaciones neuropsicológicas forenses es muy valiosa (O'Bryant y cols., 2004; Tombaugh, 1996).

2.3. Métodos derivados de la ejecución en pruebas neuropsicológicas estándar

El segundo de los métodos se basa en la detección de la simulación a partir del empleo de tests o baterías comúnmente empleadas para evaluar los distintos dominios neurocognitivos. En este sentido se pueden diferenciar dos líneas fundamentales: la que se basa en el establecimiento de *puntos de corte* con los que comparar el nivel de ejecución del sujeto, y la que se basa en los *patrones de ejecución* de los sujetos, que parece el más efectivo en cuanto a tests neurocognitivos tradicionales se refiere (Slick y cols., 1999; Meyers y Volbretch, 2003). Entre los patrones de ejecución útiles para detectar la simulación podemos destacar:

- El análisis de la ejecución del sujeto en aspectos sencillos en los que es muy improbable detectar errores (ejecución en ítems suelo)
- Comparación en la ejecución de ítems fáciles y difíciles o evaluación de las curvas de ejecución a través de múltiples ítems de diferentes grados de dificultad
- Comparación de las medidas de un dominio específico, tanto dentro de un test como entre distintos test que miden lo mismo
- Evaluación de los efectos de posición serial en tareas de aprendizaje de listas
- Examen del recuerdo, reconocimiento y discriminabilidad en tareas de recuerdo de listas
- Comparación de los índices de atención y memoria y el conocimiento semántico
- Evaluación de la ejecución en tests de memoria implícita
- Examen de la magnitud del error
- Evaluación de la configuración de las habilidades motoras
- Procedimientos estadísticos post-hoc de puntuaciones o funciones discriminantes obtenidas mediante grupos de contraste establecidos como análogos, pacientes no litigantes y probables simuladores

Se puede concluir que, a pesar de que los patrones de ejecución pueden distinguir a los simuladores en algunas pruebas, como la WMS-III (Langeluddecke y Lucas, 2003), la WMS-R (Mittenberg, Azrin, Millsaps y Heilbroner, 1993); WAIS-R (Mittenberg, Theroux-Fichera, Zielinski y Heilbroner, 1995) o la batería Halstead-Reitan (Mittenberg, Rotholc, Russell y Heilbroner, 1996), parece que los índices derivados de los tests neuropsicológicos convencionales no son lo suficientemente efectivos para identificar a los simuladores (Strauss y cols., 2002; van Gorp y cols., 1999; Tenhula y

Sweet, 1996; Trueblood y Binder, 1997), ya que carecen de la suficiente evidencia empírica debido fundamentalmente a la dificultad para el establecimiento de puntuaciones que no se solapen entre los pacientes verdaderos y los simuladores (Hilsabeck y Gouvier, 2005; O'Bryant y cols., 2004).

2.4. Análisis de los síntomas referidos por el paciente:

Se han desarrollado índices y pruebas para detectar la fabricación o exageración de los síntomas psicológicos, psiquiátricos y cognitivos referidos por el paciente. La prueba más extensamente investigada con este propósito es el MMPI, de la que se han derivado numerosos índices (ver apartado 3).

2.5. Otros

Otro tipo de métodos menos empleados y estudiados en el campo de la detección de simulación hacen referencia al empleo de medidas psicofisiológicas (especialmente el P300) o a los estudios de neuroimagen, como el llevado a cabo por Lee y cols. (2005) con resonancia magnética funcional (ver apartado 4).

Es necesario matizar que esta diferenciación de métodos es más teórica que práctica, ya que en la actualidad existen pruebas neuropsicológicas tradicionales que se basan en los métodos de las pruebas específicas y viceversa. Además, es frecuente que una misma prueba emplee distintos métodos de detección. Por todo ello, a partir de aquí simplemente distinguiremos entre tipos de pruebas (tradicionales, específicas y sintomatológicas) y no entre estrategias. Por otro lado, si bien el método de elección forzosa es el que posee mayor apoyo empírico, la superioridad de un método sobre otro vendrá determinado por las características del sujeto y de la evaluación, por lo que la estrategia más recomendada es la propuesta por la NAN, esto es, una aproximación multi-metodológica (Bush, Ruff y cols., 2005).

3. La evaluación de los síntomas referidos por el paciente:

El **Minnesota Multiphasic Personality Inventory** (MMPI; Hathaway y McKinley, 1940) y el **Minnesota Multiphasic Personality Inventory-2** (MMPI-2; Butcher, Dahlstrom, Graham, Tellegen y Kaemmer, 1989) son las pruebas más frecuentemente empleadas para evaluar factores psicosociales y de personalidad en los casos de TCE-L. Estos tests son útiles a la hora de detectar la fabricación o exageración de síntomas psicológicos o físicos de los pacientes, pero no cognitivos, ya que los “simuladores neuropsicológicos” no suelen generar perfiles inválidos en estas pruebas (Iverson y Binder, 2000), por lo que no nos detendremos demasiado en este punto. Simplemente, mencionar que existen escalas que miden la inconsistencia en las respuestas (VRIN en el MMPI y TR y CLS en el MMPI-2), así como para averiguar la tendencia a emitir siempre la respuesta “verdadera” o “falsa” (escala TRIN del MMPI 2) que invalidarían la interpretación del test. Por otro lado, estas pruebas también disponen de múltiples escalas para evaluar la exageración o invención de síntomas. En este sentido, las primeras escalas diseñadas para evaluar distorsiones de respuesta en el MMPI fueron *Lie* (L) e *Infrecuency* (F), a las que pronto siguió la escala *Correction* (K). Estas escalas se mantienen en el MMPI-2, en el que se han ido añadiendo nuevos índices como el índice *Obvious-Subtle* (O-S), la diferencia F-K, *Dissimulation* (Ds), *Backpage Infrecuency* (Fb), *Infrecuency-Psycopathology* (Fp), *Fake Bad Scale* (FBS)...

En un estudio meta-analítico, Berry y Butcher (1998) concluyen que la escala F es la que goza de un mayor apoyo en la detección de la sobre-representación de síntomas psicológicos en los casos de traumatismo. También Bagby, Buis y Nicholson (1995) concluyen que la escala F es superior a las demás en la detección de simulación, mientras Lewis, Simcox y Berry, 2002 concluyen que es la escala Fb la que obtiene resultados superiores entre todas las que evalúan validez en el MMPI-2. Sin embargo,

quizá el trabajo más destacable para nuestros propósitos sea el realizado por Larrabee (2003^b), en el que demuestra que la escala FBS es la que mejor discrimina entre simulación y daño cerebral, mientras las escalas F, Fb y Fp se muestran insensibles a la simulación del daño neuropsicológico. También Slick, Hopp, Strauss y Spellacy (1996) encontraron útil la FBS en la detección de simulación de aspectos neuropsicológicos, puesto que correlacionaba significativamente con las medidas del Victoria Symptom Validity Test, y lo hacía en mayor medida que la escala F.

Siguiendo a Lewis y cols. (2002) los problemas fundamentales que presenta la detección de simulación con el MMPI son, por un lado, que no han sido establecidas las escalas ni los puntos de corte óptimos para descubrir la simulación de síntomas, mientras por otro lado emplearlos todos aumenta considerablemente los falsos negativos, por lo que los clínicos deben elegir, de modo bastante arbitrario, entre una variedad de índices y puntos de corte que todavía no han demostrado consistentemente sus niveles de exactitud de clasificación. Además, los intentos para conseguir puntos de corte más sensibles se ven limitados por el considerable solapamiento entre las puntuaciones de simuladores y grupos diagnósticos genuinos.

Otro test frecuentemente empleado en Europa para evaluar psicopatología es el **Millon Clinical Multiaxial Inventory-Third Edition** (MCMI-III; Millon, Davis y Millon, 1997). Esta prueba incorpora cuatro escalas para identificar sesgos de respuesta: la escala V, diseñada para identificar respuestas al azar, y tres índices modificados: las escalas X, Y y Z. Una puntuación elevada en la escala X sería indicativa de personas dispuestas a manifestar sus puntos débiles y características personales embarazosas, mientras que una puntuación baja indica que es reticente a revelar cuestiones o problemas psicológicos. La escala Y pretende detectar un estilo de respuesta minimizador, y la escala Z un estilo exagerador de los problemas psicológicos,

emocionales o somáticos. Estas dos escalas no tienen puntos de corte, pero Millon sugiere que puntuaciones muy altas o muy bajas indicarían sesgos de respuesta.

La investigación actual con el MCMI-III se limita a dos estudios: el de Daubert y Metzler (2000) con población psiquiátrica, en el que las escalas X e Y mostraron una efectividad moderada, y el estudio de Schoenberg, Dorr y Morgan (2003). Estos autores realizan una investigación en la que concluyen que los índices modificados poseen una utilidad clínica muy limitada para diferenciar estudiantes análogos simuladores de pacientes psiquiátricos, concluyendo que el MCMI-III es mínimamente sensible a la simulación, por lo que recomiendan el empleo de otros test, como el MMPI-2.

El **Structured Interview of Reported Symptoms (SIRS;** Rogers, Bagby y Dickens, 1992) es quizá la aproximación objetiva más extensamente validada para la detección de síntomas psicológicos fingidos. Su formato es de entrevista estructurada e incluye una variedad de escalas que incorporan diversas estrategias para la detección del engaño (para una revisión, acudir al libro *Clinical assessment of malingering and deception*, editado por Rogers, 1997)

Los puntos fuertes de la SIRS son que en su investigación se han empleado individuos implicados en procesos forenses, que los resultados tanto de análogos como de sospechosos de simulación convergen para apoyar su utilidad, y que posee porcentajes de clasificación adecuados y bien documentados (Lewis y cols., 2002). Como contrapartida podemos destacar su tiempo de administración (una hora u hora y media), así como la necesidad de un evaluador bien entrenado en su uso para su aplicación (Smith y Burger, 1997).

Hasta donde nuestro conocimiento alcanza, la única prueba que ha sido desarrollada para evaluar la simulación de síntomas cognitivos, además de psiquiátricos, es el **Structured Interview Inventory of Malingered Symptomatology (SIMS;** Smith y

Burger, 1997). El SIMS consta de 75 ítems dicotómicos que evalúan cinco condiciones: baja inteligencia (LI), desórdenes afectivos (AF), daño neurológico (N), psicosis (P) y desórdenes amnésicos (AM). Para detectar la simulación, el SIMS emplea las siguientes estrategias: síntomas bizarros, quejas improbables y respuestas aproximadas.

Smith y Burger (1997) aplicaron la SIMS, las escalas F y K del MMPI y la escala *Faking Bad* del 16PF a 476 estudiantes análogos simuladores. Los resultados indicaron que el SIMS poseía la sensibilidad más alta (95,6%) en la detección de simulación. La puntuación total del SIMS fue la que obtuvo mejores resultados para clasificar a los sujetos como simuladores, mientras sus escalas individuales informan cualitativamente del tipo de síntomas que la persona intenta simular. A pesar de estos buenos resultados, los propios autores advierten que la finalidad del SIMS es la de proporcionar una medida de cribado o la de proporcionar una evidencia confirmatoria de simulación cuando se emplea como parte de una batería, y no la de ser empleado como único instrumento para realizar un diagnóstico de simulación.

La fiabilidad de este instrumento también se ve apoyada por un estudio de Merckelbach y Smith (2003) con una traducción alemana del SIMS en la que también emplearon estudiantes análogos simuladores y pacientes psiquiátricos, obteniendo una sensibilidad del 93% y una especificidad del 98%, a pesar de la baja incidencia de simulación en la muestra (19%).

En definitiva, de los estudios que emplean la SIMS podemos concluir que parece una adecuada herramienta de cribado en simulación, permitiendo descartar la misma o la necesidad de una evaluación más exhaustiva sin emplear un tiempo excesivo. Aún así, es necesario el apoyo de investigaciones con grupos más válidos ecológicamente (sospechosos de simulación, litigio...) para llegar a conclusiones definitivas.

4. El estudio psicofisiológico del engaño

Dado que la ejecución en los tests neuropsicológicos está sujeta a la modificación consciente de las respuestas que el sujeto puede hacer, vamos a intentar determinar la aportación que pueden hacer a la detección de la simulación las medidas psicofisiológicas, que son menos susceptibles de ser controladas por el sujeto y por tanto más objetivas. Compartimos la premisa de Vila (1996) según la cual “Detectar el engaño es detectar la intencionalidad de ocultar información deliberadamente. No existen máquinas de la verdad. A lo sumo, existen máquinas del engaño” (pg.225).

4.1. El polígrafo

La prueba del polígrafo es la aplicación psicofisiológica más extendida en la actualidad, a pesar de la escasa investigación científica existente en este campo. Existen defensores y detractores acérrimos de esta técnica, pero ambas partes carecen, generalmente, de teorías científicas que los respalden. Nuestro propósito es explicar los distintos modos de aplicación del polígrafo y explicar las conclusiones a las que ha llegado Iacono (2000) después de estudiar toda la literatura científica existente sobre el polígrafo.

Indefectiblemente unido al término polígrafo está el de la “máquina de la verdad”. Su existencia y uso es muy conocido por la población general (de hecho, hace algún tiempo un programa de televisión sometía a sus participantes a la máquina de la verdad para comprobar si eran infieles a sus parejas), en especial en EEUU, donde para conseguir o mantener un puesto de trabajo es muy habitual la utilización del polígrafo como método de selección, si bien no suele admitirse como prueba para la determinación de una sentencia en los juicios.

Tres son los principales métodos de aplicación del polígrafo: el test de preguntas relevantes-irrelevantes, el test de las preguntas control y el test del conocimiento culpable.

La primera técnica poligráfica empleada que incluía comparación de respuestas fue el test relevante-irrelevante (TRI). Este método incluye dos tipos de preguntas: las *preguntas relevantes* son aquellas determinantes para la implicación del sospechoso en el delito en cuestión, y las *irrelevantes* son preguntas del tipo: “¿Su nombre es Juan?” “¿Estamos en Granada?”. Se espera que los acusados respondan a estas preguntas correctamente, y sirven para establecer una base fisiológica con la que comparar las respuestas a las preguntas relevantes. Si las preguntas relevantes elicitan respuestas mayores que las irrelevantes, se interpreta que el individuo no está respondiendo honestamente. Si las respuestas elicidadas por ambas preguntas son equivalentes el sujeto estaría diciendo la verdad. Esta es la técnica más empleada en Norteamérica en la selección de personal. El problema obvio de esta técnica es que no puede ser comparable el impacto emocional de las preguntas control con el de las relevantes, ya que incluso una persona completamente inocente responderá fisiológicamente de forma muy distinta si le preguntan su nombre y si la pregunta es: “¿Mató usted a su vecino?”. Intuitivamente, este no parece un modo muy adecuado de determinar la veracidad de un testimonio.

El test de las preguntas control (TPC) fue ideado para intentar superar este problema. Su propuesta es la de comparar las preguntas relevantes con preguntas control diseñadas para provocar una respuesta emocional importante. Son preguntas relacionadas con el delito, pero vinculadas a actos que supuestamente todos hemos realizado alguna vez en nuestras vidas. Por ejemplo, si el caso versa sobre un robo, una pregunta control sería: “¿Ha robado usted a alguien o ha cogido algo que no es suyo

alguna vez en su vida?”. La teoría del TPC se basa en que los sujetos culpables serán más reactivos a las preguntas relevantes (“¿Robó usted el dinero del banco?”) que a las preguntas control, mientras que el patrón de los inocentes será el inverso, ya que se sentirán más implicados en sus respuestas a las preguntas control.

Una importante variación a este método consiste en el empleo de una “prueba de estimulación” previa. La finalidad de dicha prueba es la de convencer al evaluado de la infalibilidad del polígrafo. Para ello existen diversos procedimientos, como hacer que el sujeto piense en un número del 1 al 9. El examinador irá nombrando todos los números, y el sujeto deberá ir contestando “No” hasta que el número sea adivinado por la inspección de las respuestas psicofisiológicas del individuo. Otra variante consiste en que el sujeto elija una carta de una baraja, e ir haciendo preguntas “¿La carta es de oros?” “¿Es una figura?”... Hasta que el evaluador pueda indicar la carta seleccionada. Debido a que un error en estas “pruebas de estimulación” podría tener efectos nefastos (el sujeto comprobaría que el polígrafo puede fallar y no detectar el engaño, y por tanto podría no generar el patrón esperado ante las respuestas relevantes), algunos técnicos se aseguran dejando la carta elegida boca arriba entre ambos o marcando la baraja, e incluso pueden amplificar la señal del polígrafo ante el engaño para mostrársela al sujeto. La finalidad es siempre la misma: que el evaluado quede convencido de que el polígrafo detectará sus mentiras.

El tercer y último método en el que nos detendremos es el test del conocimiento culpable (TCC). En esta técnica se hace una pregunta al sospechoso para la que sólo la policía y el culpable conocen la respuesta, y se ofrecen sucesivamente múltiples respuestas plausibles. La teoría es que la respuesta correcta provocará una reacción psicofisiológica más fuerte que las demás en el culpable, mientras los inocentes tendrán reacciones similares ante todas las respuestas, sea cual sea la que elijan como válida.

Este método es el menos aplicado en el contexto forense. Un ejemplo de pregunta del TCC sería:

“¿Con qué se amenazó a la familia del secuestrado en caso de no cumplir las condiciones?

- Cortarle una oreja si se daba parte a la policía
- Matarlo de un tiro
- Dejarlo morir de hambre
- Torturarlo”

El problema fundamental que presentan todas estas técnicas es que no sabemos qué es en realidad lo que está midiendo el polígrafo: verdad vs. mentira o simplemente reacciones emocionales como el miedo a ser detectado.

Habitualmente, la actividad registrada por el polígrafo incluye la respuesta electrodermal (ya sea resistencia o conductancia), la respiración y los cambios en la presión sanguínea. En algunos casos se mide también la actividad vasomotora digital.

La aplicación del polígrafo comienza con una entrevista pre-test, seguida por la prueba poligráfica y un interrogatorio final en el que se busca la confesión del sujeto. Es necesario diferenciar entre test “adversos”, que son los administrados por las fuerzas de la ley, y test “amigables”, que son aquellos realizados a petición del abogado defensor de un sospechoso. Esta distinción es importante, porque la poca investigación existente sobre el polígrafo se ha basado en los test “adversos”, por lo que no sabemos prácticamente nada de los resultados de los test “amigables”.

En cuanto a la fiabilidad test-retest del polígrafo, Iacono (2000) opina que no existen estudios que nos permitan llegar a una conclusión. Sin embargo, la literatura parece unánime a la hora de establecer la fiabilidad interjueces muy elevada. En lo referente a la validez, hay que tener en cuenta que, igual que en el estudio de la

simulación, los estudios se dividen entre los llevado a cabo en el laboratorio (o estudios de análogos) y los estudios de campo. En cuanto al primer tipo de estudios, de cuyas limitaciones ya hemos hablado anteriormente, los estudios parecen situar su exactitud entorno a un 88%. Los estudios de campo se basan en el resultado del polígrafo de los casos que más tarde se han confesado culpables o demostrado su inocencia. Este método de estudio excluye sistemáticamente todos los casos potencialmente inválidos (tanto los sujetos que estaban mintiendo y fueron clasificados como sinceros como aquellos individuos sinceros clasificados como mentirosos). De este modo, tanto la sensibilidad como la validez de este tipo de estudios están inflados en un grado que desconocemos (Iacono, 2000).

En cuanto al test de preguntas relevantes-irrelevantes, no existe hoy en día espacio para su defensa científica, ya que se ha mostrado sistemáticamente sesgado en contra de los individuos inocentes. Los resultados del test del conocimiento culpable no son tan tajantes, y mientras los escépticos de este método sitúan su exactitud a la hora de clasificar inocentes en el 50%, sus defensores dicen que supera el 90%. De modo similar, a la hora de clasificar culpables los defensores indican que su exactitud supera el 95%, mientras sus detractores dicen que no llega al 75%. ¿De dónde proceden estas indiscutibles diferencias? Sin duda, de los criterios de selección. Como hemos visto antes, los estudios de campo están sesgados, puesto que se centran sólo en aquellos de los que se obtiene una confesión posterior. Patrick e Iacono (1991) han realizado el único estudio no sesgado en este sentido, concluyendo que los test psicofisiológicos sólo consiguen una clasificación correcta de los sujetos inocentes ligeramente por encima del azar.

Para terminar, los estudios sobre el test del conocimiento culpable apoyan unánimemente su aplicación mediante los estudios de laboratorio, en los que se

demuestra su clasificación correcta del 100% de los sujetos inocentes y del 85-90% de los culpables. Más complicado es determinar su utilidad real, ya que se deben cumplir una serie de requisitos para poder aplicar esta técnica: -se debe tener la evidencia suficiente sobre el delito para poder realizar los ítems, -la información debe ser conocida por la policía y el culpable, pero no por los sospechosos inocentes, y -los hechos demasiado oscuros para recordar no constituyen buenos ítems. Teniendo en cuenta estos requisitos, puede ser que sólo entorno al 15% de los casos puedan ser ajustables al test del conocimiento culpable.

Todas estas conclusiones de Iacono (2000) se han visto apoyadas tanto por los miembros de la Society for Psychophysiological Research (SPR) como por los miembros de la American Psychological Association (APA).

En los últimos años el polígrafo se ha venido aplicando a la detección de la simulación, y los resultados de estas modernas aplicaciones nos parecen muy interesantes y prometedores. Nos referimos, en concreto, a los estudios psicofisiológicos dirigidos a la detección de la simulación del trastorno por estrés post-traumático y sobre todo a la investigación de la detección de simulación de trastornos neuropsicológicos mediante el análisis del potencial evocado P300.

4.2. La detección psicofisiológica de la simulación del estrés post-traumático

Blanchard y cols. (1996) examinaron la respuesta psicofisiológica de pacientes con trastorno por estrés post-traumático (TEPT) secundario a accidente de tráfico ante la escucha de la grabación del evento realizada por los propios pacientes. De las distintas respuestas analizadas (tasa cardíaca, presión sanguínea y electromiograma frontal), encontraron que la variable que mejor diferenciaba a los pacientes con TEPT, pacientes con TEPT subsindrómico y un grupo control era la tasa cardíaca. En concreto, un punto de corte consistente en un aumento de 2 pulsaciones por minuto ante la escucha de la

grabación del evento obtuvo una sensibilidad del 69,4% y una especificidad del 66,7%. En un experimento similar con veteranos de guerra, Blanchard, Kolb, Pallmeyer y Gerardi (1982) consiguieron clasificar correctamente al 90,9% de los sujetos con TEPT y controles mediante la tasa cardíaca registrada ante la escucha de sonidos de guerra.

En otro estudio perfectamente controlado, Hickling, Blanchard, Mundy & Galovski (2002) entrenaron a un grupo de 6 actores en los síntomas del TEPT y los enviaron para ser evaluados por psicólogos que estaban participando en un estudio sobre el trastorno en cuestión. Ninguno de los actores fue detectado. Sin embargo, cuando se informó a los evaluadores de la existencia de simuladores (pero no de su número), 3 de los 6 simuladores fueron identificados, y cuando se permitió a los evaluadores repasar los datos de la evaluación psicológica y psicofisiológica, la clasificación correcta subió al 91%. No se encontraron diferencias entre el grupo de pacientes reales y el grupo de actores en las medidas psicológicas (si bien hay que aclarar que no administraron índices específicos de simulación, sino cuestionarios y entrevistas clínicas). Sin embargo, sí aparecieron diferencias estadísticamente significativas entre los grupos en dos variables psicofisiológicas: tasa cardíaca y presión sanguínea. Estudiando a los participantes caso a caso, la variación de la tasa cardíaca fue capaz de detectar a 4 de los 6 simuladores.

4.3. La detección de simulación mediante el empleo de potenciales evocados y neuroimagen

Las propuestas más novedosas dentro de la detección de la simulación de déficits cognitivos optan por la combinación de los procedimientos neuropsicológicos, en concreto los test de elección forzada, y las técnicas psicofisiológicas poligráficas del tipo TCC mediante la medición de potenciales evocados (especialmente del P300). En concreto, este tipo de estudios se basan en la idea de que la amplitud del P300 no está

bajo el control de las personas, por lo que ésta seguirá un patrón normal a pesar de que la persona finja un daño cerebral o no recordar cierta información. Es decir, si le presentamos a un sujeto un ítem (por ejemplo una palabra o un número) y posteriormente presentamos ítems distractores y el ítem objetivo, el P300 será mayor ante el ítem anteriormente presentado (ítem viejo o match) que ante los nuevos (o mismatch), aunque el sujeto finja no reconocerlo. Siguiendo esta lógica, Ellwagner, Tenhula, Rosenfeld y Sweet (1999) aplicaron una variante del método de Hiscock y Hiscock (1989) para su uso con potenciales evocados, el P3 Multi-Digit Memory Test, a un grupo de análogos y un grupo control. El P300 elicitada por los ítems “match” fue mayor que para los ítems “mismatch” tanto en el grupo simulador como en el control, a pesar de sus diferencias en el número de respuestas correctas: el grupo análogo contestaba incorrectamente, a pesar de que su P300 estaba indicando que reconocía el ítem presentado. Con este método, los autores consiguieron una clasificación correcta del 90,5% de los participantes. Del mismo modo, Rosenfeld y cols. (1999) encontraron que el 100% de los sujetos controles y el 87% de los análogos obtenían un P300 mayor ante los estímulos “match” que ante los “mismatch”, a pesar de sus diferencias en la ejecución del test. Resultados similares encontraron Tardiff, Barry, Fox y Johnstone (2000) con la aplicación del Warrington Recognition Memory Test.

Por otro lado, este tipo de estudios también han intentado encontrar diferencias entre las respuestas honestas y las fingidas, es decir, encontrar un perfil psicofisiológico de la mentira. La teoría subyacente a este tipo de estudios es que los procesos cognitivos relacionados con la verdad y la mentira son diferentes, y por tanto involucrarán patrones neurales distintos. Siguiendo este razonamiento, Rosenfeld y cols. (1999) encontraron un patrón en la distribución del P300 medido en distintos puntos craneales que no difería en las respuestas intencionadamente erróneas de los análogos para los estímulos

match y mismatch, concluyendo que la conciencia de la acción de fingir se convierte en el proceso dominante en las respuestas fingidas, de modo que la diferencia match-mismatch pierde importancia. Tardiff y cols. (2000) también encontraron diferencias en el P300 de los análogos comparada con la del grupo control. Esta diferencia se refiere a un segundo efecto viejo-nuevo restringido sólo a las ondas de los simuladores. Su aparición es muy temprana (100 milisegundos tras la aparición del estímulo) y su detección es máxima en el área frontal izquierda (F3 y Fz). Esta mayor activación frontal en los simuladores estaría vinculada al mayor grado de función cognitiva vinculada a la estrategia en este grupo.

En un estudio posterior, Tardiff, Barry y Johnstone (2002) aplicaron de nuevo el efecto viejo-nuevo mediante un formato de elección forzosa, pero esta vez todos los sujetos pasaban por dos condiciones: en una debían realizar el test lo mejor que podían y en otra debían simular un daño de memoria. Tanto la puntuación del test como la latencia de respuesta fueron significativamente diferentes en las dos condiciones. En la condición control, las palabras viejas (es decir, los ítems objetivo) obtuvieron ondas más positivas en las regiones frontales, lo cual parece reflejar un juicio de reconocimiento basado en la familiaridad del estímulo. En la condición de simulación, sin embargo, esta diferencia entre las palabras viejas y nuevas apareció antes y sin una distribución topográfica específica, sugiriendo la implicación de múltiples áreas corticales o de áreas cerebrales profundas. Esto posiblemente esté reflejando procesos que incluyen la inhibición de respuesta ante palabras vistas previamente, procesos de planificación para responder de acuerdo a la tarea demandada y con el mantenimiento del cálculo mental de las respuestas previas con el fin de representar un déficit creíble.

Este tipo de especulaciones sobre los posibles correlatos neurales de la simulación de déficits está siendo aclarado por los recientes estudios llevados a cabo

mediante resonancia magnética funcional. Lee y cols. (2005) han determinado en un magnífico estudio que la simulación está vinculada a las áreas prefrontales (incluida el área cingulada anterior) y parietales. Esta activación es invariable para distintos estímulos (pruebas de elección forzosa tanto de palabras como de dígitos), género o lengua materna, lo que demuestra su robustez. La relación de estas áreas con la simulación no es de extrañar si tenemos en cuenta sus funciones. Como estos autores aclaran, la corteza prefrontal está vinculada a funciones como la flexibilidad y el control cognitivo, la toma de decisiones y la regulación de respuestas emocionales, funciones todas ellas necesarias para el engaño. Además, la corteza cingulada anterior contribuye de forma sustancial a las actividades mentales volitivas, esenciales para la simulación. Por último, las regiones parietales (en concreto el giro supramarginal y el lóbulo parietal superior) se activan ante las demandas de la *working memory* o del cálculo mental complejo, activación que estaría vinculada al cálculo de la proporción de respuestas correctas e incorrectas realizada por los simuladores.

5. Variables que influyen en la detección de simulación.

A continuación vamos a exponer algunas variables que pueden modificar los resultados de la detección de la simulación, sea cual sea el método empleado para evaluarla.

5.1. El entrenamiento (“*coaching*”)

La validez de los test de simulación puede verse severamente comprometida por el entrenamiento de los abogados a sus clientes en las técnicas de detección. Este entrenamiento es una realidad constatada en la actualidad, y supone un serio problema para la neuropsicología forense. Ya en 1995 Youngjohn refería el caso de un abogado que admitió en un juicio haber entrenado a su cliente, que alegaba un daño cerebral, en

el PDRT. De modo más sistemático, Wetter y Corrigan (1995) informaban que un tercio de los estudiantes de derecho y la mitad de los abogados encuestados en su trabajo consideraban que debían entrenar a sus clientes en los indicadores de validez de los tests que pudieran administrarse durante una evaluación neuropsicológica forense. Además, <(1997) demuestra que los abogados disponen fácilmente de información sobre los test neuropsicológicos, que dicha información está muy extendida en el campo de la educación penal y que puede ser usada por los abogados para influir sobre la ejecución de sus clientes. En un estudio más reciente se ha puesto de manifiesto que el 75% de los abogados preparan a sus clientes para las evaluaciones neuropsicológicas forenses, siendo la detección de simulación un tópico que revisan frecuentemente. Así, el 29% de los abogados entrenan a sus clientes en el MMPI, el 6% lo hacen en el Portland Digit Recognition Test, y el 2% en el Dot Counting Test, Hiscock Forced Choice Test o Test de los 15-ítems de Rey (Essing, Mittenberg, Petersen, Strauman y Cooper, 2001). Como el paradigma que emplean todas estas pruebas es el de tareas sencillas que incluso pacientes con daño cerebral moderado realizan correctamente, estos autores presuponen que el entrenamiento de los clientes en esta técnica podría ser suficiente para reconocer este tipo de tests y evitar la detección, si bien esta afirmación necesita ser probada.

Además de la que puedan proveer los abogados, otra fuente fundamental de información la constituye Internet. Por este motivo, Bauer y McCaffrey (2006) han explorado la cobertura que existe en Internet de algunos de los tests más empleados en simulación con el buscador Google: el Test of Memory Malinger, el Victoria Symptom Validity Test y el Word Memory Test. De estos, el VSVT es el que ha mostrado tener más información disponible en la red, y además la primera entrada que aparece al introducir sus siglas es un ejemplo de informe completo con el nombre del test, el rango de sus puntuaciones y la clasificación de las mismas. El segundo test en

cantidad de información disponible es el TOMM, del que también se encontraron puntos de corte específicos, y del que menos información se dispone es del WMT, que parece medianamente preservado. La conclusión es que la información disponible para la población general pone en peligro la validez de estos test. Bauer y McCaffrey (2006) proponen también una serie de recomendaciones interesantes para intentar solucionar este problema: que los investigadores indiquen si los sujetos se sitúan por encima o por debajo de los puntos de corte, sin especificar la puntuación; restringir el acceso a revistas o “webs” académicas que contengan información tan sensible; no colgar en Internet ejemplos de informes; ocultar, siempre que sea posible, el nombre de los test de simulación.

Una vez establecido que los sujetos pueden disponer de información sobre los test de simulación, tanto por el entrenamiento de sus abogados como por la vía de Internet, nos queda dar respuesta a la pregunta: ¿Son capaces los simuladores de valerse de esta información y eludir la detección? Para investigar este aspecto, suelen emplearse grupos de análogos a los que se les proporciona distintos tipos de información para determinar cómo esta influye en su ejecución en los tests. Una vez más, las investigaciones arrojan resultados que parecen, en principio, contradictorios.

Por un lado, existen una serie de estudios que ponen de manifiesto que el entrenamiento no produce diferencias a la hora de realizar los tests de simulación. En este sentido, Martin, Hayes y Gouvier (1996) no encontraron diferencias entre un grupo de análogos sin entrenamiento y otro con entrenamiento específico sobre la mejor ejecución en cada uno de los tests administrados para evitar la detección (el Multi-Digit Memory Test, el Dot Counting Test y el subtest del WAIS-R Digit Span Test). Sin embargo, en este trabajo no se facilitan las instrucciones proporcionadas a los grupos,

por lo que no podemos saber hasta qué punto el entrenamiento está basado en información *específica* sobre las estrategias a emplear en las pruebas.

Inman, Vickery, Lamb, Edwards y Smith (1998) no encontraron diferencias entre los análogos sin entrenamiento y los análogos con entrenamiento sobre las consecuencias del daño cerebral en la ejecución de distintos test de simulación de elección forzosa. En la misma línea, Cato, Brewster, Ryan y Giuliano (2002) proporcionaron a un grupo de análogos información sobre los síntomas de los TCE-L y después los dividió del siguiente modo: sujetos sin entrenamiento específico, sujetos con entrenamiento específico sobre los tests y sujetos a los que además de esta información se les proporcionaban ejemplos concretos. Ninguno de los grupos de análogos fue capaz de producir resultados convincentes en el Hiscock Forced Choice Procedure, siendo detectados como simuladores. Tampoco Bender y Rogers (2004) encontraron diferencias en la ejecución del Test of Cognitive Abilities a la hora de diferenciar un grupo de análogos con información sobre el daño cerebral e instrucciones para fingir un daño plausible y otro grupo de análogos a los que se proporcionaban estrategias complejas específicas para la realización del test.

Greub y Suhr (2006) encontraron una ejecución similar en el Letter Memory Test en los análogos sin entrenamiento y en aquellos a los que se advertía de la existencia de pruebas en la batería cuya finalidad es la de detectar su engaño, especificando que son más sencillas de lo que en realidad parecen.

En cuanto a los efectos del entrenamiento en la detección de simulación de pruebas no específicas, Suhr (2002) encontró que el entrenamiento en los síntomas del daño cerebral no mejoraba la ejecución en el AVLT, de modo que los análogos con y sin entrenamiento no pudieron igualar el efecto de posición serial en esta prueba al de los pacientes con daño cerebral, que es similar al de los sujetos normales. Tampoco

Rapport, Farchione, Coleman y Axelrod (1998) pudieron constatar la mejoría de la ejecución de los análogos entrenados en los tests de funciones motoras, sino que, de hecho, la ejecución de los sujetos entrenados fue peor que la de los no entrenados, obteniendo un patrón de respuesta sin sentido fisiológico.

A pesar de todos estos hallazgos, más numerosos son los estudios que encuentran diferencias en la ejecución de los análogos en función de su entrenamiento previo a la ejecución de las pruebas. Suhr y Gunstad (2000) señalan que la simple advertencia a los análogos simuladores influye significativamente en el modo en que éstos ejecutan las tareas de elección forzosa, haciendo más difícil su detección. Frederick y Foster (1991) aplicaron un test de elección forzosa a un grupo de análogos sin entrenamiento y a otro con información específica sobre el test, y la tasa de detección fue mucho mayor en los no entrenados (a pesar de que sólo la eludió el 27% de los entrenados). También Rose, Hall, Szalda-Petree y Bach (1998) encontraron que el grupo de análogos con entrenamiento sobre los síntomas del daño cerebral más un aviso sobre “no hacer los déficits demasiado obvios” eran mejores a la hora de eludir la detección en distintas pruebas de simulación. De hecho, no todas las pruebas se mostraron igual de sensibles al entrenamiento, de modo que el Portland Digit Recognition Test (PDRT) se mostró más resistente que el 21-Item Test, el Nonverbal Forced Choice Test y que el Dot Counting Test a los efectos del entrenamiento. No obstante, otros autores concluyen que advertir sobre el empleo de tests de simulación es suficiente para que los análogos realicen el PDRT significativamente mejor que los sujetos sin entrenamiento (Suhr y Gunstad, 2000) o con entrenamiento sobre los síntomas de los TCE-L (Gunstad y Suhr, 2001). Así, parece que cuando los sujetos conocen empleo de tests específicos las pruebas de elección forzosa son reconocidas como pruebas de simulación. Esta afirmación se ve apoyada por el trabajo de Merten y

cols. (2005), en el que la mera advertencia de la inclusión de tests de simulación fue suficiente para que los sujetos identificaran la finalidad de las pruebas de elección forzosa. En el estudio de otro de los tests típicos de elección forzosa, el TOMM, también Powell, Gfeller, Hendricks y Sharland (2004) constataron que los análogos con información específica sobre el test realizaban la prueba significativamente mejor que los análogos a los que se les proporcionaba sólo entrenamiento en síntomas, si bien el TOMM resultó igual de exitoso a la hora de clasificar a ambos grupos como simuladores. El entrenamiento específico sobre los tests también ha mostrado su efectividad a la hora de eludir la detección de simulación en el Dot Counting Test y en el Test de los 15-ítems de Rey (Cato y cols., 2002), y en pruebas no específicas como el Category Test (Di Carlo, Gfeller y Oliveri, 2000).

¿A qué pueden deberse las diferencias en los resultados? Como indica Erdal (2004) el entrenamiento no se ha empleado de modo consistente en la investigación, por lo que existen numerosas variables que no están controladas en los diferentes estudios y que podrían explicar dichas inconsistencias, como el tipo de información proporcionada a los sujetos. Así, se puede informar sobre los síntomas de un trastorno, sobre los patrones de ejecución típicos de un trastorno, sobre las estrategias de los test de simulación o simplemente proporcionar una advertencia sobre el empleo de técnicas para la detección de simulación (Gunstad y Suhr, 2001). Además, no existe consenso en cuanto a lo que se refieren estos distintos tipos de información, de modo que en algún estudio entienden por advertencia un exhaustivo examen de diversas estrategias concretas para la realización del test (como sucede en el de Bender y Rogers, 2004), mientras en otros se refieren a un grupo de análogos como entrenado cuando simplemente se hace una advertencia sobre la existencia de pruebas de simulación (como en el trabajo de Greub y Suhr, 2006). En este sentido parece recomendable que

los autores incluyan en sus trabajos las instrucciones dadas a los análogos para facilitar la comparación de los resultados de las investigaciones, ya que esta práctica no siempre se sigue. Por otro lado, la cantidad de instrucciones facilitadas y el tiempo dado para estudiarlas puede afectar directamente la capacidad de los sujetos para su asimilación. Además, parece que existen diferencias entre los tests, de modo que no todos son igual de susceptibles al entrenamiento.

A modo de conclusión, parece que el entrenamiento proporcionado por los evaluadores hace que los sujetos adopten estrategias de simulación más sofisticadas. La información específica sobre los test (o sobre test además de síntomas) parece la más efectiva para evitar la detección, mientras el entrenamiento en síntomas parece carecer de utilidad. Esta afirmación se ve respaldada por el hecho de que los pacientes a los que pedimos que simulen no se muestran superiores a los análogos sin trastorno a la hora de eludir la detección (Rees, Tombaugh, Gansler y Moczynski, 1998, Strauss y cols., 2002; Vickery y cols., 2004). Los test de elección forzosa parecen los más susceptibles al entrenamiento, si bien dicho entrenamiento no es suficiente para evitar la detección, es decir, aparecen diferencias estadísticamente significativas en las puntuaciones de los análogos con vs. sin entrenamiento en este tipo de pruebas, pero esta mejora en la ejecución no es suficiente para que los sujetos eviten la detección. Otra variable que parece moderar el efecto que el entrenamiento tiene sobre la ejecución de las pruebas es el de la compensación para simular, ya que parece que cuando se ofrecen incentivos no se encuentran diferencias en la ejecución de los sujetos entrenados y los no entrenados (Erdal, 2004).

Si estas son las conclusiones que se desprenden del entrenamiento a los análogos por parte de los evaluadores, queda pendiente todavía conocer cuál es la utilidad real de la información disponible para la población general a la hora de eludir la detección. En

este sentido, Tan, Slick, Strauss y Hultsch (2002) dieron tiempo a los análogos para que prepararan su papel de pacientes con daño cerebral. El modo de preparación más frecuentemente referido por los sujetos fue Internet, seguido por conversaciones con amigos y familia y la lectura de libros y artículos. A pesar de ello, un 78% de los sujetos fue detectado como simulador con el TOMM, y un 89% con el VSVT y con el WMT. También Ress y cols. (1998) constataron la altísima sensibilidad del TOMM a pesar de dar a los análogos una semana de auto-preparación. Estos hallazgos ponen en tela de juicio las afirmaciones de Bauer y McCaffrey (2006) al cuestionar la validez de estas pruebas por la cantidad de información sobre las mismas disponible en la red. De todos modos, queda pendiente la realización de una investigación en un grupo de sujetos con los *incentivos suficientes* para simular, si bien parece que la utilidad de la información disponible en la red no es demasiada para la población general. Pensemos, por ejemplo, que los individuos no pueden acceder a la información del estudio de Bauer si no son capaces de averiguar primero el nombre o las siglas de las pruebas de simulación. En España probablemente la información disponible en la red carezca todavía más de utilidad, puesto que sólo está disponible en inglés.

Nuestra esperanza es que en un futuro lleguemos a disponer de un número tal de pruebas de simulación bien validadas que el entrenamiento deje de ser efectivo, al ser imposible tener en cuenta todo el rango posible de indicadores de validez de los que dispondrá el evaluador. Hasta ese momento, sin embargo, es necesario seguir trabajando en la construcción de nuevos instrumentos, y la mejor solución disponible hoy en día para atenuar los efectos del entrenamiento se basa en el empleo de múltiples tests e índices empíricamente validados para detectar el esfuerzo insuficiente.

5.2. Otras variables.

Además de la influencia que pueda tener en la ejecución de los tests la información proporcionada por el experimentador, algunas investigaciones han tenido en cuenta el efecto de la **información previa** de los sujetos. En este sentido, parece que el conocimiento sobre el daño cerebral, tanto si ha sido adquirido por haberlo padecido (Vickery y cols., 2004), como por ser un profesional relacionado con el campo (Strauss y cols., 2002) o por haberlo estudiado, no mejora la simulación. De hecho, parece que un mayor conocimiento neuropsicológico empeora la simulación, ya que se tiende a sobreestimar las consecuencias del daño cerebral (Erdal, 2004).

La **escolaridad** parece influir en la simulación, ya que los análogos con una educación media lo hacen peor que aquellos con más años de escolaridad, y también el **tipo de profesión** parece jugar su papel a la hora de crear déficits creíbles, encontrándose diferencias entre médicos y abogados, simulando mejor estos últimos (Schwartz, Gramling, Kerr y Morin, 1998). Cuando afirmamos que la escolaridad influye en la simulación no nos estamos refiriendo a que cuando los sujetos se enfrentan a las pruebas de simulación poniendo su máximo esfuerzo se encuentren diferencias en la ejecución de las pruebas atendiendo a la escolaridad, sino que nos referimos a que, cuando les pedimos a los sujetos que simulen, aquellos con mayor escolaridad lo hacen mejor, es decir, son más capaces de eludir la detección.

Otra variable importante a tener en cuenta es la **inteligencia**, ya que se ha demostrado que las personas más inteligentes son más capaces de evitar la detección con el MMPI-2 (Pelfrey, 2004).

Por otro lado, todavía no está claro cómo los **incentivos** modulan la simulación. Si bien se sabe que la compensación económica juega un papel fundamental en la simulación en la vida real, todavía no se conoce bien su papel en la investigación con

análogos. Erdal (2004) concluye que en los simuladores experimentales la búsqueda de compensación parece ser la motivación que lleva a una simulación más flagrante, en comparación a los controles y a los análogos cuya motivación es la de evitar la culpa de un accidente de tráfico. Hay que puntualizar, sin embargo, que dicha búsqueda de compensación era puramente imaginaria: sólo se pedía a los sujetos que supusieran que iban a juicio y que podían obtener una suculenta suma de dinero si conseguían aparecer como dañados. En estudios en los que se ofrece dinero real no se encuentran diferencias entre los análogos a los que les ofrece un incentivo económico y aquellos a los que no, llegando a la conclusión de que los incentivos no son necesarios en la investigación de simulación (Bernard, 1990). Puede que el efecto diferencial de los incentivos entre los simuladores reales y los análogos se deba a la dificultad de igualar las consecuencias económicas en uno y otro caso o que se deban más al papel que juegan los incentivos negativos (por ejemplo las consecuencias sociales de ser detectado como simulador) que a los positivos (Rogers y Cruise, 1998). Lo que parece claro es que se necesita más investigación en este campo para llegar a determinar cómo los incentivos afectan a la simulación.

Por último, es fundamental el **momento** dentro de la evaluación en el que se administran los tests. Para resaltar la importancia de esta variable, Guilmette, Whelihan, Hart, Sparadeo y Buongiorno (1996) compararon la ejecución de 100 litigantes y 40 análogos en el procedimiento de elección forzosa de Hiscock. A la mitad de los participantes les administraron la prueba al principio de la evaluación y a la otra mitad al final. Los resultados no dejan lugar a dudas: es casi tres veces más probable que los sujetos obtengan puntuaciones sospechosas de simulación cuando el test se administra al principio de la evaluación. Una explicación posible es que de este modo los sujetos no pueden comparar con otras pruebas la dificultad de la prueba, de modo que es más

fácil que sobreestimen su dificultad cuando se administra en primer lugar. Esto parece especialmente relevante para las pruebas de elección forzosa. Además, si encontramos indicios de simulación desde el comienzo de la evaluación podemos estar más alerta y administrar más índices para llegar a una conclusión más fiable sobre la motivación y esfuerzo del paciente durante la evaluación.

Una vez descritas las diversas variables que influyen en la detección de la simulación, y el análisis de los síntomas referidos por el paciente como método de detección de la simulación, así como en su detección desde el punto de vista de la psicofisiología, en el siguiente capítulo vamos a centrarnos en la revisión de las pruebas neuropsicológicas específicas y no específicas de simulación.

CAPÍTULO 5
INSTRUMENTOS EMPLEADOS EN LA
DETECCIÓN DE LA SIMULACIÓN

1. Introducción

A continuación nos vamos a centrar en dos métodos fundamentales de la detección de la simulación: el empleo de instrumentos neuropsicológicos específicos y en las pruebas neuropsicológicas no específicas, para las cuales han sido propuestos diversos índices indicativos de simulación.

En cuanto al primero de estos métodos, las pruebas específicas de simulación, podemos realizar una distinción entre (i) tests de efecto suelo, (ii) tests basados en habilidades sobreaprendidas y (iii) tests de elección forzosa, a los que dedicaremos la mayor parte de este capítulo.

En el último de los apartados realizaremos un resumen con las principales conclusiones que pueden extraerse de la revisión de los dos métodos expuestos en este capítulo.

2. Los test específicos

A continuación vamos a explicar los principales test específicos de simulación, diferenciando entre test de efecto suelo, pruebas de respuestas inusuales y tests de elección forzosa. Dentro de los test de elección forzosa, además de los test tradicionales hablaremos de las variaciones metodológicas que se le han aplicado recientemente, como los estímulos alternativos o la aplicación por ordenador. Finalizaremos este apartado con la comparación de las diferentes pruebas específicas.

2.1. Tests de efecto suelo

Estas medidas se basan en la idea de que los simuladores sobrevaloran la dificultad de algunos test, por lo que su ejecución en ellos es inferior a la obtenida por pacientes con daño cerebral severo. La prueba más empleada y estudiada es la de los **15 ítems de Rey** (Rey, 1964, citado en Lezack y cols., 2004). Este test es muy sencillo, de

modo que tanto pacientes psiquiátricos como pacientes con retraso mental lo realizan adecuadamente (Goldberg y Miller, 1986), pero las instrucciones se presentan de modo que su ejecución parezca difícil.

Según los puntos de corte propuestos por Rey, una puntuación de 9 implicaría la sospecha de exageración y una puntuación de 6 sería concluyente a la hora de establecer una colaboración dudosa por parte del paciente, siempre y cuando el nivel intelectual de los evaluados esté dentro de la normalidad. Así, a lo largo de los años las investigaciones han empleado diversos puntos de corte para esta prueba, siendo los más frecuentes 9, 7 y 6. Independientemente del punto de corte empleado, parece que la prueba obtiene una elevada especificidad, superior al 95%, en pacientes con daño cerebral severo (Taylor, Kreutzer y West, 2003), personas mayores con depresión mayor (Lee y cols., 2000), pacientes con traumatismos craneoencefálicos moderados y severos (Millis y Kler, 1995), así como con pacientes con epilepsia del lóbulo temporal y con trastornos neurológicos (Lee, Loring y Martin, 1992) y daño cerebral de diferentes etiologías (Arnett, Hammeke y Schwartz, 1995). No obstante, algunos autores opinan que la especificidad del 15-ítems no es tan buena, de modo que un elevado número de pacientes honestos serían clasificados como simuladores. En este sentido, Schretlen, Brandt, Krafft y Van Gorp (1991) advierten que los pacientes con amnesia y déficits cognitivos severos obtienen puntuaciones muy bajas en este test, y Guilmette, Hart, Giuliano y Leininger (1994) refieren lo mismo en pacientes psiquiátricos y en pacientes con daño cerebral.

A pesar de estos resultados contradictorios en cuanto a la especificidad de la prueba, existe consenso a la hora de señalar la sensibilidad de los 15-ítems como su principal problema (Schretlen y cols., 1991; Guilmette, Hart y cols., 1994; Merten y cols., 2005). Uno de los mejores trabajos realizados con esta prueba es el de Millis y

Kler (1995), puesto que es el único hasta la fecha que emplea como grupo de estudio sujetos que han puntuado significativamente por debajo del azar en una prueba de elección forzosa, encontrando una sensibilidad sólo ligeramente superior al azar para el 15-ítems de Rey. Además, parece que el establecimiento de puntos de corte más exigentes conlleva un descenso importante de la especificidad, a la vez que no produce los efectos deseados en la sensibilidad. En este sentido, Hilsabeck y Gouvier (2005) emplearon un punto de corte <12 elementos recordados, con el que sólo consiguieron detectar la mitad de los análogos, mientras casi el 20% de los pacientes con daños en memoria fueron clasificados incorrectamente como simuladores.

En un intento de arrojar luz sobre los datos contradictorios encontrados con esta prueba, Inman y Berry (2002) realizan una validación cruzada de estudios, estableciendo con un punto de corte de 9 que la especificidad del 15-ítems es del 100%, mientras su sensibilidad es tan solo del 2%. Con el mismo punto de corte de 9, Vickery, Berry, Inman, Harris y Orey (2001) encuentran en su revisión meta-analítica una sensibilidad del 43,3% y una especificidad del 92,5%. Reznik (2005) ha realizado recientemente otro meta-análisis en el que concluye que con un punto de corte de 9 la sensibilidad de la prueba es del 36%, y su especificidad del 85%. Con un punto de corte de 8 la especificidad aumenta al 92%, pero la sensibilidad desciende bruscamente al 9%, y con un punto de corte de 7 y exceptuando retraso mental, la sensibilidad es del 10% y la especificidad del 92%.

Para concluir, podemos decir que, a pesar de que en muchos casos los trabajos no son comparables por emplear metodologías y puntos de corte diferentes, el 15-ítems de Rey parece una prueba sensible a algunos daños neurológicos y con una sensibilidad inaceptablemente baja, de modo que sólo permite detectar los simuladores con las estrategias más descaradas (Cato y cols., 2002).

Una asignatura pendiente en los estudios con el 15-ítems es el de determinar la sensibilidad de otras puntuaciones, ya que un análisis cualitativo o la utilización de otras medida podría aumentar la utilidad de esta prueba; por ejemplo, algunos estudios indican que los simuladores tienden a añadir ítems nuevos (Guilmette, Hart y cols., 1994), o que pueden ser útiles los errores disléxicos o los errores gestalticos (Lee y cols., 2000). Para un resumen de los principales resultados encontrados con el 15-Ítems de Rey, ver Tabla 5.

2.2. Medidas de habilidades sobreaprendidas:

El **Dot Counting Test** (Rey, 1941, citado en Lezack y cols., 2004) es la prueba basada en habilidades sobreaprendidas más extensamente estudiada. Comparte con el 15-ítems muchos aspectos, además de su autor. De hecho, podemos decir que son paralelas en muchos sentidos: las dos poseen una buena especificidad pero una sensibilidad muy cuestionada, en ambas se han propuesto numerosos índices y puntos de corte que dificultan la comparación de estudios y en las dos pruebas las propuestas de Lezack (1995) han sido las más seguidas e investigadas, sin obtener muy buenos resultados.

En esta prueba el tiempo para contar los puntos que necesitan los sujetos honestos incrementa gradualmente a medida que aumenta el número de puntos. Lezack (1995) establece que una seria desviación en este patrón, es evidencia de que el sujeto no se está enfrentando a la prueba con un esfuerzo adecuado. Además, el número total requerido para contar los puntos agrupados debe ser inferior al de la condición no-agrupada. Cuando el tiempo en ambas condiciones es el mismo o se invierte, se sospecha que existe un sesgo en la respuesta del sujeto (Lezack, 1995).

Tabla 5. Principales resultados de las investigaciones con el test de los 15-items de Rey

Estudio	G	Tipo de sujetos	N	Es	En	Inc	Inf	Regla de clasificación	Sens	Espec
Schretlen, Brandt, Krafft y Van Gorp (1991)	S	Análogos que fingen amnesia y locura	69	Si	No	Si	No	Punto de corte de 9 elementos recordados	13%	64% 100%
	S	Sospechosos de simulación	7							
	C	Pacientes neurológicos y neuropsiquiátricos	148							
	C	Voluntarios y universitarios	80							
Lee, Loring y Martin (1992)	S	Pacientes en litigio	16					Punto de corte de 7 elementos recordados	96% 95%	
	C	Epilepsia del lóbulo temporal	100							
	C	Trastornos neurológicos	40							
Guillette, Hart, Giuliano y Leininger (1994)	S	Estudiantes análogos	20	No	No	No	No	Punto de corte de 7 elementos recordados	5%	60% 80%
	C	Pacientes con daño cerebral	20							
	C	Pacientes con daño psiquiátrico	20							
Greiffenstein, Baker y Gola (1994)	S	SPC sospechosos de simulación	43	Si	No	Si	No	Punto de corte de 9 elementos recordados	62%	93%
	C	TCE-L	33							
Arnett, Hammeke y Schwartz (1995)	S	Estudiantes análogos	25	Si	No	No	Si	Punto de corte de 7 elementos recordados	64%	96%
	C	Pacientes neurológicos	25							
Millis y Kler (1995)	S	Sospechosos de simulación	7					Punto de corte de 7 elementos recordados	57%	100%
	C	Traumatismos moderados y severos	7							
Lee y cols. (2000)		Ancianos con depresión mayor	64					Punto de corte de 9 elementos recordados		100%
Inman y Berry (2002)	S	Análogos estudiantes con y sin daño cerebral	44	Si	No	Si	Si	Punto de corte de 9 elementos recordados	2%	100%
	C	Estudiantes con y sin daño cerebral	48							
Taylor, Kreutzer y West (2003)	C	Daño cerebral severo	5					Punto de corte de 15 elementos recordados		100%
Hilsabeck y Gouvier (2005)	S	Análogos	30	Si	No	No	Si	Punto de corte de 12 elementos recordados	50%	81%
	C	Pacientes déficit de memoria	11							
Greub y Suhr (2006)	S	Pacientes con TCE-L	49	Si	Si	No	Si	Punto de corte de 9 elementos recordados	17%	100%
	C	con TCE-L y estudiantes	53							

Nota: G= grupo; S= grupo empleado como simulador; C= grupo control; Es= escenario; En= entrenamiento; Inc= incentivo; Inf= informe posterior; SPC= Síndrome Post-Connocional; TCE-L= traumatismo craneoencefálico leve; Sens= sensibilidad; Espec= especificidad.

Algunas investigaciones concluyen que las recomendaciones de Lezack no permiten diferenciar a los simuladores. En esta dirección apunta el trabajo de Greiffenstein y cols. (1994), que además tiene la ventaja de emplear como grupo experimental a sujetos sospechosos de simulación. Tampoco Rose y cols. (1998) encontraron ningún índice útil en el Dot Counting para la detección de simulación, y Frederick, Sarfaty, Johnston, y Powel (1994) obtuvieron sólo una sensibilidad del 5% empleando como variable el tiempo mayor para contar las tarjetas agrupadas.

Un apoyo a las recomendaciones de Lezack lo encontramos en el trabajo de Binks, Gouvier y Waters (1997). De todas formas, estos autores encuentran todavía más útil para diferenciar a los simuladores el número de errores cometidos en la prueba, aunque no proponen ningún punto de corte para el empleo de esta variable. Existen otras investigaciones que apoyan el empleo de esta variable: Lee y cols. (2000) encuentran una especificidad del 100% en ancianos con depresión mayor empleando un punto de corte de tres errores en esta prueba. Cato y cols. (2002) encontraron que un punto de corte de 6 errores conseguía una identificación correcta del 67,9% de los simuladores, si bien este porcentaje decaía drásticamente al 12,5% entre los simuladores entrenados, por lo que el Dot Counting parece una prueba especialmente sensible al entrenamiento. Otro estudio que emplea esta variable es el de Hilsabeck y Gouvier (2005), que con un punto de corte >4 errores obtienen muy buena especificidad, pero una sensibilidad muy baja, detectando sólo al 23% de los análogos.

Quizá uno de los mejores estudios sobre el Dot Counting sea el realizado por Boone, Lu, Back y cols. (2002), quienes aplicaron el Dot Counting a 100 pacientes sospechosos de simulación y a 251 sujetos englobados en 9 grupos clínicos diferentes. En este estudio se encontraron diferencias significativas entre los dos grupos respecto a las siguientes puntuaciones: media de tiempo empleado para contar las tarjetas con puntos

agrupados, ratio (división entre el tiempo de los ítems agrupados y no agrupados), número de errores y, sobre todo, puntuación combinada (media de tiempo de ítems agrupados más no agrupados más errores), obteniendo una buena sensibilidad y especificidad, con excepción del grupo de las demencias.

El Dot Counting cuenta con una versión por ordenador cuyos resultados parecen confirmar los de su versión de lápiz y papel, es decir, una buena especificidad (97,4%) y una sensibilidad que no supera el nivel del azar. Además de la exactitud de la respuesta, parece que la latencia podría ser útil para identificar a los simuladores (que responden significativamente más lentos que los controles), aunque son necesarios más trabajos que respalden estas afirmaciones.

Sobre este test podemos concluir que los criterios típicos propuestos por Lezack han mostrado una sensibilidad muy baja, si bien parece que el empleo de otros índices (como los propuestos por Boone, Lu, Back y cols., 2002) podrían aumentar notablemente la utilidad de la prueba, aunque es necesaria más investigación para determinar la sensibilidad de estas variables. En la Tabla 6 se presentan los principales resultados encontrados con el Dot Counting Test.

Otra prueba muy reciente que se basa en habilidades sobreaprendidas es el **test de la b**. Este test consiste en una tarea de reconocimiento y discriminación de letras diseñada para detectar simulación. Se supone que señalar las letras b de entre diversos distractores es una tarea extremadamente sencilla, de modo que incluso pacientes con daño cerebral o neurológico pueden realizar este test adecuadamente. A medida que el test avanza a través de sus 15 páginas, los estímulos se hacen cada vez más pequeños, de modo que incrementa su dificultad aparente. Sus autores señalan que este test presenta dos ventajas sobre las pruebas basadas en la simulación de problemas de memoria:

Tabla 6. Principales resultados de las investigaciones con el Dot Counting Test

Estudio	G	Tipo de sujetos	N	Es	En	Inc	Inf	Regla de clasificación	Sens	Espec
Greiffenstein, Baker y Gola (1994)	S	SPC sospechosos de simulación	43	Si	No	Si	No	Propuestas de Lezack (consultar texto). No encuentran diferencias significativas entre los grupos.		
	C	TCE-L	33							
Binks, Gouvier y Waters (1997)	S	Estudiantes análogos	62	Si	Si	No	No	6 variables sin propuesta de puntos de corte: - Tiempo para contar las tarjetas no agrupadas menos agrupadas		
	C	Pacientes neurológicos	31					- Número total de errores - Declive linear en la curva de latencia de respuesta de los ítems no-agrupados - Declive de la curva de latencia de respuesta de los agrupados - Suma de las desviaciones de la linealidad en las presentaciones agrupadas y no-agrupadas - Veces en las que el tiempo para las tarjetas agrupadas excede el tiempo de las no-agrupadas correspondientes		
Rose, Hall, Szalda-Petree y Bach (1998)	S	Estudiantes análogos	60	Si	No	No	Si	- Respuestas correctas agrupadas	9%	
	C	Pacientes con daño cerebral	30					- Respuestas correctas no-agrupadas - Tiempo total agrupadas - Tiempo total no-agrupadas - Tendencia a la inversión en las agrupadas - Tendencia a la inversión en las no-agrupadas		96%
Cato, Brewster, Ryan y Giuliano (2002)	S	Estudiantes análogos	28	Si	Si	No	Si	Punto de corte de 6 errores	12,5%	
	C	Pacientes con TCE-L	21							
Strauss y cols. (2002)	S	Análogos sin experiencia en TCE, análogos con experiencia profesional sobre TCE y pacientes	34	Si	No	No	No	Número de respuestas correctas en tarjetas agrupadas	50%	
	C	Sujetos sin experiencia en TCE, con experiencia profesional sobre TCE y pacientes	40							97,4%

Nota: G= grupo; S= grupo empleado como simulador; C= grupo control; Es= escenario; En= entrenamiento; Inc= incentivo; Inf= informe posterior; SPC= Síndrome Post-Connomonal; TCE-L= traumatismo craneoencefálico leve; Sens= sensibilidad; Espec= especificidad.

- la lectura y el reconocimiento de letras tiende a estar preservado en pacientes con significativos daños en memoria
- los sujetos que fingen un daño cognitivo pueden ser selectivos en las habilidades que simulan, por lo que pueden decidir no simular en test de memoria y sí en otras habilidades como el reconocimiento de letras

En el manual del test de la b (Boone, Lu y Herzberg, 2002) consta que fue validado empleando dos grupos de participantes: un grupo sospechoso de simulación verificado por medidas de esfuerzo diferentes al test de la b, y otro grupo con diversos problemas y diagnósticos clínicos (personas mayores sin diagnóstico, personas mayores con depresión, pacientes con esquizofrenia, pacientes con daño cerebral moderado o severo, pacientes con accidentes cerebro-vasculares y un grupo con problemas de aprendizaje). Esta prueba se mostró válida y fiable en la detección del esfuerzo sospechoso en ámbitos de litigio. En su estudio de validez, el test de la b discriminó consistentemente a los participantes en litigio sospechosos de simulación y los siete grupos de pacientes con quejas y diagnósticos variados. La puntuación E-score, que combina medidas de velocidad y exactitud de la ejecución, se mostró como el discriminador más fiable. En el manual se muestran distintos puntos de corte y rangos interpretativos para minimizar los falsos positivos. Para todos los grupos de comparación con un esfuerzo normal, los puntos de corte recomendados consiguen una especificidad de al menos el 85% y un poder predictivo positivo de al menos el 60%.

Empleando los mismos grupos expuestos arriba en otro estudio, Boone y cols. (2000) obtienen que el grupo sospechosos de simulación comete más errores de omisión y comisión que el resto de los grupos, con excepción de los pacientes con daño cerebrovascular derecho. Además, los sospechosos de simulación necesitan más tiempo para completar la tarea que todos los demás, con exclusión de los pacientes con daño

cerebrovascular y de los ancianos. Un punto de corte >2 errores de comisión produjo una sensibilidad de 76,5% y una especificidad del 82,6% para el resto de los grupos de comparación combinados. Para los errores de omisión y para el tiempo se hallaron valores de sensibilidad más bajos, si bien la especificidad se mantuvo alta. Mediante un análisis cualitativo de los tipos de errores de comisión se obtuvo que un punto de corte >0 para los errores tipo “d” identificaba correctamente el 79% de los sospechosos de simulación, con una especificidad de al menos 80% para todos los grupos de comparación. Los autores concluyen también que el test de la b no se ve influido por la presencia de daños neurológicos, psiquiátricos o problemas de aprendizaje, así como tampoco mantiene asociaciones con variables demográficas.

Teniendo en cuenta las investigaciones llevadas a cabo por los autores del test de la b parece que esta tarea se muestra como una herramienta muy útil para la detección de simulación, si bien a día de hoy todavía no ha estado sujeta a investigaciones independientes que lo demuestren.

2.3. Tests de elección forzosa

Las pruebas de elección forzosa o *Symptom Validity Testing* se basan en la teoría de la distribución binomial. El ejemplo más sencillo que suele emplearse para entenderlo es el lanzamiento de una moneda, en la que la probabilidad de que salga cara es exactamente la misma de que salga cruz. Así, en un test de elección forzosa con dos alternativas, la posibilidad de contestar correctamente (incluso al azar, sin haber visto los estímulos) es del 50%, por lo que una ejecución significativamente por debajo del nivel del azar indica que el sujeto conoce la respuesta correcta e *intencionadamente* elige la incorrecta o, dicho de otro modo, que existe una motivación deliberada para realizar las pruebas pobremente.

Los tests de elección forzosa tienen en realidad una carga de memoria muy baja,

de modo que incluso los pacientes con un daño cerebral significativo, con daños cognitivos o pacientes psiquiátricos, obtienen puntuaciones casi perfectas (Guilmette, Hart y cols., 1994). La premisa de este procedimiento es que los simuladores presentarán un grado de daño implausible debido a la dificultad aparente de estas pruebas. Quizá la principal ventaja del empleo de la probabilidad binomial es que no se ve afectada por los porcentajes de base.

El primer ejemplo de la utilización de este método se lo debemos a Grosz y Zimmerman (1965), que lo emplearon con un paciente con “ceguera histórica”. Estos autores desarrollaron una tarea de discriminación visual en la que el paciente tenía que responder cada vez que se producía un sonido si dos estímulos eran diferentes. Las repuestas del sujeto (que había confesado a una cuidadora que no quería volver a trabajar, sino vivir de la seguridad social) cayeron significativamente por debajo del azar.

Pankratz, Fausti y Peed (1975) lograron descubrir de modo muy similar a un paciente que simulaba sordera. Dado el éxito de la técnica, Pankratz (1983) la adaptó para utilizarla con pacientes sospechosos de simular problemas de memoria. La prueba consistía en dos luces (una blanca y otra roja) y los pacientes debían contestar cuál se había encendido tras una tarea distractora. Binder y Pankratz (1987) emplearon un procedimiento muy similar con otro paciente sospechoso de simular problemas de memoria: empleando como estímulos un bolígrafo negro y un lápiz amarillo, el paciente debía recordar cuál de los dos se le había presentado con anterioridad. El problema de estos métodos, a pesar de su utilidad clínica, es su falta de estandarización. Con el propósito de solucionar este problema surge el **Digit Memory Test (DMT)**, creado en 1989 por Hiscock y Hiscock.

El DMT consta de tres bloques de 24 ítems. En cada ensayo se presenta al sujeto un

número de cinco dígitos seguido por una prueba de reconocimiento con dos alternativas. El intervalo entre la presentación del estímulo objetivo y la tarea de reconocimiento se va incrementando, siendo de 5 segundos en el primer bloque, de 10 en el segundo y de 15 en el tercero. Esta prueba mejora el diseño de Pankratz por varios motivos: en primer lugar, el empleo de números de cinco dígitos aumenta la plausibilidad de que el test mida memoria, es decir, disminuye su transparencia. En segundo lugar, el aumento del intervalo entre la presentación y el reconocimiento crea la ilusión de incremento de la dificultad o lo que es lo mismo, aumenta su dificultad aparente (Bianchini, Mathias y Greve, 2001).

El problema de las respuestas significativamente por debajo del nivel del azar es que, a pesar de su elevadísimo poder predictivo positivo, carece de la adecuada sensibilidad, por lo que un excesivo número de simuladores no serían detectados. Para solucionar este problema, los investigadores han propuesto puntos de corte por encima de dicho nivel, para lo cual se han seguido dos aproximaciones (Bianchini, Mathias y Greve, 2001): la aproximación estadística, que consiste en el establecimiento de puntos de corte arbitrarios, y la aproximación normativa, que establece “suelos empíricos” basados en la ejecución más baja obtenida por un grupo de contraste con algún daño documentado. Atendiendo a los trabajos que siguen estas aproximaciones, podemos concluir que es innecesario el criterio de significación por debajo del azar para detectar simulación.

Siguiendo esta lógica, Guillmette, Hart y Giuliano (1993) aplicaron el DMT a pacientes con daño cerebral, pacientes psiquiátricos y dos grupos de no-pacientes, uno de ellos de análogos simuladores. A pesar de que el grupo de análogos simuladores obtuvo una puntuación inferior a la de los otros grupos, sólo el 34% puntuó significativamente por debajo del azar. Además, el porcentaje de respuestas correctas

fue muy elevado en los otros grupos: 96% entre los pacientes con daño cerebral, 98% en los pacientes psiquiátricos y 100% en los controles. Teniendo esto en cuenta, estos autores concluyen que tan sólo algunos errores (3 o 4 para daños menos severos) deberían alertar al examinador sobre la posibilidad de que el sujeto no está poniendo un nivel de motivación o esfuerzo adecuado. Por ello, proponen un punto de corte del 90% de respuestas correctas para sospechar de simulación (consiguiendo así una clasificación correcta del 90% de los sujetos), y del 75% para estar casi seguro de que la ejecución no se debe a un daño cerebral, sino a la exageración de la sintomatología.

Otro ejemplo del establecimiento de un punto de corte por encima del azar en esta prueba lo encontramos en el estudio de Prigatano, Smason, Lamb y Bortz (1997), en el que concluyen que los sujetos con traumatismo cranoencefálico, con amnesia severa secundaria a un daño en el lóbulo temporal, los pacientes afásicos y los pacientes con disfunción del lóbulo temporal obtienen un porcentaje de respuestas correctas del 95% o incluso superiores. Sólo la demencia tipo Alzheimer obtiene unos niveles de ejecución similares a los de los simuladores (excepto en el bloque 3 del test, en el que sí aparecen diferencias estadísticamente significativas). Una ventaja que posee este trabajo sobre los anteriores es que demuestra la utilidad del DMT con sospechosos de simulación, y no con análogos.

También se ha constatado el aumento de la sensibilidad con un mínimo decremento en la especificidad al emplear puntos de corte por encima del nivel del azar con versiones abreviadas del DMT (Guilmette, Hart y Giuliano, 1993; Cato, Brewster, Ryan y Giuliano, 2002; Vickery y cols., 2004). En recientes investigaciones se ha realizado un meta-análisis (Vickery y cols., 2001) en la que la prueba ha obtenido muy buenos resultados, obteniendo con un punto de corte del 90% de aciertos una sensibilidad del 83,4% y una especificidad del 95,1%. La prueba también ha sido

sometida a una validación cruzada (Inman y Berry, 2002) en la que también se validó dicho punto de corte en la versión abreviada del DMT. Para un resumen de los principales resultados de la investigación con esta prueba (ver Tabla 7).

El sucesor del DMT es el **Portland Digit Recognition Test** (Binder, 1990). Esta prueba consiste en 72 ítems de reconocimiento de dígitos y su administración requiere aproximadamente 45 minutos. Utilizando una presentación de estímulos auditiva, se leen a los sujetos 5 dígitos a razón de un dígito por segundo. Acto seguido, el sujeto cuenta en alto hacia atrás hasta que el evaluador lo interrumpe con la presentación visual de la prueba de reconocimiento en la que hay un ítem distractor y uno objetivo. Los 18 primeros ensayos el sujeto cuenta desde 20 hacia atrás durante 5 segundos, en el segundo bloque de 18 ensayos cuenta de 50 hacia atrás durante 15 segundos, y en el tercer y cuarto bloque de 18 ensayos debe contar hacia atrás desde 100 durante 30 segundos. Esta tarea distractora entre la presentación del estímulo y su reconocimiento es la principal diferencia entre el DMT y el PDRT. Después del primer y segundo ensayo se informa a los sujetos de que el test se va haciendo progresivamente más difícil, y a los sujetos que han contestado correctamente 12 ítems o más se les dice que lo están haciendo muy bien. Los ítems de los dos primeros bloques se clasifican como fáciles y los de los bloques 3 y 4 como difíciles. Dichos ítems de los bloques difíciles parecen especialmente útiles a la hora de detectar la exageración de déficits (Binder, 1993).

El primer estudio en el que se emplea esta prueba data de 1990 y, en él, Binder presenta varios casos únicos en los que pacientes con TCE-L obtienen puntuaciones por debajo del nivel del azar en el PDRT.

Tabla 7. Principales resultados de la investigación con el Digit Memory Test

Estudio	G	Tipo de sujetos	N	Es	En	Inc	Inf	Regla de clasificación	Sens	Espec
Guilmette, Hart y Giuliano (1993)	S	Análogos, principalmente estudiantes	29	No	No	No	No	Punto de corte de 90% de respuestas correctas	90%	90%
	C	Pacientes con daño cerebral	20							
Guilmette, Hart y Giuliano (1994)	S	Estudiantes análogos	20	No	No	No	No	Punto de corte de 90% de respuestas correctas	85%	
	C	Pacientes con daño cerebral	20							100%
	C	Pacientes con daño psiquiátrico	20							95%
Prigatano, Smason, Lamb y Bortz (1997)	S	Sospechosos de simulación	10					Punto de corte de 95% de respuestas correctas	97%	
	C	Diversos grupos de daño cerebral	33							89%
Cato, Brewster, Ryan y Giuliano (2002)	S	Estudiantes análogos	28	Si	Si	No	Si	Punto de corte de 10 errores	70%	
	C	Pacientes con TCE-L	21							
Inman y Berry (2002)	S	Análogos estudiantes con y sin daño cerebral	44	Si	No	Si	Si	Punto de corte de 90% de respuestas correctas	64%	
	C	Estudiantes con y sin daño cerebral	48							100%
Vickery y cols. (2004)	S	Análogos con daño cerebral	23	Si	Si	Si	Si	Punto de corte de 90% de respuestas correctas	56,5%	
	S	cerebral	23						54,5%	
	C	Análogos voluntarios	23							95,5%
	C	Pacientes con daño cerebral	23							100%
		Voluntarios normales								

Nota: G= grupo; S= grupo empleado como simulador; C= grupo control; Es= escenario; En= entrenamiento; Inc= incentivo; Inf= informe posterior; SPC= Síndrome Post-Comnacional; TCE-L= traumatismo craneoencefálico leve; SPC= Síndrome Post-Comnacional; TCE-L= traumatismo craneoencefálico leve; Sens= sensibilidad; Espec= especificidad.

En estudios posteriores se llega a la conclusión de que el PDRT está relacionado con la búsqueda de compensación económica, puesto que los sujetos que persiguen dicha compensación obtienen puntuaciones significativamente inferiores a los pacientes con daños más severos sin incentivos económicos (Binder y Willis, 1991; Binder, 1993). De hecho, entre los grupos que buscan compensación, o no se encuentra relación entre la severidad del daño y la puntuación en el PDRT (Binder y Willis, 1991) o los sujetos con disfunciones más severas realizan la prueba mejor que los sujetos con TCE-L (Binder, 1993).

Empleando el criterio por debajo del azar, el 12% de los sujetos con TCE-L y el 3% de los pacientes con disfunción cerebral (ambos buscando compensación) obtuvieron puntuaciones indicativas de simulación. Sin embargo, mediante el empleo de un punto de corte basado en el “suelo empírico” (19 para los ítems fáciles, 18 para los difíciles y 39 para el total) el 33% de los TCE-L y el 18% de los pacientes con disfunción cerebral puntuaron como simuladores (Binder, 1993). Así, una vez más parece constatarse que las puntuaciones por debajo del azar constituyen un criterio demasiado exigente e innecesario, teniendo en cuenta que el establecimiento de puntos de corte más altos aumenta la sensibilidad de las pruebas de simulación sin afectar demasiado a su especificidad.

Greiffenstein y cols. (1994) estudiaron una versión abreviada del PDRT (que contenía 27 ítems) en un grupo con TCE-L en litigio sospechoso de simulación, al que compararon con un grupo con TCE y otro con SPC, obteniendo unos elevados porcentajes de clasificación correctos (sensibilidad del 66% y especificidad del 94% con un punto de corte del 63% de respuestas correctas).

Tabla 8. Principales resultados en la investigación con el Portland Digit Recognition Test

Estudio	G	Tipo de sujetos	N	Es	En	Inc	Inf	Regla de clasificación	Sens	Espec
Binder y Willis (1991)	S	TCE-L en búsqueda de compensación	46					Punto de corte de 54% ítems correctos (<39)		
	C	Pacientes con daño cerebral	38							100%
Binder (1993)	S	Pacientes con daño cerebral en búsqueda de compensación	33					Punto de corte de 54% ítems correctos (<39)		
	C	Pacientes con daño cerebral	67							100%
Greiffenstein, Baker y Gola (1994)	S	SPC sospechosos de simulación	43	Si	No	Si	No	Punto de corte de 58%/63% ítems correctos	66/75%	
	C	SPC	30							94/78%
Rose, Hall, Szalda-Petree y Bach (1998)	S	Estudiantes análogos	30	Si	Si	No	Si	Punto de corte de 54% ítems correctos (<39)	47%	
	C	Pacientes con daño cerebral	28							89%
	S	cerebral	29	Si	No	No	Si		83%	
C	Estudiantes análogos Voluntarios	30								100%

Nota: G= grupo; S= grupo empleado como simulador; C= grupo control; Es= escenario; En= entrenamiento; Inc= incentivo; Inf= informe posterior; SPC= Síndrome Post-Convulsional; TCE-L= traumatismo craneoencefálico leve; Sens= sensibilidad; Espec= especificidad.

En el meta-análisis realizado por Vickery y cols. (2001), en el que se recogen los trabajos anteriores, el PDRT demostró poseer una moderada sensibilidad (44%) pero una especificidad muy elevada (97,3%). Esta moderada sensibilidad parece deberse al empleo del punto de corte propuesto por Binder (54% de respuestas correctas), ya que puntos de corte más laxos conducen a un notable aumento en la sensibilidad (ver, por ejemplo, los resultados de Greiffenstein y cols., 1994). Para ver los principales resultados de la investigación con el PDRT, ver Tabla 8.

A partir del desarrollo del DMT y del PDRT la investigación dentro de los tests de elección forzosa ha estado encaminada a la aplicación de variaciones sobre estos procedimientos para disminuir su transparencia y así incrementar su sensibilidad. Para ello, los autores han recurrido a la utilización de estímulos alternativos, a las aplicaciones por ordenador, o a otro tipo de variaciones, que exponemos a continuación:

2.3.1. Estímulos Alternativos:

El **Test of Memory Malingering** (Tombaugh, 1996) es probablemente el SVT que cuenta en la actualidad con más investigación a sus espaldas. La prueba está formada por dos ensayos de aprendizaje y un ensayo de retención. Durante los dos ensayos de aprendizaje se le muestran al sujeto 50 dibujos de objetos comunes y posteriormente se le presentan láminas con dos dibujos y se le pide que reconozca el que ha visto previamente. Después de cada respuesta el examinador informa al paciente sobre si su elección ha sido o no correcta. El ensayo de retención es optativo y se utiliza para corroborar el resultado de los ensayos de aprendizaje 15 minutos después de la aplicación de estos. Esta parte opcional se aplica mostrando al sujeto directamente las láminas de reconocimiento, sin mostrar previamente los estímulos objetivo. La corrección del test es muy sencilla, otorgándose un punto por cada respuesta correcta.

El Test of Memory Malingering (TOMM) es especialmente útil en la detección de

simulación por varios motivos (Tombaugh, 1996; Rees, Tombaugh, Gansley y Moczynski, 1998):

1. Al mostrar una lista tan larga de estímulos, los sujetos simuladores interpretan que la tarea es muy complicada y por tanto la realizan intencionadamente mal.
2. La presentación de los 50 dibujos le otorga una elevada validez aparente como test de memoria y aprendizaje, es decir, no aparenta ser un test de simulación.
3. El feedback dado a los sujetos tras cada respuesta aumenta la motivación de los pacientes incrementando su ejecución, produciendo en los simuladores el efecto contrario.

Además, de los estudios de Tombaugh (1996) se desprende que incluso los pacientes con un daño sustancial obtienen puntuaciones elevadas en el TOMM (los pacientes con demencia obtienen una puntuación de respuestas correctas por encima del 92% en el segundo ensayo), por lo que una puntuación inferior a 45 en el ensayo 2 o en el ensayo de retención indica que es probable que el sujeto esté simulando. Al ser una prueba de reconocimiento entre dos estímulos, el TOMM también permite emplear como regla de decisión las puntuaciones por debajo del azar. Así, las puntuaciones por debajo de 18 indican que el sujeto está eligiendo intencionadamente los dibujos incorrectos.

Diversas investigaciones han demostrado que el TOMM es insensible a los efectos de la edad (Constantinou y McCaffrey, 2003; Tombaugh, 1996; Rees, Tombaugh y Boulay, 2001; Teichner y Wagner, 2004), la educación (Constantinou y McCaffrey, 2003; Rees y cols., 2001), al daño cerebral traumático (Tombaugh, 1996; Rees y cols., 2001; Haber y Fichtenberg, 2006) y a la demencia (Tombaugh, 1996), a trastornos psiquiátricos (Duncan, 2005; Gierok, Dickson y Cole, 2005) a la ansiedad (Ashendorf, Constantinou y McCaffrey, 2004) al dolor inducido en laboratorio (Etherton, Bianchini,

Greve y Ciota, 2005) y a la depresión (Ashendorf y cols., 2004; Rees y cols., 2001; Yanez, Fremouw, Tennant, Strunk y Coker, 2006).

Esta prueba ha demostrado su eficacia en análogos (Powell, Gfeller, Hendricks y Sharland, 2004; Etherton y cols., 2005; Rees y cols., 1998), en pacientes en litigio (Constantinou y cols., 2005; Rees y cols., 1998; Gavett, O'Bryant, Fisher y McCaffrey, 2005; Greve y Bianchini, 2006) o que buscan obtener una compensación económica (Moore y Donders, 2004; Haber y Fichtenberg, 2006), y en diseños de grupos conocidos (Greve, Bianchini y Doane, 2006). Además, parece insensible a los efectos del entrenamiento tanto cuando es proveído por el investigador (Powell y cols., 2004) como cuando se incita a los análogos a que se entrenen ellos mismos para no ser detectados y se les proporciona tiempo para ello (Rees, Tombaugh, Gansler y Moczynski, 1998; Tan, Slick, Strauss y Hultsch, 2002). Asimismo, el TOMM no se ve afectado por el conocimiento sobre las consecuencias de los TCE (ya que se ha demostrado que los pacientes con traumatismos no son mejores a la hora de eludir la detección), ha mostrado su efectividad en su administración por ordenador (en cuyo caso la latencia de respuesta también se encontró útil para la detección de simuladores), y también al ser administrado dentro de una batería neuropsicológica amplia (Rees y cols., 1998).

La utilidad el punto de corte de 45 en el Ensayo 2 ha recibido apoyo de modo consistente en las diversas investigaciones. Sin embargo, se han propuesto algunas modificaciones y recomendaciones con respecto a su aplicación tradicional. Por ejemplo, parece que es adecuado interrumpir el test cuando los sujetos obtienen una puntuación de 45 o superior en el ensayo 1 (Gavett y cols., 2005). También parece el ensayo de retención no debería ser tan "optativo", ya que conlleva la no-detección del 3% de los pacientes litigantes (Greve y Bianchini, 2006) y se ha mostrado particularmente efectivo para la clasificación de los simuladores (Greve y cols., 2006).

Así, parece que el ensayo de retención puede ser obviado cuando se administran otras medidas de elección forzosa, pero nunca cuando es el único SVT de la evaluación. Greve y cols. (2006) han propuesto puntos de corte para el ensayo 1, así como alternativas al punto de corte de 45 en el ensayo 2, por considerarlo conservador. Este estudio es especialmente interesante porque proporciona la sensibilidad, especificidad y poder predictivo positivo para diversos porcentajes de base (de 0,10 a 0,50). Otra investigación que encuentra conservador el punto de corte de 45 es la de Haber y Fichtenberg (2006), que enfatizan la utilidad del empleo de un rango de errores de 2 a 5 en los TCE-L, lo que además atenuaría los posibles efectos del entrenamiento de pacientes.

Gracias a sus buenos resultados, parece que sólo es necesario extremar las precauciones en la administración del TOMM con pacientes muy severamente dañados (Greve y cols., 2006) y en los casos de demencia, en los que esta prueba no se ha mostrado eficaz ni siquiera con un punto de corte de 40 (Teichner y Wagner, 2004).

Por otro lado, en las investigaciones con el TOMM se han incluido grupos étnicos y minoritarios (Gierok y cols., 2005; Moore y Donders, 2004), pero nunca se han estudiado dichos grupos per se. Existen dos artículos con población española: en uno de ellos se estudian sólo pacientes con TCE con una motivación adecuada (Ramírez y cols., 2004), lo que nos permite conocer la especificidad de la prueba en nuestro país (83,3%), pero no su sensibilidad, y en el otro se explora el comportamiento de la prueba en grupos de prevalencia diferencial (Vilar y cols., In Press). Ver Tabla 9 para un resumen de los principales resultados de las investigaciones con el TOMM.

Otro test que emplea estímulos alternativos, en este caso palabras, es el **21-Item Memory Test** (Iverson, Franzen y McCracken, 1991). Esta prueba consiste en una lista de 21 sustantivos que se lee al evaluado en voz alta, seguida por una tarea de recuerdo

Tabla 9. Principales resultados en la investigación con el Test of Memory Malingering

Estudio	G	Tipo de sujetos	N	Es	En	Inc	Inf	Regla de clasificación	Sens	Espec
Rees, Tombaugh, Gansler y Moczynski (1998)	S C C	TCE análogos Estudiantes controles TCE controles	8 10 10	Si	Si* Si Si	Si Si Si	Si Si Si	Punto de corte de 45 en el ensayo 2	100% 100% 96%	100% 96%
Rees, Tombaugh y Boulay (2001)	C	Pacientes con depresión mayor	26					Punto de corte de 45 en el ensayo 2		100%
Tan, Slick, Strauss y Hultsch (2002)	S C	Estudiantes análogos Estudiantes controles	25 27	Si	Si* Si	No Si	Si Si	Punto de corte de 45 en el ensayo 2	80%	96%
Powell, Gfeller, Hendricks y Sharland (2004)	S C	Estudiantes análogos Estudiantes controles	52 28	Si	Si Si	No Si	Si Si	Punto de corte de 45 en el ensayo 2	94,2%	100%
Ramírez, Chivirella-Garrido, Caballero, Ferricampos y Noé-Sebastián (2004)	C	Pacientes españoles con TCE, en su mayoría graves	30					Punto de corte de 45 en el ensayo 2		83,3%
Teichner y Wagner (2004)	C C C	Ancianos normales Ancianos con daño cognitivo Ancianos con demencia	21 36 21					Punto de corte de 45 en el ensayo 2		100% 95% 24%
Vickery y cols. (2004)	S S C C	Análogos con daño cerebral Análogos voluntarios Pacientes con daño cerebral Voluntarios normales	23 23 23 23	Si	Si Si Si Si	Si Si Si Si	Si Si Si Si	Punto de corte de 45 en el ensayo 2	52,2% 43,5%	100% 100%

Duncan (2005)	C	Pacientes psicóticos sin problemas de concentración	21						Punto de corte de 45 en el ensayo 2	100%
	C	Pacientes psicóticos con problemas de concentración	29							96%
Etherton, Bianchini, Greve y Ciota (2005)	S	Estudiantes análogos	20	Si	No	No	Si		Punto de corte de 45 en el ensayo 2	85%
	C	Estudiantes con dolor inducido en laboratorio	20							100%
Gierok, Dickson y Cole (2005)	C	Pacientes psiquiátricos	20						Punto de corte de 45 en el ensayo 2	95%
Greve, Bianchini y Doane (2006)	C	Pacientes con problemas de memoria	22						Punto de corte de 45 en el ensayo 2	42%
	C	TCE no SDN	79						Punto de corte de 45 en el ensayo de retención	57%
	S	TCE sospechosos de SDN	41						Punto de corte de 47 en el ensayo de retención	61%
	S	TCE probable SDN	32							95%
	S	TCE definitiva SDN	9							
Haber y Fichtenberg (2006)	S	TCE-L sospechosos de simulación	17						Punto de corte de 45 en el ensayo 2	76%
	C	TCE con déficit en memoria visual	22							100%
Yanez, Fremouw, Tennant, Strunk y Coker (2006)	C	Pacientes con depresión mayor	20						Punto de corte de 45 en el ensayo 2	90%
									Punto de corte de 45 en el ensayo de retención	95%

Nota: G= grupo; S= grupo empleado como simulador; C= grupo control; Es= escenario; En= entrenamiento; Inc= incentivo; Inf= informe posterior; SPC= Síndrome Post-Comocional; TCE-L= traumatismo craneoencefálico leve; Sens= sensibilidad; Espec= especificidad; SDN= simulación de disfunción neurocognitiva.; Si*= autoentrenamiento, entrenamiento no proporcionado por los evaluadores sino buscado por los participantes.

libre. A continuación se realiza una tarea de elección forzosa, en la que el sujeto debe diferenciar entre dos palabras (una objetivo y un distractor) cuál de ellas estaba en la lista leída originalmente. Una de las ventajas que posee esta prueba es que su administración requiere sólo 5 minutos. Los resultados con este test indican que el empleo de puntuaciones significativamente por debajo del azar (<9) consiguen una excelente especificidad, pero una sensibilidad inaceptablemente pobre con pacientes forenses sospechosos de simulación (Inman, Vickery, Lamb, Edwards y Smith, 1998) y análogos (Iverson y Franzen, 1996; Inman y cols., 1998), lo cual se ha visto apoyado por un meta-análisis (Vickery y cols., 2001) en el que la prueba obtiene un 100% de especificidad, pero una sensibilidad del 22%. Parece que puntos de corte menos exigentes (<13) podrían solucionar este problema, ya que Iverson y Franzen (1996) encuentran que el 80% de los análogos son detectados con este criterio, mientras la especificidad se mantiene en el 100%. Sin embargo, estos hallazgos no han sido corroborados, ya que en una investigación posterior Rose y cols. (1998) detectan con el mismo punto de corte tan sólo el 42% de los análogos, mientras un 32% de pacientes con daño cerebral serían incorrectamente clasificados como sospechosos de simulación.

En definitiva, el 21-Item Test es una prueba que posee una excelente especificidad cuando sus hallazgos se basan en puntuaciones por debajo del azar, por lo que puede ser útil su empleo siempre que sea administrada en combinación con otras medidas dentro de una batería, dado que es una prueba poco costosa en cuanto a tiempo. La adopción de puntos de corte más laxos podría significar un aumento en su sensibilidad, pero queda pendiente comprobar si la especificidad se ve o no resentida, ya que los hallazgos son contradictorios (ver Tabla 10).

Tabla 10. Principales resultados en la investigación con el 21-Item Test

Estudio	G	Tipo de sujetos	N	Es	En	Inc	Inf	Regla de clasificación	Sens	Espec
Iverson y Franzen (1996)	S	Análogos estudiantes y pacientes psiquiátricos	40	Si	Si	No	Si	Punto de corte <9/<13 aciertos	22/80%	
	C	Estudiantes, pacientes psiquiátricos y pacientes con daño de memoria	60							100%
Inman y cols. (1998)	S	Sospechosos de simulación	19					Punto de corte <9 aciertos	5%	100%
	C	Pacientes neurológicos	21						5%	100%
	S	Estudiantes análogos	100	Si	Si	Si	Si			
	C	Pacientes neurológicos	32							
	S	Análogos voluntarios	54						22%	100%
	C	Pacientes neurológicos y psiquiátricos	66							100%

Nota: G= grupo; S= grupo empleado como simulador; C= grupo control; Es= escenario; En= entrenamiento; Inc= incentivo; Inf= informe posterior; Sens= sensibilidad; Espec= especificidad.

2.3.2. Administración por ordenador:

El **Victoria Symptom Validity Test** (VSVT; Slick, Hopp, Strauss y Thompson, 1997) es una versión abreviada y computerizada del DMT. Este test de reconocimiento consiste en la presentación de tres bloques de 16 números de 5 dígitos, cada uno de ellos seguido por la presentación de un distractor junto al número previamente presentado, que el sujeto debe reconocer. La dificultad aparente del test incrementa mediante dos estrategias: en primer lugar, el intervalo de retención va aumentando con los bloques (5, 10 y 15 segundos), además de las instrucciones que indican al sujeto que “podría” encontrar la siguiente sección más difícil. En segundo lugar, el test incluye ítems “fáciles” (aquellos en los que el estímulo objetivo y el distractor no comparten números, por ejemplo 83649 y 25170) e ítems difíciles (en los que el número objetivo y el distractor rotan el segundo y tercer dígito o el tercero y el cuarto, por ejemplo 58379 y 58739). Además, el VSVT tiene en cuenta la latencia de respuesta, ya que se ha puesto de manifiesto que las personas que exageran sus déficits tienden a exagerar el tiempo necesario para realizar su elección, y también considera los sesgos relacionados con la tendencia a responder a los ítems de la derecha o de la izquierda. Esta prueba se ha mostrado insensible a los efectos de la edad, género, educación y hallazgos de neuroimagen (Macciocchi, Seel, Alderson y Godsall, 2006), así como a diversos trastornos neurológicos y psiquiátricos (Slick y cols., 1994; Inman y cols., 1998; Grote y cols., 2000; Slick y cols., 2003).

Slick, Hopp, Strauss y Spellacy (1996) aplicaron el VSVT a cuatro grupos de sujetos (un grupo de comparación, un grupo de análogos simuladores, un grupo de pacientes que buscan compensación económica (N=206) y un grupo de pacientes que no la buscan). Teniendo en cuenta los errores cometidos en el test se establece una clasificación de tres niveles: inválido (8 aciertos o menos de 24 posibles para los ítems

fáciles o difíciles), cuestionable (puntuaciones de 9 a 15) o válido (de 16 a 24 aciertos), que ha mostrado una validez convergente adecuada y una excelente validez divergente, ya que el VSVT no correlaciona con los tests cognitivos en general y con los de memoria en particular, y por lo tanto esta prueba no se ve afectada por el nivel cognitivo del paciente. Según los autores de la prueba, las puntuaciones en el rango “cuestionable” deben considerarse como indicadores de una probable exageración de síntomas, que ha de ser confirmada o desmentida por otro tipo de evidencia (Slick y cols., 1996). Siguiendo los puntos de corte propuestos en el manual, se han encontrado unos niveles de sensibilidad y especificidad muy elevados (Tan, Slick, Strauss y Hultsch, 2002; Strauss y cols., 2000), mostrándose especialmente útil el nivel de ejecución en los ítems difíciles (Strauss y cols., 2000; Strauss y cols., 2002).

Si la distinción entre puntuaciones válidas, cuestionables e inválidas basadas en la probabilidad binomial han sido empleadas en el manual del test para minimizar los posibles falsos positivos, es ampliamente conocido su bajo poder de detección (por ejemplo, sólo el 15% de los estudiantes análogos en el estudio de Slick y cols., (1994), el 32% en el de Tan y cols., (2002) y el 45% en el de Strauss y cols. (2000) puntuaron en el rango inválido, por debajo del nivel del azar). Además, también se ha constatado que los pacientes con amnesia anterógrada o daños severos en memoria obtienen puntuaciones perfectas o casi perfectas en el VSVT (Slick y cols., 2003). Por estos motivos algunos autores han decidido estudiar criterios normativos, ya que, como Slick y cols. (1994) reconocen, conducen a unos niveles de clasificación correcta mucho más elevados que el criterio probabilístico.

Siguiendo esta línea, Grote y cols. (2000) estudiaron un grupo de pacientes en litigio con quejas de memoria y un grupo de pacientes con epilepsia intratable sin búsqueda de compensación económica. Mientras que el 93,3% de los pacientes

epilépticos contestaban bien el 90% o más de los ítems difíciles, sólo el 35,8% de los sujetos litigantes lo superaban. En cuanto a la respuesta de latencia, estos autores le encuentran poca utilidad, dados los buenos resultados de el porcentaje de respuestas correctas, si bien aconsejan el empleo de un punto de corte de 4 segundos para los ítems difíciles, ya que el 90% del los sujetos epilépticos los superaban, comparado con el 44% de los sujetos litigantes. También Inman y cols. (1998) encuentran muy adecuado el empleo del 90% de respuestas correctas en el VSVT para diferenciar a pacientes y análogos, consiguiendo una clasificación correcta del 96%.

Macciocchi y cols. (2006) investigaron la ejecución de pacientes con TCE severos en fase aguda en el VSVT, encontrando que esta población tiene un 99% de probabilidades de obtener puntuaciones superiores al 90% de aciertos en esta prueba. Por otro lado, las latencias de respuesta de los pacientes se situaron muy próximas al rango inválido propuesto por el manual, por lo que esta variable no parece útil para determinar el esfuerzo en pacientes con traumatismos severos. Basándose en sus hallazgos de “suelo empírico”, estos autores realizan una propuesta de clasificación alternativa a la establecida por el manual de la prueba: las puntuaciones probablemente válidas estarían entre 44 y 48 respuestas correctas, las puntuaciones cuestionables entre 38 y 43 y las inválidas se corresponderían con 37 o menos respuestas correctas.

Tomando en conjunto las investigaciones anteriores, podemos concluir que el VSVT es una prueba que ha demostrado una excelente especificidad siguiendo las clasificaciones propuestas en el manual, si bien estas propuestas parecen excesivamente conservadoras. La sensibilidad de la prueba se ve notablemente incrementada, sin suponer una reducción importante de su especificidad, al adoptar puntos de corte basados en datos normativos. Especialmente prometedor se ha mostrado el 90% de respuestas correctas (ya sea en el total de las respuestas o en los ítems difíciles) que

Tabla 11. Principales resultados en la investigación con el Victoria Symptom Validity Test

Estudio	G	Tipo de sujetos	N	Es	En	Inc	Inf	Regla de clasificación	Sens	Espec
Slick, Hopp, Strauss y Spellaey (1996)	S	Estudiantes análogos	42	No	No	No	No	Puntuaciones significativamente por debajo del nivel del azar en ítems fáciles o difíciles	39%	100%
	C	Pacientes neurológicos	32							
Inman y cols. (1998)	S	Estudiantes análogos	100	Si	Si	Si	Si	90% de respuestas correctas totales	98%	91%
	C	Pacientes neurológicos	32							
Strauss y cols. (2000)	S	Estudiantes análogos	20	Si	No	No	No	Puntuaciones significativamente por debajo del nivel del azar en ítems difíciles (≤ 8)	45%	100%
	C	Estudiantes	21							
Grote y cols. (2000)	S	Pacientes en búsqueda de compensación	53					90% de ítems difíciles correctos		
	C	Pacientes epilépticos	30							93,3%
Tan, Slick, Strauss y Hultsch (2002)	S	Estudiantes análogos	25	Si	Si*	No	Si	Puntuaciones significativamente por debajo del nivel del azar en ítems fáciles o difíciles	32%	100%
	C	Estudiantes controles	27							
Slick y cols. (2003)	C	Pacientes amnésicos	6					90% de ítems difíciles correctos		100%
Macciocchi y cols. (2006)	C	Pacientes con TCE severos	71					90% de respuestas correctas totales		96%

Nota: G= grupo; S= grupo empleado como simulador; C= grupo control; Es= escenario; En= entrenamiento; Inc= incentivo; Inf= informe posterior; TCE= traumatismo craneoencefálico; Si*= autoentrenamiento, entrenamiento no proporcionado por los evaluadores sino buscado por los participantes.

entronca con investigaciones anteriores, como la de Guilmette y cols. (1993) con el DMT o Tombaugh con el TOMM (1996). La principal carencia del VSVT es que, si bien ha demostrado su especificidad en pacientes con diferentes trastornos y su sensibilidad a la hora de detectar análogos simuladores y pacientes en litigio, todavía no disponemos de ningún estudio que investigue el comportamiento de esta prueba en diseños de grupos conocidos, si bien se considera que, gracias a la investigación que lo sustenta, el VSVT cumple los criterios Daubert para su admisibilidad como evidencia científica (Thompson, 2002). Los principales resultados con el VSVT se presentan en la Tabla 11.

El **Computerized Assessment of Response Bias** (Allen, Conder, Green y Cox, 1997) es otra prueba que se administra por ordenador y que emplea dígitos como estímulos. En el **Computerized Assessment of Response Bias (CARB)** los intervalos distractores son de 1,5, 2,5 y 3,5 segundos por bloque, y en ellos el sujeto debe realizar una tarea distractora (contar de 20 hacia atrás). Pasar el CARB supone obtener una puntuación total igual o superior al 89% de respuestas correctas. Una de las grandes ventajas de este test es que permite una terminación temprana cuando los sujetos están realizando correctamente la tarea, lo que reduce el tiempo de administración y el tedio de los sujetos que ponen un esfuerzo adecuado en la evaluación.

Algunos estudios han puesto de manifiesto que los pacientes con traumatismos craneoencefálicos moderados y severos, así como pacientes con diversas enfermedades neurológicas (Allen, Iverson y Green, 2002; Green e Iverson, 2001) realizan adecuadamente el CARB. Por otro lado, los individuos con daño cerebral leve o cuestionable en litigio obtienen peores puntuaciones que los grupos con daños más severos, lo que es interpretado como una exageración de síntomas (Green e Iverson, 2001; Green, Rohling, Lees-Haley y Allen, 2001). La utilidad del CARB para detectar

esfuerzo insuficiente en pacientes en litigio (N=326) ha sido recientemente replicada (Gervais, Rohling, Green y Ford, 2004). También se han realizado estudios con análogos favorables a la CARB (Iverson, Ropacki, Allen y Elias, 2001), y parece adecuado para la detección de un esfuerzo subóptimo en fibromialgia (Gervais y cols., 2001) y depresión (Rohling, Green, Allen e Iverson, 2002).

El principal problema del CARB lo constituyen sus elevadas desviaciones típicas, que hacen que en ocasiones exista un solapamiento entre daño cerebral y simulación. Por otro lado, todavía no disponemos de investigaciones independientes a los autores de la prueba que repliquen los hallazgos con la CARB, por lo que consideramos que carece todavía del apoyo suficiente para recomendar su uso.

La **versión computadorizada del Portland Digit Recognition Test**, estudiada por Rose, Hall, Szalda-Petree y Bach (1998), ha demostrado una exactitud en la clasificación muy elevada, tanto empleando los puntos de corte propuestos por Binder para el número total de respuestas correctas como empleando también la latencia de respuesta o mediante el empleo de funciones discriminantes. El problema es que los datos no han sido validados, por lo que se necesita más trabajo antes de poder establecer conclusiones sobre la utilidad de esta prueba, especialmente en cuanto a los tiempos de reacción (Bianchini, Mathias y Greve, 2001).

2.3.3. Otras variaciones:

Además de el empleo de estímulos alternativos y la administración por ordenador, los test de elección forzosa han sido sometidos a otro tipo de modificaciones en los últimos años, entre las que podemos diferenciar: el aumento del número de alternativas y de respuestas correctas posibles; el diseño de pruebas que no tienen la apariencia de tests de memoria y; la combinación de distintos métodos de detección de simulación combinados con el método de elección forzosa.

Entre las pruebas que emplean más de dos alternativas está el **Letter Memory Test** (Inman, Vickery, Lamb, Edwards y Smith, 1998) que se administra por ordenador y, como su propio nombre indica, en lugar de emplear números sus ítems están compuestos por letras. En concreto, consiste en 45 estímulos contruidos con las 10 primeras letras consonantes del abecedario. Un ensayo comienza con la presentación de un estímulo objetivo, durante 5 segundos. Después de un intervalo de 5 segundos el sujeto tiene que discriminar el estímulo objetivo de los distractores. El Letter Memory Test (LMT) está compuesto por nueve bloques en los que se va incrementando el número de estímulos a recordar (de tres a cinco letras) así como el número de distractores (de 2 a 4). Inman y cols. (1998) han establecido el “suelo empírico” de esta prueba a partir de la ejecución de pacientes con trastornos neurológicos en el 93% de aciertos. Empleando este punto de corte, la especificidad (83-95%) y sensibilidad (100%) de esta prueba han sido muy elevadas, tanto en poblaciones de análogos estudiantes y voluntarios como en pacientes forenses sospechosos de simulación. Otras investigaciones también han constatado el buen funcionamiento de la prueba con análogos estudiantes y con daño cerebral, corroborando la excelente especificidad de la prueba (100%), si bien la sensibilidad encontrada ha sido algo inferior a la referida por Inman y cols. (1998), situándola entre el 73% (Inman y Berry, 2002) y el 87% (Vickery y cols., 2004). Por otro lado, la prueba ha mostrado su eficacia dentro de una batería neuropsicológica amplia (Vickery y cols., 2004; Greub y Suhr, 2006), y los análogos a los que se advierte de la existencia de al menos una prueba de simulación en la evaluación no realizan mejor el LMT (Greub y Suhr, 2006), por lo que parece que esta prueba posee una buena validez aparente.

Tabla 12. Principales resultados en la investigación con el Letter Memory Test

Estudio	G	Tipo de sujetos	N	Es	En	Inc	Inf	Regla de clasificación	Sens	Espec
Inman y cols. (1998)	S	Sospechosos de simulación	19					Punto de corte <93% de respuestas correctas	95%	100%
	C	Pacientes neurológicos	21							
	S	Estudiantes análogos	100	Si	Si	Si	Si		83%	100%
	C	Pacientes neurológicos	32							
Inman y Berry (2002)	S	Análogos voluntarios	54						84%	
	C	Pacientes neurológicos y psiquiátricos	66							100%
Vickery y cols. (2004)	S	Análogos estudiantes con y sin daño cerebral	44	Si	No	Si	Si	Punto de corte <93% de respuestas correctas	73%	100%
	S	Análogos con daño cerebral	23	Si	Si	Si	Si		87%	
	C	Análogos voluntarios	23					Punto de corte <93% de respuestas correctas	82,6%	87%
Greub y Sulhr (2006)	S	Pacientes con TCE-L	49	Si	Si	No	Si	Punto de corte <93% de respuestas correctas	76%	
	C	Pacientes con TCE-L y estudiantes	53							96%

Nota: G= grupo; S= grupo empleado como simulador; C= grupo control; Es= escenario; En= entrenamiento; Inc= incentivo; Inf= informe posterior; TCE-L= traumatismo craneoencefálico leve; Sens= sensibilidad; Espec= especificidad.

En resumen, el LMT es una prueba que ha demostrado de modo consistente su excelente especificidad y, a pesar de que su sensibilidad se ha encontrado más variable en los distintos estudios, en todos ha sido muy adecuada (ver Tabla 12). Quizá la principal carencia es que sólo uno de los estudios ha analizado la prueba en sospechosos de simulación (Inman y cols., 1998), empleándose en el resto de las investigaciones análogos, por lo que los hallazgos han de ser replicados, si bien todo apunta a que es un test muy prometedor.

Otro test que emplea más de dos estímulos alternativos es el **Amsterdam Short-Term Memory Test** (ASTM; Schagen, Schmand, de Sterke y Lindeboom, 1997). La prueba consiste en un cuadernillo con 30 ítems más 2 de práctica. Cada ítem ocupa tres páginas: en la primera aparecen cinco palabras de una categoría semántica (medias, falda, camiseta, jersey, abrigo) que el sujeto debe leer en voz alta y memorizar. La segunda página presenta una tarea distractora, que consiste en una suma o una resta simple sin tiempo límite (por ejemplo, $27+15=$). En la tercera página aparecen cinco palabras de la misma categoría semántica que antes (traje, medias, falda, jersey, pijama). El sujeto debe indicar qué tres palabras aparecían en la primera serie, y se le da feedback sobre el número de palabras acertadas. La puntuación total es el número de respuestas correctas, siendo 90 la máxima puntuación. Un aspecto importante es que las palabras han sido elegidas de tablas de categorías, perteneciendo las cinco palabras presentadas en la primera página a las frecuencias más altas, mientras las dos palabras adicionales pertenecen a las frecuencias más bajas de dicha categoría semántica.

Schagen y cols. (1997) emplearon un grupo de pacientes con daño cerebral moderado-severo, análogos simuladores constituidos por los familiares de dichos pacientes (por considerar que estos están bien instruidos sobre las consecuencias del daño) y un grupo de controles. Empleando un punto de corte <86 , el ASTM clasificó

correctamente a todos los sujetos.

Entre las principales ventajas de este test, sus autores señalan las siguientes:

- el nivel del azar es más difícil de detectar para los sujetos que en otros SVT
- al haber siempre al menos una respuesta correcta, se incrementa la posible tendencia a simular
- la carga de memoria del test es mínima por dos motivos: no existe interferencia, ya que cada ítem emplea una categoría diferente. Además, la elección de las palabras basada en sus frecuencias aumenta las probabilidades de elegir la correcta, incluso cuando esta se ha olvidado realmente
- la tarea distractora aumenta la dificultad aparente de la prueba

El ASTM ha sido validado recientemente en un estudio con análogos, en el que se consiguió una sensibilidad y especificidad del 100%. En este estudio se comprobó que, a pesar de que muchos análogos identificaban el test como una posible medida de simulación, no lograban puntuar por encima del punto de corte de la prueba para evitar la detección.

Esta prueba nos parece una de las propuestas más interesantes para la detección de simulación, pero sorprendentemente ha estado sujeta a un número muy escaso de publicaciones, por lo que todavía es pronto para posicionarse en cuanto a su validez como medida del esfuerzo de los sujetos.

Una prueba que también emplea más de dos alternativas es el **Word Memory Test** (WMT; Green, Allen y Astner, 1996), que emplea palabras como estímulos alternativos. Se administra por ordenador, e incluye múltiples medidas de esfuerzo y memoria. Consiste en 20 pares de palabras simples (perro/gato, hombre/mujer...) que el paciente debe aprender. La lista se presenta dos veces y cada par de palabras se presenta durante 6 segundos. En el reconocimiento inmediato se muestran pares en los que sólo

una de las palabras estaba en la lista original (perro/conejo), que el sujeto debe identificar. Tras un intervalo de 30 minutos se aplica el reconocimiento demorado. A esto le sigue un procedimiento de elección forzosa en el que al sujeto se le presenta una palabra y debe elegir la otra del par entre ocho opciones. Después se administra la tarea de pares asociados, en la que el evaluador dice la primera palabra de cada par y el sujeto debe responder la palabra asociada a ella. Por último se administra el recuerdo libre, en cuyo subtest el sujeto debe recordar todas las palabras de la lista sin importar el orden ni el emparejamiento, y se repite tras una demora de 20 minutos.

Sus autores afirman que el Word Memory Test (WMT) puede emplearse, además de como un test de esfuerzo, como un test de memoria verbal, mediante la adición de dos subtests. Sin embargo, todavía no se disponen de datos normativos suficientes con este propósito interpretativo, por lo que se ha sugerido que ha de ser empleado con otras medidas estándar de memoria hasta que dispongamos de más datos sobre sus propiedades psicométricas (Wynkoop y Denney, 2001).

Lo que ha sido más estudiado es el comportamiento de la prueba como medida indicativa de simulación. En este sentido, Green, Iverson y Allen (1999) encontraron que los pacientes litigantes obtenían puntuaciones más bajas en el WMT que pacientes con lesiones más severas. Además, los pacientes con hallazgos en neuroimagen obtenían puntuaciones correctas superiores al 90% en las medidas de esfuerzo. Por otro lado, los análogos obtienen puntuaciones muy bajas en el WMT (Iverson, Green y Gervais, 1999). Atendiendo a estos hallazgos, Tan y cols. (2002) obtuvieron un 100% de sensibilidad y especificidad entre estudiantes controles y análogos estableciendo como punto de corte el 82,5% de respuestas correctas, si bien sólo el 8% de los análogos puntuaron significativamente por debajo del azar. Por otro lado, el WMT parece una prueba especialmente útil para detectar esfuerzo insuficiente en litigantes

(Gervais y cols., 2004).

En resumen, el WMT parece una herramienta prometedora en la detección de simulación, pero necesita un mayor apoyo empírico para recomendar su empleo rutinario en las evaluaciones forenses.

El **Nonverbal Forced Choice Test** (Frederick y Foster, 1991) es uno de los pocos test de simulación que no toma la apariencia de una prueba de memoria, sino que se basa en la habilidad de resolución de problemas, mediante un formato de elección forzosa con dos alternativas. Se construyó a partir del Test of Nonverbal Intelligence (Brown, Sherbenou y Johnsen, 1982), y su interpretación se basa en tres índices: la puntuación total, el declive de la ejecución en función de la dificultad de los ítems y la proporción de consistencia (basado en los pares contestados correctamente en el mismo nivel de dificultad en relación al número de respuestas correctas totales posibles). Si la proporción de declive×consistencia es mayor de $-.0041$ y la puntuación se encuentra por debajo del nivel del azar, el sujeto es sospechoso de simulación (Frederick y Foster, 1991; Frederick, Sarfaty, Johnston y Powel, 1994). Esta prueba ha demostrado su utilidad con análogos, pacientes forenses y pacientes neuropsicológicos en litigio (Bianchini, Mathias y Greve, 2001). Rose y cols. (1998) han cuestionado la utilidad del Nonverbal Forced Choice Test, puesto que siguiendo las indicaciones de los autores del test encuentran una sensibilidad del 33% y una especificidad del 75%, de modo que la clasificación total correcta de esta prueba no supera el nivel del azar. En definitiva, los estudios que han empleado esta prueba necesitan ser replicados para que se pueda determinar su utilidad real.

El **Word Completion Memory Test** (WCMT; Hilsabeck, LeCompte, Marks y Grafman, 2001) se diferencia de los test tradicionales de simulación en que no parece una prueba de memoria explícita, sino que se basa en los principios de facilitación y

memoria implícita. Ofrece tres puntuaciones: inclusión (I), exclusión (E) y la diferencia entre estas dos puntuaciones (R). Hilsabeck y Gouvier (2005) demostraron que tanto los pacientes con déficits de memoria como los sujetos controles obtenían puntuaciones R e I significativamente mayores que las de los análogos, mientras sus puntuaciones E eran significativamente menores que las de los análogos. Empleando un punto de corte <9 para la puntuación R ningún paciente o control fue incorrectamente clasificado como simulador (100% de especificidad), mientras la sensibilidad fue del 93,3% entre los análogos. Además, el WCMT mostró una adecuada validez convergente, divergente y test-retest, mostrándose como una herramienta muy prometedora, si bien está pendiente de ser sometida a más investigación.

Otra de las tendencias de los estudios más actuales y avanzados en simulación parece ser la de mejorar la exactitud de la clasificación a través de la combinación de estrategias de detección. Con este propósito fue diseñado el **Multidimensional Investigation of Neuropsychological Dissimulation** (MIND, Thomas y Wanlass, 1994), que emplea como estrategias el formato de elección forzada, las discrepancias en tareas de recuerdo y reconocimiento, la consistencia a través de medidas repetidas, y la identificación de patrones de ejecución inconsistentes con el daño cerebral. La prueba se divide en seis partes, siendo las tres primeras idénticas a las seis últimas excepto porque varía la presentación del orden de las tarjetas y porque se indica al sujeto que podría encontrar la segunda parte más difícil, cuando en realidad es exactamente igual que la primera. El sujeto tiene que realizar, pues, tres tareas diferentes. La primera tarea, que conforma las partes I y IV, consiste en decir si unas tarjetas contienen un número de elementos par o impar en 16 tarjetas con ítems no-agrupados y cuatro con ítems agrupados. Se registra la respuesta (par/impar) y el tiempo empleado. Lo esperado es que el tiempo para contar los ítems agrupados sea inferior, que aumente a medida que

aumenten los iconos presentados en las tarjetas, y que el tiempo empleado para responder a cada tarjeta y el tiempo total sea igual o disminuya de la parte I a la IV debido al efecto práctica.

Las partes II y V consisten en recordar tantos iconos como sea posible de los presentados en la parte I y IV en cinco minutos. Las partes III y VI consiste en enseñar al sujeto pares de iconos para que reconozca cuál se le ha presentado con anterioridad en la primera parte de la evaluación. El protocolo proporciona ocho índices de puntuación diferentes obtenidos de la combinación de estas tareas.

Holmquist y Wanlass (2002) realizaron una validación del instrumento empleando como grupos a sujetos normales, análogos simuladores y pacientes con daño cerebral de moderado a severo. El MIND parece ser un instrumento sensible a déficits funcionales específicos del daño neurológico de moderado a severo, detectando problemas de memoria y atención típicos de estos grupos de sujetos. Utilizando las restantes seis variables predictoras de las ocho iniciales se consiguió una clasificación correcta del 68% de los sujetos, consiguiendo reducir el número de falsos negativos al 10%. Clasificando a los sujetos en dos grupos (simuladores y daño cerebral) se alcanzó una exactitud en la clasificación del 82%, por lo que parece un instrumento prometedor para identificar individuos exagerando déficits neuropsicológicos de leves a moderados, aunque necesita más investigación para probar su utilidad.

El **Test of Cognitive Abilities** (TOCA, Rogers, 1996), es una prueba que examina tanto la simulación como la habilidad cognitiva de los sujetos basándose en diferentes estrategias: efecto suelo, magnitud del error, curvas de ejecución, Symptom Validity Test y tiempo de respuesta. El TOCA es una prueba administrada por ordenador, y emplea un formato de elección múltiple en el que el sujeto elige dos de cuatro respuestas posibles para completar una secuencia eligiendo el final y el principio

correctos. Está dividida en tres secciones: la Sección de Secuenciación, la Sección de Diseños y la Sección de Frases.

En el estudio de validación inicial del TOCA, Bender y Rogers (2004) emplearon dos grupos de análogos simuladores (con y sin advertencia sobre el empleo de métodos de detección de simulación) a los que compararon con pacientes con daño cerebral y con controles no dañados. La magnitud del error fue el método más efectivo a la hora de diferenciar simuladores de pacientes (clasificación correcta del 94% de los sujetos); moderadamente efectivas fueron las estrategias de efecto suelo y tiempo de reacción, apareciendo la estrategia SVT como la menos efectiva debido a su pobre sensibilidad. La hipótesis de si múltiples estrategias de detección mejoran los niveles de clasificación no pudo ser efectivamente probada debido al efecto techo de la magnitud del error, si bien la función discriminante compuesta por la magnitud del error y el tiempo de respuesta alcanzaron un nivel de clasificación correcta del 96%. Este estupendo resultado necesita ser replicado.

2.4. ¿Cuáles son los mejores tests específicos para detectar la simulación?

Por desgracia, disponemos de muy pocos estudios que comparen diversas pruebas específicas para poder determinar la superioridad de unas sobre otras. De estos pocos estudios, la mayoría se han centrado en comprobar la superioridad de diversas pruebas de elección forzosa sobre el test de los 15-ítems de Rey o el Dot Counting. Por eso, una de las pocas conclusiones que podemos extraer es que ambas pruebas son inferiores a aquellas basadas en el método de elección forzosa. Ambas pruebas se han mostrado inferiores al Hiscock Forced Choice Test (Cato y cols., 2002), al Word Completion Memory Test y al Multi-Digit Memory Test (Hilsabeck y Gouvier, 2005). Además, el 15-ítems de Rey también se ha mostrado inferior al Digit Memory Test (Vickery y cols., 2001; Inman y Berry, 2002), al Portland Digit Recognition Test

(Greiffenstein y cols., 1994), al, al Letter Memory Test (Inman y Berry, 2002; Greub y Suhr, 2006) y al Amsterdam Short-Term Memory Test (Merten y cols., 2005). Por su parte, el Dot Counting Test se ha mostrado inferior al Portland Digit Recognition Test, 21-Item Test y al Nonverbal Forced Choice Test (Rose y cols., 1998), así como al Victoria Symptom Validity Test.

En definitiva, tanto el 15-ítems de Rey como el Dot Counting se han mostrado inferiores a todas las pruebas de elección forzosa con las que han sido comparadas.

Menos numerosos y concluyentes son los estudios que comparan diversas pruebas de elección forzosa entre sí. Lo que parece claro es la inferioridad del 21-Item Test en comparación con las pruebas que siguen su mismo método: el Portland Digit Recognition Test (Rose y cols., 1998; Vickery y cols., 2001), el Digit Memory Test (Vickery y cols., 2001; Inman y Berry, 2002), el Letter Memory Test (Inman y Berry, 2002; Inman y cols., 1998) y el Victoria Symptom Validity Test (Inman y cols., 1998).

Con respecto a las otras pruebas de elección forzosa, la investigación no parece tan concluyente. A continuación se expone un esquema de las conclusiones de diversos estudios (el símbolo “>” debe entenderse en este esquema como “superior a”):

- Digit Memory Test > Portland Digit Memory Test (Vickery y cols., 2001)
- Word Memory Test > Computerized Assessment of Response Bias > Test of Memory Malinger (Gervais y cols., 2004)
- Letter Memory Test > Digit Memory Test > Test of Memory Malinger (Vickery y cols., 2004)
- Word Memory Test > Victoria Symptom Validity Test > Test of Memory Malinger (Tan y cols., 2002)
- Victoria Symptom Validity Test y Letter Memory Test equiparables (Inman y cols., 1998)

Teniendo en cuenta el conjunto de los datos, consideramos recomendable el empleo del Digit Memory Test o de su equivalente por ordenador, el Victoria Symptom Validity Test, puesto que además de mostrarse más efectivos que otros test de elección forzosa, gozan del suficiente apoyo en la literatura como para recomendar su empleo. Por otro lado, parece que tanto el Portland Digit Recognition Test como el Test of Memory Malinger se muestran ligeramente inferiores a estos, pero debido a que sistemáticamente han recibido el apoyo de numerosos estudios, su empleo también nos parece adecuado. El Word Memory Test, el Letter Memory Test y el Computerized Assessment of Response Bias parecen instrumentos realmente prometedores, pero todavía no disponen del suficiente apoyo empírico. Por último, no el empleo del 21-Item Test no parece demasiado útil, obteniendo resultados comparables a los del 15-ítems o al Dot Counting Test.

2.5. El empleo conjunto de diversos test específicos de simulación:

A pesar de la insistencia en la necesidad de emplear diversas medidas para detectar simulación, son realmente escasos los trabajos que exploran la eficacia de la aplicación conjunta de varias pruebas específicas. De las pocas publicaciones existentes se deduce que al emplear funciones discriminantes que engloben diversas medidas se incrementa la exactitud de las clasificaciones, tanto si se trata de medidas del mismo test como de pruebas diferentes. Por ejemplo, cuando en la versión por ordenador del Portland Digit Recognition Test se tiene en cuenta también su latencia de respuesta se incrementa el poder de clasificación de la prueba (Rose y cols., 1995). Lo mismo sucede con el Victoria Symptom Validity Test (VSVT) cuando al resultado de los ítems difíciles se añade el de los fáciles o cuando se combina con otra prueba, como el TOMM, consiguiéndose en ambos casos una correcta asignación del 100% de los casos estudiados (Tan y cols., 2002). Otros estudios que indican la superioridad del empleo

conjunto de distintas pruebas en comparación con los resultados individuales de las mismas son el de Strauss y cols. (2002) con el VSVT, Reliable Digit Span y Dot Counting Test, el de Cato y cols. (2002) con el Recognition Memory Test, el Dot Counting, el Test de elección forzosa de Hiscock y el Test de los 15-ítems de Rey o el de Hilsabeck y Gouvier (2005) con el Word Completion Memory Test, el 15-ítems de Rey, el Dot Counting Test, el Multi-Digit Memory Test y el Recognition Memory Test.

Sin embargo, hay que tener en cuenta que no se trata de aplicar irreflexivamente todas las medidas de simulación de las que dispongamos, sino que deberemos intentar emplear aquellas que no posean una correlación demasiado elevada para no ser redundantes en nuestra evaluación. De lo contrario, no podremos emplear los resultados de las distintas pruebas aplicadas como evidencia convergente (Nelson y cols., 2003).

3. Los tests neuropsicológicos no específicos de simulación:

En los últimos años se han estudiado numerosos índices dentro de las pruebas neuropsicológicas tradicionales para ser empleados como indicativos de simulación. A continuación exponemos algunas de las pruebas cuyos índices cuentan, en la actualidad, con un mayor número de investigaciones.

El **California Verbal Learning Test** (CVLT; Delis, Kramer, Kaplan y Ober, 1987) es uno de los test neuropsicológicos tradicionales que más investigación ha suscitado en cuanto a la evaluación del esfuerzo insuficiente. Uno de los trabajos fundamentales en esta área es el de Millis, Putnam, Adams y Ricker (1995), que estudiaron la utilidad del CVLT en la detección de simulación con un grupo de pacientes con daño cerebral moderado-severo y un grupo sospechoso de simulación por haber obtenido puntuaciones en el rango del azar en un test de elección forzosa. Estos autores lograron una sensibilidad y especificidad del 91% mediante una función lineal

discriminante que incluía la puntuación del recuerdo total (ensayos 1-5), el recuerdo a largo plazo con claves y la discriminabilidad, y una sensibilidad y especificidad del 96% mediante el empleo de una función cuadrática. A pesar de estos excelentes resultados, no se encontraron diferencias significativas con respecto al empleo de puntos de corte para las variables aisladas de la prueba, que alcanzaron una clasificación correcta del 93% de los casos con la discriminabilidad y del 89% con los aciertos del reconocimiento. El recuerdo a largo plazo con claves y el número total de palabras recordadas (aciertos ensayo 1-5) obtuvieron resultados inferiores pero también satisfactorios (87% y 83%, respectivamente). Por tanto, especialmente la discriminabilidad parece una estrategia prometedora para detectar simulación, ya que con ella se obtuvo una exactitud de clasificación igual al de las funciones multivariantes más complejas. Posteriormente, Millis y Putnam (1997, citado en Curtis, Greve, Bianchini y Brennan, 2006) realizaron una ecuación de regresión logística con diversas variables del CVLT (primacía, pendiente, reconocimiento y sesgo de respuesta) que alcanzó una especificidad del 93% y una sensibilidad del 87%.

Los estudios de Millis han sido replicados varias veces, consiguiendo por norma general sensibilidades inferiores a las referidas en los trabajos originales. Por ejemplo, Sweet, Wolfe y cols. (2000), con un diseño de grupos conocidos obtuvieron sensibilidades inferiores a las planteadas en el estudio de 1995, si bien mediante el ajuste de los puntos de corte consiguieron incrementarlas. Otro trabajo que ha estudiado las propuestas de Millis es el de Curtis y cols., (2006) que, diferenciando los grupos mediante la propuesta de Slick y cols. (1999) y manteniendo una especificidad del 95%, no han obtenido una sensibilidad del 50% en ninguna de las puntuaciones estudiadas (46% en la regresión logística, 31% en la función discriminante, y 38% en el reconocimiento, que es la variable aislada del CVLT que sigue obteniendo mejores

resultados).

Ashendorf, O'Bryant y McCaffrey (2003) estudiaron la especificidad de las propuestas de Millis y cols. (1995, 1997) con personas mayores, estableciendo niveles superiores al 90% en la función discriminante, ecuación de regresión logística, discriminabilidad y recuerdo a largo plazo con claves, y ligeramente inferior al 90% en el reconocimiento (89,8%) y en el recuerdo total (82,7%).

Otros trabajos interesantes con el CVLT son el de Martens, Donders y Millis (2001), que han demostrado la utilidad de la prueba en pacientes con TCE en litigio, y el de Demakis (1999), que explora las diferencias cualitativas de los análogos a la hora de realizar el test, concluyendo que los simuladores son más inconsistentes a la hora de recordar las mismas palabras a través de los cinco ensayos de la prueba que los sujetos honestos.

La segunda edición del CVLT ha incluido una prueba de reconocimiento de elección forzosa opcional diseñada específicamente para evaluar el nivel de esfuerzo de los sujetos. En su manual se refiere que en torno al 90% de la muestra evaluada obtiene puntuaciones perfectas (16 correctas de 16 posibles) en esta tarea, y ninguno de los sujetos explorados comete dos errores. Así, los autores han propuesto que puntuaciones <15 sean consideradas como indicativas de una ejecución inválida del test (Delis y cols., 2000). Este índice se ha mostrado sensible a los efectos de la búsqueda de compensación económica (Moore y Donders, 2004), por lo que parece prometedor, pudiendo suponer una mejora sobre su edición predecesora.

En resumen, el CVLT ha mostrado una especificidad muy elevada en la detección de simulación. Tanto las funciones multivariadas como las variables aisladas del CVLT, en especial el reconocimiento, se han mostrado útiles, si bien la sensibilidad que alcanzan puede describirse como moderada en la mayoría de los trabajos.

Meyers, Morrison y Miller (2001) estudiaron el potencial de la tarea de reconocimiento del **Rey Auditory Verbal Learning Test** (AVLT; Rey, 1964) como índice de simulación. Para ello, establecieron el suelo empírico de esta tarea (puntuación por encima de la cual puntúan todos los pacientes con traumatismo severo) en un punto de corte de 9 aciertos de 15 posibles. Así, con una especificidad del 100% asegurada, detectaron al 50% de los análogos, y el 12% de los pacientes con TCE-L en litigio obtuvieron puntuaciones indicativas de un esfuerzo insuficiente.

Suhr y Gunstad (2000) emplearon distintas variables derivadas del AVLT: aprendizaje vs. reconocimiento (variable empleada anteriormente por Suhr, Tranel, Wefel y Barrash (1997) basada en el fallo en el reconocimiento a los 30 minutos de palabras recordadas al menos tres veces en los ensayos de aprendizaje), inconsistencia de memoria (variable empleada por Barrash, Suhr y Manzel (1998) basada en una ejecución pero en el recuerdo en el intervalo de 30 minutos que en el de 60), reconocimiento vs. recuerdo (porcentaje de palabras reconocidas a los 30 minutos menos porcentaje de palabras recordadas en el mismo intervalo), y extensión del aprendizaje (número de palabras recordadas en el ensayo 5 menos palabras recordadas en el ensayo 1). Estableciendo puntos de corte para una especificidad máxima, el fallo de al menos uno de los cuatro indicadores alcanzó una detección entre el 74 y el 85% de los análogos.

Suhr (2002) estudió con éxito el efecto de posición serial en la simulación con el AVLT, descubriendo que mientras los sujetos controles y los pacientes obtenían curvas de posición serial normales, los simuladores obtenían puntuaciones más bajas en primacía durante los ensayos de aprendizaje. Además, en el recuerdo demorado, los análogos obtenían puntuaciones inferiores de recencia y en la porción media de la lista.

En el manual de la **Figura Compleja de Rey** (*Rey Complex Figure Test*, RCFT; Meyers y Meyers, 1995; adaptación española: de la Cruz, M.V., Seisdedos, N. y Cordero, A., 1997) se presentan los denominados “patrones de errores de memoria” (PEM), clasificados como de atención, codificación, almacenamiento, recuperación y normal/otros (posteriormente se descubrió que la codificación podría ser una variante del PEM de atención, por lo que generalmente ambos se engloban bajo la denominación PEM de atención). Meyers, Bayless y Meyers (1996) descubrieron una consistente relación entre PEMs y estatus funcional (a mayor daño funcional, menor PEM), encontrándose PEMs de recuperación y atención sólo en sujetos muy dañados. Así, Meyers y Volbrecht (1999) estudiaron la utilidad de los PEMs para predecir simulación, consiguiendo diferenciar sujetos con traumatismo leve litigantes y no litigantes, ya que sólo entre los primeros se producían PEMs de atención (definido por una puntuación en el recuerdo a los tres minutos, a los 30 minutos y en el reconocimiento inferior a una puntuación T de 24) y de recuperación (definido por una bajada mayor de 3 puntuaciones T en el reconocimiento con respecto al recuerdo a los 3 minutos). Además, en los estudios sucesivos, el 100% de los sujetos sospechosos de simulación estudiados y el 80% de los análogos obtuvieron dichos PEMs de atención o recuperación.

Lu, Boone, Cozolino y Mitchell (2003) intentaron replicar estos hallazgos, pero sin éxito, ya que mediante el PEM de atención la sensibilidad hallada fue muy baja (26%), mientras que el PEM de recuperación no consiguió un buen nivel de clasificación ni de los simuladores ni de los pacientes con problemas de memoria como no simuladores. Estos autores también han estudiado diversas variables del RCFT para la detección de simulación, en concreto: la puntuación en la copia, en el recuerdo inmediato, en el reconocimiento y los falsos positivos en la tarea de reconocimiento.

Dichos índices no se mostraron demasiado sensibles para diferenciar a los sujetos con esfuerzo insuficiente, ya que los valores con una especificidad adecuada (superior al 90%) se asociaron a sensibilidades no superiores al 50%, mientras los intentos para incrementar la sensibilidad llevaban asociadas disminuciones en la especificidad a niveles inaceptables. Un análisis cualitativo reveló que el grupo sospechoso de simulación cometía más errores de reconocimiento “atípicos” que los grupos de pacientes con daños en memoria (ítems 1, 4, 6, 10, 11, 16, 18 y 21). Teniendo en cuenta todo lo anterior, los autores proponen el empleo de una puntuación combinada [puntuación en copia+(reconocimiento-errores atípicos en el reconocimiento)×3], con el que establecen diversos puntos de corte y consiguen una especificidad del 94% y una sensibilidad del 74%, y por tanto superior al resto de las medidas exploradas en este estudio.

En definitiva, parece que los distintos índices propuestos como indicadores de simulación en el Test de la Figura Compleja de Rey obtienen sensibilidades bajas, con excepción de los errores atípicos, variable pendiente de más investigación para poder determinar su utilidad.

El **STROOP, Test de colores y palabras** (Golden, C.J., 1994) también ha sido estudiado como indicador de simulación. Osimani, Alon, Berger, y Abarbanel (1997), aplicando una versión por ordenador de la prueba, encontraron que los análogos obtenían tiempos de reacción significativamente prolongados con respecto a los pacientes con daño cerebral, un porcentaje de errores incrementado y un efecto STROOP ausente o invertido (es decir, tiempos de reacción más largos en la condición congruente que en la incongruente). Dado que esta alteración del efecto STROOP nunca está presente en déficits cognitivos orgánicos, estos autores lo proponen como un patrón característico de la simulación.

El STROOP también ha sido estudiado en la detección de la simulación del estrés post-traumático mediante el paradigma emocional. En este sentido, un grupo de análogos actores entrenados para simular fue capaz de producir un enlentecimiento general en la latencia de respuesta, de modo similar a como lo hacen los verdaderos pacientes. Sin embargo, no fueron capaces de modular la latencia de respuesta en función del contenido del estímulo, patrón que caracteriza a los pacientes con estrés post-traumático.

Uno de los primeros y más influyentes estudios con el **Wisconsin Card Sorting Test** (WCST; Heaton, 1981) sobre la detección de simulación se lo debemos a Bernard, McGrath y Houston (1996), que diseñaron funciones discriminantes para diferenciar a un grupo sospechoso de simulación de pacientes con TCE, con daños neurológicos mixtos y de un grupo control normal, obteniendo una buena especificidad (entre el 92 y el 100%) pero una sensibilidad variable dependiendo de los grupos (entre el 58 y el 100%). Donders (1999) replicó la buena especificidad de esta fórmula, obteniendo sólo un 5% de falsos positivos.

Suhr y Boyer (1999) realizaron dos regresiones logísticas con esta prueba empleando las variables número de categorías y fallo para mantener el set: una para diferenciar análogos y controles (consiguiendo una sensibilidad del 71% y una especificidad del 87%), y otra para diferenciar pacientes sospechosos de simulación y pacientes con daño cerebral (con una sensibilidad del 82% y una especificidad del 93%).

Posteriormente, Miller, Donders y Suhr (2000) aplicaron las fórmulas de Bernard y de Suhr a un grupo de pacientes con TCE que habían obtenido puntuaciones indicativas de esfuerzo insuficiente atendiendo a tests específicos de simulación, y ninguna de las dos fórmulas consiguió identificar a ninguno de estos sospechosos de

simulación, por lo que los autores concluyen que hay que extremar la precaución al emplear dichas fórmulas debido a su baja sensibilidad. Inman y Berry (2002) también estudiaron la fórmula de Bernard y cols. (1996) realizando una validación cruzada en la que hallaron que, estableciendo una especificidad del 100%, esta función obtenía una sensibilidad extremadamente pobre (9%).

Otro estudio que estudia las fórmulas de Bernard y de Suhr es el de Greve y Bianchini (2002). Estos autores se centran en su especificidad, concluyendo que la fórmula de Suhr y Boyer (1999) obtiene un número inaceptablemente alto de falsos positivos, mientras la de Bernard y cols. (1996) mantiene una especificidad adecuada en los casos de TCE leves y moderados, pero no con TCE severos, accidentes cerebrovasculares y estudiantes.

Ashendorf y cols. (2003) también cuestionan la utilidad de estas fórmulas, ya que encuentran una especificidad inaceptablemente baja entre personas mayores, llamando la atención sobre el efecto que la edad tiene en el WCST, que debe ser tenido en cuenta para el empleo de índices indicativos de simulación en esta prueba.

King, Sweet, Sherer, Curtiss y Vanderploeg (2002), realizaron una regresión logística consistente en el número de categorías completadas, número de fallos al mantener el set y porcentaje de respuestas a nivel conceptual, consiguiendo una clasificación correcta de los TCE entre el 82 y el 99%, mientras permitió una detección del 70% de los sospechosos de simulación. Estos autores también aplican las fórmulas de Bernard y Suhr; la primera de ellas consiguió una clasificación entre el 73 y el 95% de los pacientes con TCE, detectando al 63% de los sospechosos de simulación, mientras la de Suhr diferenció entre un 75 y un 85% de los pacientes con traumatismos, detectando al 59% de los sospechosos como simuladores. En definitiva, en este estudio se desprenden unas especificidades variables dependiendo de la cronicidad y severidad

del traumatismo, pero que se sitúan en unos niveles aceptables, mientras la sensibilidad se sitúa en unos niveles bajos, excepto la propuesta de los autores del artículo, que se sitúa en niveles moderados.

Tomados en conjunto, de estos trabajos podemos concluir que el WCST es una prueba que se ve muy influida por factores como la edad o la severidad de los traumatismos, por lo que la especificidad de las fórmulas propuestas como indicativas de simulación en esta prueba son muy variables dependiendo de los grupos de estudio. Por tanto, podría ser útil para la detección del esfuerzo insuficiente en casos de TCE-L, si bien su sensibilidad tampoco se ha mostrado consistentemente elevada a lo largo de las investigaciones, por lo que los datos de esta prueba han de ser interpretados con cautela.

El **Recognition Memory Test** (RMT; Warrington, 1984) fue diseñado para evaluar la memoria para caras y palabras presentadas visualmente. Cada subtest (memoria de reconocimiento para palabras –RMW- y para caras –RMF) consiste en 50 estímulos objetivo y 50 distractores presentados en un formato de elección forzosa de dos alternativas, por lo que el test puede ser interpretado en el contexto de la teoría de la probabilidad. Así, una puntuación por debajo de 20 se considera significativamente por debajo del azar. Siguiendo esta lógica, Millis y Putnam (1994) encontraron que la media en la ejecución de este test entre pacientes con TCE-L en búsqueda de compensación económica se encontraba por debajo de dicho nivel del azar tanto en el subtest de palabras como en el de caras.

Iverson y Franzen (1998) también evaluaron la eficacia del RMT en la detección de simulación empleando pacientes neuropsicológicos con y sin daño de memoria y un grupo de análogos. En este estudio emplearon una función discriminante con la que lograron clasificar correctamente al 100% de los sujetos, y un 96,7% en la validación

cruzada. También proponen puntos de corte (por debajo de 38 para el RMW y 26 en el RMF) con los que clasificaron correctamente el 95% y el 65% de los análogos respectivamente. Por último, también analizaron los niveles de ejecución al azar, descubriendo que el 35% de los análogos puntuaban significativamente por debajo del azar, y el 55% en el rango del azar, mientras que casi todos los pacientes (entre el 90 y el 95%) puntuaban significativamente por encima de dicho nivel.

Contradiendo estos resultados, Cato y cols. (2002) no encuentran diferencias en la ejecución del RMT entre varios grupos de análogos, estudiantes controles y pacientes con TCE-L, ya que estos últimos obtienen puntuaciones muy bajas, por lo que estos autores desaconsejan el empleo de este test como indicador de simulación. Sin embargo, sí que se encuentran diferencias estadísticamente significativas entre la ejecución de los análogos y de los sujetos controles en este test (Hilsabeck y Gouvier, 2005), consiguiéndose una clasificación correcta del 100% de los sujetos sin daño, con una sensibilidad del 87% el subtest de palabras y del 67% el de caras en la detección de análogos.

En definitiva, el RMT, especialmente el subtest de palabras, parece un instrumento muy prometedor en la detección de simulación, ya que ha probado su eficacia tanto en análogos como en sujetos en litigio, si bien queda todavía por aclarar el nivel de especificidad de la prueba.

La simulación también ha intentado ser detectada mediante los tests que evalúan la función motora. Se han realizado diversos estudios con el **Finger Tapping Test** en los que esta prueba no se ha mostrado muy fiable (Greiffenstein, Baker y Gola, 1996; Trueblood y Schmidt, 1993). En un trabajo más reciente, Larrabee (2003^a) estableciendo un punto de corte de 63 para la suma de la mano dominante y la no-dominante, consiguió clasificar correctamente al 40% de los sujetos clasificados bajo la

etiqueta de “simulación definitiva” de Slick y cols. (1999), y al 93,5% de los pacientes con traumatismos moderados-severos. También Arnold y cols. (2005) encuentran una sensibilidad moderada en el Finger Tapping, siendo la puntuación más sensible a la hora de diferenciar a un grupo sospechoso de simulación de distintos grupos diagnósticos la media de la mano dominante. Uno de los hallazgos curiosos de estos autores es que las diferencias de género asociadas a este test se mantienen también entre los sujetos simuladores. Así, el Finger Tapping parece más eficaz a la hora de detectar mujeres simuladoras (con un punto de corte de 28 la especificidad se sitúa en el 92% y la sensibilidad en el 61% entre las mujeres, mientras el mejor punto de corte para los hombres es de 35, con el que la especificidad es del 90% y la sensibilidad del 50%). Según estos autores, esta prueba requiere del empleo de distintos puntos de corte para diferentes grupos de comparación, ya que si se emplea el mismo se obtiene un número muy alto de falsos positivos, lo que quizá pueda explicar el mal resultado de esta prueba en investigaciones anteriores.

Greiffenstein, Baker y Gola (1996), basándose en la teoría de que las tareas motoras que implican mayor integración sensorio-motora son más sensibles al daño cerebral, estudian si el patrón encontrado en las pruebas que miden esta área es útil para detectar simuladores. De acuerdo con la teoría, el grupo con TCE obtuvo un resultado en el Grip Strength mayor que en el Finger Tapping y en este mayor que en el Pegboard, mientras el grupo con SPC en búsqueda de compensación obtuvo el patrón de puntuaciones contrario. Sin embargo, este resultado no pudo ser replicado en trabajo de Rapport, Farchione, Coleman y Axelrod (1998), ya que estos autores encontraron, por un lado, que los análogos reproducen en estas pruebas el mismo patrón que los pacientes con TCE, mientras que en sujetos controles estudiantes se encuentra el patrón denominado “no fisiológico”, que implica la obtención de mejores puntuaciones en

tareas motoras más demandantes. La explicación es que, puesto que estos sujetos son universitarios, tienen un gran desarrollo del sistema motor fino. Lo curioso es que este mismo “patrón no fisiológico” se encuentra también entre los análogos que han recibido entrenamiento, pero no entre los análogos no entrenados. En definitiva, el análisis del patrón motor con la finalidad de establecer esfuerzo insuficiente no ha recibido el suficiente apoyo.

3.1. Baterías neuropsicológicas:

La **Wechsler Adult Intelligence Scale-R** (WAIS; Weschler, 1981) es una de las baterías neuropsicológicas más utilizadas, y consecuente de las que más investigación se desprende en general y sobre el tópico de la simulación en particular (Heaton, Smith, Lehman y Vogt, 1978; Binder y Willis, 1991; Martin y Franzen, 1992; Mittenberg, Azrin, Millsaps y Heilbronner, 1993; Mittenberg, Theroux-Fichera, Heilbronner, y Zielinski, 1995; Trueblood, 1994; Trueblood y Schmidt, 1993). Uno de los hallazgos más replicados es el de la utilidad del Reliable Digit Span a la hora de diferenciar simulación y daño cerebral. Este índice fue propuesto por Greiffenstein y cols. (1994) y consiste en la suma del número de dígitos que el sujeto repite sin error en los dos ensayos en dígitos hacia delante y en dígitos hacia atrás. Empleando un punto de corte de 7, estos autores consiguieron identificar una proporción bastante elevada del grupo de sospechosos de simulación (68-70%). El problema es que un número nada despreciable de pacientes fueron clasificados como simuladores con el empleo de este índice (el 27% de los pacientes con TCE y el 11% de los pacientes con SPC). Meyers y Volbretch (1998) realizaron una validación cruzada de este índice, clasificando incorrectamente como simuladores sólo al 4% de los pacientes con TCE-L, mientras casi la mitad de los pacientes con TCE-L en litigio fueron identificados como simuladores. También Larrabee (2003^a) ha estudiado este índice, obteniendo resultados

similares en cuanto a la clasificación de pacientes (el 93,5% de los pacientes con TCE moderados/severos fueron correctamente clasificados), pero detectando sólo el 50% de los sujetos clasificados en el nivel de “simulación definitiva” (atendiendo a la clasificación de Slick y cols., 1999). Bajo nuestro punto de vista, hay que ser cauto a la hora de emplear este índice, puesto que su especificidad no parece lo suficientemente elevada, mientras que su sensibilidad ofrece bastantes dudas.

Otro modo de estudiar la clasificación de sujetos simuladores en el WAIS es mediante el empleo de funciones discriminantes. En este sentido, Mittenberg y cols. (1995) mediante un análisis basado en la diferencia Vocabulario-Digit Span fueron capaces de clasificar correctamente al 79% de los sujetos con daño cerebral no-litigantes y al 71% de los análogos. Además, dicha función discriminante se empleó con éxito en la validación cruzada de diversos grupos de simuladores.

Otras investigaciones no han obtenido resultados tan positivos. Por ejemplo, Schwartz y cols. (1998) no encontraron diferencias estadísticamente significativas entre pacientes con daño cerebral y análogos en los cuatro subtest del WAIS estudiados (Vocabulary, Similarities, Picture arrangement y Digit Symbol). Meyers, Galinsky y Volbretch (1999) encuentran una tendencia de los pacientes con TCE severos a puntuar por encima de los TCE-L en litigio, sin embargo estas diferencias no son estadísticamente significativas. Inman y Berry (2002) tampoco encontraron diferencias estadísticamente significativas entre simuladores y no-simuladores en el subtest Información, mientras la sensibilidad del Digit Span y del Digit Symbol encontrada fue muy baja (27% y 2% respectivamente).

En definitiva, parece que el empleo de funciones discriminantes en el WAIS es útil a la hora de detectar simulación, mientras que el establecimiento de puntos de corte en los distintos subtest no se ha mostrado demasiado efectivo, puesto que o bien un

porcentaje de pacientes honestos son incorrectamente clasificados o bien el nivel de sensibilidad es demasiado bajo.

Otra de las baterías más extensamente empleadas en neuropsicología es la **Wechsler Memory Scale** (Wechsler, 1945). Según Schwartz y cols. (1998), de todos los subtest que componen la prueba, los únicos que permiten diferenciar análogos de pacientes reales son la Reproducción Visual Inmediata y el Recuerdo Demorado, por lo que esta batería no parece especialmente útil en la detección de simulación. En cuanto a la Wechsler Memory Scale-Revised (WMS-R; Wechsler, 1987), diversas investigaciones han puesto de manifiesto que el empleo de funciones discriminantes es útil en esta prueba. Por ejemplo, Mittenberg, Azrin, Millsaps y Heilbronner (1993), mediante una función discriminante con ocho medidas del WMS-R consiguieron una clasificación correcta del 91% de los pacientes con daño cerebral y los análogos. Bernard (1990) consiguió la clasificación correcta del 84% de los análogos y los sujetos control mediante una función discriminante con tan sólo dos medidas (Memoria Figurativa y Representación Visual I), que descendió al 74% en la validación cruzada. Sin embargo, esta función discriminante conlleva un inaceptable número de falsos positivos (en torno al 25%), por lo que debe emplearse con precaución.

Iverson y Franzen (1996) diseñaron una medida específica para detectar simulación dentro de la WMS-R: el suplemento de elección forzosa del subtest de Memoria Lógica, que se administra inmediatamente después de la prueba de recuerdo libre. Para la tarea de reconocimiento se crearon dos historias muy similares a la del subtest, y presentando partes de la historia el sujeto debía decidir cuál de las dos pertenecía a la historia original. Empleando un punto de corte <18 , los autores identificaron correctamente al 100% de los pacientes con daños en memoria y al 100% de los sujetos controles, mientras la identificación correcta de los análogos fue del 85%.

En cuanto a la Wescher Memory Scale-III (Weschler, 1997^a), Langeluddecke y Lucas (2003) encontraron diferencias significativas entre sospechosos de simulación y sujetos no simuladores en todos los subtests e índices de esta batería, obteniendo los simuladores puntuaciones excesivamente bajas comparadas con las puntuaciones generalmente encontradas en los TCE leves e incluso comparándolas con las puntuaciones de TCE moderados-severos. Empleando el punto de corte por debajo de 43 para el Reconocimiento Auditivo Demorado y por debajo de 18 en el Reconocimiento de Palabras de la lista II, consiguieron identificar al 80% de los simuladores, mientras que sólo un porcentaje muy bajo de los TCE severos (<10%) obtuvieron puntuaciones inferiores a los puntos de corte establecidos para la simulación de los TCE leves.

Tomando todos estos datos en conjunto, parece que entre las baterías de memoria desarrolladas por Weschler, la WMS-III es la que se muestra más prometedora para la detección de simulación. En general, los subtest de la WMS y WMS-R no se han mostrado muy efectivos, mientras las funciones discriminantes propuestas parecen superiores, si bien hay que tomarlas con precaución, puesto que algunas poseen una especificidad bastante baja. El índice diseñado por Iverson y Franzen (1996) parece muy prometedor, pero sus resultados necesitan ser replicados.

Otra batería que se ha mostrado útil en la detección de simulación es la **Halstead-Reitan (HRB)**. Mittenberg, Rotholc, Russell y Heilbroner (1996) validaron una función discriminante con 8 puntuaciones de la batería (el Category Test, la puntuación de tiempo y de memoria del Tactual Performance Test, el Seashore Rhythm Test, Speech-sounds Perception Test, Trails A y B, Finger Tapping, Sensory Suppressions y Finger-tip Number Writing) para determinar perfiles de respuesta inválidos, consiguiendo una proporción de discriminación total del 88.75% de los

pacientes con TCE y análogos, con un 16,2% de falsos positivos y un 6,2% de falsos negativos. McKinzey y Russell (1997) realizaron una validación cruzada, llegando a la conclusión de que la fórmula de Mittenberg y cols. (1996) para la detección de la simulación en la HRB necesita ser mejorada, ya que clasificó incorrectamente como simuladores al 27% del grupo de pacientes.

Otra puntuación compuesta que se ha mostrado útil en la detección de simulación es la propuesta por Trueblood y Schmidt (1993), basado en puntuaciones de 45 o superiores en la escala de déficit neuropsicológico general, ya que dichas puntuaciones se asocian en los pacientes con TCE-L con puntuaciones por debajo del azar en test de elección forzada.

Una mención aparte dentro de esta batería la merecen los estudios sobre el Category Test. Tenhula y Sweet (1996) resaltan que, puesto que el Category Test es básicamente una tarea de elección forzada, las puntuaciones por debajo del azar podrían emplearse para detectar simulación. Además, la variación en la dificultad de los ítems, así como las múltiples habilidades cognitivas implicadas en la prueba, podría dificultar la correcta valoración por parte de los simuladores tanto de la prueba como de sus ítems individuales. Estos autores comprueban la ejecución en el Category Test de pacientes con daño cerebral, análogos y estudiantes controles en diferentes variables, logrando niveles de clasificación elevados en los errores en el subtest VII, en los ítems fáciles, errores totales, número de criterios excedidos y, especialmente, en los errores en los subtest I y II, que con un punto de corte >1 lograron una clasificación correcta del 92% de los sujetos.

La utilidad de estas variables fue replicada por DiCarlo, Gfeller y Oliveri (2000), encontrando resultados muy similares a los referidos por Tenhula y Sweet (1996). El porcentaje de análogos detectados osciló entre el 77% y el 93%, siendo de nuevo la

regla de decisión de más de un error en los subtest I y II el índice de simulación más preciso (si bien cuando los análogos fueron entrenados el nivel de detección descendió drásticamente del 93 al 59%), y el número total de errores el menos eficaz. Además, la especificidad del número de errores en los subtest I y II fue perfecta, mientras en los otros índices los falsos positivos oscilaron entre el 10% y el 30%, por lo que han de ser interpretados con más precaución. En este sentido hay que puntualizar que, puesto que el grupo de pacientes en este estudio había sufrido traumatismos leves-moderados, es probable que con pacientes más severos el porcentaje de falsos positivos se incremente notablemente.

Esta suposición se ve apoyada por el estudio de Forrest, Allen y Goldstein (2004), que encuentran que el 3% de los pacientes con daño cerebral severo cometen más de 5 errores en los subtest I y II, por lo que este índice no debe emplearse como indicativo de simulación en los casos de daño cerebral masivo. Exceptuando estos casos, este índice vuelve a mostrarse el más sensible a la detección de simulación, ya que fue la única variable en la que se produjeron diferencias estadísticamente significativas entre estudiantes controles, análogos y pacientes con daño cerebral documentado. Además, estos autores proponen una función discriminante (basada en los errores en el subtest I, errores totales y errores en el subtest II) que identificó correctamente al 73% de los análogos y al 91% de los pacientes, mientras la variable número de errores totales empleada aisladamente queda descartada como indicativa de simulación por estos autores.

Otro subtest también estudiado por separado es el Trail Making Test (TMT). O'Bryant, Hilsabeck, Fisher y McCaffrey (2003) realizaron recientemente un estudio en el que se pone de manifiesto que los errores del TMT no son útiles en la detección de simulación, y que el tiempo total para completar tanto la parte A como la B, así como el

ratio, a pesar de ser significativamente inferiores en los sospechosos de simulación, sólo detectan a un 63% y 45% respectivamente, por lo que la utilidad clínica de estos indicadores es limitada. Además, dado que este instrumento es uno de los más sensibles al daño genuino, puede ser muy difícil determinar cuándo la ejecución es pobre debido a un bajo esfuerzo, a un daño neurológico, o a una mezcla de las dos. Por tanto, las puntuaciones del TMT deben ser interpretadas con mucha cautela como marcadores de simulación en pacientes con traumatismo leve en litigio.

En definitiva, tanto las funciones discriminantes que emplean diversas medidas de la batería Halstead-Reitan como los subtests estudiados aisladamente parecen poco útiles en la detección de simulación, con excepción del índice de errores en los subtest I y II del Category Test.

La batería **Memory Assessment Scales** (MAS; Williams, 1991) consiste en 12 subtests que evalúan varios aspectos de la atención y memoria. El MAS incluye medidas tanto para información visual como verbal, recuerdo y reconocimiento, y componentes inmediatos y demorados. Ross, Krukowski, Putnam y Adams (2003), evaluaron la efectividad del MAS para detectar simulación seleccionando sus subtests más importantes (*span* verbal, *span* visual, reconocimiento visual –inmediato y demorado) y tres de los índices de la lista de aprendizaje (adquisición, recuerdo demorado con claves y reconocimiento). Emplearon un grupo de sujetos con daño leve en litigio sospechosos de simulación por haber puntuado en el rango del azar en el Recognition Memory Test, y sujetos con daño cerebral moderado-severo sin incentivos externos para simular. En este estudio se encontró que los índices de *span* visual y verbal eran muy útiles para detectar esfuerzo incompleto, obteniendo una sensibilidad del 90,5% y una especificidad del 81%. Todos los índices de la lista de aprendizaje obtuvieron también relativamente buenos resultados, en particular el recuerdo demorado

de la lista con claves, con una clasificación correcta total del 73%. Los tests de reconocimiento basados en la elección forzosa, a pesar de conseguir una clasificación de los grupos por encima del azar, no aumentaban la predicción de los subtest de *span* visual y verbal. Sin embargo, como sus propios autores indican, los puntos de corte obtenidos en este estudio han de ser interpretados con la debida precaución, ya que la prevalencia empleada en esta investigación fue del 50%.

Otro reciente estudio sobre el MAS se lo debemos a O'Bryant, Duff, Fisher y McCaffrey (2004). En él se pretendía averiguar la efectividad de los patrones de ejecución y de puntos de corte para diferenciar sujetos con daño traumático cerebral en litigio sospechosos de simulación (según sus puntuaciones en el TOMM o en el 15 ítems de Rey) y sujetos no sospechosos de simulación. Aunque los sujetos del primer grupo obtenían una ejecución globalmente inferior, los perfiles de ejecución para cada grupo eran similares. Los puntos de corte de todas las puntuaciones del MAS clasificaban correctamente del 70% al 80% del grupo en litigio. Sin embargo, la utilización de varios puntos de corte, utilizados serialmente, obtenían mejores índices de clasificación (por ejemplo, si el recuerdo de la lista obtenía una clasificación correcta del 78% de los sospechosos de simulación y el *span* visual del 77%, ambos puntos de corte usados serialmente conseguían una clasificación correcta del 96%). Los autores concluyen que, si bien una baja ejecución en el MAS no es por sí misma indicativa de simulación, ofrece una información útil para ser usada en conjunto con otros test específicos de simulación y con información de fuentes secundarias y entrevistas clínicas.

El empleo de esta batería se muestra prometedor, pero los resultados han de ser replicados.

3.2. El empleo de diversas medidas no específicas en la detección de simulación:

Igual que sucede con las pruebas específicas, son muy escasos los estudios que evalúan la eficacia *conjunta* de varios índices no específicos en la detección de simulación. En conjunto, los resultados nos informan de la superioridad del empleo de diversos índices sobre el empleo de medidas únicas. Algunos ejemplos que lo demuestran se presentan a continuación:

Meyers, Galinsky y Volbrecht (1999), empleando cuatro test de forma conjunta (Juicio de Orientación de Líneas, Token Test, un test de escucha dicótica y otro de elección forzosa), obtuvieron un 100% de especificidad y un 95% de sensibilidad a la hora de separar litigantes, no litigantes y análogos empleando como criterio el fallo en al menos una de las pruebas.

Meyers y Volbrecht (2003), en un extenso estudio de 796 sujetos, consiguieron una sensibilidad del 83% y una especificidad del 100% en la detección de simulación, estableciendo esta a partir del fallo de dos o más medidas en nueve tests neuropsicológicos (Figura Compleja de Rey, Digit Span, Juicio de Orientación de Líneas, Token Test, Sentence Repetition, Auditorial Verbal Learning Test, un test de elección forzosa y la puntuación estimada del Finger Tapping).

Iverson y Franzen (1996), empleando como criterio la puntuación por debajo del punto de corte establecido en al menos una de cinco pruebas (el Test de los 16-Items, el 21-Item Test, Digit Span y Logical Memory (WMS-R), así como un cuestionario autobiográfico) consiguió una clasificación correcta del 100% de los controles y de los pacientes con problemas de memoria, así como del 92,5% de los análogos.

Larrabee (2003^a), empleando como criterio el fallo en dos pruebas no específicas (Benton Visual Form Discrimination, Finger Tapping, Reliable Digit Span, Wisconsin

Card Sorting Test y MMPI-II) consiguió una clasificación correcta del 89% de los pacientes con TCE moderado-severo y del 87,5% de los sujetos clasificados como simuladores definitivos según los criterios de Slick y cols. (1999).

4. Conclusiones sobre las pruebas específicas y no específicas de simulación:

Empezando por los test no específicos, podemos concluir que el empleo de puntos de corte suele ser fructífero a la hora de diferenciar análogos y sujetos controles, si bien es muy difícil encontrar puntos de corte en estos tests que sean sensibles a la simulación sin etiquetar a pacientes con daño cerebral como simuladores (Suhr y Gunstad, 2000). La solución por la que suele optarse es la de establecer puntos de corte que eviten los falsos positivos al estar basados en una elevada especificidad, pero esto conlleva en estas pruebas sensibilidades muy pobres, de modo que bastantes estudios refieren que estas pruebas son inefectivas a la hora de diferenciar la simulación del daño legítimo (Slick y cols., 1999).

Por norma general, en estos test no-específicos se encuentran mejores resultados empleando funciones discriminantes, que además son más resistentes al efecto del entrenamiento o *coaching* (Suhr y Gunstad, 2000). El problema es que la mayoría de los estudios carecen de replicación y validaciones cruzadas (Slick y cols., 1999), y cuando se replican no suelen ofrecer los resultados esperados.

Por otro lado, la mayoría de los estudios sobre los tests tradicionales aplicados a la detección de simulación han sido realizados con análogos, por lo que su generabilidad es limitada. Relativamente pocos estudios emplean diseños de grupos conocidos, como pacientes litigantes versus no litigantes con daño cerebral leve. Por todos estos motivos, los tests tradicionales no se consideran con el peso suficiente para su aceptación en el contexto médico-legal.

Una de las ventajas de los test no específicos es que permiten determinar la validez de evaluaciones neuropsicológicas administradas previamente en las que no se incluyó ninguna medida específica de simulación (Mittenberg, Aguila-Puentes, Patton, Canary y Heilbroner, 2002), aunque el juicio clínico basado en los resultados de estos test es menos que satisfactorio (Rose, Hall, Szalda-Petree y Bach, 1998).

Otra de las principales ventajas que suele adjudicarse a los tests tradicionales es que cumplen una doble tarea, proporcionando información sobre el estado neuropsicológico del paciente y proporcionando índices de simulación sin suponer un coste de tiempo extra (Curtis, Greve, Bianchini y Brennan, 2006), mientras las pruebas específicas sólo sirven para detectar simulación, y no tienen utilidad para otra interpretación neuropsicológica (Meyers y Volbretch, 2003).

Como contrapartida, de los tests no específicos no podemos deducir la intencionalidad del sujeto (Slick y cols., 1999), mientras que si un sujeto obtiene puntuaciones significativamente por debajo del azar en un test específico, no nos quedan demasiadas explicaciones alternativas viables al comportamiento intencional del sujeto (Rogers, Harrel y Liff, 1993).

El principal problema del criterio de puntuación significativamente por debajo del azar es que ha mostrado sistemáticamente una sensibilidad inaceptablemente baja (Bender y Rogers, 2004; Gervais, Rohling, Green y Ford, 2004; Guilmette, Hart y Giuliano, 1993; Guilmette, Hart y cols., 1994; Greiffenstein y cols., 1994; Hiscock, Branham y Hiscock, 1994; Holmquist y Wanlass, 2002; Martin, Bolter, Todd, Gouvier y Nicolls, 1993; Martin, Hayes y Gouvier, 1996; Rees, Tombaugh, Gansler y Moczynski, 1998; Rose, Hall y Szalda-Petree, 1995; Slick, Hopp, Strauss, Hunter y Pinch, 1994), por ello los tests específicos de simulación disponen de puntos de corte con los que se consigue una elevada sensibilidad sin disminuir apenas la especificidad.

Entre los principales inconvenientes de las medidas específicas de simulación podemos citar que los pacientes que están poniendo su máximo esfuerzo en la realización de las pruebas pueden sentirse molestos con la realización de las tareas de elección forzosa, tanto porque consideren que insulta su inteligencia como por lo aburridos que son, lo que puede llevar a que en algún momento los pacientes exhiban un esfuerzo insuficiente en estos tests (Lezack, 1995). En este sentido son muy interesantes los trabajos que versan sobre la posible terminación temprana de los tests cuando los sujetos los están realizando correctamente, o las pruebas que incluyen dicha posibilidad, como el CARB.

Por otro lado, los sujetos que simulan no lo hacen necesariamente de un modo consistente, sino que intentan fingir distintos tipos de daño, lo que se detectaría mejor mediante técnicas específicas a la disfunción más que por una medida global de simulación (Meyers y Volbretch, 2003). Sin embargo, a este respecto hay que señalar que los dos únicos trabajos con los que contamos en la actualidad que estudian las estrategias que emplean las personas a la hora de simular señalan que fingir problemas de memoria es la estrategia más empleada por los sujetos (Iverson, 1995; Tan y cols., 2002).

También se ha señalado que los individuos conscientes de la posibilidad de la evaluación de simulación reconocen el formato de elección forzosa como un intento de detección, si bien este reconocimiento no influye en la capacidad de detección de las pruebas.

La efectividad de las pruebas específicas y no específicas ha sido raramente comparada, y siempre las pruebas específicas han demostrado su superioridad (Inman y Berry, 2002; Schagen y cols., 1997; Greiffenstein y cols., 1994; Iverson y Binder, 2000; O'Bryant y cols., 2003; Tan y cols., 2002).

A pesar de dicha superioridad, nos gustaría recordar que una puntuación normal en una medida de simulación no elimina la posibilidad de simulación debido a que la sensibilidad típicamente asociada a estas medidas no es perfecta (Bianchini, Mathias y Greve, 2001). De ahí la necesidad de emplear múltiples medidas en la detección de simulación, tan defendida por los autores (Nies y Sweet, 1994; Vickery y cols., 2001, 2004; O'Bryan y cols., 2004; Forrest y cols., 2004; Meyers y cols., 2001). Una ventaja de la que disponemos es que no tenemos la obligación de elegir entre un tipo de medidas u otras, por lo que nos podemos valer de ambas, con sus ventajas e inconvenientes, para así poder basar nuestras conclusiones en el mayor número de fuentes posibles, lo que incrementa la seguridad de nuestro juicio.

CAPÍTULO 6
OBJETIVOS GENERALES

El objetivo general de esta tesis es investigar la detección de simulación de déficits cognitivos en pacientes que han sufrido un TCE-L. Para ello, se estudiarán fundamentalmente algunas de las pruebas específicas de simulación más empleadas o que poseen un mayor apoyo empírico en la literatura. Por otro lado, se averiguará la capacidad de otro tipo de medidas para detectar simulación, como los test neuropsicológicos no específicos de simulación, la psicofisiología y la neuroimagen funcional. Este trabajo se ha dividido en tres estudios.

El objetivo general del **estudio 1** será el de adaptar y comprobar las propiedades psicométricas de las pruebas específicas de simulación seleccionadas para determinar si es pertinente o no su aplicación en población española. Concretamente, en primer lugar en este estudio se realizará una adaptación y una doble traducción cruzada (“*back translation*”) de las instrucciones de los test. En segundo lugar, teniendo en cuenta la importancia que se ha venido atribuyendo al entrenamiento (“*coaching*”) en el estudio de la detección de la simulación (ver Capítulo 4, apartado 5.1: El entrenamiento), estudiaremos el efecto del entrenamiento mediante las instrucciones sobre la ejecución en los test de simulación en un grupo de análogos. Además, compararemos nuestros resultados con los hallazgos norteamericanos y estableceremos el nivel de clasificación correcta de los análogos como sujetos simuladores por las distintas pruebas de simulación.

Hemos decidido comenzar la investigación empleando un grupo de análogos debido a que este tipo de diseños son muy valiosos, especialmente en los estudios iniciales para el desarrollo de las pruebas, que es precisamente el estadio en el que nos encontramos actualmente en nuestro contexto. Por otro lado, en la investigación de la simulación generalmente se emplea como grupo de comparación estudiantes con una ejecución normal (Bianchini, Mathias y Greve, 2001), por lo que la mayoría de lo que sabemos

hoy en día sobre la detección de la simulación se lo debemos a la investigación llevada a cabo con análogos, lo que pone de manifiesto la importancia que se le concede a este tipo de diseños. Sin embargo, para dotar a los estudios de utilidad clínica, las investigaciones deberían incluir como grupos de contraste pacientes reales (Rogers, 1997^b). Esto es así debido a que la cuestión diagnóstica crítica en la investigación en simulación es si una determinada ejecución es debida al daño cerebral o a una distorsión intencionada en la ejecución, no si somos capaces de diferenciar una ejecución simulada de una ejecución normal. Atendiendo a esto han sido seleccionados los grupos del estudio 2.

Una vez que conozcamos que las adaptaciones y las propiedades psicométricas de los tests cumplen los requisitos para ser aplicados en nuestro contexto, en el **estudio 2** comprobaremos la utilidad de dichas pruebas para la detección de simulación en nuestro contexto. Para ello, investigaremos si las pruebas específicas de simulación seleccionadas son capaces de diferenciar entre distintos grupos, valiéndonos de diseños mixtos, ya que la mejor validación de las medidas de simulación consiste en una combinación de estudios de análogos y de grupos conocidos (Rogers y Bender, 2000), puesto que de este modo se maximiza tanto la validez interna (grupo de análogos) como la externa (comparación de grupos conocidos). Este objetivo general será desglosado en cinco objetivos específicos.

En el primero de ellos estudiaremos la influencia de las variables demográficas y del tiempo transcurrido desde el traumatismo sobre los tests específicos de simulación, para descartar que dichas pruebas estén influidas por variables ajenas a la simulación.

En el segundo de los objetivos de este estudio 2 averiguaremos si las distintas pruebas específicas de simulación son capaces de diferenciar entre los “grupos de prevalencia diferencial”. Este objetivo se abordará utilizando el diseño de grupos de

prevalencia diferencial más un grupo de análogos: grupo de búsqueda de compensación (BC), grupo de NO búsqueda de compensación (NBC) y grupo de análogos (AN).

El tercer objetivo consiste en averiguar si las distintas pruebas específicas de simulación son capaces de diferenciar entre “grupos conocidos”. Este objetivo se abordará utilizando un diseños de “grupos conocidos” más un grupo de análogos: grupo de no búsqueda de compensación (NBC), grupo de búsqueda de compensación no sospechoso de simulación (BC-NSS), grupo de búsqueda de compensación sospechoso de simulación (BC-SS) y grupo de análogos (AN).

En el objetivo 4 obtendremos los puntos de corte de las pruebas de simulación que optimicen la especificidad para las distintas variables. Debido a la importancia que tiene en el contexto de la detección de simulación el no clasificar erróneamente a los pacientes como simuladores, hemos optado por obtener los puntos de corte vinculados a una especificidad mínima del 90%, a pesar del detrimento en la sensibilidad que esto pueda suponer.

Por último, en el quinto objetivo de este estudio 2 compararemos nuestros resultados con los publicados en la literatura para la población norteamericana, buscando de este modo un apoyo más a la validez y consistencia de nuestros hallazgos.

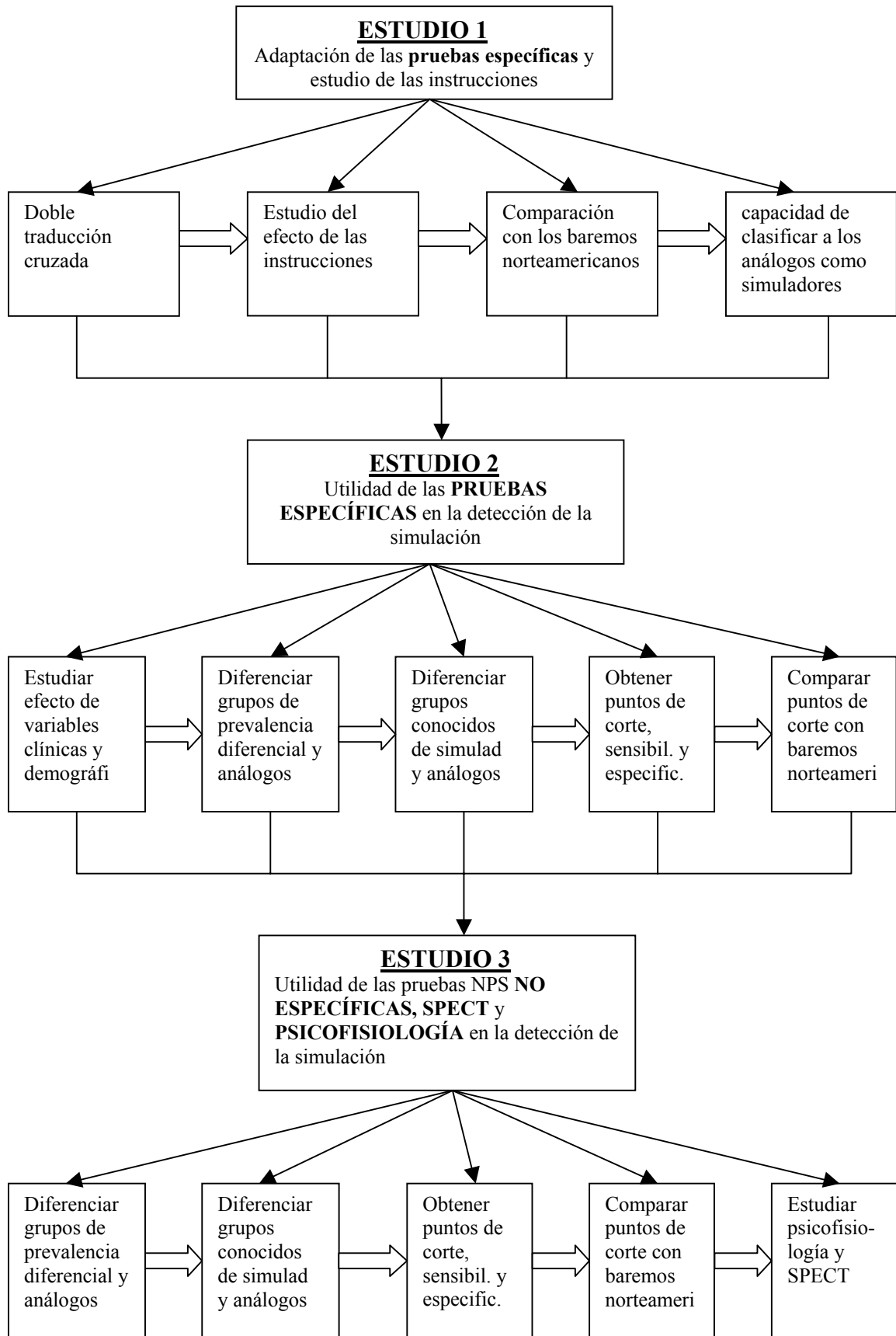
Una vez comprobada la utilidad del empleo de las pruebas neuropsicológicas específicas de simulación, en el **estudio 3** examinaremos la aportación de otro tipo de métodos para la detección de la simulación, específicamente las pruebas neuropsicológicas tradicionales no específicas de simulación, las variables psicofisiológicas y la neuroimagen funcional. Los cuatro primeros objetivos de este estudio son paralelos a los objetivos 2-4 del estudio 2 referidos a las pruebas específicas. De este modo, en primer lugar averiguaremos si las las pruebas neuropsicológicas no específicas de simulación son capaces de diferenciar entre los

grupos de prevalencia diferencial y a continuación estudiaremos si son capaces de diferenciar entre “grupos conocidos”. Obtendremos los puntos de corte para las distintas variables de los tests neuropsicológicos no específicos que optimicen la especificidad, y compararemos nuestros resultados con los publicados en la literatura para la población norteamericana.

El quinto objetivo del estudio 3 consistirá en comprobar la utilidad de diversas medidas psicofisiológicas de conductancia y pulso empleadas en combinación con una medida específica en la detección de simulación.

Por último, en el objetivo 6 se determinará la utilidad de la SPECT-PCR a la hora de diferenciar pacientes honestos de personas sospechosas de simulación.

La secuencia de los estudios y sus respectivos objetivos se puede visualizar en el siguiente diagrama de flujo:



CAPÍTULO 7

ESTUDIO 1: ADAPTACIÓN Y ESTUDIO PRELIMINAR DE LAS PRUEBAS ESPECÍFICAS DE SIMULACIÓN

1. Introducción:

Si la determinación del esfuerzo insuficiente por parte de un paciente es un tema siempre complicado y conflictivo, estas dificultades se multiplican cuando la evaluación se realiza en nuestro contexto, en el que la investigación sobre simulación parte de cero. Sólo disponemos dos estudios realizados con población española: el primero de ellos investiga el Test of Memory Malingering aplicado a pacientes con TCE con una motivación adecuada para la realización del test (Ramírez y cols., 2004), lo que nos permite conocer la especificidad de esta prueba concreta en nuestro contexto (83,3%), pero no podemos realizar ningún tipo de inferencia sobre su sensibilidad. En el segundo de los estudios se explora la ejecución en distintas pruebas específicas de simulación mediante el empleo de un diseño mixto, en el que se incluyen grupos de prevalencia diferencial y un grupo de análogos. Sin embargo, este estudio todavía no ha sido publicado (Vilar, In Press).

Es debido a este “partir de cero” que nuestro primer objetivo ha sido el de adaptar las pruebas que vamos a emplear en la investigación y comprobar sus propiedades psicométricas. Por otro lado, para aumentar el rigor de nuestro estudio hemos empleado un grupo de análogos siguiendo todas las recomendaciones de Rogers (1997^b) (ver Capítulo 3, apartado 5: La investigación de la simulación), y hemos estudiado el efecto de las instrucciones aportadas al grupo con la finalidad, por un lado, de crear un grupo de “simuladores sofisticados”, y por otro para intentar arrojar un poco de luz sobre una de las críticas más repetidas en la investigación sobre los tests específicos de simulación: su vulnerabilidad ante el entrenamiento. Para finalizar, buscando un apoyo a la validez de las pruebas empleadas que permita sustentar de un modo más consistente los hallazgos que se produzcan de aquí en adelante, realizaremos una comparación de

nuestros datos con los obtenidos en los manuales de los tests y, cuando esto no sea posible, con los disponibles en la literatura.

2. Objetivos e hipótesis:

Como hemos mencionado en el capítulo anterior, el objetivo general del **estudio 1** será el de adaptar y comprobar las propiedades psicométricas de las pruebas específicas de simulación seleccionadas para determinar si es pertinente o no su aplicación en población española. Este objetivo general se concreta en los siguientes objetivos específicos:

- **Objetivo 1:** Adaptar y realizar la “*back-translation*” de las instrucciones de los tests.
- **Objetivo 2:** Estudiar el efecto de las instrucciones sobre la ejecución en los tests de simulación en un grupo de análogos.
- **Objetivo 3:** Analizar las propiedades psicométricas de los tests.
- **Objetivo 4:** Establecer el nivel de clasificación correcta de los análogos como sujetos simuladores por las distintas pruebas de simulación.
- **Objetivo 5:** Comparar nuestros resultados con los hallazgos norteamericanos.

De los objetivos anteriores se derivan las siguientes hipótesis:

Hipótesis 1: El tipo de instrucciones afectará a los resultados de los análogos en los tests de simulación, de modo que el grupo con las instrucciones con menos información realizarán peor las pruebas de simulación, es decir, serán más fácilmente detectados.

Hipótesis 2: Las principales variables de los tests de simulación presentarán una elevada correlación entre sí.

Hipótesis 3: Las principales variables de los tests de simulación presentarán una baja correlación con las variables de otros tests neuropsicológicos convencionales.

Hipótesis 4: Las pruebas de simulación conseguirán clasificar una elevada proporción de los análogos como simuladores.

Hipótesis 5: No existirán diferencias entre nuestros datos y los publicados en los manuales de las pruebas o en literatura.

3. Metodología:

3.1. Participantes:

En este estudio participaron 54 estudiantes de 4º curso de Psicología (7 hombres y 47 mujeres) con conocimientos neuropsicológicos. La media de edad del grupo era de 22,93 años (Sd= 1,60), y la escolaridad media de 16,83 años (Sd= 0,42).

Para la comprobación de la hipótesis 1 se dividió a los análogos en dos grupos de forma aleatoria. Uno de los grupos recibió a la hora de enfrentarse a la realización de los tests un tipo de instrucciones (A), y el otro grupo recibió unas instrucciones diferentes (B). El grupo de las instrucciones tipo A estaba compuesto por 26 sujetos, con una media de 23,08 años de edad (Sd=1,6 años) y una escolaridad media de 16,92 años (Sd=0,27 años), y el grupo que recibió las instrucciones del tipo B tenía una media de 22,88 años de edad (Sd=1,63) y de 16,75 años de escolaridad (Sd=0,51).

3.2. Instrumentos:

Las pruebas específicas de simulación aplicadas a los análogos fueron las siguientes:

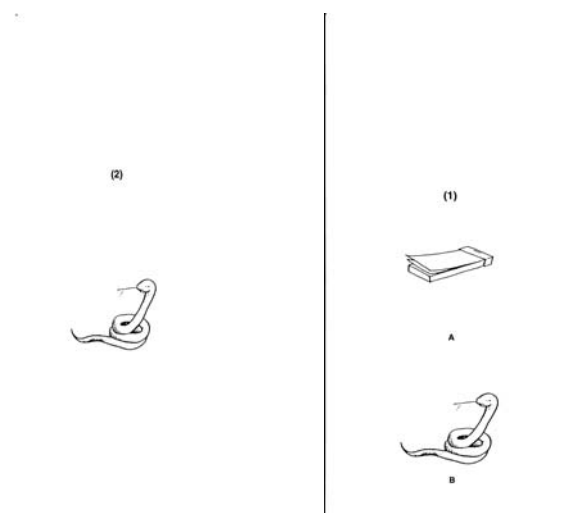
- *Victoria Symptom Validity Test* (VSVT; Slick, Hopp, Strauss y Thompson, 1997)

Este test se administra por ordenador y consiste en la presentación de tres bloques de 16 números cada uno. En primer lugar se le presenta al sujeto un número de cinco dígitos en la pantalla, que el sujeto debe memorizar. La presentación del número va seguida de un intervalo de retención, seguido por la presentación de dos estímulos:

un distractor y el número previamente presentado, que el sujeto debe reconocer y responder lo más rápidamente posible. La dificultad aparente del test se incrementa mediante dos estrategias: en primer lugar, el intervalo de retención va aumentando con los bloques (5, 10 y 15 segundos), y las instrucciones le indican al sujeto que “podría” encontrar la siguiente sección más difícil. En segundo lugar, el test incluye ítems “fáciles” (aquellos en los que el estímulo objetivo y el distractor no comparten números, por ejemplo 83649 y 25170) e ítems difíciles (en los que el número objetivo y el distractor rotan el segundo y tercer dígito o el tercero y el cuarto, por ejemplo 58379 y 58739). El VSVT tiene en cuenta la latencia de respuesta, ya que se ha puesto de manifiesto que las personas que exageran sus déficits tienden a exagerar el tiempo necesario para realizar su elección, y también considera los sesgos relacionados con la tendencia a responder a los ítems de la derecha o de la izquierda.

- *Test of Memory Malingering* (TOMM; Tombaugh, 1996).

Esta prueba está formada por dos ensayos de aprendizaje y un ensayo de retención. Durante los dos ensayos de aprendizaje se le muestran al sujeto 50 dibujos de objetos comunes que posteriormente deberá reconocer entre dos alternativas. El ensayo de retención es optativo y se utiliza para corroborar el resultado de los ensayos de aprendizaje 15 minutos después de la aplicación de estos. Esta parte opcional se aplica mostrando al sujeto directamente las láminas de reconocimiento, sin mostrar previamente los estímulos objetivo. La corrección del test es muy sencilla, y se realiza otorgando un punto por cada respuesta correcta.



Ejemplo de ítem del TOMM.

- *The b test* (Boone, Lu y Herzberg, 2002)

El test de la b es una tarea de reconocimiento y discriminación de letras. Se supone que señalar las letras *b* de entre diversos distractores es una tarea extremadamente sencilla, ya que el reconocimiento y discriminación de letras se considera una habilidad sobreaprendida, de modo que incluso pacientes con daño cerebral o neurológico pueden realizar este test adecuadamente. La prueba está formada por un cuadernillo de 15 páginas en las que el sujeto debe rodear las letras *b* que aparecen intercaladas entre distractores distintos. A medida que el test avanza, los estímulos se hacen cada vez más pequeños, de modo que incrementa su dificultad aparente. A modo de ejemplo, a continuación se presenta la primera línea del primer estímulo de la prueba:

b d b p q d q p b

- *Dot Counting Test* (Boone, Lu, Back y cols., 2002).

En esta prueba la tarea del sujeto consiste en contar, lo más rápidamente posible, el número de puntos que se le presentan en unas tarjetas. Existen dos tipos de tarjetas: aquellas en las que los números aparecen de modo disperso, y en las que los puntos

aparecen de modo agrupado. Tradicionalmente se consideraba que existían indicios de simulación cuando el tiempo para contar los puntos agrupados excedía el de los no agrupados, pero recientemente se han expuesto otras medidas que parecen más fiables, como por ejemplo la puntuación combinada (Boone, Lu, Back y cols., 2002).

- *Test de los 15-ítems de Rey* (Rey, 1964)

En esta prueba se le presenta al sujeto una lámina con 15 estímulos. En las instrucciones la tarea se presenta como complicada, ya que se enfatiza que son “15 estímulos diferentes”. A continuación se retira la lámina y el sujeto debe dibujar todo lo que recuerde. En realidad se trata de series de estímulos muy fácilmente memorizables (A B C; 1 2 3; a b c; I II III; círculo, cuadrado y triángulo), por lo que una puntuación inferior a 9 se considera sospechosa de simulación (Rey, 1964; Lezack y cols., 2004).

3.3. Las instrucciones de los instrumentos y su doble traducción cruzada (“back translation”):

El primer punto que fue necesario solucionar para poder aplicar estas pruebas en nuestra población fue la de realizar una adecuada traducción de las instrucciones. Para ello se llevó a cabo una doble traducción cruzada “back-translation” de las instrucciones que aparecían en los manuales del VSVT, el TOMM y el Test de la b, y de las instrucciones del Test de los 15-Ítems de Rey tal como aparecen en Goldberg y Miller (1986) (ver Apéndice 1). Las instrucciones originales de todas las pruebas están recogidas en el Apéndice 1.1. Dichas instrucciones fueron traducidas del inglés al español por una traductora cuya lengua materna es el español (ver Apéndice 1.2). El resultado de esta traducción fue enviado a otra traductora cuya lengua materna es el inglés para que realizara una nueva traducción al inglés de esta versión en castellano (ver Apéndice 1.3). Ambas versiones en inglés (Apéndice 1.1 y 1.3) fueron revisadas por una persona bilingüe para comprobar que no existían discrepancias. Debido a que la

opinión de este experto fue que las discrepancias eran mínimas, consideramos que las instrucciones recogidas en el Apéndice 1.2 constituyen una buena adaptación al castellano de las instrucciones originales.

3.4. Procedimiento:

El grupo de análogos fue reclutado de la asignatura de Neuropsicología Aplicada. Siguiendo las recomendaciones propuestas por Rogers (1997^b), a los participantes se les proporcionaban unas instrucciones específicas en las que se ofrecía un escenario en el que podían identificarse. Las explicaciones verbales dadas por el experimentador presentaban la evaluación como un desafío al sujeto para ver si serían capaces de simular un daño en una situación real. Se tuvieron en cuenta los incentivos tanto positivos (puntos extra en la materia) como negativos (vergüenza social), y se especificaba que la simulación debía ser lo suficientemente creíble como para evitar la detección. Además, en las instrucciones se especificaban algunos de los síntomas más comunes del trastorno que los sujetos debían simular, pero sólo en la mitad de los casos se proporcionaron, además, instrucciones sobre métodos específicos de las pruebas de simulación. Por lo tanto, existían dos tipos de instrucciones, A y B (ver Apéndice 2) que se otorgaban de modo aleatorio a los sujetos, puesto que ellos mismos elegían uno de entre todos los sobres cerrados y barajados que contenían los dos tipos de instrucciones.

Se daba a los participantes todo el tiempo que ellos consideraban necesario para la lectura y comprensión de las instrucciones, así como para el diseño de una estrategia a seguir durante la evaluación, y a continuación se administraban las pruebas de simulación arriba explicadas. Tras la evaluación los participantes cumplimentaban un informe (ver Apéndice 3) con la finalidad de comprobar el recuerdo de las instrucciones y su comprensión, así como la implicación de los participantes en la simulación. Dicha implicación fue insuficiente en 6 casos, ya que reconocieron no haber seguido las

instrucciones de simular “nunca” o “casi nunca” durante la evaluación, por lo que de los 60 participantes iniciales consideramos sólo los datos de 54 sujetos. La duración aproximada de todo el proceso (instrucciones, evaluación y cuestionario posterior) era de una hora y cuarto.

3.5. Variables:

El criterio seguido a la hora de elegir las variables de las pruebas de simulación fue, principalmente, el de recoger toda la información relevante. Las variables seleccionadas son las que poseen una mayor significación clínica y ofrecen una mayor y mejor información. A continuación exponemos las variables atendiendo a la prueba a la que pertenecen:

- *Victoria Symptom Validity Test* (VSVT; Slick, Hopp, Strauss y Thompson, 1997)

- Número de aciertos en los ítems fáciles
- Número de aciertos en los ítems difíciles
- Número de aciertos totales (suma de aciertos en los ítems fáciles y difíciles)

- *Test of Memory Malingering* (TOMM; Tombaugh, 1996).

- Ensayo 1= Número de aciertos totales en el primer ensayo
- Ensayo 2= Número de aciertos totales en el segundo ensayo
- Retención= Número de aciertos totales en el ensayo de retención

- *The b test* (Boone, Lu y Herzberg, 2002)

- E-score= $[(\text{total de errores "d"} + \text{total de errores de comisión}) \times 10] + \text{Total errores de omisión} + \text{media del tiempo para cada estímulo}$
- Errores d= número total de letras d marcadas en el test
- Errores de comisión= cualquier marca en el test que no se corresponda con una letra b o una d.
- Errores omisión= Número de letras b no marcadas

- Tiempo= Media del tiempo (en segundos) por estímulo. Se calcula dividiendo los segundos necesarios para realizar el test entre 15.

- *Dot Counting Test* (Boone, Lu, Back y cols., 2002)

- Puntuación combinada= Media de tiempo para ítems no agrupados + media de tiempo para ítems agrupados + número de errores
- Tiempo ítems agrupados= Media del tiempo necesario para contar los puntos de las tarjetas agrupadas
- Ratio= Media del tiempo necesario para contar los puntos de las tarjetas agrupadas dividida entre la media del tiempo necesario para contar los puntos de las tarjetas no-agrupadas
- Errores= Número total de errores (suma de los errores en las tarjetas agrupadas y en las no-agrupadas)

- Test de los 15-ítems de Rey (Rey, 1964):

La variable seleccionada para este test ha sido el número total de elementos recordados.

3.6. Análisis estadísticos

Para estudiar si existían diferencias entre el tipo de instrucciones (hipótesis 1) se realizó un análisis de *t* de *Student* para muestras independientes. Para el estudio psicométrico (hipótesis 2 y 3) se realizó un alfa de Cronbach y correlaciones de Pearson. Finalmente, se realizaron diferencia de medias, usando la *F* de Snedecor entre las medias de los baremos y las obtenidas en nuestra muestra.

4. Resultados:

4.1. El Efecto de las instrucciones (Hipótesis 1):

Para estudiar el efecto del tipo de información proporcionada a los análogos sobre su ejecución en los tests de simulación se realizaron seis pruebas t , siendo la VI el tipo de instrucciones (A o B) y las VD las puntuaciones correspondientes a las principales variables de las pruebas: aciertos en los ítems difíciles y aciertos totales en el VSVT, aciertos en el Ensayo 2 y en el ensayo de Retención en el TOMM, e-score en el Test de la b , puntuación combinada en el Dot Counting y número de aciertos totales en el test de los 15-ítems de Rey. Los resultados mostraron que no existían diferencias estadísticamente significativas entre el tipo de instrucciones recibidas y la puntuación obtenida en el test para ninguna de las variables estudiadas (ver Tabla 13).

Tabla 13. Medias, desviaciones típicas, pruebas t y significación de los tests de simulación atendiendo al tipo de instrucciones recibidas por los sujetos.

Variables	Instrucciones A		Instrucciones B		t	p
	Media	Dt	Media	Dt		
VSVT ítems difíciles	10,41	3,61	11,55	4,69	-0,96	0,341
TOMM ensayo 2	35,16	7,87	35,25	9,07	-0,03	0,970
Test b e-score	238,81	201,04	347,95	302,65	-0,89	0,374
Dot Counting combo	18,90	8,38	18,07	6,68	0,37	0,714
15-ítems recuerdo total	12,68	2,59	12,11	3,17	0,70	0,485

Nota: Dt= desviación típica; combo= puntuación combinada; VSVT= Victoria Symptom Validity Test; TOMM= Test of Memory Malingering.

4.2. Propiedades psicométricas de las pruebas (Hipótesis 2 y 3):

Para estudiar la fiabilidad, se han realizado análisis de Alfa de Cronbach para los tests (VSVT y Test de la b) y las variables en los que se podía realizar. Debido a que este tipo de análisis se realizan para evaluar la consistencia entre los ítems de una prueba, sólo se pudo obtener para aquellas variables dependientes que se obtenían a

partir de varios ítems. Así, se obtuvo el Alfa de Cronbach para el VSVT (ítems fáciles, ítems difíciles e ítems totales) y para el Test de la b (errores *d*, errores de omisión y errores de comisión). En el caso del VSVT, los resultados mostraron un Alfa entre 0,840 y 0,913, y en el caso del Test de la b el Alfa osciló entre 0,810 y 0,864 (ver Tabla 14).

Tabla 14. Alfa del VSVT y del Test de la b

Test	Variable	Alfa
VSVT	Ítems fáciles	0,840
	Ítems difíciles	0,898
	Ítems totales	0,913
Test de la b	Errores d	0,852
	Errores o	0,864
	Errores c	0,810

Para estudiar la validez se realizaron correlaciones de Pearson entre las pruebas de simulación (validez convergente) y de las pruebas de simulación con otras pruebas neuropsicológicas (validez divergente). Las pruebas seleccionadas para el estudio de la validez divergente fueron el ensayo cuarto del Cinco Dígitos (flexibilidad cognitiva), el primer ensayo del Test de Trazado Oral (velocidad de búsqueda) y la Tarea de Sesgo Cognitivo (toma de decisiones). Estas pruebas fueron seleccionadas porque miden diferentes aspectos, tanto simples como complejos de las funciones cognitivas y porque fueron empleadas en un estudio paralelo. En el caso de la validez convergente, los resultados mostraron correlaciones estadísticamente significativas, cuyos valores se situaron entre 0,24 y 0,48. En cuanto a la validez divergente, las correlaciones oscilaron entre 0,01 y 0,23, no alcanzando ninguna de ellas la significación estadística (ver Tabla 15).

Tabla 15. Correlación de Pearson para estudiar la validez convergente y divergente de las pruebas de simulación

	Validez Convergente			Validez Divergente			
	15-Ítems	Dot Counting	Test de la b	TOMM	OMT1	FD4	CBT
VSVT	,289*	,446**	-,366**	,426**	-,119	,027	,075
TOMM	,458**	-,484**	-,410**		,024	-,017	,231
Test de la b	-,165	,318			,056	,195	-,081
Dot counting	-,244				,018	,076	-,167
15-Ítems					-,211	-,087	-,124

Nota: 15-Ítems= Test de los 15-Ítems de Rey; VSVT= Victoria Symptom Validity Test; TOMM= Test of Memory Malinger; OMT1= primer ensayo del Oral Making Test; FD4= cuarto ensayo del Five Digit Test; CBT= Cognitive Bias Task; *= La correlación es significativa al nivel 0,005; **= La correlación es significativa al nivel 0,001

4.3. Porcentaje de análogos clasificados correctamente (Hipótesis 4):

Una vez comprobada la validez convergente y divergente de las pruebas específicas de simulación, procedimos a comprobar su validez clínica.

Cuando hablamos de validez clínica nos estamos refiriendo al grado en el que podemos confiar en que los tests específicos de simulación clasifican a los sujetos que están intentando fingir un daño como simuladores. Para averiguar el porcentaje de análogos clasificados como simuladores por las principales variables de los test categorizamos la variable número de ítems correctos del VSVT (estableciendo como indicativa de simulación una puntuación inferior a 16, interpretada como una ejecución cuestionable en el manual), la puntuación total en el ensayo de retención en el TOMM (tomando como punto de corte una puntuación de 45, especificada en el manual), la variable e-score del Test de la b (tomando como punto de corte una puntuación de 90, propuesta por el manual para los casos de traumatismo), la puntuación combinada del Dot Counting (tomando como punto de corte indicativo de simulación una puntuación de 17, propuesta por Boone, Lu, Back y cols., 2002) y el número total de elementos recordados en el Test de los 15-ítems de Rey (tomando como punto de corte un número

de elementos recordados inferior a 9, propuesto por Rey (1964) como sospechoso de simulación).

Mediante la categorización de dichas variables, el VSVT clasificó correctamente al 84,6% de los simuladores, el TOMM al 87% y el Test de la b al 79,6%. Menos eficaz se mostró el Dot Counting, que detectó sólo a la mitad de los análogos como simuladores (50%), y particularmente bajo fue el porcentaje de simuladores detectado por el Test de los 15-ítems de Rey (9,4%).

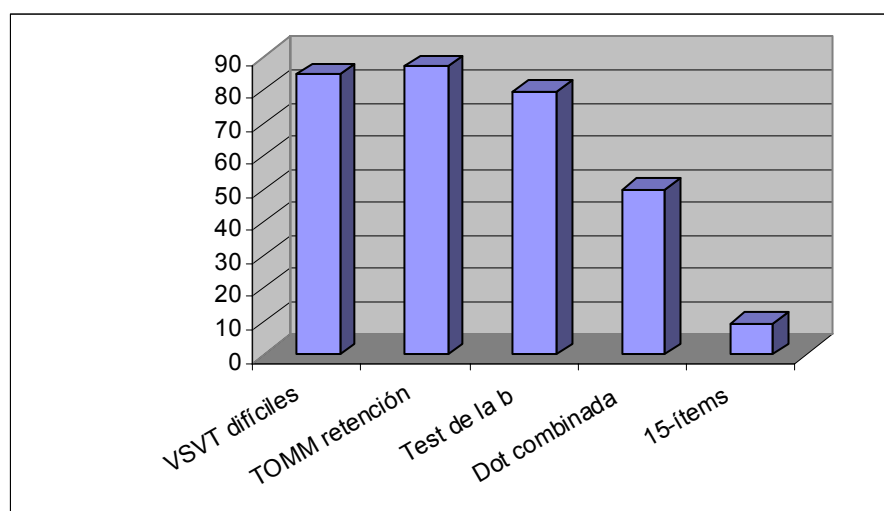


Figura 1. Porcentajes de clasificación de los análogos como simuladores

4.4. Comparación de los resultados de nuestra muestra con los de las muestras originales (Hipótesis 5):

Se procedió a estudiar en qué medida nuestros resultados eran similares a los de la población de origen. Para ello se realizaron 7 comparaciones de las medias de las principales variables de los test obtenidas en nuestro estudio y las referidas en los manuales de las pruebas y, en el caso del Test de los 15-ítems de Rey, con las puntuaciones publicadas en un citado artículo sobre esta prueba (Arnett y cols., 1995). Los resultados mostraron que, para el VSVT, no existen diferencias estadísticamente significativas (ver Tabla 16).

Tabla 16. Comparación de las puntuaciones en el Victoria Symptom Validity Test con las referidas en el manual

	Estudio actual (n= 54)		Muestra manual (n= 43)		<i>p</i>
	Media	Dt	Media	Dt	
Puntuación ítems fáciles	20,13	4,42	20,30	4,35	0,850
Puntuación ítems difíciles	11,01	4,17	10,95	6,06	0,954
Puntuación total	31,13	6,83	31,26	9,05	0,935

Nota: Dt= desviación típica

En cuanto al TOMM, los resultados encontrados en nuestra muestra de análogos tampoco mostraron diferencias estadísticamente significativas con respecto a las obtenidas en el manual de la prueba (ver Tabla 17).

Tabla 17. Comparación de las puntuaciones en el Test Of Memory Malingering con las referidas en el manual

	Estudio actual (n= 54)		Muestra manual (n= 27)		<i>p</i>
	Media	Dt	Media	Dt	
Ensayo 1	31,27	6,99	32,50	7,50	0,468
Ensayo 2	35,44	8,55	35,30	9,40	0,946
Retención	32,92	9,99	30,90	10,60	0,372

Nota: Dt= desviación típica

Con respecto al Test de la b, en su manual no se emplea un grupo de análogos, por lo que no nos ha sido posible realizar la comparación. Tampoco en el estudio original del que hemos extraído las variables empleadas para el Dot Counting se empleó grupo de análogos. En las Tablas 18 y 19 se muestran las medias y desviaciones típicas obtenidas por nuestro grupo en ambas pruebas.

Tabla 18. Medias y desviaciones típicas obtenidas por los análogos en el Test de la b

	e-score	Errores d	Errores comisión	Errores omisión	Tiempo
Media	313,09	11,75	9,59	60,72	38,92
Dt	259,10	18,10	13,13	42,39	15,64

Nota: Dt= desviación típica

Tabla 19. Medias y desviaciones típicas obtenidas por los análogos en el Dot Counting Test

	Puntuación Combinada	Tiempo ítems agrupados	Ratio	Errores
Media	18,56	6,40	1,41	3,86
Dt	7,45	3,10	0,52	2,89

Nota: Dt= desviación típica

Por último, el grupo de análogos de nuestro estudio tampoco presentó diferencias estadísticamente significativas con respecto a las del grupo de análogos empleado por Arnett y cols. (1995) en el Test de los 15-ítems de Rey (ver Tabla 20).

Tabla 20. Comparación de las puntuaciones en el Test de los 15-ítems de Rey con las referidas por Arnett y cols. (1995)

	Estudio actual (n= 54)		Muestra Arnett y cols. (n= 49)		<i>p</i>
	Media	Dt	Media	Dt	
Ítems totales	12,43	2,89	12,14	2,57	0,593

Nota: Dt= desviación típica

5. Discusión:

En este estudio hemos realizado, en primer lugar, una adaptación de las pruebas de simulación que decidimos estudiar, para lo cual llevamos a cabo una doble traducción cruzada (“*back translation*”). A continuación, estudiamos el efecto del entrenamiento (“*coaching*”) en los análogos mediante el empleo de diferentes instrucciones, descubriendo que no influía en la ejecución de las pruebas de simulación. Por otro lado, hemos realizado un estudio preliminar de la fiabilidad y la validez de las pruebas, que ha puesto de manifiesto que el VSVT, el TOMM, el Test de la b y el Test 15-Ítems de Rey poseen una adecuada validez convergente y una excelente validez divergente. Una vez concluidos estos pasos previos, se estudió la ejecución de un grupo de análogos sofisticados en las pruebas específicas de simulación. Los resultados han mostrado (con excepción del Test 15-Ítems de Rey) una clasificación correcta de los análogos como simuladores muy adecuada y, además, se ha encontrado una enorme

similitud entre los resultados de nuestra muestra y los publicados sobre las muestras norteamericanas originales.

El efecto de las instrucciones en el grupo de análogos

El tema de las instrucciones ha sido especialmente cuidado en nuestra investigación debido a que uno de los principales problemas que suele señalarse de los tests específicos de simulación es su vulnerabilidad al entrenamiento. En este sentido, hemos dividido de modo aleatorio a los análogos en dos grupos, de modo que uno recibía instrucciones sobre los síntomas típicos del Síndrome Post-Convulsional (SPC) y el otro recibía, además, instrucciones específicas sobre los tests (ver Apéndice 2). Sorprendentemente, y en contra de lo que suele encontrarse en la literatura (Gunstad y Suhr, 2001; Merten y cols., 2005; Powell, Gfeller, Hendricks y Sharland, 2004; Rose, Hall, Szalda-Petree y Bach, 1998; Surh y Gunstad, 2000), no se han encontrado diferencias estadísticamente significativas entre los dos grupos para ninguna de las variables estudiadas. Este resultado puede tener distintas explicaciones: en primer lugar, podría ser que el mero entrenamiento sobre los síntomas del trastorno a simular mejore la ejecución en los tests de simulación. Sin embargo, esta explicación nos parece poco plausible, ya que va en contra de lo que suele encontrarse en las publicaciones (por ejemplo, Inman y cols., 1998). Una segunda explicación es que el simple hecho de advertir a los sujetos sobre no hacer los déficits demasiado obvios es suficiente para que los sujetos mejoren su ejecución en los tests específicos de simulación, y, dado que esta advertencia estaba incluida en los dos tipos de instrucciones, este sea el motivo por el que no se han encontrado diferencias entre los dos grupos de análogos. Esto sería congruente con lo encontrado en otros trabajos, como en el de Greub y Suhr, (2006), Rose y cols., (1998) o Suhr y Gunstad, (2000). Una tercera explicación posible es que, debido a que las instrucciones específicas que hemos proporcionado son complejas, los

sujetos no serían capaces de integrarlas de modo que les permitan alcanzar una mejor ejecución de los tests de simulación y, por tanto, ser más difíciles de detectar (Bender y Rogers, 2004). Otra explicación alternativa a esta falta de diferencias entre los análogos entrenados en síntomas y los que también han recibido entrenamiento específico sobre los tests específicos de simulación, es que este tipo de pruebas no son sensibles al entrenamiento. Si bien esta no es la postura generalmente aceptada, otras investigaciones apoyan que el entrenamiento específico sobre las pruebas no mejora la ejecución de los análogos en los tests específicos de simulación (Cato y cols., 2002; Martin y cols., 1996). Por último, también es posible pensar que, debido a que nuestro grupo de análogos posee una formación en cuanto a la evaluación y sintomatología del daño cerebral, el tipo de instrucciones proporcionadas no haya influido en su ejecución en la evaluación debido a que su conocimiento previo era superior al proporcionado por las instrucciones. En este sentido, los grupos estarían igualados en cuanto a su conocimiento, y por lo tanto no se encontraron diferencias entre sus resultados. No obstante, y de modo contrario a la intuición, hay investigaciones que ponen de manifiesto que las personas con mayores conocimientos neuropsicológicos son peores simuladores de daño cerebral (Erdal, 2004), lo que invalidaría esta propuesta.

Sea cual sea la verdadera razón de la ausencia de diferencias entre los grupos, consideramos que podemos concluir que, debido al rigor metodológico en la composición de nuestro grupo de análogos, podemos denominarlos como análogos “sofisticados”. Además, el hecho de que sean estudiantes universitarios, y probablemente “intelectualmente superiores” a los pacientes generalmente evaluados por traumatismos craneoencefálicos (Martin y cols., 1996), estos participantes podrían ser menos susceptibles a la dificultad aparente de estas pruebas, que es el punto en el que radica su efectividad. Así, puede argumentarse que las pruebas específicas de

simulación que demuestren su efectividad con este tipo de análogos, serán igual de sensibles, o incluso más, cuando se apliquen a la población con TCE-L, menos “sofisticados” (Bianchini, Mathias y Greve, 2001).

Teniendo en cuenta todo lo mencionado hasta este momento, parece que no es prematuro concluir que los resultados obtenidos por nuestro grupo de análogos poseerán una adecuada validez externa, por lo que nos proveerán con una valiosa información sobre la utilidad de las pruebas de simulación en nuestro contexto.

Propiedades psicométricas de las pruebas

Estamos plenamente de acuerdo con Boone, Lu y Herzberg (2002) cuando señalan que el término fiabilidad no tiene mucho sentido en el contexto de la detección de simulación. La fiabilidad se refiere a la habilidad de un test de medir un rasgo o habilidad estable de modo consistente a través del tiempo y las situaciones. Claramente, esto no es aplicable a ninguna prueba de simulación, puesto que estos no miden ningún rasgo o habilidad estable. A pesar de tener en cuenta estas consideraciones, hemos decidido averiguar si existe consistencia entre todos los ítems de una misma prueba. Los resultados han mostrado que dicha consistencia es muy elevada (tanto en el VSVT como en el Test de la b).

En cuanto a la validez convergente, entre las distintas pruebas de simulación, los resultados han mostrado, de modo general, que existen relaciones estadísticamente significativas entre los distintos tests, lo que está indicando que estas pruebas miden lo mismo, si bien las correlaciones encontradas han sido moderadas. Este hallazgo es realmente interesante, puesto que, si bien numerosos estudios recomiendan el empleo de diferentes medidas de esfuerzo durante la evaluación neuropsicológica (Forrest y cols., 2004; Meyers y cols., 2001; Nies y Sweet, 1994; O'Bryan y cols., 2004; Vickery y cols., 2001, 2004), en el pasado se ha obviado el estudio de las correlaciones entre las

pruebas de simulación. Esto es importante ya que, si las pruebas obtienen correlaciones muy altas, no deberían emplearse como evidencia convergente (Nelson y cols., 2003). Sin embargo, nuestros resultados muestran que los diferentes tests parecen constituir fuentes de información relacionadas, pero independientes. Las correlaciones varían bastante, ya que van de la ausencia de significación a una significación moderada (siempre inferiores a 0,5), resultados comparables a la única exploración que, hasta donde nuestro conocimiento alcanza, se ha realizado hasta el momento sobre este tema (Nelson y cols., 2003). Esto indica que las pruebas de simulación aquí estudiadas ofrecen información no redundante y, por tanto, constituyen fuentes de información relativamente independientes sobre la validez de la evaluación. Estos hallazgos apoyan de modo consistente la recomendación realizada tan frecuentemente sobre el empleo de múltiples medidas de simulación en las evaluaciones forenses.

Por otro lado, también hemos estudiado la validez divergente entre las pruebas específicas de simulación y otras pruebas neuropsicológicas. La ausencia de correlaciones estadísticamente significativas nos indica que las pruebas de simulación no están relacionadas con los niveles reales del funcionamiento cognitivo de los sujetos (Boone, Lu y Herzberg, 2002; Slick y cols., 1997; Tombaugh, 1996). Nuestros resultados indican que esto es así por lo menos en lo que respecta a aspectos tales como velocidad de búsqueda, flexibilidad cognitiva o toma de decisiones. Esta ausencia de correlaciones entre las pruebas neuropsicológicas tradicionales, que los sujetos realizaron poniendo su máximo esfuerzo, y las pruebas de simulación, que los análogos realizaron en la condición de simulación, es más que razonable, si tenemos en cuenta que las pruebas específicas de simulación no miden ninguna habilidad cognitiva, sino la motivación del evaluado.

Clasificación de los análogos como simuladores

Una vez que hemos realizado una adaptación satisfactoria de las pruebas específicas de simulación al castellano, que hemos seleccionado un grupo de análogos altamente sofisticado y válido, y que hemos comprobado que las pruebas de simulación poseen una adecuada fiabilidad y validez, todavía está pendiente el punto más importante de este estudio preliminar. Nos estamos refiriendo a la capacidad de las pruebas de simulación para identificar a los análogos como simuladores. En este sentido, los resultados han mostrado una clasificación muy buena del VSVT, que consiguió detectar al 84,6% de los simuladores. Estos resultados son muy similares al 82% encontrado por Straus y cols., (2002), al 81% de Slick y cols. (1996); al 88% de Tan y cols. (2002) o al 81,5% referido por el manual (Slick y cols., 1997). Resultados similares aunque ligeramente superiores fueron alcanzados por el TOMM, que consiguió identificar al 87% de los análogos. Resultados equiparables los encontramos en el 80% referido por Tan y cols. (2002), el 85% de Etherton y cols. (2005), el 94% de Powell y cols. (2004), el 88% de Tombaugh y cols. (1998) o el 93% referido en el manual (Tombaugh, 1996). También el Test de la b obtiene resultados similares a los obtenidos por el VSVT y por el TOMM, detectando al 79,6% de los análogos, si bien estos resultados no pueden ser comparados con los de la literatura ya que, en las investigaciones con este test, no han sido empleados análogos, por lo que esta parece la primera vez que se demuestra la utilidad de este test mediante el empleo de análogos.

Menos eficaz se mostró el Dot Counting, que detectó sólo a la mitad de los análogos como simuladores (50%). Este porcentaje coincide con el 50% referido por Strauss y cols. (2002) o con el 48% de Binks y cols. (1997). Sin embargo, hay que resaltar que ninguna de estas dos investigaciones emplea, como nosotros, la variable puntuación combinada como criterio de decisión. Igual que sucedía con el Test de la b,

no existe hasta el momento ningún estudio que haya probado la utilidad de dicha puntuación combinada con análogos, por lo que nuestro resultado no puede ser comparado de modo adecuado con ninguna investigación. De todos modos, la obtención de los mismos resultados en distintas investigaciones con el Dot Counting, a pesar de emplear distintas variables, nos lleva a pensar que los resultados de esta prueba en la detección de simulación son moderados.

Finalmente, el porcentaje de simuladores detectado por el Test de los 15-ítems de Rey fue particularmente bajo (9,4%). Este bajo poder de detección coincide en gran medida con otras investigaciones (en torno al 5% de Arnett y cols., 1995 o al 12% referido por Schretlen y cols., 1991). Una posible explicación a estos pobres resultados del Test 15-ítems de Rey es que, puesto que en nuestro estudio hemos empleado estudiantes de 4º curso de Psicología, y esta prueba es ampliamente conocida, es probable al menos algunos de los análogos de nuestro estudio la conozcan, y que por tanto sean capaces de eludir la detección. Esta hipótesis se vio confirmada mediante la aplicación del informe que se administró a los participantes al finalizar la evaluación, en el que una elevada proporción de los estudiantes afirmó conocer esta prueba.

En resumen, el VSVT, TOMM y Test de la b obtienen muy buenos resultados a la hora de clasificar a los análogos como simuladores, mientras el Dot Counting obtiene un nivel de clasificación que no supera el nivel del azar, y el Test 15-Ítems de Rey no consigue detectar a los análogos.

Comparación de los resultados de nuestra muestra con los de las muestras originales

Una vez comprobado que los resultados de la clasificación de los análogos por las distintas pruebas eran similares a los encontrados en la literatura, para finalizar este estudio preliminar decidimos comprobar si las puntuaciones medias de nuestro grupo de análogos eran comparables a las puntuaciones obtenidas por los análogos

norteamericanos. Es sorprendente que los resultados de ambas muestras son casi idénticos, tanto para las medias como para las desviaciones típicas. Esto parece apuntar que las pruebas específicas de simulación no se ven afectadas por variables culturales, sino que parecen más bien pruebas “universales”. Es un poco prematuro extraer conclusiones definitivas, puesto que, a pesar de las recomendaciones de la NAN (2005) sobre la necesidad de estudiar las pruebas de simulación para poder ser aplicadas en contextos diferentes para el que fueron diseñadas, todavía no disponemos de investigaciones sobre este punto. Se han llevado a cabo algunos estudios en los que en el grupo de estudio se incluían algunos participantes pertenecientes a minorías étnicas (Gierok y cols., 2005; Moore y Donders, 2004), o se han desarrollado tests para otras poblaciones, como es el caso del Amsterdam Short-Term Memory Test (Schagen y cols., 1997), sin embargo, hasta donde nosotros hemos podido comprobar, no se han llevado a cabo estudios que comparen directamente la ejecución de poblaciones pertenecientes a diferentes culturas. En este sentido, nuestros datos parecen muy reveladores, aunque es necesario seguir investigando al respecto.

En resumen, los datos obtenidos en este estudio indican que la adaptación y las propiedades psicométricas de los tests neuropsicológicos objeto de estudio en esta Tesis Doctoral son adecuados, mostrando en el grupo de análogos un comportamiento muy parecido al mostrado en los baremos originales en población norteamericana. Una vez obtenidos estos datos, se procedió a estudiar su utilidad clínica en el estudio 2.

CAPÍTULO 8

ESTUDIO 2: UTILIDAD DE LAS PRUEBAS

ESPECÍFICAS DE SIMULACIÓN

1. Introducción:

Una vez que hemos comprobado que las adaptaciones realizadas son satisfactorias, vamos a comprobar la utilidad de las pruebas específicas de simulación en nuestro contexto. Como ha quedado explicado anteriormente (Capítulo 3, apartado 5: La investigación de la simulación), disponemos de diversos métodos para investigar el esfuerzo insuficiente. En este estudio hemos decidido emplear un diseño mixto con la finalidad de valernos de las ventajas que ofrecen los distintos métodos. Para ello, en primer lugar hemos realizado una comparación entre “grupos de prevalencia diferencial” (Rogers, 1997^b), para averiguar si los tests de simulación detectan diferencias entre un grupo en el que, a priori, no existen simuladores (puesto que se trata de un grupo de pacientes en el que no existe búsqueda de compensación económica), y otro grupo en el que se supone que existe una elevada probabilidad de que los sujetos decidan simular (ya que se trata de individuos con traumatismos leves y sin hallazgos en la neuroimagen, que a los 6 meses presentan quejas persistentes y que persiguen la obtención de una compensación económica -están en procesos civiles de litigio, están siendo evaluados por la mutua...).

Criterio de pertenencia al grupo de simulación.

Por otro lado, hemos empleado un diseño de “grupos conocidos” (Rogers, 1997^b), mediante la diferenciación dentro del grupo de búsqueda de compensación de individuos sospechosos de simulación. Para la composición de este grupo los sujetos debían puntuar como simuladores en al menos dos pruebas específicas de simulación independientes a la que esté siendo estudiada en cada momento (es decir, si estamos estudiando el VSVT, la puntuación en el rango inválido en este test no puede ser empleada como criterio para ser incluido en el grupo sospechoso de simulación, sino

que los pacientes tienen que obtener puntuaciones sospechosas en otras dos de las pruebas específicas empleadas en la investigación).

Por último, con el objetivo de aumentar la validez externa, hemos incluido un grupo de análogos (Rogers, 1997^b) que, como ha quedado explicado en el estudio 1, podemos denominar como “sofisticados” ya que poseen conocimientos neuropsicológicos y se les han proporcionado instrucciones encaminadas a hacerlos más resistentes a la detección.

2. Objetivos e Hipótesis:

Una vez comprobado que las adaptaciones y las propiedades psicométricas de los tests específicos de simulación cumplen los requisitos para ser aplicados en nuestro contexto, en el **estudio 2** vamos a establecer la utilidad de dichas pruebas para la detección de simulación en nuestro país. Para ello investigaremos si las pruebas específicas de simulación seleccionadas son capaces de diferenciar entre distintos grupos, valiéndonos de diseños mixtos. Este objetivo general se desglosa en cinco objetivos específicos:

- **Objetivo 1:** Influencia de las variables demográficas y del tiempo transcurrido desde el traumatismo sobre los tests específicos de simulación.
- **Objetivo 2:** Averiguar si las distintas pruebas específicas de simulación son capaces de diferenciar entre los “grupos de prevalencia diferencial”.
- **Objetivo 3:** Establecer el nivel de clasificación correcta de los “grupos de prevalencia diferencial” como sujetos simuladores o no simuladores por las distintas pruebas de simulación.
- **Objetivo 4:** Averiguar si las distintas pruebas específicas de simulación son capaces de diferenciar entre “grupos conocidos”.

- **Objetivo 5:** Establecer el nivel de clasificación correcta de los “grupos conocidos” como sujetos simuladores o no simuladores por las distintas pruebas de simulación.
- **Objetivo 6:** Obtener los puntos de corte que optimicen la especificidad para las distintas variables.
- **Objetivo 7:** Comparar nuestros resultados con los publicados en la literatura para la población norteamericana.

De los objetivos anteriores se han derivado las siguientes hipótesis:

- **Hipótesis 1:** Los tests específicos de simulación no se verán afectados por variables demográficas ni por el tiempo transcurrido desde el traumatismo hasta el momento de la evaluación.
- **Hipótesis 2:** Se encontrarán diferencias entre los grupos de prevalencia diferencial de simulación en la ejecución en los test de simulación, de modo que el grupo con una prevalencia menor (grupo NBC) tendrá una ejecución mejor que la del grupo con una elevada prevalencia (grupo BC), que será igual al grupo AN.
- **Hipótesis 3:** Las pruebas de simulación conseguirán clasificar una elevada proporción del grupo BC y de los análogos como simuladores, mientras el grupo NBC será clasificado por las pruebas de simulación como no simuladores.
- **Hipótesis 4:** La ejecución en los test de simulación será diferente entre los “grupos conocidos”, de modo que el grupo NBC será igual al grupo BC-NSS, el grupo BC-SS será igual al grupo AN, y los grupos NBC y BC-NSS obtendrán una ejecución distinta a la de los grupos BC-SS y AN.
- **Hipótesis 5:** Las pruebas de simulación conseguirán clasificar una elevada proporción del grupo BCSS y de los análogos como simuladores, mientras los

grupos NBC y BCNSS serán clasificados por las pruebas de simulación como no simuladores.

- **Hipótesis 6:** Manteniendo una elevada especificidad, los puntos de corte de las pruebas de simulación llevarán asociada una elevada sensibilidad.
- **Hipótesis 7:** No se encontrarán diferencias estadísticamente significativas entre las puntuaciones encontradas en nuestro estudio y las publicadas por la literatura en las pruebas de simulación.

3. Metodología:

3.1. Participantes:

En este estudio participaron 65 pacientes diagnosticados de TCE-L (duración de la pérdida de conciencia igual o inferior a 30 minutos, Glasgow Coma Scale entre 13 y 15, y amnesia postraumática inferior a 24 horas) que presentaban el denominado síndrome postconmocional (SPC). La cuantificación del SPC se realizó mediante la aplicación del Cuestionario Rivermead (ver Apéndice 2). Ninguno de los pacientes obtuvo hallazgos en la TAC, ni presentó alteraciones psiquiátricas o consumo de sustancias premórbidos (debido a estos motivos fueron excluidos del estudio 11 participantes, por lo que el tamaño final de la muestra fue $n=54$). Estos dos últimos aspectos fueron explorados mediante una entrevista clínica con el paciente y, en los casos en los que consideró necesario, con algún familiar cercano. Todos los pacientes fueron evaluados al menos 6 meses después del TCE-L, y su edad estaba comprendida entre los 18 y los 55 años, con la finalidad de incluir sólo pacientes mayores de edad en los que no es esperable la existencia de alteraciones neuropsicológicas o neurofuncionales debidas al deterioro adscrito a la edad.

Para el diseño de prevalencia diferencial (Hipótesis 2) estos pacientes fueron divididos en dos grupos; el grupo de no-búsqueda de compensación (NBC) no estaba

implicado en ningún proceso del que pudiera obtener algún tipo de beneficio por las consecuencias de su traumatismo. Estaba formado por 16 hombres y 14 mujeres (N=30), con una media de edad de 32,50 años (Sd=13,67) y una media de 9,30 años de escolaridad (Sd= 3,45). La media del tiempo transcurrido desde el TCE hasta el momento de la evaluación fue de 263,13 días (Sd= 378,66). El grupo de búsqueda de compensación (BC) estaba constituido por 17 hombres y 7 mujeres (n=24) implicados en procesos de los que esperaban obtener algún tipo de beneficio económico por las consecuencias de su traumatismo. Este grupo tenía una media de 35,62 años de edad (Sd= 10,74) y de 8,87 años de escolaridad (Sd= 2,78), y la media de los días transcurridos desde el TCE hasta la evaluación fue de 172,04 (Sd= 507,26 días).

Para el diseño de grupos conocidos (hipótesis 3) el grupo BC fue dividido, a su vez, en dos grupos: un grupo de búsqueda de compensación no sospechoso de simulación (BC-NSS), constituido por 8 hombres y 6 mujeres (N= 14) con una media de edad de 35,92 años (Sd= 10,88), una media de escolaridad de 9,14 años (Sd= 3,50), y una media de tiempo transcurrido desde el traumatismo hasta la evaluación de 287,78 días (Sd= 448,2), y un grupo de búsqueda de compensación sospechoso de simulación. Este grupo estaba formado por aquellos pacientes que puntuaron en el rango indicativo de simulación en dos o más de las pruebas específicas empleadas en la investigación. Estaba constituido por 8 hombres y 2 mujeres (N= 10), con una media de edad de 35,20 años (Sd= 11,10), una escolaridad de 8,50 años (Sd= 1,35) y una media de tiempo transcurrido desde el traumatismo de 490,00 días (Sd= 583,88).

En este estudio también participaron los 54 análogos descritos en el estudio 1.

3.2. Instrumentos:

Los instrumentos específicos de simulación empleados en este estudio fueron los mismos que los utilizados en el estudio 1. Estos instrumentos se administraron en el

conjunto de una amplia batería neuropsicológica, cuyas pruebas serán descritas en el capítulo siguiente.

3.3. Procedimiento:

El reclutamiento de los pacientes con TCE se realizó retrospectivamente, a través de un listado de casos de TCE leve atendidos por los servicios médicos desde enero de 2003 hasta abril de 2004. En todos estos pacientes se había realizado una exploración de TAC sin que ninguno presentara alteraciones relevantes (TAC negativo). Mediante una entrevista telefónica se exploró la existencia del síndrome post-conmocional y si estaba presente el paciente era citado para una evaluación exhaustiva. De los 385 pacientes que componían la lista el 32% fue excluido por no cumplir los criterios de inclusión referidos a la edad. No fue posible contactar con el 10% de los pacientes debido a errores en el número telefónico o a cambios de domicilio, y el 3% no pudo participar en el estudio debido a factores geográficos, dado que no pertenecían a la provincia de Granada. Otro 10% de los participantes fue excluido por abuso de sustancias o presencia de enfermedades psiquiátricas. El 30% de los pacientes de la lista no presentaban ningún tipo de sintomatología, por lo que tampoco cumplían los criterios de inclusión para el estudio, y la tasa de renuncia fue cercana al 0,5%. El 14% restante se corresponde con los 65 pacientes evaluados.

El protocolo de estudio era explicado claramente a cada paciente y tras la obtención de su consentimiento informado se procedía a su evaluación. La metodología de exploración ha estado dividida en dos partes desarrolladas en días diferentes: en la primera se realizaba una prueba de neuroimagen funcional (SPECT-PCR) y en la segunda se administraba la batería neuropsicológica. La entrevista y la batería neuropsicológica tenían una duración aproximada de dos horas y media, motivo por el cual transcurrida aproximadamente una hora y media se realizaba un descanso de unos

20 minutos. En ocasiones puntuales y cuando las características particulares del caso así lo requerían se realizaban más períodos de descanso o se empleaban dos sesiones para realizar la evaluación.

El procedimiento seguido con el grupo de análogos ha quedado descrito en el Estudio 1.

3.4. Variables:

Las variables empleadas en este estudio fueron las mismas que las descritas en el estudio 1, excluyendo la medida de Retención del TOMM (por ser esta parte del test opcional) e incluyendo tres nuevas variables del VSVT:

- TR ítems fáciles= Media de tiempo empleado para responder a los ítems fáciles.
- TR ítems difíciles= Media de tiempo empleado para responder a los ítems difíciles.
- TR total= Media de tiempo empleado para responder a todos los ítems de la prueba.

3.5. Análisis estadísticos

En primer lugar, se realizaron correlaciones de Spearman entre distintas variables demográficas y las principales variables de los tests de simulación. En segundo lugar, debido a que no se cumplía el supuesto de homogeneidad de las varianzas, se realizó un análisis no paramétrico para “n” grupos independientes utilizando el estadístico de U de Mann-Whitney, tanto en el diseño de “grupos de prevalencia diferencial” como en el de “grupos conocidos”, para conocer si existían diferencias globales entre los grupos en la ejecución de las distintas variables de las pruebas. Si estas diferencias eran significativas, se realizaron análisis a posteriori no paramétricos comparando los grupos por pares (estadístico Kruskal-Wallis).

A continuación, para conocer la significación clínica de las diferencias encontradas en el análisis anterior, se obtuvieron los tamaños del efecto utilizando la *delta* de Cohen de acuerdo con la formulación de Zackzanis (2001). La fórmula utilizada fue la siguiente:

$$\text{Delta de Cohen} = \frac{\text{Media}_1 - \text{Media}_2}{\text{Media}_{Dt1 \text{ y } Dt2}}$$

Finalmente, para comprobar la utilidad diagnóstica de las variables se realizaron análisis de curvas ROC de las cuales se obtuvieron los puntos de corte, la sensibilidad y la especificidad de cada variable. Para aumentar su utilidad clínica, se obtuvieron los distintos puntos de corte, su valor predictivo positivo y negativo ajustado a las prevalencias teóricas utilizadas en la literatura como son el 15%, el 30% y el 45%.

4. Resultados:

4.1. Influencia de las variables demográficas y del tiempo transcurrido desde el traumatismo sobre los tests de simulación (Hipótesis 1):

En primer lugar procedimos a averiguar si las variables demográficas (edad, escolaridad y sexo) y el tiempo transcurrido desde el traumatismo hasta el momento de la evaluación influían sobre las puntuaciones obtenidas por los pacientes del grupo NBC en las pruebas de simulación. Para ello se realizaron correlaciones de Spearman entre dichas variables (edad, escolaridad, sexo y tiempo transcurrido desde el traumatismo) y las principales variables de las pruebas de simulación (número total de aciertos en el VSVT, aciertos en el ensayo 2 del TOMM, e-score en el Test de la b, puntuación combinada en el Dot Counting y número de elementos recordados en el Test de los 15-Ítems de Rey).

Los resultados mostraron que las variables estudiadas no están relacionadas con las puntuaciones obtenidas en las pruebas de simulación, excepto en el caso de del Test

15-ítems de Rey, que está relacionado con la escolaridad de los pacientes, si bien de un modo bastante marginal (ver Tabla 21).

Tabla 21. Correlaciones de Spearman entre las variables demográficas y el tiempo transcurrido desde el traumatismo con las puntuaciones en las pruebas de simulación

Test (variable)	edad	escolaridad	género	Tiempo TCE
VSVT (ítems totales)	0,022	0,205	-0,115	-0,084
TOMM (ensayo 2)	-0,116	0,146	-0,018	-0,129
Test b (e-score)	0,082	-0,208	-0,185	0,175
Dot Counting (Combo)	0,141	-0,141	0,093	0,148
15-Ítems (recuerdo total)	0,274	-0,301*	0,091	-0,053

Nota: Tiempo TCE= tiempo transcurrido desde la ocurrencia del traumatismo craneoencefálico; VSVT= Victoria Symptom Validity Test; TOMM= Test of Memory Malingering; Combo= puntuación combinada; * La correlación es significativa al nivel 0,05 (bilateral).

4.2. Análisis de los “grupos de prevalencia diferencial” (Hipótesis 2)

En segundo lugar procedimos a averiguar si las pruebas específicas de simulación diferenciaban entre grupos con distintas probabilidades de simulación. Para ello se realizaron 18 análisis no paramétricos para 3 grupos independientes (Kruskal-Wallis): seis para el VSVT, dos para el TOMM, cinco para el Test de la b, cuatro para el Dot Counting y uno para el Test de los 15-Ítems de Rey. En todos los casos la VI fue el grupo diferencial [grupo de no-búsqueda de compensación (NBC) vs. grupo de búsqueda de compensación (BC) vs. grupo de análogos (AN)]. Las VD para el VSVT fueron: número de ítems fáciles correctos, número de ítems difíciles correctos, número de ítems totales correctos, media del tiempo de reacción para los ítems fáciles, media del tiempo de reacción para los ítems difíciles y media del tiempo de reacción total. Las VD para el TOMM fueron: número total de aciertos en el ensayo 1 y número total de aciertos en el ensayo 2. Para el Test de la b se incluyeron las variables: e-score, número de errores d, número de errores de comisión, número de errores de omisión y media del tiempo para completar los 15 estímulos. En el caso del Dot Counting se emplearon

como VD la puntuación combinada, la media del tiempo empleado para contar los ítems agrupados, el número de errores y el ratio, y en el Test de los 15-Ítems de Rey la VD seleccionada fue el número de elementos recordados. En los casos en los que se encontraron diferencias estadísticamente significativas en el análisis anterior se procedió a realizar comparaciones dos a dos (grupos NBC vs. BC, grupos NBC vs. AN y grupos BC vs. AN) mediante el análisis de Mann-Whitney.

En el caso del VSVT, los resultados mostraron que existían diferencias estadísticamente significativas entre los tres grupos para todas las variables estudiadas: número de aciertos en los ítems fáciles [$\chi^2(2)= 22,10; p<0,000$], número de aciertos en los ítems difíciles [$\chi^2(2)= 41,91; p<0,000$], número de aciertos totales [$\chi^2(2)= 43,49; p<0,000$], media del tiempo de reacción para los ítems fáciles [$\chi^2(2)= 13,96; p<0,001$], media le tiempo de reacción para los ítems difíciles [$\chi^2(2)= 6,77; p<0,034$] y media del tiempo de reacción para los ítems totales [$\chi^2(2)= 7,37; p<0,026$]. En cuanto a las comparaciones a posteriori por pares, en el caso de los ítems fáciles se encontró que el rendimiento del grupo NBC fue estadísticamente superior al grupo BC y este estadísticamente superior al grupo AN, de acuerdo con la hipótesis planteada. En el caso de los ítems difíciles y de los ítems totales se encontró que el rendimiento del grupo AN fue estadísticamente inferior a los otros dos. Por último, en cuanto a los tiempos de reacción (tanto de los ítems fáciles como de los difíciles y los totales) se encontró que el grupo NBC era estadísticamente inferior a los grupos BC y AN (ver Tabla 22).

En el TOMM se encontraron diferencias estadísticamente significativas entre los grupos en las dos variables estudiadas: número de aciertos en el ensayo 1 [$\chi^2(2)= 60,88; p<0,000$] y número de aciertos en el ensayo 2 [$\chi^2(2)= 60,32; p<0,000$]. En ambos casos, las comparaciones a posteriori mostraron que el grupo NBC fue estadísticamente

superior a los otros dos, seguido por el grupo de BC y finalmente por el grupo AN, que fue el que obtuvo un menor número de aciertos (ver Tabla 22).

En cuanto al Test de la b, se encontraron diferencias estadísticamente significativas entre los grupos para todas las variables excepto para el tiempo: e-score [$\chi^2(2)= 41,16; p<0,000$], errores tipo d [$\chi^2(2)= 29,44; p<0,000$], errores de comisión [$\chi^2(2)= 34,64; p<0,000$] y errores de omisión [$\chi^2(2)= 22,55; p<0,000$]. Los análisis a posteriori mostraron que en el caso del e-score se cumplió el patrón esperado (grupo NBC significativamente inferior al grupo BC y este significativamente inferior al grupo AN), mientras en el caso del número de errores tipo *d* el grupo NBC obtuvo una puntuación significativamente inferior a los otros dos grupos (BC y AN). Tanto en el caso de los errores de omisión como en el de los de comisión el grupo AN obtuvo una puntuación significativamente superior a los otros dos grupos (NBC y BC) (ver Tabla 22).

En el caso del Dot Counting, se encontraron diferencias estadísticamente significativas entre los grupos para las cuatro variables estudiadas: puntuación combinada [$\chi^2(2)= 29,61; p<0,000$], media del tiempo necesario para responder a los ítems agrupados [$\chi^2(2)= 34,01; p<0,000$], número de errores [$\chi^2(2)= 6,35; p<0,042$] y ratio [$\chi^2(2)= 22,35; p<0,000$]. Las comparaciones a posteriori mostraron que, en la puntuación combinada, el grupo AN obtuvo una puntuación estadísticamente superior a los otros grupos (NBC y BC). En cuanto a la media del tiempo necesario para contar los ítems agrupados, el grupo NBC se mostró significativamente inferior a los otros dos grupos (BC y AN). Para el número total de errores, las comparaciones a posteriori mostraron que el grupo AN cometió significativamente más errores que el grupo NBC, mientras que no se encontraron diferencias estadísticamente significativas entre el grupo

Tabla 22. Medias, desviaciones típicas y análisis a posteriori para los grupos de prevalencia diferencial en las pruebas específicas de simulación

Prueba	Variables	Grupo NBC			Grupo BC			Grupo AN			X^2 (Kruskall-Wallis)	P	Comparaciones a posteriori (Mann-Whitney)
		Media	Dt		Media	Dt		Media	Dt				
VSVT	Ítems fáciles	23,73	0,73	22,06	2,93	4,42	20,13	4,42	22,10	0,000	NBC>BC>AN		
	Ítems difíciles	21,36	2,49	16,53	6,98	4,17	11,01	4,17	41,91	0,000	(NBC=BC)>AN		
	Ítems totales	45,10	2,6	38,6	9,06	6,83	31,13	6,83	43,49	0,000	(NBC=BC)>AN		
	TR ítems fáciles	1,73	1,53	2,98	1,93	3,29	3,03	3,29	13,96	0,001	NBC<(BC=AN)		
	TR ítems difíciles	2,6	1,67	4,71	3,46	3,31	3,98	3,31	6,77	0,034	NBC<(BC=AN)		
	TR ítems total	2,17	1,31	3,85	2,65	6,33	3,96	6,33	7,37	0,026	NBC<(BC=AN)		
TOMM	Ensayo 1	47,86	2,66	39,62	8,2	6,99	31,27	6,99	60,88	0,000	NBC>BC>AN		
	Ensayo 2	49,80	0,76	43,62	8,39	8,55	35,44	8,55	60,32	0,000	NBC>BC>AN		
Test de la b	E-score	58,02	20,85	153,98	169,55	259,10	313,09	259,10	41,16	0,000	NBC<BC<AN		
	Errores d	0,20	0,61	4,58	7,82	18,10	11,75	18,10	29,44	0,000	NBC<(BC=AN)		
	Errores c	0,16	0,46	3,37	9,14	13,13	9,59	13,13	34,64	0,000	(NBC=BC)<AN		
	Errores o	22,13	17,03	31,20	26,80	42,39	60,72	42,39	22,55	0,000	(NBC=BC)<AN		
	Tiempo	33,29	9,59	44,86	26,63	15,64	38,92	15,64	3,61	0,164			
Dot Counting	Combo	10,54	2,72	14,95	8,55	7,45	18,56	7,45	29,61	0,000	(NBC=BC)<AN		
	Tiempo ítems agrupados	2,74	1,22	5,45	3,47	3,10	6,40	3,10	34,01	0,000	NBC<(BC=AN)		
15-Ítems de Rey	Errores	2,28	1,71	2,54	1,99	2,89	3,86	2,89	6,35	0,042	BC=(NBC,AN); NBC<AN		
	Ratio	2,25	0,91	1,41	0,40	0,52	1,41	0,52	22,35	0,000	NBC>(BC=AN)		
	Ítems totales	12,11	3,94	10,15	4,13	2,89	12,43	2,89	5,59	0,061			

Nota: AN= análogos; BC= Búsqueda de compensación; Combo= puntuación combinada; Dt= desviación típica; Errores c= errores comisión; Errores o= errores omisión; NBC= No Búsqueda de Compensación; TOMM= Test of Memory Malinger; TR= tiempo de reacción; VSVT= Victoria Symptom Validity Test.

BC y los otros dos (NBC y AN). Por último, para la puntuación ratio, los análisis mostraron una puntuación superior para el grupo NBC con respecto a los otros dos (BC y AN) (ver Tabla 22).

Por último, el Test de los 15-Ítems de Rey fue la única de las pruebas aplicadas en la que no se encontraron diferencias estadísticamente significativas entre los grupos [$\chi^2(2)= 5,59; p<0,061$] (ver Tabla 22).

Estudio de los tamaños del efecto:

Para realizar un análisis más preciso de la relevancia estadística de los hallazgos se calcularon los tamaños del efecto de los análisis anteriores empleando el estadístico Delta de Cohen. En el estudio de los grupos NBC y BC, la mejor variable será aquella que mejor diferencie entre los grupos, y por tanto la que lleve asociado un Delta de Cohen mayor. Los resultados mostraron que esta variable era el Ensayo 1 del TOMM ($\delta= 1,51$), mientras que la variable que menos diferencia entre los grupos, y que por tanto obtiene peores resultados, es la variable errores del Dot Counting ($\delta= 0,14$) (ver Tabla 23).

En cuanto a la diferenciación de los grupos NBC y AN, la variable que mostró mejores resultados de acuerdo con la hipótesis planteada fue, nuevamente, el Ensayo 1 del TOMM ($\delta= 3,43$), y la variable que menos diferencias encontró entre estos grupos fue el número de elementos totales recordados en el Test de los 15-Ítems de Rey ($\delta= 0,09$) (ver Tabla 23).

Por último, para diferenciar entre el grupo BC y el grupo AN, la mejor variable será aquella que menos diferencias encuentre entre los grupos, y por tanto la que lleve asociado un Delta de Cohen menor. La variable Ratio del Dot Counting fue la que menos diferencias encontró entre estos dos grupos, y la variable Ensayo 1 del TOMM fue la que más diferencias detectó entre los grupos BC y AN ($\delta= 1,09$) (ver Tabla 23).

Tabla 23. Delta de Cohen de las variables de los tests específicos de simulación para los grupos de prevalencia diferencial

Prueba	Variables	NBC vs. BC	NBC vs. AN	BC vs. AN
VSVT	Ítems fáciles	0,912	1,398	0,525
	Ítems difíciles	1,020	3,108	0,990
	Ítems totales	1,114	2,962	0,940
	TR ítems fáciles	0,722	0,539	0,019
	TR ítems difíciles	0,822	0,554	0,215
	TR ítems total	0,848	0,468	0,024
TOMM	Ensayo 1	1,517	3,438	1,099
	Ensayo 2	1,350	3,084	0,965
Test de la b	E-score	1,007	1,822	0,742
	Errores d	1,039	1,234	0,553
	Errores c	0,668	1,387	0,558
	Errores o	0,413	1,298	0,853
	Tiempo	0,638	0,446	0,281
Dot Counting	Combo	0,782	1,577	0,451
	Tiempo ítems agrupados	1,155	1,694	0,289
	Errores	0,140	0,686	0,540
	Ratio	1,282	1,174	0,000
15-Ítems de Rey	Ítems totales	0,485	0,093	0,649

Nota: AN= análogos; BC= búsqueda de compensación; c= comisión; Combo= puntuación combinada; NBC= no búsqueda de compensación; o= omisión; TOMM= Test of Memory Malingering; TR=tiempo de reacción; VSVT= Victoria Symptom Validity Test.

4.3. Clasificación de los “grupos de prevalencia diferencial” (Hipótesis 3):

Una vez comprobado que los tests específicos de simulación son capaces de diferenciar entre los grupos de prevalencia diferencial, procedimos a comprobar el porcentaje de sujetos simuladores y no simuladores que las pruebas encontraban en cada uno de estos grupos. Para ello realizamos cinco análisis de contingencia cruzando la variable grupo (NBC, BC y AN) y la clasificación de simulación (simulador vs. no simulador) obtenida mediante la categorización de las principales variables de los test de simulación siguiendo los siguientes criterios: VSVT ítems difíciles (las puntuaciones inferiores a 16, consideradas en el manual como cuestionables por encontrarse en el

rango de las respuestas al azar, fueron categorizadas como indicativas de simulación), TOMM Ensayo 2 (las puntuaciones inferiores a 45 fueron categorizadas como indicativas de simulación, siguiendo las recomendaciones del manual), e-score del Test de la b (las puntuaciones superiores a 90 fueron consideradas indicativas de simulación, ya que este es el punto de corte indicado en el manual para los casos de traumatismo), puntuación combinada del Dot Counting (las puntuaciones superiores o iguales a 17 fueron categorizadas como indicativas de simulación, siguiendo la propuesta de Boone, Lu, Back y cols., 2002), puntuación total en el Test de los 15-Ítems de Rey (las puntuaciones inferiores a 9 fueron consideradas indicativas de simulación, siguiendo la propuesta de Lezack y cols., 2004).

Los resultados mostraron que existían asociaciones estadísticamente significativas entre los grupos y la clasificación de simulador vs. no simulador en la variable ítems difíciles del VSVT [$\chi^2(2) = 42,52; p < 0,000$], clasificando correctamente al 95% de los pacientes que no buscaban compensación como no simuladores, mientras el 84,6% de los análogos fueron correctamente clasificados como simuladores. En el grupo de pacientes buscadores de compensación, el 31,3% obtuvieron puntuaciones indicativas de simulación en esta variable (ver Tabla 24 y Figura 2).

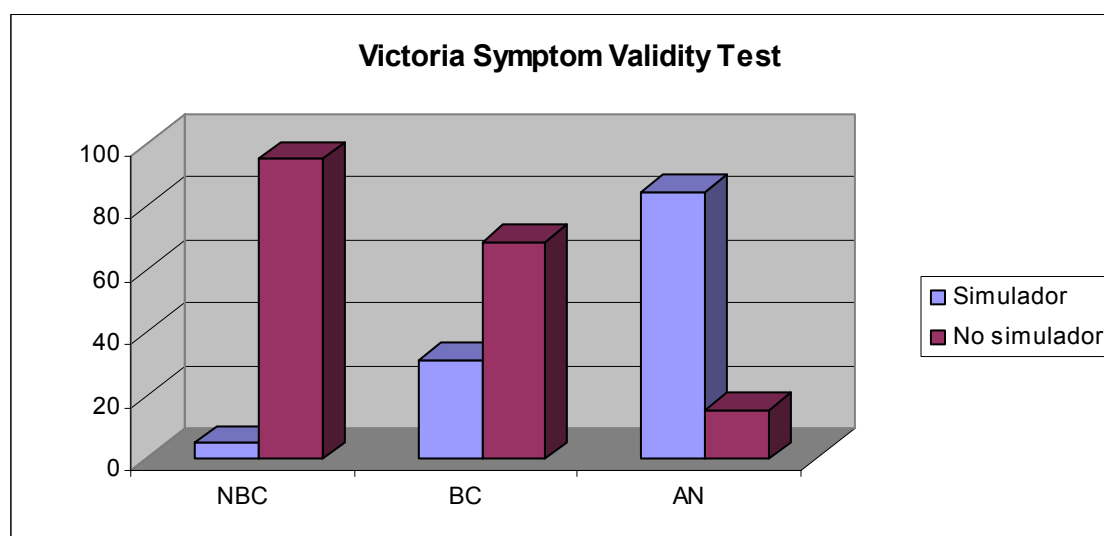


Figura 2: Porcentaje de sujetos clasificados como simuladores y no simuladores por el Victoria Symptom Validity Test en los “grupos de prevalencia diferencial”

También se encontraron asociaciones estadísticamente significativas entre los grupos y la clasificación de simulación en la variable Ensayo 2 del TOMM [$\chi^2(2)=56,74$; $p<0,000$], clasificando correctamente como no simuladores al 100% de los pacientes del grupo NBC y como simuladores al 85,2% del grupo AN y al 45,8% del grupo BC (ver Tabla 24 y Figura 3).

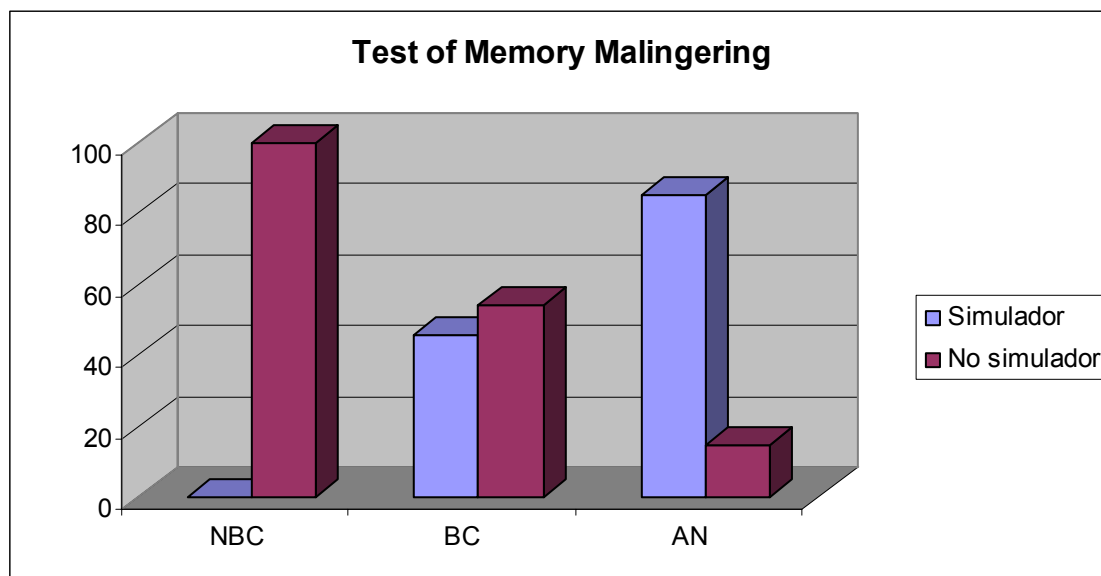


Figura 3: Porcentaje de sujetos clasificados como simuladores y no simuladores por el Test of Memory Malingering en los “grupos de prevalencia diferencial”

En la variable e-score del Test de la b también se encontraron asociaciones estadísticamente significativas entre los grupos y la clasificación de simulación [$\chi^2(2)=35,93$; $p<0,000$], clasificando correctamente al 90% de los pacientes del grupo NBC como no simuladores y al 77,8% del grupo AN como simuladores. En cuanto al grupo BC, el 45,8% obtuvo puntuaciones consideradas indicativas de simulación (ver Tabla 24 y Figura 4).

Los resultados mostraron asociaciones estadísticamente significativas entre los grupos y la clasificación de simulación en la puntuación combinada del Dot Counting [$\chi^2(2)=20,10$; $p<0,000$], clasificando correctamente a la totalidad de los pacientes del

grupo NBC como no simuladores y al 47,8% del grupo AN como simuladores. También fueron clasificados como simuladores el 22,7% del grupo BC (ver Tabla 24 y Figura 5).

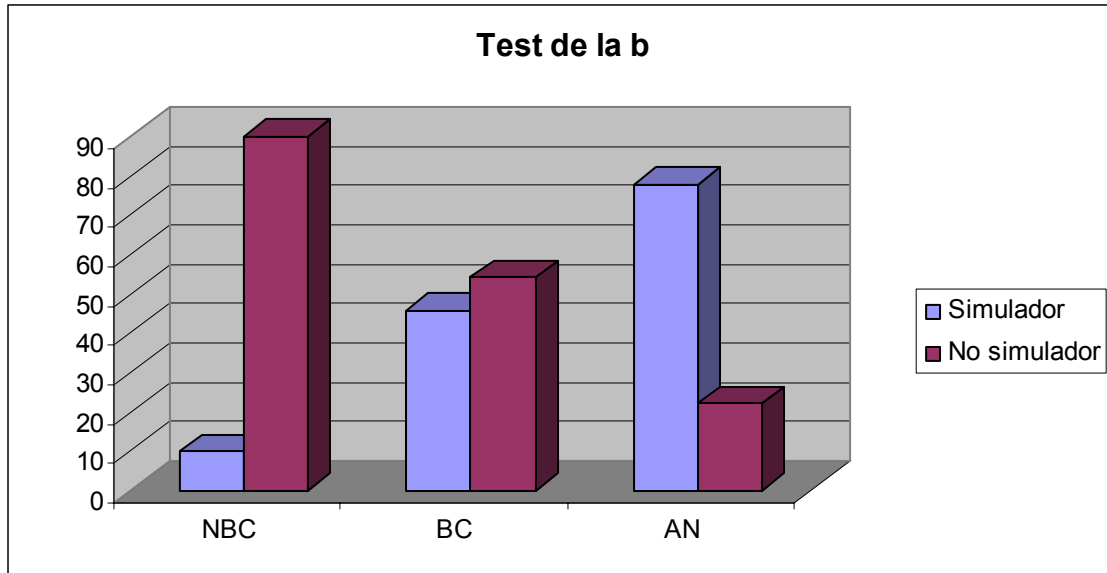


Figura 4: Porcentaje de sujetos clasificados como simuladores y no simuladores por el Test de la b en los “grupos de prevalencia diferencial”

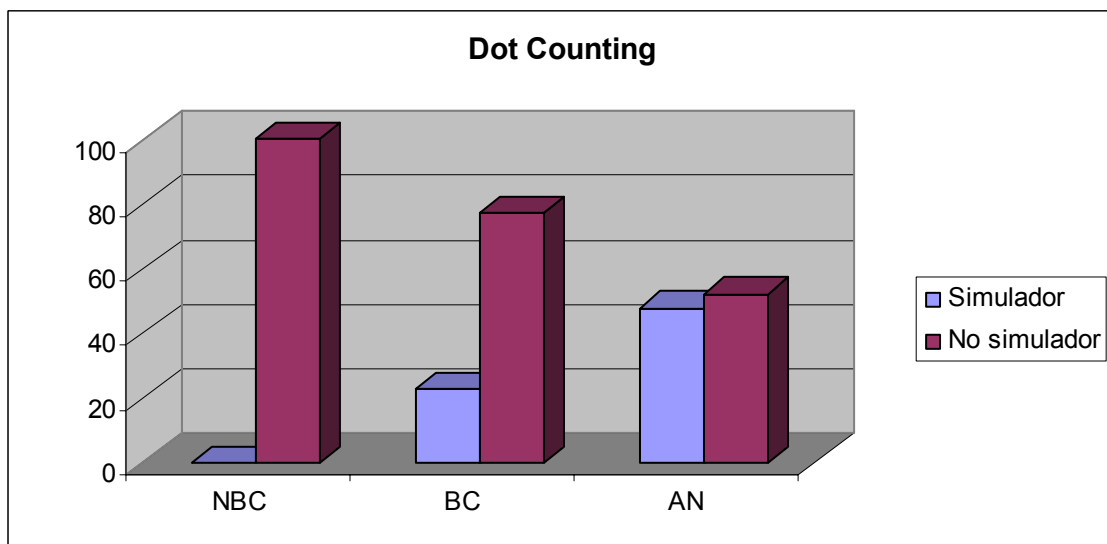


Figura 5: Porcentaje de sujetos clasificados como simuladores y no simuladores por el Dot Counting en los “grupos de prevalencia diferencial”

Por último, también se encontraron asociaciones estadísticamente significativas entre los grupos y la clasificación de simulación en la variable número de elementos totales recordados del Test de los 15-Ítems de Rey [$\chi^2(2) = 7,10; p < 0,029$], clasificando

correctamente el 85,2% de los pacientes NBC, y como simuladores al 9,4% del grupo AN y al 35% del grupo BC (ver Tabla 24 y Figura 6).

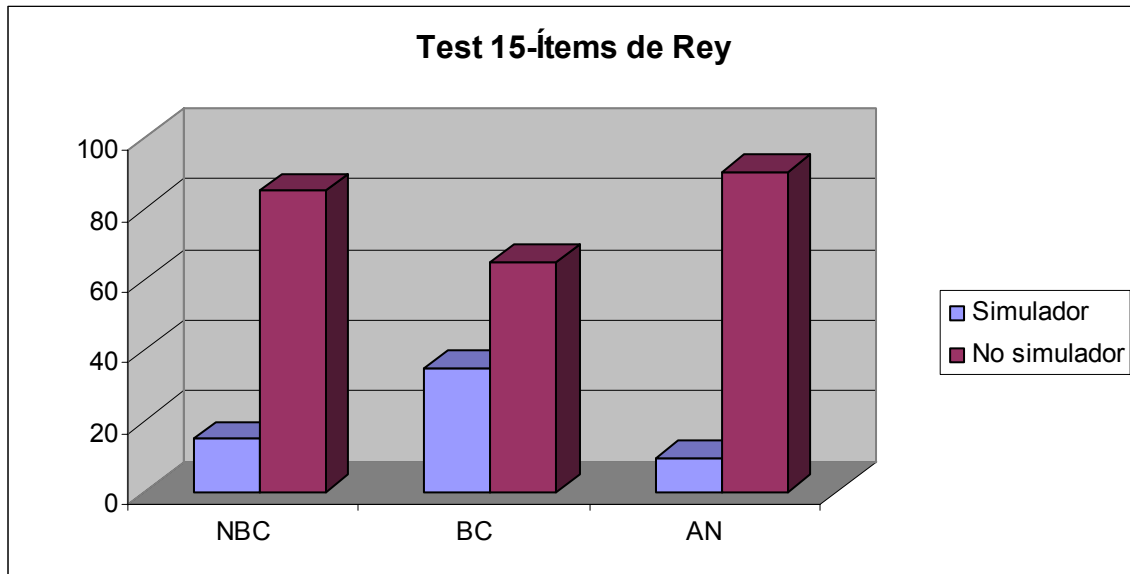


Figura 6: Porcentaje de sujetos clasificados como simuladores y no simuladores por el Test 15-Ítems de Rey en los “grupos de prevalencia diferencial”

Tabla 24. Porcentajes de sujetos correctamente clasificados por las pruebas de simulación en los grupos de prevalencia diferencial.

Prueba (variable)	Grupo NBC	Grupo BC	Grupo AN	X^2	p
	% No simulación	% Simulación	% Simulación		
VSVT (ítems difíciles)	95%	31,3%	84,6%	42,52	0,000
TOMM (Ensayo 2)	100%	45,8%	85,2%	56,74	0,000
Test de la b (e-score)	90%	45,8%	77,8%	35,93	0,000
Dot Counting (Combo)	100%	22,7%	47,8%	20,10	0,000
15-Ítems (total)	85,2%	35%	9,4%	7,10	0,029

Nota: NBC= no búsqueda de compensación; BC= búsqueda de compensación; AN= análogos; VSVT= Victoria Symptom Validity Test; TOMM= Test of Memory Malingering; Combo= puntuación combinada.

4.4. Análisis de “grupos conocidos” (Hipótesis 4):

Una vez hemos comprobado que los tests específicos de simulación son capaces de diferenciar entre grupos de prevalencia diferencial y de clasificarlos correctamente,

se procedió a comprobar si eran capaces de distinguir, realmente, a los simuladores. Para ello, empleamos cuatro grupos: un grupo de no-búsqueda de compensación (NBC), un grupo con búsqueda de compensación pero no sospechoso de simulación (BCNSS), un grupo con búsqueda de compensación y sospechoso de simulación (BCSS) y por último un grupo de análogos (AN).

Siguiendo la estructura del objetivo anterior, se realizaron 18 análisis no paramétricos para 4 grupos independientes (Kruskal-Wallis): seis para el VSVT, dos para el TOMM, cinco para el Test de la b, cuatro para el Dot Counting y uno para el Test de los 15-Ítems de Rey.

En todos los casos la VI fue el grupo diferencial [grupo de no-búsqueda de compensación (NBC) vs. grupo de búsqueda de compensación no sospechoso de simulación (BCNSS) vs. grupo de búsqueda de compensación sospechoso de simulación (BCSS) vs. grupo de análogos (AN)]. Las VD para el VSVT fueron: número de ítems fáciles correctos, número de ítems difíciles correctos, número de ítems totales correctos, media del tiempo de reacción para los ítems fáciles, media del tiempo de reacción para los ítems difíciles y media del tiempo de reacción total. Las VD para el TOMM fueron: número total de aciertos en el ensayo 1 y en el ensayo 2. Para el Test de la b se incluyeron las variables: e-score, número de errores d, número de errores de comisión, número de errores de omisión y media del tiempo para completar los 15 estímulos. En el caso del Dot Counting se emplearon como VD la puntuación combinada, la media del tiempo empleado para contar los ítems agrupados, el número de errores y el ratio, y en el Test de los 15-Ítems de Rey la VD seleccionada fue el número de elementos recordados. En los casos en los que se encontraron diferencias estadísticamente significativas en el análisis anterior se procedió a realizar comparaciones dos a dos (grupos NBC vs. BCNSS, grupos NBC vs. BCSS, grupos

NBC vs. AN, grupos BCNSS vs. BCSS, grupos BC-NSS vs. AN, y grupos BCSS vs. AN) mediante el análisis de Mann-Whitney.

En cuanto al VSVT, los resultados mostraron que existían diferencias estadísticamente significativas entre los cuatro grupos para todas las variables estudiadas: número de aciertos en los ítems fáciles [$\chi^2(2)= 25,59; p<0,000$], número de aciertos en los ítems difíciles [$\chi^2(2)= 51,84; p<0,000$], número de aciertos totales [$\chi^2(2)= 52,72; p<0,000$], media del tiempo de reacción para los ítems fáciles [$\chi^2(2)= 18,86; p<0,000$], media le tiempo de reacción para los ítems difíciles [$\chi^2(2)= 18,83; p<0,013$] y media del tiempo de reacción para los ítems totales [$\chi^2(2)= 12,21; p<0,007$] (ver Tabla 25).

En el TOMM se encontraron diferencias estadísticamente significativas entre los grupos en las dos variables estudiadas: número de aciertos en el ensayo 1 [$\chi^2(2)= 68,99; p<0,000$] y número de aciertos en el ensayo 2 [$\chi^2(2)= 72,42; p<0,000$] (ver Tabla 25).

También en el Test de la b se encontraron diferencias estadísticamente significativas entre los grupos para todas las variables estudiadas: e-score [$\chi^2(2)= 53,80; p<0,000$], errores tipo d [$\chi^2(2)= 43,37; p<0,000$], errores de comisión [$\chi^2(2)= 39,68; p<0,000$], errores de omisión [$\chi^2(2)= 24,83; p<0,000$] y tiempo [$\chi^2(2)= 9,87; p<0,020$] (ver Tabla 25).

En el caso del Dot Counting, se encontraron diferencias estadísticamente significativas entre los grupos para las cuatro variables estudiadas: puntuación combinada [$\chi^2(2)= 35,55; p<0,000$], media del tiempo necesario para responder a los ítems agrupados [$\chi^2(2)= 41,14; p<0,000$], número de errores [$\chi^2(2)= 10,33; p<0,016$] y ratio [$\chi^2(2)= 27,42; p<0,000$] (ver Tabla 25).

Tabla 25. Medias, desviaciones típicas y análisis a posteriori para los “grupos conocidos” en las pruebas específicas de simulación

Prueba	Variable	Grupo NBC		Grupo BCNSS		Grupo BCSS		Grupo AN		X ² (Kruskal-Wallis)	p	Análisis a posteriori (Mann-Whitney)
		Media	Sd	Media	Sd	Media	Sd	Media	Sd			
VSVT	Ítems fáciles	23,73	0,73	23,57	0,78	20,75	3,53	20,13	4,42	25,59	0,000	NBC>(BCSS=AN) NBC=BCNSS; BCNSS=BCSS; BCNSS>AN
	Ítems difíciles	21,36	2,49	21,57	3,25	12,12	6,40	11,01	4,17	51,84	0,000	(NBC=BCNSS)>(BCSS=AN)
	Ítems totales	45,10	2,60	45,14	2,07	32,87	8,72	31,13	6,83	52,72	0,000	(NBC=BCNSS)>(BCSS=AN)
TR ítems fáciles		1,73	1,53	1,96	1,19	3,88	2,07	3,03	3,29	18,86	0,000	NBC<(BCSS=AN) NBC=BCNSS; BCNSS=BCSS; BCNSS=AN
		2,60	1,67	3,47	3,05	5,79	3,61	3,98	3,31	10,83	0,013	NBC<(BCSS=AN) NBC=BCNSS; BCNSS<BCSS; BCNSS=AN
TR ítems total		2,17	1,31	2,71	2,11	4,84	2,80	3,96	6,33	12,21	0,007	NBC<(BCSS=AN) NBC=BCNSS; BCNSS<BCSS; BCNSS=AN
		47,86	2,66	43,87	5,72	31,12	5,19	31,27	6,99	68,99	0,000	NBC>(BCSS=AN); NBC>BCNSS; BCNSS>(BCSS=AN)
TOMM	Ensayo 1	49,80	0,76	48,12	3,30	34,62	8,33	35,44	8,55	72,42	0,000	NBC>(BCSS=AN); NBC>BCNSS; BCNSS>(BCSS=AN)
	Ensayo 2	58,02	20,85	73,66	5,66	314,63	211,47	313,09	259,10	53,80	0,000	(NBC=BCNSS)<(BCSS=AN)
Test de la b	E-score	0,20	0,61	1,00	2,55	11,75	9,95	11,75	18,10	43,37	0,000	(NBC=BCNSS)<(BCSS=AN)
	Errores c	0,16	0,46	0,25	0,57	9,62	14,38	9,59	13,13	39,68	0,000	(NBC=BCNSS)<(BCSS=AN)

Prueba	Variable	Grupo NBC		Grupo BCNSS		Grupo BCSS		Grupo AN		X ² (Kruskall- Wallis)	p	Análisis a posteriori (Mann- Whitney)
		Media	Dt	Media	Dt	Media	Dt	Media	Dt			
Test de la b	Errores o	21,13	17,03	24,81	22,31	44,00	31,82	60,72	42,39	24,83	0,000	NBC<(BCSS=AN); NBC=BCNSS; BCNSS=BCSS; BCNSS<AN (NBC=BCNSS)<BCSS; NBC=AN; BCNSS=AN; BCSS>AN
	Tiempo	33,29	13,56	37,14	13,56	60,30	39,09	38,92	15,64	9,87	0,020	
Dot Counting	Combo	10,54	2,72	11,60	6,16	18,98	9,55	18,56	7,45	35,55	0,000	(NBC=BCNSS)<(BCSS=AN)
	Tiempo agrupados	2,74	1,22	3,83	2,22	7,39	3,80	6,40	3,10	41,14	0,000	(NBC=BCNSS)<(BCSS=AN)
15-Ítems de Rey	Errores	2,28	1,71	1,75	1,60	3,50	2,06	3,86	2,89	10,33	0,016	BCNSS<(BCSS=AN); NBC=BCNSS; NBC=BCSS; NBC<AN
	Ratio	2,25	0,91	1,61	0,40	1,17	0,26	1,41	0,52	27,42	0,000	NBC<(BCNSS, BCSS, AN); BCNSS<(BCSS=AN) NBC=BCNSS=AN;
	Ítems totales	12,11	3,94	11,54	3,50	8,44	4,39	12,43	2,89	8,08	0,044	NBC>BCSS; BCNSS=BCSS; BCSS<AN

Nota: AN= análogos; BCNSS= Búsqueda de compensación no sospechoso de simulación; BCSS= Búsqueda de compensación sospechoso de simulación; Combo= puntuación combinada; Dt= desviación típica; Errores c= errores comisión; Errores o= errores omisión; NBC= No Búsqueda de Compensación; TOMM= Test of Memory Malingering; TR= tiempo de reacción; VSVT= Victoria Symptom Validity Test.

Por último, el Test de los 15-Ítems de Rey también encontró diferencias estadísticamente significativas entre los grupos para el número total de elementos recordados [$\chi^2(2)= 8,08; p<0,044$] (ver Tabla 25).

Debido a la complejidad y al número de comparaciones realizadas a posteriori, se procedió a agrupar los resultados de la Tabla 15 atendiendo a los patrones de significación clínica. En primer lugar se comprobó en cuáles de las VD el rendimiento del grupo NBC era igual al del grupo BC-NSS. Los resultados mostraron que la condición se cumplía para las siguientes variables: ítems fáciles, ítems difíciles, ítems totales, TR ítems fáciles, TR ítems difíciles y TR ítems totales del VSVT; e-score, errores d, errores c, errores o y tiempo del Test de la b; puntuación combinada y tiempo para los ítems agrupados del Dot Counting; número de elementos recordados para el Test de los 15-Ítems de Rey. En total, 14 de las 18 variables cumplieron el primer patrón propuesto, lo que supone un 78% (ver Tabla 26).

En segundo lugar, se comprobó en cuáles de las VD el rendimiento del grupo BC-SS era igual al del grupo AN. Los resultados mostraron que esta condición se cumplía para las variables: ítems fáciles, ítems difíciles, ítems totales, TR ítems fáciles, TR ítems difíciles y TR ítems totales del VSVT; e-score, errores tipo d, errores de comisión y errores de omisión del Test de la b; puntuación combinada, tiempo para los ítems agrupados, número de errores y ratio para el Dot Counting, lo cual supone un que 16 de las 18 variables posibles, o el 89% de las variables, cumplió el segundo patrón esperado (ver Tabla 26).

En tercer lugar, se procedió a comprobar cuáles de las VD eran capaces de diferenciar entre el grupo BC-NSS y el grupo BC-SS. Los resultados mostraron que las variables que eran capaces de diferenciar al grupo sospechoso de simulación eran las siguientes: ítems difíciles, ítems totales, TR ítems difíciles y TR ítems totales en el

Tabla 26. Patrones esperados en los análisis a posteriori para los “grupos conocidos” en las pruebas específicas de simulación

Patrón esperado	Variables que cumplen el patrón esperado	% de variables que cumplen el patrón esperado
NBC=BCNSS	VSVT: ítems fáciles, ítems difíciles, ítems totales, TR ítems fáciles, TR ítems difíciles, TR ítems totales Test de la b: e-score, errores d, errores comisión, errores omisión, tiempo. Dot Counting: puntuación combinada, tiempo para los ítems agrupados. Test de los 15-Ítems de Rey: número de elementos recordados.	78% (14/18)
BCSS=AN	VSVT: ítems fáciles, ítems difíciles, ítems totales, TR ítems fáciles, TR ítems difíciles, TR ítems totales Test de la b: e-score, errores d, errores comisión, errores omisión. Dot Counting: puntuación combinada, tiempo para los ítems agrupados, número de errores, ratio.	89% (16/18)
BCNSS≠BCSS	VSVT: ítems difíciles, ítems totales, TR ítems difíciles, TR ítems totales TOMM: ensayo 1, ensayo 2. Test de la b: e-score, errores d, errores comisión, tiempo. Dot Counting: puntuación combinada, tiempo para los ítems agrupados, número de errores, ratio.	78% (14/18)
(NBC=BCNSS) ≠ (BCSS=AN)	VSVT: ítems difíciles, ítems totales. Test de la b: e-score, errores d, errores comisión. Dot Counting: puntuación combinada, tiempo para los ítems agrupados.	39% (7/18)

Nota: AN= análogos; BCNSS= búsqueda de compensación no sospechoso de simulación; BCSS= búsqueda de compensación sospechoso de simulación; NBC= no búsqueda de compensación; TR= tiempo de reacción; VSVT= Victoria Symptom Validity Test.

VSVT; ensayo 1 y ensayo 2 del TOMM; e-score, errores tipo d, errores de comisión y tiempo en el Test de la b; puntuación combinada, tiempo para los ítems agrupados, número de errores y ratio en el Dot Counting. Por tanto, 14 de las 18 variables (o lo que es lo mismo, el 78% de las variables), fueron capaces de diferenciar entre los sujetos que buscan compensación a aquellos que son sospechosos de simulación (ver Tabla 26).

Por último, se comprobó cuáles de las VD cumplían todos los patrones esperados, esto es, qué variables identificaban al grupo NBC igual al grupo BC-NSS, pero estadísticamente diferentes de los grupos BC-SS y AN, iguales entre sí. Los resultados mostraron que para las variables: ítems difíciles e ítems totales del VSVT, e-score, errores tipo d y errores de comisión del Test de la b, y puntuación combinada y tiempo para los ítems agrupados en el Dot Counting se cumplían todos los supuestos. Esto supone un 39% del total de las variables (7/18) (ver Tabla 26).

Estudio de los tamaños del efecto:

Para realizar un análisis más preciso de la relevancia estadística de los hallazgos se calcularon los tamaños del efecto de los análisis anteriores empleando el estadístico Delta de Cohen. En el estudio de los grupos NBC y BCNSS, la mejor variable será aquella que mejor diferencie entre los grupos, y por tanto la que lleve asociado un Delta de Cohen mayor. Los resultados mostraron que esta variable era el e-score del Test de la b ($\delta= 1,17$), y la variable que menos diferencias halló entre estos grupos fue el número de ítems totales acertados en el VSVT ($\delta= 0,01$) (ver Tabla 27).

En el estudio del Delta de Cohen para diferenciar los grupos NBC y BCSS, la variable que más diferencie será la que obtenga los mejores resultados. Dicha variable fue el Ensayo 1 del TOMM ($\delta= 4,26$). La variable que menos diferencias encontró entre estos grupos, y que por tanto obtuvo los peores resultados, fue el número de errores del Dot Counting ($\delta= 0,64$) (ver Tabla 27).

Entre los grupos NBC y AN, la variable que más diferenciaba fue, de nuevo, el Ensayo 1 del TOMM ($\delta= 3,43$), y la que menos el número de elementos recordados en el Test de los 15-Ítems de Rey ($\delta= 0,09$), siendo el resultado más adecuado el mayor delta de Cohen (ver Tabla 27).

Tabla 27. Delta de Cohen de las variables de los tests específicos de simulación para los grupos conocidos

Prueba	Variabl e	NBC vs. BCNSS	NBC vs. BCSS	NBC vs. AN	BCNSS vs. BCSS	BCNSS vs. AN	BCSS vs. AN
VSVT	Ítems fáciles	0,211	1,399	1,398	1,308	1,323	0,155
	Ítems difíciles	0,073	2,078	3,108	1,958	2,846	0,210
	Ítems totales	0,017	2,160	2,962	2,274	3,148	0,223
	TR	0,169	1,194	0,539	1,177	0,477	0,317
	ítems fáciles	0,368	1,208	0,554	0,696	0,160	0,523
	TR						
	ítems difíciles						
	TR	0,315	1,299	0,468	0,867	0,296	0,192
ítems total							
TOMM	Ensayo 1	0,952	4,264	3,438	2,337	1,982	0,024
	Ensayo 2	0,827	3,339	3,084	2,321	2,140	0,097
Test de la b	E-score	1,179	2,209	1,822	2,219	1,808	0,006
	Errores d	0,506	2,187	1,234	1,720	1,041	0,000
	Errores c	0,174	1,274	1,387	1,253	1,363	0,002
	Errores o	0,187	0,936	1,332	0,709	1,110	0,450
Dot Counting	Tiempo	0,283	1,026	0,385	0,879	0,121	0,781
	Combo	0,238	1,375	1,577	0,939	1,022	0,049
	Tiempo agrupados	0,633	1,852	1,694	1,182	0,966	0,286
	Errores	0,320	0,647	0,686	0,956	0,939	0,145
	Ratio	0,977	1,846	1,174	1,333	0,434	0,615
15-Ítems de Rey	Ítems totales	0,153	0,881	0,093	0,785	0,278	1,096

Nota: AN= análogos; BCNSS= búsqueda de compensación no sospechoso de simulación; BCSS= búsqueda de compensación sospechoso de simulación; c= comisión; Combo= puntuación combinada; NBC= no búsqueda de compensación; o= omisión; TOMM= Test of Memory Malinger; TR=tiempo de reacción; VSVT= Victoria Symptom Validity Test.

Para diferenciar entre los grupos BCNSS y BCSS, la mejor variable será aquella que mejor diferencie entre los grupos. El número de ítems totales recordados en el VSVT fue la variable que obtuvo mejores resultados ($\delta=2,27$), mientras el tiempo de reacción para los ítems difíciles del mismo test fue la variable que menos diferenció entre estos dos grupos ($\delta= 0,69$) (ver Tabla 27).

En cuanto a la distinción entre los grupos BCNSS y AN, la variable que más diferenciaba era el número de ítems totales del VSVT ($\delta= 3,14$), y la que menos el tiempo de reacción para los ítems difíciles, también del VSVT ($\delta= 0,16$), siendo más adecuado un mayor delta de Cohen (ver Tabla 27).

Por último, en la diferenciación entre los grupos BCSS y AN, la mejor variable será aquella que obtenga un delta de Cohen menor. Esta variable fue el número de errores d del Test de la b ($\delta=0,00$), mientras que la variable que más diferencia entre estos dos grupos, y que por tanto peores resultados alcanza, es el número de elementos recordados en el Test de los 15-Ítems de Rey ($\delta=-1,09$) (ver Tabla 27).

4.5. Clasificación de los “grupos conocidos” (Hipótesis 5):

Una vez comprobado que los tests específicos de simulación son capaces de diferenciar entre los “grupos conocidos”, procedimos a comprobar el porcentaje de sujetos simuladores y no simuladores que las pruebas encontraban en cada uno de estos grupos. Para ello realizamos cinco análisis de contingencia cruzando la variable grupo (NBC, BCNSS, BCSS y AN) y la clasificación de simulación (simulador vs. no simulador) obtenida mediante la categorización de las principales variables de los test de simulación siguiendo los criterios ya expuestos en el apartado: “Clasificación de los grupos de prevalencia diferencial” (pág. 226).

Los resultados mostraron que existían asociaciones estadísticamente significativas entre los grupos y la clasificación de simulador vs. no simulador en la

variable ítems difíciles del VSVT [$\chi^2(3) = 48,89; p < 0,000$], clasificando correctamente como no simuladores al 95% del grupo NBC y al 100% del grupo BCNSS, y como simuladores al 84,6% del grupo AN y al 62,5% del grupo BCSS (ver Tabla 28 y Figura 7).

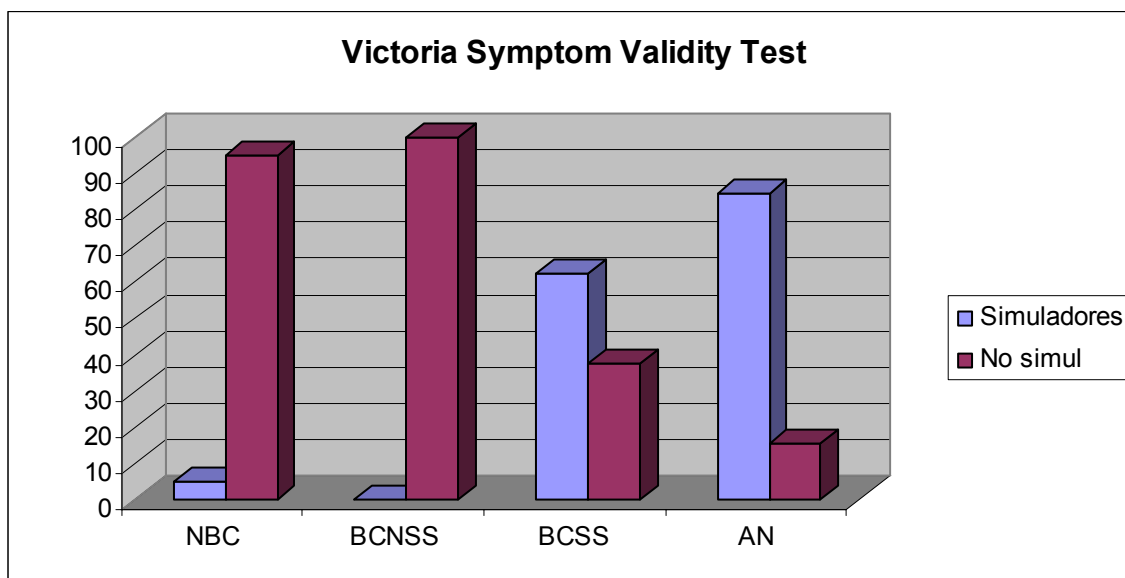


Figura 7: Porcentaje de sujetos clasificados como simuladores y no simuladores por el Victoria Symptom Validity Test en los “grupos conocidos”

También se encontraron asociaciones estadísticamente significativas entre los grupos y la clasificación de simulador en la variable Ensayo 2 del TOMM [$\chi^2(3) = 70,87; p < 0,000$], clasificando correctamente como no simuladores a la totalidad del grupo NBC y al 81,3% del grupo BCNSS, y como simuladores al 85,2% del grupo AN y al 100% del grupo BCSS (ver Tabla 28 y Figura 8).

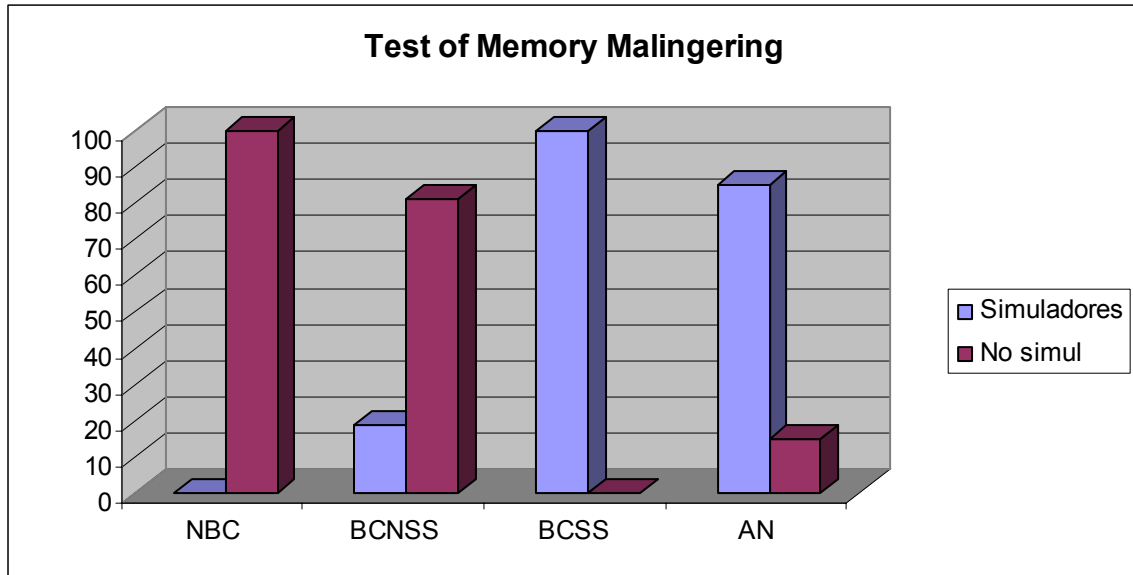


Figura 8: Porcentaje de sujetos clasificados como simuladores y no simuladores por el Test of Memory Malingering en los “grupos conocidos”

En lo que respecta a la variable e-score del Test de la b, nuevamente se encontraron asociaciones estadísticamente significativas entre los grupos y la clasificación de simulador [$\chi^2(3) = 50,03; p < 0,000$], clasificando correctamente como no simuladores al 90% del grupo NBC y al 81,3% del grupo BCNSS, mientras fueron correctamente identificados como simuladores el 77,8% del grupo AN y el 100% del grupo BCSS (ver Tabla 28 y Figura 9).

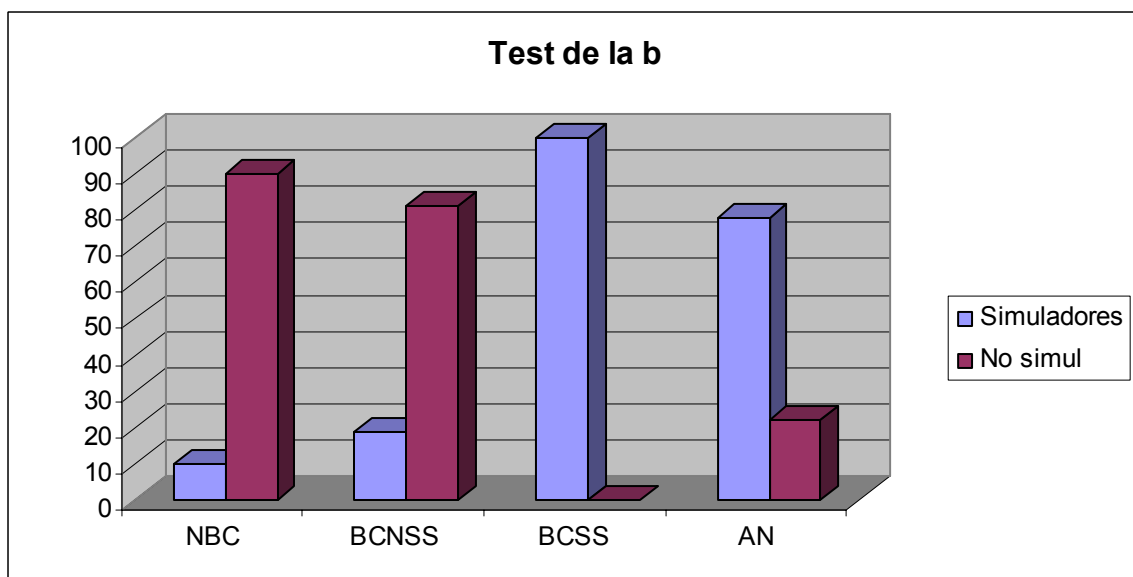


Figura 9: Porcentaje de sujetos clasificados como simuladores y no simuladores por el Test de la b en los “grupos conocidos”

En la puntuación combinada del Dot Counting también aparecieron asociaciones estadísticamente significativas entre los grupos y la clasificación de simulador [$\chi^2(3) = 22,81; p < 0,000$], clasificando como no simuladores a la totalidad del grupo NBC y al 91,7% del grupo BCNSS, mientras el 47,8% del grupo AN y el 40% del grupo BCSS fueron identificados como simuladores (ver Tabla 28 y Figura 10).

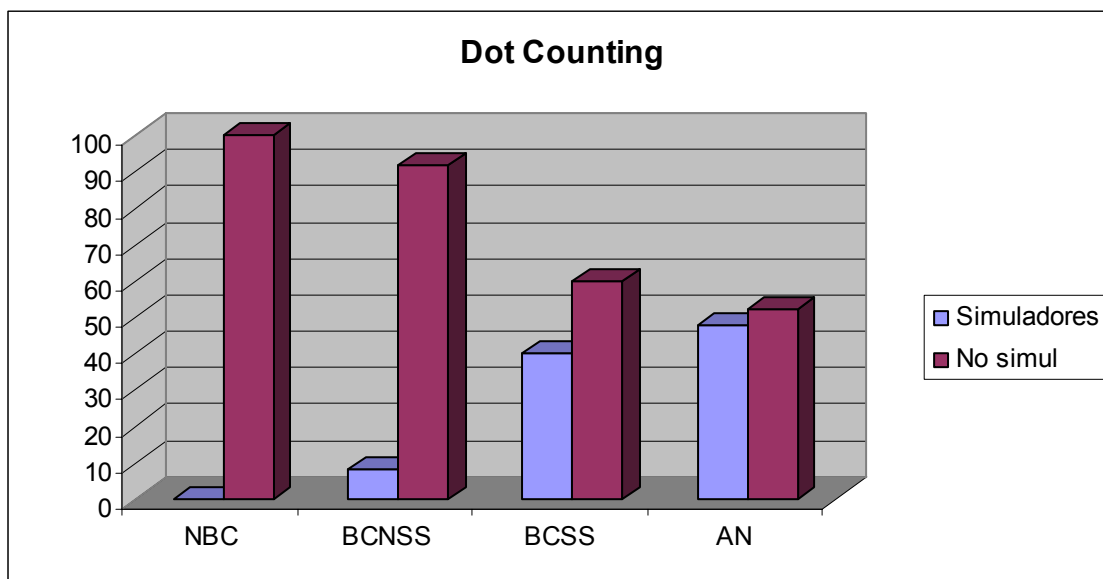


Figura 10: Porcentaje de sujetos clasificados como simuladores y no simuladores por el Dot Counting en los “grupos conocidos”

Por último, los resultados mostraron que existían asociaciones estadísticamente significativas entre los grupos y la clasificación de simulador para la variable número total de elementos recordados del Test 15-Ítems de Rey [$\chi^2(3) = 12,24; p < 0,007$], clasificando como no simuladores al 85,2% del grupo NBC y al 81,8% del grupo BCNSS, identificando correctamente como simuladores al 9,4% del grupo AN y al 55,6% del grupo BCSS (ver Tabla 28 y Figura 11).

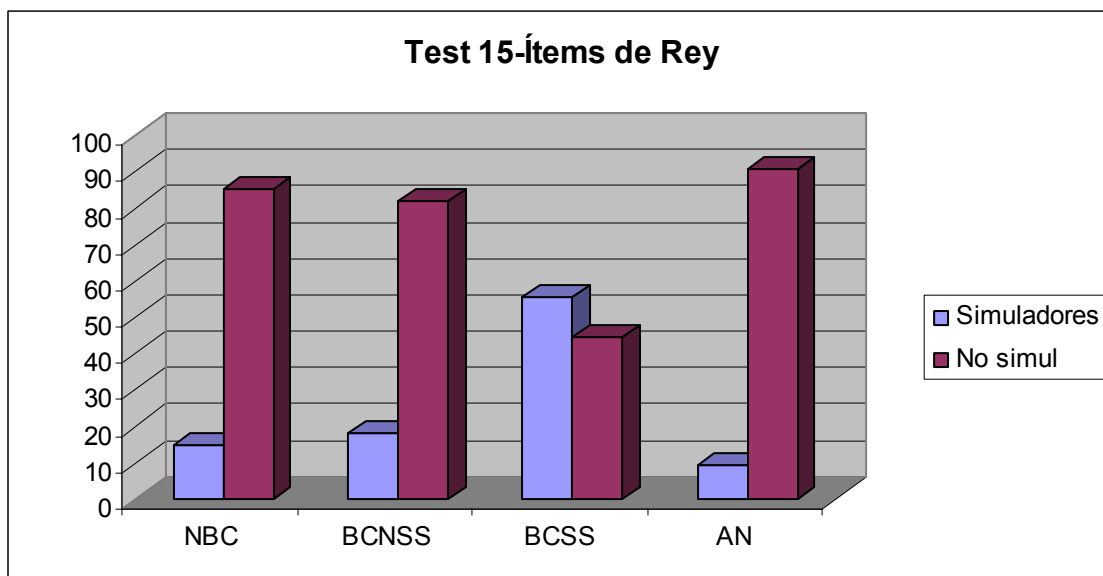


Figura 11: Porcentaje de sujetos clasificados como simuladores y no simuladores por el Test de los 15-Ítems de Rey en los “grupos conocidos”

Tabla 28. Porcentajes de sujetos correctamente clasificados por las pruebas de simulación en los grupos conocidos.

Prueba (variable)	Grupo NBC	Grupo BCNSS	Grupo BCSS	Grupo AN	X^2	p
	% No simulación	% No simulación	% Simulación	% Simulación		
VSVT (ítems difíciles)	95%	100%	62,5%	84,6%	48,89	0,000
TOMM (Ensayo 2)	100%	81,3%	100%	85,2%	70,87	0,000
Test de la b (e-score)	90%	81,3%	100%	77,8%	50,03	0,000
Dot	100%	91,7%	40%	47,8%	22,81	0,000
Counting (Combo)						
15-Ítems (total)	85,2%	81,8%	55,6%	9,4%	12,24	0,007

Nota: NBC= no búsqueda de compensación; BC= búsqueda de compensación; AN= análogos; VSVT= Victoria Symptom Validity Test; TOMM= Test of Memory Malingering; Combo= puntuación combinada.

4.6. Sensibilidad, especificidad y puntos de corte (Hipótesis 6):

Una vez comprobado que los tests específicos de simulación son capaces de diferenciar a los sujetos sospechosos de simulación, se procedió a obtener los posibles puntos de corte para las distintas variables con sus respectivas sensibilidades y especificidades. Dicha obtención de los puntos de corte se realizó sólo en aquellas

variables para las que se encontraron diferencias estadísticamente significativas entre los grupos en el objetivo anterior.

Debido al problema que suponen los falsos positivos en el estudio de la detección de la simulación y siguiendo las recomendaciones de la literatura (Greve y Bianchini, 2004) se decidió utilizar como criterio de estudio de los puntos de corte una especificidad superior al 90%.

Así, se procedió a estudiar los puntos de corte que diferenciaban entre pacientes con TCE-L no simuladores (tanto no buscadores de compensación como buscadores de compensación que no obtuvieron puntuaciones indicativas de simulación en las pruebas específicas) y el grupo sospechoso de simulación. Para ello se realizaron 18 análisis de curvas ROC: 6 para el VSVT, 2 para el TOMM, 5 para el Test de la b, 4 para el Dot Counting test y una para el Test de los 15-Ítems de Rey.

En el VSVT los resultados mostraron una sensibilidad moderada-alta asociada a un punto de corte ≤ 16 para los ítems difíciles y ≤ 38 para el número de ítems totales. La variable ítems fáciles sólo alcanzó una sensibilidad del 50% con un punto de corte ≤ 20 , y las variables basadas en el tiempo de reacción obtuvieron unas sensibilidades inaceptablemente bajas (ver Tabla 29).

En cuanto al TOMM, el punto de corte ≤ 37 para el ensayo 1 y ≤ 44 para el ensayo 2 se encontraron asociados a una sensibilidad del 100% (ver Tabla 29).

La principal variable del Test de la b, el e-score, también obtuvo una sensibilidad perfecta con un punto de corte $> 95,2$, igual que la variable errores tipo d con un punto de corte > 1 . El resto de las variables (errores de comisión, errores de omisión y tiempo) obtuvieron sensibilidades pobres (ver Tabla 29).

Tabla 29. Sensibilidad y especificidad asociada a los puntos de corte de las distintas variables para diferenciar entre pacientes y sospechosos de simulación.

Prueba	Variable	Punto de corte	Sensibilidad	IC 95%	Especificidad	IC 95%	Área ROC
VSVT	Ítems fáciles	<=20	50,0	16,0 - 84,0	100,0	87,5 - 100,0	0,763
	Ítems difíciles	<=16	62,5	24,7 - 91,0	92,9	76,5 - 98,9	0,937
	Ítems totales	<=38	75,0	35,0 - 96,1	100,0	87,5 - 100,0	0,944
	TR ítems fáciles	>4,47	25,0	3,9 - 65,0	96,2	80,3 - 99,4	0,851
	TR ítems difíciles	>10,13	25,0	3,9 - 65,0	100,0	86,7 - 100,0	0,817
	TR ítems total	>7,3	25,0	3,9 - 65,0	100,0	86,7 - 100,0	0,841
	TOMM	Ensayo 1	<=37	100,0	69,0 - 100,0	97,7	87,9 - 99,6
Ensayo 2		<=44	100,0	69,0 - 100,0	97,7	87,9 - 99,6	0,989
Test de la b	E-score	>95,2	100,0	69,0 - 100,0	97,7	87,9 - 99,6	0,991
	Errores d	>1	100,0	69,0 - 100,0	95,5	84,5 - 99,3	0,993
	Errores c	>1	30,0	7,0 - 65,2	95,5	84,5 - 99,3	0,745
	Errores o	>57	40,0	12,4 - 73,6	95,5	84,5 - 99,3	0,767
Dot Counting	Tiempo	>47,2	50,0	18,9 - 81,1	93,2	81,3 - 98,5	0,845
	Combo	>15,16	55,6	21,4 - 86,0	97,5	86,8 - 99,6	0,799
	Tiempo agrupado	>5,12	77,8	40,1 - 96,5	92,5	79,6 - 98,3	0,912
	Errores	>4	33,3	7,9 - 69,9	95,0	83,0 - 99,2	0,713
	Ratio	<=1,18	55,6	21,4 - 86,0	95,0	83,0 - 99,2	0,865
15-Ítems de Rey	Ítems totales	<=5	22,2	3,5 - 59,9	92,1	78,6 - 98,2	0,744

Nota: Área ROC= área bajo la curva ROC; Combo= puntuación combinada; Errores c= errores de comisión; Errores o= errores de omisión; IC= intervalo de confianza; TOMM= Test of Memory Malinger; TR= tiempo de reacción; VSVT= Victoria Symptom Validity Test.

El Dot Counting alcanzó una elevada sensibilidad con un punto de corte $>5,12$ para la variable tiempo para contar los ítems agrupados, y sensibilidades moderadas para las variables puntuación combinada y ratio, con unos puntos de corte $>15,16$ y $\leq 1,18$ respectivamente. La variable número de errores obtuvo una sensibilidad muy baja (ver Tabla 29).

Por último, el Test de los 15-Ítems de Rey obtuvo una sensibilidad muy pobre con un punto de corte ≤ 5 (ver Tabla 29).

Además, para todas las variables el área bajo la curva ROC fue muy elevada, correspondiendo la más baja al Test de los 15-Ítems de Rey, que no obstante se situó en 0,744.

En segundo lugar, con la finalidad de dotar a nuestro hallazgos de una mayor utilidad clínica, se procedió a obtener los valores predictivos positivos (VP+) y los valores predictivos negativos (VP-) asociados a distintas prevalencias de base (“*base rate*”) de simulación para las principales variables que pueden emplearse como índices de simulación en cada test específico de simulación empleado en este estudio.

Para ello, y siguiendo las recomendaciones de Greve y Bianchini (2004) se han seleccionado distintas prevalencias de base que cubren el espectro de porcentajes de simulación más frecuentemente referidos. De esta forma, se seleccionó 15% como criterio de baja prevalencia (Trueblood y Schmidt, 1993; Rogers y cols., 1994), 30% como un criterio de prevalencia media (Binder, 1993; Constantinou y cols., 2005) y 45% como un criterio de una elevada prevalencia de simulación (Gervais y cols., 2004; Larrabee, 2003^a; Mittenberg, Patton y cols., 2002; Youngjohn y cols., 1995), de modo similar a los análisis realizados por Boone, Lu y Herzberg (2002) en el manual del Test de la b. Para cada una de estas prevalencias se proponen varios puntos de corte, de

modo que el neuropsicólogo pueda decidir entre mayor o menor certeza a la hora de emitir un juicio clínico (ver Anexo 1).

4.7. Estabilidad de los hallazgos con las curvas ROC

Debido a que el número de sospechosos de simulación (grupo BCSS) era reducido ($n=10$), se procedió a realizar un nuevo análisis para estudiar la estabilidad de los hallazgos encontrados con las curvas ROC en un grupo de mayor tamaño como es el grupo de análogos ($n=54$). Para este nuevo análisis se seleccionaron aquellas variables que había demostrado en los análisis anteriores que eran estadísticamente iguales entre el grupo de BCSS y el grupo de análogos (AN) y que, al mismo tiempo, eran diferentes a los grupo de NBC y BCNSS. De esta forma, se podría comprobar en un grupo de mayor tamaño (AN), que se comportaba clínicamente igual que el grupo de simuladores, si los puntos de corte, sensibilidad, especificidad y área bajo las curvas ROC encontrados para dicho grupo eran estables.

Los resultados mostraron que los puntos de corte, especificidad y área bajo la curva ROC eran idénticos o muy similares para los grupos BCSS y AN (ver Tabla 30). Sin embargo, han aparecido ciertas diferencias en cuanto a la sensibilidad encontrada en ambos grupos, siendo en algunas variables superior para el grupo AN (ítems difíciles y totales del VSVT, errores de comisión del Test de la b y puntuación combinada del Dot Counting) y en otras variables superior para el grupo BCSS (e-score y errores d del Test de la b, tiempo empleado para contar los ítems agrupados del Dot Counting) (ver Tabla 30).

Tabla 30. Puntos de corte, sensibilidad, especificidad y área bajo la curva ROC de los grupos BCSS y AN.

Prueba	Variable	Punto de corte		Sensibilidad		Especificidad		Área ROC	
		BCSS n=10	AN N=54	BCSS n=10	AN N=54	BCSS n=10	AN N=54	BCSS n=10	AN N=54
VSVT	Ítems difíciles	<=16	<=16	62,5	88,5	92,9	92,9	0,937	0,982
	Ítems totales	<=38	<=38	75,0	88,5	100,0	100,0	0,944	0,986
Test de la b	e-score	>95,2	>95,2	100,0	75,5	97,7	97,7	0,991	0,913
	Error d	>1	>1	100,0	58,5	95,5	95,5	0,993	0,831
	Error c	>1	>1	30,0	62,3	95,5	95,5	0,745	0,837
Dot Counting	Combo	>15,16	>14,14	55,6	69,6	97,5	90,0	0,799	0,866
	Tiempo agrupados	>5,12	>4,94	77,8	65,2	92,5	90,0	0,912	0,866

Nota: Área ROC= área bajo la curva ROC; BCSS= búsqueda de compensación sospechosos de simulación; AN= análogos; Combo= puntuación combinada; Errores c= errores de comisión; VSVT= Victoria Symptom Validity Test.

4.8. Comparación con los resultados publicados en la literatura (Hipótesis 7):

Como un apoyo más a la consistencia de nuestros datos, y en un intento de respaldar el empleo de las pruebas en una población para la que no fueron originalmente diseñadas, se procedió a comparar nuestros resultados con los referidos en los manuales de los tests para todas las variables y grupos en los que fue posible. No se pudieron realizar estas comparaciones para el Dot Counting ni para el Test de los 15-Ítems de Rey por falta de datos en la literatura. Puesto que estos resultados ya han sido expuestos para el grupo de los análogos en el Estudio 1, no los repetiremos aquí de nuevo.

Para estudiar en qué medida nuestros resultados eran similares a los de la población de origen se realizaron 11 comparaciones de las medias de las variables de los test obtenidas en nuestro estudio y las referidas en los manuales de las pruebas. Para realizar un análisis más preceiso de la relevancia estadística de los hallazgos se calcularon los tamaños de estos análisis empleando el estadístico Delta de Cohen.

Los resultados mostraron que, para el VSVT, existen diferencias estadísticamente significativas en todas las variables estudiadas para el grupo BC. En

cuanto al grupo de pacientes sin motivación para simular (grupo NBC), sólo se encontraron diferencias estadísticamente significativas en los ítems difíciles, pero no en los ítems fáciles ni en los totales (ver Tabla 31).

Mediante el análisis de los Deltas de Cohen se comprobó que, en la diferenciación entre los grupos BC y los grupos NBC españoles y norteamericanos, los resultados oscilaron entre $\delta= 0,64$ y $\delta= 0,20$ (ver Tabla 31).

Tabla 31. Comparación de las puntuaciones de nuestra muestra y las referidas en el manual para el Victoria Symptom Validity Test.

	BC muestra actual ^a Media (Dt)	BC muestra manual ^b Media (Dt)	<i>p</i>	Delta de Cohen (δ)	NBC muestra actual ^c Media (Dt)	NBC muestra manual ^d Media (Dt)	<i>p</i>	Delta de Cohen (δ)
Ítems fáciles	22,06 (2,93)	23,33 (1,97)	0,005	0,51	23,73 (0,73)	23,53 (1,19)	0,43	0,20
Ítems difíciles	16,53 (6,98)	20,17 (4,80)	0,001	0,61	21,36 (2,49)	22,63 (1,79)	0,02	0,59
Ítems totales	38,6 (9,06)	43,50 (6,08)	0,000	0,64	45,10 (2,60)	46,16 (2,59)	0,11	0,40
TR ítems fáciles	2,98 (1,93)	2,34 (1,19)	0,021	0,41	1,73 (1,53)			
TR ítems difíciles	4,71 (3,46)	3,52 (2,10)	0,016	0,48	2,6 (1,67)			

Nota: ^a $n=25$; ^b $n= 205$; ^c $n= 19$; ^d $n= 32$; BC= búsqueda de compensación; NBC= no búsqueda de compensación; Dt= desviación típica.

En cuanto a los resultados en el TOMM, las puntuaciones del grupo sospechoso de simulación de nuestra muestra y las referidas por el manual no son estadísticamente diferentes. Tampoco se encuentran diferencias en la ejecución del grupo de pacientes sin motivación para simular en el Ensayo 2, pero sí en el Ensayo 1. Los Deltas de Cohen se situaron entre $\delta= 0,72$ y $\delta= 0,16$ (ver Tabla 32).

Tabla 32. Comparación de las puntuaciones de nuestra muestra y las referidas en el manual para el Test of Memory Malinger.

	NBC muestra actual ^a Media (Dt)	NBC muestra manual ^b Media (Dt)	<i>p</i>	Delta de Cohen (δ)	BCSS muestra actual ^c Media (Dt)	BCSS muestra manual ^d Media (Dt)	<i>p</i>	Delta de Cohen (δ)
Ensayo 1	47,86 (2,66)	45,50 (5,10)	0,042	0,60	31,12 (5,19)	25,30 (10,80)	0,138	0,72
Ensayo 2	49,80 (0,76)	49,50 (1,10)	0,275	0,32	34,62 (8,33)	32,80 (13,40)	0,716	0,16

Nota: ^a $n=30$; ^b $n= 17$; ^c $n= 10$; ^d $n= 11$; NBC= no búsqueda de compensación; BCSS= búsqueda de compensación sospechosos de simulación; Dt= desviación típica.

Por último, para el Test de la b, los resultados mostraron que no existían diferencias estadísticamente significativas en ninguna de las variables estudiadas entre los grupos de pacientes que no buscan compensación de nuestra muestra y la del manual, así como tampoco aparecen diferencias entre los sospechosos de simulación de ambas poblaciones. Los Deltas de Cohen se situaron entre $\delta= 0,60$ y $\delta= 0,02$ (ver Tabla 33).

Tabla 33. Comparación de las puntuaciones de nuestra muestra y las referidas en el manual para el Test de la b.

	NBC muestra actual ^a Media (Dt)	NBC muestra manual ^b Media (Dt)	<i>p</i>	Delta de Cohen (δ)	BCSS muestra actual ^c Media (Dt)	BCSS muestra manual ^d Media (Dt)	<i>P</i>	Delta de Cohen (δ)
e-score	58,02 (20,85)	58,5 (18,7)	0,934	0,02	314,63 (211,47)	612,5 (1069,6)	0,383	0,46
Err c	0,16 (0,46)	0,3 (0,6)	0,355	0,60	9,62 (14,38)	34,4 (76,6)	0,311	0,54
Err o	21,13 (17,03)	20,6 (12,2)	0,905	0,03	44,50 (31,82)	54,8 (50,3)	0,528	0,25
Tiempo total	499,35 (203,4)	530,4 (206,4)	0,601	0,15	904,50 (586,35)	992,0 (692,9)	0,701	0,13

Nota: ^a $n=30$; ^b $n= 20$; ^c $n= 10$; ^d $n= 91$; NBC= no búsqueda de compensación; BCSS= búsqueda de compensación sospechosos de simulación; Dt= desviación típica.

5. Discusión:

En el Estudio 2 intentamos poner de manifiesto la utilidad de diversas pruebas específicas de simulación para ser aplicadas en población española. Los resultados han mostrado que las pruebas de simulación no se ven afectadas por variables demográficas como la edad, la escolaridad o el género, ni tampoco por el tiempo transcurrido desde la ocurrencia del traumatismo hasta la evaluación. Una excepción fue el Test 15-Ítems de Rey, en el que se encontró una correlación con la escolaridad. Además, estas pruebas encuentran diferencias estadísticamente significativas entre los grupos de prevalencia diferencial (NBC, BC y AN), con excepción del Test 15-Ítems de Rey. De modo general, podemos decir que las pruebas específicas no presentan problemas a la hora de

diferenciar pacientes sin incentivos para simular (grupo NBC) y sujetos que sabemos a ciencia cierta que están simulando (grupo AN). En cuanto al estudio de los “grupos conocidos” (NBC, BCNSS, BCSS y AN), los resultados muestran que todas las variables estudiadas permiten diferenciar entre estos grupos, señalando los resultados del tamaño del efecto y la clasificación de simulación que las pruebas investigadas permiten identificar a un grupo sospechoso de simulación (grupo BCSS). En este estudio también se han diferenciado algunas variables que, manteniendo una elevada especificidad (superior al 90%), ofrecen una sensibilidad adecuada, y que además han demostrado su consistencia. Por último, no se han encontrado diferencias entre los resultados de este estudio y los publicados para las poblaciones originales. En conjunto, este estudio parece apoyar la utilidad de las pruebas específicas de simulación en nuestro contexto.

En primer lugar, se ha estudiado el posible efecto que ciertas variables como la edad, la escolaridad, el género o el tiempo transcurrido desde el evento traumático pudieran tener sobre los tests específicos de simulación. Esto es especialmente importante ya que, en gran medida, la habilidad de las pruebas para detectar simulación depende de su insensibilidad a otras variables como las ya mencionadas (Tombaugh, 1996). Desde el punto de vista teórico, si estas pruebas están midiendo esfuerzo, y no habilidad, no parece existir ninguna razón para que los tests de simulación varíen sistemáticamente con la edad, la educación, el género o variables relacionadas con el daño neurológico (Boone, Lu y Herzberg, 2002). Los resultados mostraron que, como era esperable, las variables estudiadas no afectan los resultados del VSVT, TOMM, Test de la b y Dot Counting, lo que coincide con la literatura, en la que, o no se encuentra correlación entre ejecución en estas pruebas de simulación y variables demográficas o neurológicas (Constantinou y McCaffrey, 2003; Grote y cols., 2000; Haber y

Fichtenberg, 2006; Macciocchi y cols., 2006; Tombaugh, 1996; Rees y cols., 2001; Teichner y Wagner, 2004), o la significación clínica de estas relaciones es despreciable (Boone, Lu y Herzberg, 2002; Boone, Lu, Back y cols., 2002; Slick y cols., 1997). Estos resultados no se cumplen, sin embargo, para el Test 15-Ítems de Rey, en el que sí se encuentran correlaciones entre su ejecución y la escolaridad de los pacientes (resultados también encontrados por Arnett y cols., 1995 y Schretlen y cols., 1991), lo que constituye el primer problema de esta prueba.

Diseño de grupos de prevalencia diferencial

En segundo lugar, se procedió a estudiar cómo se comportan las pruebas específicas de simulación mediante un diseño de prevalencia diferencial.

Los resultados mostraron que todas las variables estudiadas de los diferentes tests, excepto la variable tiempo del Test de la b y la variable número de elementos recordados del Test 15-Ítems de Rey, permitían diferenciar entre los grupos de pacientes con TCE que buscan compensación, los pacientes que no buscan compensación, y el grupo de análogos. En los análisis a posteriori se encuentran, de modo sistemático, diferencias estadísticamente significativas en la ejecución del grupo NBC y el grupo AN, de modo que los pacientes y los sujetos que intentan fingir un daño cognitivo realizan este tipo de pruebas de modo muy diferente, correspondiendo la mejor ejecución al grupo NBC, que obtiene puntuaciones casi perfectas en estas pruebas. Así, parece que, por un lado, los tests específicos de simulación son insensibles al daño cognitivo, y que no están midiendo la disfunción o la habilidad de los pacientes, mientras que, por otro lado, nos permiten diferenciar a estos de los individuos que sabemos a ciencia cierta que están simulando (grupo AN), lo que parece estar apoyando la validez de constructo de estas pruebas.

Sin embargo, las comparaciones a posteriori no son tan claras cuando se trata de diferenciar al grupo BC, ya que, por un lado, en aproximadamente la mitad de las variables este grupo se comporta de modo significativamente diferente al grupo AN, e igual que el grupo NBC, mientras que en aproximadamente la otra mitad se comporta de modo significativamente diferente al grupo NBC e igual que el grupo AN.

Las diferencias encontradas entre los grupos NBC y BC irían en la línea de lo esperado, indicando que las puntuaciones obtenidas en los tests de simulación se asocian a la presencia de incentivos económicos por el daño, y no al daño cerebral en sí. Algunas de las variables aquí estudiadas parecen, por tanto, sensibles a la hora de detectar a individuos con un alto riesgo de presentar un esfuerzo insuficiente. Así, la presencia de incentivos económicos parece una fuerte variable moderadora de la respuesta a los tests específicos de simulación después de un TCE-L, como tantas veces se ha señalado en la literatura (Binder y Willis, 1991; Binder, 1993; Binder y Rohling, 1996).

No obstante, como acabamos de mencionar previamente, en el mismo número de variables se aprecian diferencias estadísticamente significativas entre los grupos BC y AN. Una explicación posible para el hecho de que los análogos se diferencien significativamente del grupo BC (hallazgo que también aparece en el trabajo de Slick, Hopp, Strauss y Spellacy de 1996 con el VSVT) es que los estudios análogos carecen de validez externa. El miedo a ser detectado, la motivación para simular y la ocurrencia real de un evento traumático son difícilmente igualables en análogos (Haines y Norris, 1995), por lo que de aquí podrían venir las diferencias encontradas. Sin embargo, como se ha comentado ampliamente en el estudio 1, hemos seguido todas las recomendaciones propuestas por los autores más relevantes en la materia para aumentar el rigor de la investigación y conseguir un grupo de análogos sofisticados que nos

permita obtener una buena validez externa y extraer conclusiones de sus resultados (Rogers, 1997^b), por lo que esta conclusión nos parece prematura.

Otra posible explicación a la falta de diferencias entre el grupo BC y el grupo AN reside en el problema fundamental que presenta el seleccionar el grupo de elevada prevalencia basándonos en que tienen un incentivo financiero para presentar un esfuerzo sospechoso durante la evaluación: que la presencia de una motivación para simular no implica automáticamente la producción de un esfuerzo sospechoso (Haber y Fichtenberg, 2006) o, dicho de otro modo, que entre los buscadores de compensación existen simuladores y no simuladores (Bianchini, Mathias, Greve, Houston y Crouch, 2001). Es totalmente plausible que algunos miembros de este grupo estén esforzándose todo lo posible en la ejecución de las pruebas, lo que explicaría las diferencias encontradas entre los grupos BC y AN, al existir una heterogeneidad de la motivación real de los miembros del grupo BC frente a la homogeneidad del grupo AN.

Con independencia de que, como acabamos de mencionar, en algunos casos se encuentran diferencias estadísticamente significativas y en otras ocasiones no, el patrón que reflejan las puntuaciones medias de los grupos es siempre la misma. Esto puede apreciarse analizando las principales medidas propuestas en la literatura como indicativas de simulación para cada test (ítems difíciles en el VSVT; ensayo 2 en el TOMM, e-score en el Test de la b y puntuación combinada en el Dot Counting), ya que la puntuación del grupo NBC se aleja del punto de corte indicativo de simulación en todas estas variables, mientras la media del grupo AN cae definitivamente en el rango indicativo de simulación. Las puntuaciones encontradas en el grupo BC en estas variables son intermedias entre ellas, de modo que, o están ligeramente por debajo del punto de corte indicativo de simulación (TOMM y Test de la b) o próximas a él (VSVT y Dot Counting).

Esta tendencia también se ve reflejada en los resultados del análisis de los tamaños del efecto, mediante los cuales se puede apreciar que las diferencias entre los grupos NBC y AN son elevadas y notables, mientras que las diferencias entre los grupos NBC y BC son también importantes, pero inferiores, y siendo bajas las diferencias encontradas entre los grupos BC y AN.

La constatación más evidente de que el grupo BC está constituido tanto por simuladores como por no simuladores la obtenemos de los porcentajes observados en la clasificación de las principales variables de los distintos tests de simulación. Mediante la categorización de simulación vs. no simulación tomando como puntos de corte los indicados en los manuales, comprobamos cómo el grupo NBC es un grupo de no simuladores, el grupo AN es un grupo de simuladores, y el grupo BC es un grupo mixto, en el que entre el 23 y el 46% de los sujetos (dependiendo de la prueba) obtienen puntuaciones indicativas de simulación.

En definitiva, si el diseño de prevalencia diferencial se ha mostrado útil (puesto que las pruebas parece que diferencian entre sujetos con y sin incentivos para simular), tienen una limitación fundamental, y es que no nos permiten cuántos ni cuáles de los sujetos están, realmente, simulando (Rogers, 1997^b). Por tanto, parece que no es adecuado denominar a un grupo sospechoso de simulación atendiendo simplemente a la existencia de incentivos económicos, sino que debemos adoptar criterios externos más estrictos, en la línea de lo propuesto por Slick y cols. (1999).

Diseño de grupos conocidos

Los diseños de grupos conocidos son fundamentales en la investigación de la simulación, ya que son los que poseen una mayor relevancia clínica (Rogers, 1997^b). La mayor evidencia sobre la utilidad de una prueba de simulación se basa en su poder para discriminar entre grupos cuyo esfuerzo ha sido establecido por otros medios (Boone, Lu

y Herzberg, 2002), como hemos realizado en este estudio para la composición del grupo sospechoso de simulación (BCSS). Además, la inclusión del grupo de pacientes buscadores de compensación no simuladores (BCNSS) es muy útil en estas investigaciones, ya que la ausencia de diferencias entre este grupo y el grupo NBC nos permitirá concluir que las puntuaciones por debajo de los puntos de corte indicativas de simulación no pueden ser explicadas por motivos externos distintos a la simulación como el estrés debido al litigio.

Otro de los puntos fuertes de este tipo de diseños es que, a diferencia de los diseños de prevalencia diferencial, nos permite conocer la prevalencia de base de la simulación en nuestro contexto. Si bien estos datos han de ser tomados con precaución, ya que es necesaria una mayor investigación al respecto para poder extraer conclusiones definitivas, parece que, adoptando un criterio bastante estricto (búsqueda de compensación económica más una puntuación indicativa de simulación en dos o más pruebas específicas de simulación bien estudiadas) la prevalencia de la simulación entre los casos de traumatismos craneoencefálicos leves en nuestro contexto está en torno al 41% (ya que 10 de los 24 sujetos en litigio aparecen como sospechosos de simulación). Esta es la primera vez que se exponen resultados concernientes a la posible prevalencia de la simulación en nuestro contexto, si bien los resultados parecen consistentes, ya que están en consonancia con los trabajos más recientes y más sólidos realizados sobre la prevalencia de la simulación en TCE-L con población norteamericana (39% según Mittenberg, Patton y cols., 2002; 40% según Larrabee, 2003^a). Por tanto, y a la espera de posteriores investigaciones que confirmen nuestros hallazgos, no parece descabellado adoptar una prevalencia de simulación en torno al 40% en los casos de TCE-L.

Los resultados del diseño de grupos conocidos han mostrado que todas las variables estudiadas son capaces de diferenciar de modo estadísticamente significativo

entre estos grupos. Los análisis a posteriori mostraron que, de modo general, los grupos se comportan del siguiente modo: no se encuentran diferencias estadísticamente significativas entre los grupos NBC y BCNSS (en 14 de las 18 variables estudiadas), ni entre los grupos BCSS y AN (en 16 de las 18 variables), mientras que las diferencias sí que son estadísticamente significativas cuando se comparan los grupos BCNSS y BCSS (en 14 de las 18 variables), resultados que se ven apoyados por los del tamaño del efecto. Gracias a esta comparación entre los grupos BCSS y BCNSS podemos obtener la demostración más convincente de la validez discriminativa de los tests que estamos estudiando (Bianchini, Mathias y Greve, 2001).

Para mostrar mejor lo que está sucediendo vamos a poner un ejemplo concreto. En el ensayo 2 del TOMM, en la comparación de los grupos de prevalencia diferencial, nos encontramos dos grupos claramente polarizados: el grupo NBC obtiene puntuaciones casi perfectas (49,8 de 50 posibles), mientras el grupo AN obtiene puntuaciones muy inferiores a estas (35,44) situándose claramente por debajo del punto de corte de 45 indicativo de simulación. En cuanto al grupo BC, obtiene una puntuación de 43,62, por lo que cae ligeramente por debajo del punto de corte de simulación, situándose en el medio de las puntuaciones del grupo NBC y del grupo AN. Si posteriormente estudiamos este grupo BC mediante un estudio de grupos conocidos obtenemos que, dentro de este grupo, existen en realidad dos grupos de sujetos: un grupo motivado para realizar las pruebas correctamente, y otro grupo que probablemente esté simulando. Siguiendo con el ejemplo del ensayo 2 del TOMM, mediante el diseño de grupos conocidos puede apreciarse claramente como la porción de sujetos adecuadamente motivados (grupo BCNSS) obtiene puntuaciones iguales al grupo NBC (48,12 frente a 49,80), que se encuentran muy alejadas del punto de corte indicativo de simulación, mientras que el grupo BCSS obtiene puntuaciones muy

similares a las del grupo AN (34,62 frente a 35,44), y que se encuentran claramente en el rango indicativo de simulación. Este es el patrón encontrado en la práctica totalidad de las variables estudiadas (ver Tabla 15). Este es el patrón encontrado en todas las pruebas.

Análisis por pruebas

Analizando las pruebas por separado, parece que las mejores variables para diferenciar a sujetos sospechosos de simulación se corresponden, en el caso del VSVT, con los ítems difíciles y totales, resultados apoyan lo encontrado en la investigación con sujetos buscadores de compensación y con análogos (Grote y cols., 2002; Slick y cols., 1996; Tan y cols., 2002; Strauss y cols., 2002). Esta superioridad de los ítems difíciles sobre los fáciles puede ser debida a que el empleo de estímulos más complejos incrementa la transparencia o dificultad aparente de la prueba, de modo que se incrementa su sensibilidad (Bianchini, Mathias y Greve, 2001). Los tiempos de reacción no parecen ofrecer muy buenos resultados, ya que el grupo AN obtiene puntuaciones similares a las de los pacientes por lo que, como se señala en el manual (Slick y cols., 1996) o en otras investigaciones (Macciocchi y cols., 2006), las decisiones clínicas sobre la simulación nunca deberían estar basadas en este tipo de puntuaciones.

En cuanto al TOMM, además de los buenos resultados encontrados en el ensayo 2 (esperables por lo publicado hasta el momento: Constantinou y cols., 2005; Etherton y cols., 2005; Gavett y cols., 2005; Greve y Bianchini, 2006; Powell y cols., 2004; Rees y cols., 1998), son sorprendentes los excelentes hallados para el ensayo 1, referidos también por Greve y cols. (2006), por lo que parece que esta variable se muestra prometedora y merece ser investigada más a fondo en el futuro. Hay que tener en cuenta, de todos modos, que si bien la carga de memoria de esta prueba es mínima, es muy probable que la correcta realización del ensayo 1 requiera una mayor demanda

cognitiva que el ensayo 2. En este sentido, los excelentes resultados que hemos encontrado se refieren sólo a población con TCE-L, por lo que es posible que otro tipo de pacientes más dañados obtengan puntuaciones en el ensayo 1 que se solapen con las de los simuladores. Sea como fuere, parece claro que hay una interesante vía de investigación abierta para comprobar la utilidad real de esta variable.

En cuanto al Test de la b y al Dot Counting, las variables que se han mostrado superiores coinciden con las publicadas en la literatura (e-score: Boone, Lu y Herzberg, 2002; y puntuación combinada: Boone, Lu, Back y cols., 2002). La superioridad de estas variables es lógica si tenemos en cuenta que ambas son puntuaciones compuestas de las diferentes medidas del test, por lo que es normal que obtengan mejores resultados que el resto de las puntuaciones aisladas.

En el Test de los 15-Ítems de Rey no se han encontrado los resultados esperados, ya que tanto los análogos como los sujetos sospechosos de simulación (BCSS) realizan esta prueba como los pacientes no simuladores (BCNSS y NBC), confirmándose de nuevo los malos resultados de esta prueba. Una posible explicación a estos resultados es la selección de la variable empleada (número de elementos recordados), que si bien es la más empleada (Lezack y cols., 2004), no parece ser la más sensible. En este sentido, sería muy interesante continuar la investigación empleando diversas variables que se han propuesto en diversos trabajos y que todavía no han sido validadas (errores de comisión, propuestos por Guilmette, Hart y cols., 1994; errores disléxicos o errores gestalticos, propuestos por Lee y cols., 2000).

En cuanto a la clasificación de los sujetos como simuladores o no simuladores, de modo general podemos decir que el VSVT, TOMM y Test de la b obtienen muy buenos resultados, siendo los del Dot Counting intermedios y bastante negativos los del Test 15-Ítems de Rey. Uno de los aspectos que llaman la atención es que existes ciertas

diferencias en la exactitud de la clasificación de los AN y de los BCSS, siendo la detección superior en estos últimos. Una de las posibles explicaciones es que el grupo de AN es más difícil de detectar debido al entrenamiento que han recibido. De este modo, nuestras expectativas de conseguir un grupo de análogos “sofisticados” se verían sobrepasadas. Lo que parece claro es que, antes de aceptar un resultado, este debe ser sometido no sólo a estudios con análogos (como suele hacerse generalmente), sino que ha de buscarse también el apoyo de investigaciones con “simuladores reales”.

Otro dato que es preciso señalar es que, especialmente en cuanto al TOMM y al Test de la b, existen ciertas discrepancias en la clasificación como no simuladores de los pacientes NBC y los BCNSS, cometiéndose un porcentaje mayor de errores entre estos últimos. Este hallazgo podría deberse a que en el grupo BCNSS existen, realmente, sujetos simuladores. Consideramos que esto es bastante probable ya que, a la hora de componer el grupo sospechoso de simulación, algunos sujetos obtenían puntuaciones indicativas de simulación en dos pruebas, siendo una de ellas el TOMM o el Test de la b. Estos casos no eran incluidos en el grupo BCSS a la hora de estudiar estas pruebas, sino en el grupo BCNSS, por lo que la capacidad de ambos tests para clasificar a los pacientes en litigio como no simuladores podría estar infravalorada en nuestro estudio.

Es curioso que la detección de simulación es superior en el TOMM que en el VSVT, teniendo en cuenta que ambas pruebas son de elección forzosa y siguen principios muy similares. Una explicación posible a estas diferencias está relacionada con los estímulos empleados por las pruebas (números en el VSVT y dibujos en el TOMM), que podrían estar vinculados a sensibilidades distintas (Gervais y cols., 2004). Otra posibilidad es que las diferencias en los niveles de detección entre ambas pruebas sean debidas al tipo de administración de ambas, ya que se ha comprobado que las

pruebas administradas por ordenador (como el VSVT) son menos sensibles que las administradas en formato papel (como el TOMM) (Bolan, Foster, Schmand y Bolan, 2002). Una tercera posibilidad es que la menor detección por parte del VSVT sea debida a que el criterio seguido en esta prueba para clasificar a un sujeto como sospechoso de simulador es más estricto que el seguido en el TOMM, ya que mientras en este se emplea una puntuación inferior al 90% de las respuestas correctas, en el VSVT es necesario que el sujeto obtenga una puntuación inferior al 62% de las respuestas correctas para que se considere sospechoso de simulación.

En relación a la exigencia en los criterios hay que señalar que, de haber tomado la significación por debajo del azar, el TOMM sólo habría detectado al 2% de los análogos como simuladores, y no habría detectado a ningún paciente sospechoso de simulación como simulador. El VSVT, por su parte, hubiera identificado al 22% de los análogos y al 25% de los pacientes del grupo sospechoso de simulación. Así, mientras que el TOMM se muestra superior al VSVT en la detección de sospechosos de simulación, el VSVT es superior detectando ejecuciones por debajo del azar, y por tanto simuladores “definitivos”, lo cual también es realmente importante. Este resultado también se ha confirmado en otro trabajo (Tan y cols., 2002). Con estos resultados se está confirmando, una vez más, que el criterio de puntuación por debajo del azar para clasificar a los sujetos como simuladores es demasiado exigente e innecesario (Bender y Rogers, 2004; Gervais y cols., 2004; Guilmette, Hart y Giuliano, 1993; Guilmette, Hart y cols., 1994; Greiffenstein y cols., 1994; Hiscock, Branham y Hiscock, 1994; Holmquist y Wanlass, 2002; Martin, Bolter, Todd, Gouvier y Nicolls, 1993; Martin, Hayes y Gouvier, 1996; Rose, Hall y Szalda-Petree, 1995; Slick, Hopp, Strauss, Hunter y Pinch, 1994).

Por su parte, el Dot Counting no alcanza el 50% a la hora de detectar análogos o sospechosos de simulación, mientras el Test 15-Ítems de Rey detecta al 55,6% de los sospechosos de simulación, pero no alcanza el 10% en los análogos. Esta diferencia entre la detección de análogos y sospechosos de simulación parece apoyar nuestra hipótesis inicial de que es posible que los estudiantes conozcan la prueba. No obstante, la detección de sospechosos sigue siendo bastante baja, y además esta prueba comete entre un 15 y un 20% de falsos positivos. Este elevado porcentaje de pacientes clasificados como simuladores podría ser debido a que mediante las instrucciones de esta prueba se está instando a los sujetos a que la realicen mal. Hay que tener en cuenta que el 15-Ítems de Rey se administraba en la última parte de la evaluación, por lo que ya había sido establecido un adecuado *rappor*t con los pacientes. De este modo, podría ser que los pacientes se vean sugestionados por las instrucciones relacionadas con la dificultad de la prueba, de modo que estarían realizándola mal como respuesta a nuestra incitación, lo cual parece éticamente bastante cuestionable (Rogers y Cavanaugh, 1983). A pesar de este hipotético problema, los resultados encontrados en este estudio apoyan los referidos por otras investigaciones en cuanto a la baja sensibilidad del 15-Ítems de Rey (Guilmette, Hart y cols., 1994; Merten y cols., 2005; Millis y Kler, 1995; Schretlen y cols., 1991), y también en cuanto a los problemas en su especificidad (Guilmette, Hart y cols., 1994; Schretlen y cols., 1991).

Teniendo en cuenta que la evidencia convergente de análogos y de grupos conocidos es el proceso recomendado para aceptar un procedimiento de detección de simulación como válido (Rogers, 1997^b), podemos concluir que las pruebas VSVT, TOMM y Test de la b cumplen los criterios necesarios para ser aceptados y empleados en nuestro contexto.

Sensibilidad, especificidad y valores predictivos

En cuarto lugar, en este Estudio 2 procedimos a averiguar los puntos de corte para las diferentes variables estudiadas tomando como criterio la obtención de una especificidad mínima del 90%. Hemos optado por este criterio por las graves consecuencias que un falso diagnóstico de simulación tienen para los sujetos, tanto financieras como ocupacionales y personales (emocionales, interpersonales...). Así, hemos decidido basarnos en la especificidad para el establecimiento de los puntos de corte, a pesar de las sabidas consecuencias que esto tiene en el descenso de la sensibilidad (Greve y Bianchini, 2004). No hemos tomado una especificidad del 100% porque consideramos que, puesto que estamos proponiendo numerosas medidas indicativas de simulación, es prácticamente imposible que un paciente que exhiba una motivación adecuada obtenga puntuaciones indicativas de simulación en dos o más variables que poseen una especificidad superior al 90%.

En este sentido, son destacables los excelentes resultados del TOMM (tanto en el ensayo 1 como en el 2) y del Test de la b (e-score y Errores *d*), que obtienen sensibilidades muy elevadas. Los resultados también han sido muy positivos para el VSVT, especialmente para el número de ítems totales, y para el tiempo de los ítems agrupados del Dot Counting. Estos excelentes resultados se ven confirmados por las áreas bajo la curva ROC, que para todas estas variables ha sido superior a 0,9. Destacables también, pero en el sentido contrario, son los malos resultados de las variables que tienen en cuenta el tiempo de reacción en el VSVT (si bien estos resultados son esperables, como ha sido comentado anteriormente) y especialmente los resultados del Test de los 15-Ítems de Rey, cuya sensibilidad se sitúa en el 22%, inferior al 36% referido en el meta-análisis de Reznick (2005), pero superior a la encontrada en otras investigaciones (Arnett y cols., 1995; Schretlen y cols., 1991).

Además, siguiendo las recomendaciones de Greve y Bianchini (2004) en el Anexo 1 se presentan distintos puntos de corte para cada variable partiendo de una especificidad del 100% y vinculados a distintas probabilidades de base (baja=15%, media=30% y alta=45% prevalencia de simulación), de modo que establecemos un continuum de poder predictivo que parte del 100% y va descendiendo, con el objetivo de maximizar la utilidad clínica de nuestros hallazgos. De este modo, un neuropsicólogo que se enfrente a la detección de simulación podría, basándose en nuestros resultados, decidir entre los distintos puntos de corte dependiendo del nivel de confianza que quiera tener en sus hallazgos (desde la extrema seguridad que un poder predictivo positivo ofrece de que un resultado positivo está reflejando un sesgo de respuesta por parte del paciente a expensas de la capacidad de detección, hasta el grado que quiera arriesgarse a cometer algún falso positivo pero optimizando la detección de sesgos de respuesta).

Estabilidad de los hallazgos

En cuanto a la estabilidad de los resultados encontrados con las curvas ROC, los resultados han mostrado que al repetir los análisis con un grupo mayor (el grupo de análogos) los puntos de corte obtenidos, así como las especificidades y el área bajo la curva ROC eran idénticos o muy similares. Por tanto, esta ausencia de diferencias entre el grupo sospechoso de simulación y el grupo de análogos nos hace concluir que nuestros hallazgos son consistentes.

Sin embargo, sí se han encontrado diferencias referentes a la sensibilidad de las pruebas en ambos grupos, lo cual podría achacarse a problemas en cuanto a la validez externa o generabilidad del grupo de análogos (Rogers, 1997^b), debido a diferentes factores: en primer lugar, la motivación externa para simular de un modo convincente parece difícilmente comparable entre los análogos y los casos clínicos (Tombaugh, 1997). Por ejemplo, el miedo a ser detectado o la ocurrencia real de un evento

traumático son difícilmente comparables en ambos casos (Haines y Norris, 1995). Sin embargo, uno de los puntos fuertes de este estudio está relacionado con el control de diversas variables encaminadas a reducir estos problemas de validez externa, siguiendo las recomendaciones de Rogers (1997^a) para mejorar los diseños de este tipo de investigaciones. Así, hemos tenido en cuenta tanto los incentivos positivos contingentes a la simulación exitosa (mediante una puntuación extra en una asignatura si los estudiantes conseguían eludir la detección), como los incentivos negativos, que raramente son tenidos en cuenta en este tipo de estudios (Bender y Rogers, 2004). De hecho, parece que los incentivos negativos influyen más en la ejecución de los análogos que los incentivos positivos (Rogers y Cruise, 1998). Los incentivos se han tenido en cuenta tanto en el escenario propuesto a los sujetos (beneficios económicos vs. sanción criminal según el resultado del juicio) como atendiendo al éxito de los participantes (puntuación extra vs. exposición pública de los nombres de los “peores” simuladores - que, por supuesto, no fue llevada a cabo). Por otro lado, se ha evaluado el grado de implicación de los participantes en el estudio mediante una entrevista posterior, se ha proporcionado un escenario adecuado, y se han ofrecido instrucciones precisas enfatizando que la simulación de los déficits debía de ser creíble.

Teniendo en cuenta todas estas razones, nos parece poco probable pensar que las diferencias encontradas en la sensibilidad de las pruebas entre los sospechosos de simulación y los análogos sean debidas a las diferencias inherentes a los grupos, aunque no deja de ser una explicación plausible. Por otro lado, dichas diferencias intergrupales conllevarían una mayor detección (y por tanto una mayor sensibilidad) del grupo de análogos, mientras que nuestros resultados a este respecto no son claros, detectando algunas variables una mayor proporción de análogos y otras variables una mayor proporción de pacientes sospechosos de simulación. Además, las diferencias sólo son

sustanciales en el caso del Test de la b, y en sus variables marginales, no en su medida principal (el e-score), por lo que quizás en esta prueba haya que ser cauto a la hora de extrapolar los resultados obtenidos con análogos a poblaciones clínicas. En este sentido hay que recordar que en el manual de esta prueba no se han empleado análogos, ni tampoco en la investigación realizada hasta la fecha. Por tanto, a pesar de que el Test de la b ha probado su utilidad en la diferenciación de grupos conocidos, hay que tener en cuenta las recomendaciones de los autores relativas a la necesidad de apoyo convergente en la investigación con grupos conocidos y con análogos para aceptar una prueba como útil (Rogers, 1997^b). Así, parece recomendable estudiar en profundidad en trabajos futuros el comportamiento de esta prueba en análogos.

Comparación con los resultados publicados en población norteamericana

En cuarto y último lugar, en el Estudio 2 hemos procedido a comprobar las puntuaciones de nuestra muestra con las de las muestras de origen. La única prueba en la que aparecen diferencias estadísticamente significativas entre los grupos (español y norteamericano) es en el VSVT. Sin embargo, observando las medias y desviaciones típicas de las distintas medidas (ver Tabla 22) podemos comprobar que se trata de un artefacto estadístico debido, muy probablemente, a la diferencia en el tamaño de las muestras. De hecho, mediante el análisis del Delta de Cohen podemos apreciar claramente que no existen diferencias entre ambas muestras. La similitud existente en todas las medidas de todas las pruebas entre nuestra población y la población norteamericana constituye un importante apoyo a nuestros datos. Este hallazgo parece indicar que los factores culturales podrían no afectar a la ejecución de las pruebas específicas de simulación, como suele suceder en el resto de los tests neuropsicológicos (Puente y Pérez, 2000^{a,b}). En conclusión, considerando también el resto de los

resultados, podemos concluir que las pruebas específicas de simulación estudiadas son útiles para su aplicación en nuestro contexto.

En resumen, los datos obtenidos en este segundo estudio indican que las pruebas neuropsicológicas específicas para la detección de la simulación obtienen unos buenos resultados en sensibilidad y especificidad, especialmente para las principales variables dependientes del VSVT, el TOMM y el Test de la b.

CAPÍTULO 9

ESTUDIO 2: ESTUDIO DE OTROS PROCEDIMIENTOS EN LA DETECCIÓN DE SIMULACIÓN

1. Introducción:

Una vez comprobada la utilidad de las pruebas específicas para detectar sospechosos de simulación entre los pacientes, vamos a proceder a examinar otro tipo de procedimientos propuestos para la detección de la simulación. En especial, nos centraremos en algunos tests tradicionalmente empleados en la evaluación neuropsicológica, ya que diversos estudios han puesto de manifiesto que algunas de estas medidas constituyen indicadores fiables del esfuerzo insuficiente (Bernard y cols., 1996; Demakis, 1999; Iverson y Franzen, 1998; King y cols., 2002; Martens y cols., 2001; Meyers y Volbretch, 1999; Millis y cols., 1995; Millis y Putnam, 1997; Moore y Donders, 2004; Osimani y cols., 1997; Suhr, 2002; Suhr y Boyer, 1999; Suhr y Gunstad, 2000). Típicamente, los individuos simuladores tienden a ejecutar este tipo de pruebas de modo similar a como lo hacen los análogos. Sin embargo, suele ser difícil establecer puntos de corte con los que no se solapen los simuladores y los verdaderos pacientes, de modo que se consiga no etiquetar erróneamente a los pacientes honestos como simuladores. A pesar de esta gran limitación, los tests neuropsicológicos tradicionales presentan importantes ventajas, como no consumir tiempo extra en la evaluación o permitir evaluar el esfuerzo de una evaluación previa en la que no se han aplicado pruebas específicas, que hacen que su estudio sea interesante. Por otro lado, siempre nos proporcionarán una fuente de información adicional a los tests específicos que no debemos desaprovechar, teniendo en cuenta la complejidad que supone el proceso de decisión vinculado a la determinación de la existencia de simulación.

Por otro lado, trataremos de averiguar la posible utilidad de la psicofisiología en la detección de la simulación de aspectos cognitivos. Como ya ha quedado mencionado en otro lugar de este trabajo (ver Capítulo 4, apartado 4.1: El polígrafo) el empleo del polígrafo para la detección del engaño es un tema extremadamente controvertido, que

sigue dando lugar hoy en día a las más acaloradas discusiones (Vila, 1996). El empleo de los potenciales evocados vinculados a la aplicación de test específicos en la detección de simulación parece un método realmente prometedor, según algunos estudios han puesto de manifiesto (Ellwagner y cols., 1999; Rosenfeld y cols., 1999; Tardiff y cols., 2000, 2002), y ha sido apoyado también por estudios de neuroimagen (Lee y cols., 2005). Paralelamente, el empleo del polígrafo se ha empleado con éxito para evaluar la simulación del trastorno por estrés post-traumático (Blanchard y cols., 1996; Hickling y cols., 2002). Sin embargo, hasta donde nuestro conocimiento alcanza, no ha sido comprobada la utilidad de medidas psicofisiológicas simples, como el pulso o la conductancia, cuando estas se emplean conjuntamente con tests específicos en la detección de simulación de déficits cognitivos. Así, en este tercer estudio llevamos a cabo una investigación preliminar para explorar este aspecto.

Por último, realizaremos otro estudio exploratorio encaminado a dilucidar si existe algún patrón en los hallazgos de la SPECT-PCR que permita diferenciar a pacientes con síndrome post-conmocional de sujetos que, según nuestros resultados neuropsicológicos, están fingiendo el trastorno.

2. Objetivos e Hipótesis:

En el **estudio 3** exploraremos la utilidad de algunos tests neuropsicológicos no específicos de simulación, de la psicofisiología y de la neuroimagen funcional para la detección de la simulación. Concretamente, los objetivos de este tercer estudio serán:

- **Objetivo 1:** Averiguar si las pruebas neuropsicológicas no específicas de simulación son capaces de diferenciar entre los grupos de prevalencia diferencial.

- **Objetivo 2:** Averiguar si las distintas pruebas neuropsicológicas no específicas de simulación son capaces de diferenciar entre “grupos conocidos”.
- **Objetivo 3:** Obtener los puntos de corte para las distintas variables de los tests neuropsicológicos no específicos que optimicen la especificidad.
- **Objetivo 4:** Comparar nuestros resultados con los publicados en la literatura para la población norteamericana.
- **Objetivo 5:** Comprobar la utilidad de diversas medidas psicofisiológicas empleadas en combinación con una medida específica en la detección de simulación.
- **Objetivo 6:** Comprobar la utilidad de la SPECT-PCR a la hora de diferenciar pacientes honestos de personas sospechosas de simulación.

De los objetivos anteriores se derivaron las siguientes hipótesis:

- **Hipótesis 1:** Se encontrarán diferencias entre los grupos de prevalencia diferencial de simulación en la ejecución en los test neuropsicológicos, de modo que el grupo con una prevalencia menor (grupo NBC) tendrá una ejecución significativamente mejor que la del grupo con una elevada prevalencia (grupo BC), que será igual al grupo AN.
- **Hipótesis 2:** La ejecución en los test neuropsicológicos será diferente entre los “grupos conocidos”, de modo que el grupo NBC será igual al grupo BC-NSS, el grupo BC-SS será igual al grupo AN, y los grupos NBC y BC-NSS obtendrán una ejecución distinta a la de los grupos BC-SS y AN.
- **Hipótesis 3:** Manteniendo una elevada especificidad, los puntos de corte de las pruebas neuropsicológicas no específicas de simulación llevarán asociados una sensibilidad moderada (inferior al 60%).

- **Hipótesis 4:** La SPECT-PCR cerebral permitirá diferenciar a los sujetos análogos y sospechosos de simulación de los pacientes no simuladores.
- **Hipótesis 5:** Las medidas psicofisiológicas asociadas a la ejecución de una prueba específica de simulación serán diferentes cuando los sujetos opten por simular, de modo que:
 - o Se apreciará una mayor activación en los ensayos en los que los participantes decidan mentir (errores), en comparación con los ensayos en los que los participantes decidan contestar correctamente (aciertos).
 - o Se apreciará una mayor activación cuando los participantes estén simulando que cuando no estén simulando.

3. Metodología:

3.1. Participantes:

En este estudio participó el grupo de 54 pacientes empleado en el capítulo anterior, más un nuevo grupo de análogos. Este grupo nuevo estaba compuesto por 30 estudiantes de 1^{er} y 4^o curso de Psicología, con una media de 20,92 años de edad (Sd= 3,08), y una media de 13,92 años de escolaridad (Sd= 2,35). Como criterios de exclusión de este grupo se emplearon el embarazo (por motivos de protección radiológica) y el haber sufrido algún TCE o daño neurológico previo.

3.2. Instrumentación:

Las pruebas específicas de simulación fueron las mismas empleadas en el Estudio 1 y en el Estudio 2. La batería incluía, además, las siguientes pruebas:

- Test de la figura compleja de Rey (Rey, 1964). Esta prueba es una medida de memoria episódica con material visual que consiste en la copia, reproducción inmediata a los 3 minutos, reproducción demorada a los 30 minutos y reconocimiento de una figura compleja.

- Test de atención d2 (Brickenkamp, 1962). Este test permite obtener una gran cantidad de información sobre el proceso atencional del paciente, como la velocidad del procesamiento, la atención selectiva y sostenida, el control atencional e inhibitorio, concentración... La tarea del sujeto consiste en la cancelación de todas las letras d que encuentre a lo largo de 14 filas y que posean dos rayas en cualquier posición con respecto a la letra (dos arriba, dos abajo o una arriba y otra abajo).
- Test de fluidez semántica y fonética. Esta prueba está considerada una medida de fluidez, de memoria semántica y de función ejecutiva, y consiste en decir en un tiempo de 1 minuto por condición, nombres de animales y frutas (fluidez semántica) y palabras que empiecen por F, A y S (fluidez fonética).
- Subtest de Letras y Números del Wechsler Adult Intelligence Test-III (WAIS-III) (Wechsler, 1997^b). Este test de nueva inclusión en el WAIS es una medida directa de la Memoria de Trabajo o Memoria Operativa. Esta prueba consiste en la presentación de secuencias de letras y números, con incremento consecutivo de tamaño, ante las cuales el paciente debe agrupar verbalmente los números y las letras y ordenarlos siguiendo un criterio numérico o alfabético creciente.
- Test de flexibilidad cognitiva Cambios (Seisdedos, 2000). Este instrumento permite medir la capacidad del sujeto para concentrarse atendiendo a la vez a varias condiciones cambiantes y la flexibilidad para analizar, de un modo holístico, si se cumplen o no los cambios pedidos y en qué momento han dejado de cumplirse, aspectos que están directamente vinculados a la flexibilidad cognitiva.
- *Finger Tapping Test* (Reitan y Wolfson, 1995^b). Esta prueba consiste en medir la velocidad con que se presiona una palanca con el dedo índice durante 5 intervalos por mano de 10 segundos cada uno. Esta prueba es de amplio uso y no está comercializada

por ninguna empresa española, usándose los baremos de Spreen y Strauss (1998) por su mínima carga cultural.

- Test de colores y palabras STROOP (Golden, 1978). Esta prueba es una medida atencional de concentración o inhibición de respuestas atencionales secundarias.

- Test Auditivo-Verbal España-Complutense (TAVEC; Benedet y Alexandre 1998). Esta prueba es una medida de memoria episódica para material verbal que consiste en la administración repetida de una lista de 16 palabras en distintas condiciones de aprendizaje, recuerdo libre inmediato y demorado, recuerdo con claves inmediato y demorado y reconocimiento. Esta prueba permite la descripción del funcionamiento del sistema de memoria del sujeto. Esta es la versión española del ampliamente conocido California Verbal Learning Test.

- Cámara Picker Prism 3000, de tres detectores, dotada de colimación neurofocal de alta resolución, en matriz de 128x128, muestreo angular cada 2°, 35 segundos por paso, 75 kcts/paso.

- Polígrafo: se utilizó un polígrafo marca PowerLab 800 con un preamplificador Cibertec Biosig de tres canales. La conductancia dérmica se registró a través del preamplificador Conductance PGR y el pulso sanguíneo con el Frequency CFR. Para el registro de la actividad electrodermal se utilizaron dos electrodos In Vivo Metrics de tamaño estándar fijados a la piel mediante discos adhesivos de doble cara y gel electrolítico (gel hipoconductor salínico –marca Jonhson & Jonhson K-Y Jelly- con una concentración del cloruro sódico en torno a 0.29 y 100 g de agua). Para el registro del pulso sanguíneo se utilizó un transductor fotoeléctrico.

- Sistema informatizado: El registro computerizado se realizó con una tarjeta analógico-digital. La tarjeta estaba conectada a un ordenador PC- Pentium que controlaba el

polígrafo. El programa de registro de las variables utilizado fue el Chart v3.4.2. para Windows. Para el análisis de las mismas se utilizó la versión Chart v5 para Windows.

3.3. Procedimiento:

El procedimiento seguido con los pacientes ha quedado descrito en el estudio anterior. Con respecto a los análogos, el reclutamiento se realizó mediante la explicación del experimento en una clase de 1^{er} y 4^o curso de Psicología, siendo voluntaria la participación. Debido a que no se encontraron diferencias entre el tipo de instrucciones proporcionadas a los participantes en el Estudio 1, optamos por emplear en este estudio las instrucciones más detalladas (ver Apéndice 2.2). El tipo de incentivo proporcionado a estos análogos fue de tipo económico, recibiendo 60 euros por la participación en el experimento. La evaluación se realizó en dos días, empleándose el primero para la evaluación psicofisiológica y neuropsicológica, con una duración aproximada de dos horas y media, y el segundo para la realización de la SPECT-PCR, con una duración de una hora. En ambos casos los participantes firmaban un consentimiento informado, en el que se informaba extensamente de los procedimientos (ver Anexo X). Tres de los sujetos fueron descartados del estudio debido a hallazgos en la neuroimagen; en dos de estos casos la etiología probable fueron traumatismos sufridos en la infancia de los que no habían informado previamente, y en el tercer caso dicha etiología es desconocida.

Para la realización de la SPECT-PCR la exploración se realizaba siguiendo el protocolo al uso en el Servicio de Medicina Nuclear del Hospital Universitario “Virgen de las Nieves”, que exponemos brevemente a continuación.

El paciente es invitado a permanecer echado en condiciones de reposo neurosensorial en una habitación libre de ruidos y con una luz tenue. Tras unos quince

minutos aproximadamente se procede a la inyección del trazador Tecnecio-99m·Etilindiol-Di-Cisteinato Dímero (ECD), en dosis acorde al peso del paciente (20-25 mCi). Tras un aclaramiento de 60 minutos se realiza la adquisición de las imágenes mediante la cámara Picker Prism 3000. Las imágenes iniciales (*raw data*) son posteriormente reconstruidas (algoritmos de retro-reconstrucción), filtradas (Filtro Butterworth; orden: 6.0, punto de corte: 0.3), corregidas por atenuación volumétrica (método de Chang) y reorientadas en un plano virtual Órbito-Meatal (OM) hasta la obtención de imágenes en ejes ortogonales de planos transversal, coronal y sagital con un grosor de corte de 2.2 mm. En condiciones genéricas puede referirse la exploración como Tomografía Neuroespecífica de Alta Resolución.

Las imágenes procesadas definitivas son valoradas por al menos dos especialistas en Medicina Nuclear con especial atención al patrón de distribución del trazador (valoración global).

Para la evaluación psicofisiológica procedimos del siguiente modo: los participantes llegaban al laboratorio y se les explicaba en qué iba a consistir la tarea. A continuación se procedía a colocar los electrodos y se verificaba que los registros psicofisiológicos eran correctos. Entonces comenzaban a realizar el Victoria Symptom Validity Test en un ordenador portátil (marca Acer, TravelMate 291LMI). Cada vez que los sujetos seleccionaban una respuesta el ratón enviaba al polígrafo una señal de 5 voltios que quedaba maracada en el canal nº tres del polígrafo. Cuando los participantes concluían la tarea, se les instaba a realizar de nuevo la tarea, pero olvidando las instrucciones de simular que habían recibido, es decir, se pedía que realizaran la tarea lo mejor que podían (condición control). En esta segunda realización del test los sujetos sólo realizaban el primer bloque de ensayos del VSVT. En total el tiempo medio utilizado para la tarea fue de 40 minutos.

3.4. Variables:

Atendiendo a la prueba a la que pertenecen, las variables seleccionadas han sido:

- *Test de la Figura Compleja de Rey (Rey, 1964):*

- Copia: se corresponde con la puntuación obtenida al sumar los aciertos en la fase de copia de la figura.
- Recuerdo inmediato: se corresponde con la puntuación que el sujeto obtiene en la reproducción de memoria que realiza de la figura 3 minutos después de la realización de la copia
- Recuerdo demorado: se corresponde con la puntuación que el sujeto obtiene en la reproducción de memoria que realiza de la figura 30 minutos después de la realización de la copia
- Reconocimiento: inmediatamente después de la realización del recuerdo demorado, el sujeto ha de reconocer entre una serie de distractores los elementos que pertenecen a la figura. El número de elementos correctamente identificados constituyen la variable reconocimiento.
- Falsas alarmas: esta variable la constituyen las elecciones erróneas de la tarea de reconocimiento.

- *Test de atención d2 (Brickenkamp, 1962)*

- TR: Se calcula sumando el número de la columna a la que el sujeto consigue llegar en las 14 filas de la prueba. Esta variable nos da información sobre la velocidad de procesamiento y la cantidad de trabajo realizado.
- TA: Es el número total de aciertos del sujeto de la prueba, y nos proporciona una medida de la precisión del procesamiento.

- Errores de omisión: Es el número de letras *b* no marcadas. Es una medida de control atencional, del cumplimiento de una regla, de la precisión de la búsqueda visual y de la calidad de la actuación del sujeto en la tarea.
- Errores de comisión: Es el número de estímulos marcados que no se corresponden con letras *b*. Esta variable se relaciona con el control inhibitorio, el cumplimiento de una regla, la precisión de la búsqueda visual, la minuciosidad y la flexibilidad cognitiva.
- TOT: Se calcula restando el número total de errores (omisión más comisión) a la puntuación TR. Proporciona una medida de control atencional e inhibitorio y de la relación entre la velocidad y precisión de los sujetos.
- CON: Se calcula restando a la puntuación TA el número de errores de comisión. Es una medida de la concentración del sujeto.
- VAR: Se calcula restando a la columna mayor a la que el sujeto ha conseguido llegar en la prueba la columna menor. Es, por tanto, una medida de la variación fluctuación en el modo de trabajar. Una puntuación extremadamente alta en esta variable puede estar relacionada con una falta de motivación.

- *Test de fluidez semántica y fonética.*

- Fluidez fonética: Se calcula mediante la suma de todas las palabras que el sujeto consigue emitir que comiencen por la letra F, A y S, concediéndole un minuto de tiempo para cada letra.
- Fluidez semántica- Animales: Es el número total de animales que el sujeto consigue nombrar en un minuto de tiempo.
- Fluidez semántica- Frutas: Consiste en el número total de frutas que el sujeto consigue nombrar en un minuto.

- *Subtest de Letras y Números del Wechsler Adult Intelligence Test-III (WAIS-III; Wechsler, 1997).*

- Total: número de series de letras y números que el sujeto consigue repetir ordenadamente.

- *Test de flexibilidad cognitiva Cambios (Seisdedos, 2000).*

- Total: Número total de respuestas correctas del sujeto.

- *Finger Tapping Test (Reitan y Wolfson, 1995^b).*

- Mano dominante: Esta variable es la media del número de veces que el sujeto consigue pulsar la palanca con el dedo índice de la mano dominante durante un mínimo de cinco ensayos de 10 segundos de duración cada uno.
- Mano No-dominante: es la media del número de veces que el sujeto consigue pulsar la palanca con el dedo índice de su mano no dominante durante un mínimo de cinco ensayos de 10 segundos de duración cada uno.

- *Test de colores y palabras STROOP (Golden, 1978).*

- Palabras: Número de palabras que el sujeto consigue leer en 45 segundos.
- Colores: Número de colores que el sujeto consigue denominar en 45 segundos.
- PalxCol: Número de palabras que el sujeto consigue leer en la condición de interferencia.
- Interferencia: Se calcula mediante la siguiente fórmula:

$$\text{Interferencia} = \text{PalxCol} - (\text{Colores} \times \text{Palabras} / \text{Colores} + \text{Palabras})$$

- *Test Auditivo-Verbal España-Complutense (Benedet y Alexandre, 1998).*

- Ensayo 1: Número de palabras que el sujeto consigue recordar después del primer ensayo.
- Total ensayos: Suma del número de palabras que el sujeto ha recordado a lo largo de los cinco ensayos de aprendizaje.

- RL-CP: Recuerdo libre a corto plazo.
- RL-LP: Recuerdo libre a largo plazo.
- Reconocimiento: Número de aciertos en la prueba de reconocimiento.
- Discriminabilidad: Se calcula mediante la siguiente fórmula a partir de la prueba de reconocimiento: $1 - (\text{Falsos positivos} + \text{omisiones} / 44) \times 100$

- *Valoración Global de la SPECT-PCR*: La exploración de la neuroimagen se clasifica de modo ordinal, estableciéndose las categorías desde 0= sin hallazgos significativos, normal; hasta el valor 5= claramente patológico.

- *Conductancia eléctrica de la piel*. La conductancia dérmica se registró mediante el paso de una pequeña corriente externa continua y voltaje constante aplicada a través de dos electrodos colocados en la prominencia hipotenar de la mano no dominante (una zona de la palma de la mano donde hay actividad sudorípara de las glándulas ecrinas) y expresada en microSiemens. La tasa de muestreo fue de 1000 muestra por segundo. La variable utilizada para el análisis de la conductancia fue la media del valor máximo (definido como el valor del mayor punto de datos en la selección). La selección escogida para los análisis fueron:

- El *Tiempo de Reacción (TR)* de cada sujeto para cada uno de los ensayos del VSVT.
- El intervalo desde la desaparición del número objetivo y la presentación de las dos alternativas de respuesta (de 5, 10 ó 15 segundos) o lo que hemos denominado *intervalo entre estímulos (IEE)*.
- Los *dos segundos (2 segundos)* que transcurren desde que los participantes responden hasta que aparece un nuevo objetivo.

- *Pulso sanguíneo*. El pulso sanguíneo se registró con un transductor fotoeléctrico colocado en la falange distal del dedo índice de la mano no dominante. La tasa de

muestreo fue de 1000 muestra por segundo. El tiempo seleccionado para los análisis fue el mismo que para la conductancia (TR, IEE y 2 segundos). Las variables utilizadas para el análisis del pulso sanguíneo fueron:

- Tasa o *BPM*: Promedio de la tasa (pulsos por minuto, en inglés, BPM) de los ciclos de la selección.
- *Altura cíclica*: el promedio de la diferencia entre los máximos y los mínimos en los ciclos seleccionados (medida en Voltios).

Con el propósito de facilitar la comprensión de los intervalos seleccionados para el análisis de las variables, a continuación se presenta una figura con un ejemplo de un ensayo completo de la prueba VSVT en la que se han marcado los distintos intervalos con los que hemos trabajado (ver Figura 12)

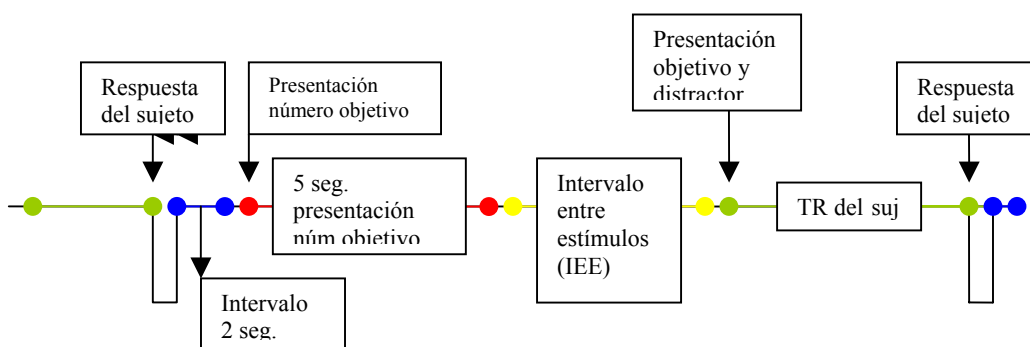


Figura 12. Ejemplo de un ensayo completo del Victoria Symptom Validity Test

3.5. Análisis estadísticos

En este estudio se los análisis estadísticos realizados son paralelos a los del estudio 2. Así, debido a que no se cumplía el supuesto de homogeneidad de las varianzas, se realizó un análisis no paramétrico para “n” grupos independientes utilizando el estadístico de U de Mann-Whitney, tanto en el diseño de “grupos de prevalencia diferencial” como en el de “grupos conocidos”, para conocer si existían diferencias globales entre los grupos en la ejecución de las distintas variables de las

pruebas. Si estas diferencias eran significativas, se realizaron análisis a posteriori no paramétricos comparando los grupos por pares (estadístico Kruskal-Wallis).

A continuación, para conocer la significación clínica de las diferencias encontradas en el análisis anterior, se obtuvieron los tamaños del efecto utilizando la *delta* de Cohen de acuerdo con la formulación de Zakzanis (2001), empleando la misma fórmula que en el estudio 2.

Además, para comprobar la utilidad diagnóstica de las variables se realizaron análisis de curvas ROC de las cuales se obtuvieron los puntos de corte, la sensibilidad y la especificidad de cada variable. Para aumentar su utilidad clínica, se obtuvieron los distintos puntos de corte, su valor predictivo positivo y negativo ajustado a las prevalencias teóricas utilizadas en la literatura como son el 15%, el 30% y el 45%.

Por otro lado, se realizó un análisis de ANOVA univariada siendo la VI el grupo (pacientes vs. sospechosos de simulación vs. análogos) y la VD la puntuación resultante de cuantificar los resultados de la SPECT-PCR cerebral.

Por último, se realizaron modelos lineales generales de medidas repetidas con 3 niveles (aciertos, errores e intervalo entre estímulos) para las distintas variables psicofisiológicas en el grupo de análogos.

4. Resultados:

4.1. Análisis de los “grupos de prevalencia diferencial” en las pruebas neuropsicológicas no específicas de simulación (Hipótesis 1):

En primer lugar procedimos a averiguar si las pruebas específicas de simulación diferenciaban entre grupos con distintas probabilidades de simulación. Para ello se realizaron 29 análisis no paramétricos para 3 grupos independientes (Kruskal-Wallis): cinco para la Figura Compleja de Rey, siete para el Test d2, tres para la fluidez, uno

para la prueba Letras y Números, uno para el Cambios, dos para el Finger Tapping, cuatro para el STROOP y seis para el TAVEC.

En todos los casos la VI fue el grupo diferencial [grupo de no-búsqueda de compensación (NBC) vs. grupo de búsqueda de compensación (BC) vs. grupo de análogos (AN)]. Las VD para la Figura Compleja de Rey fueron: copia, recuerdo inmediato, recuerdo demorado, reconocimiento y falsas alarmas. En cuanto al Test d2, las VD empleadas fueron: TR, TA, errores de omisión, errores de comisión, TOT, CON y VAR. Las VD para las pruebas de fluidez fueron: número de palabras totales emitidas en la fluidez fonética, número de animales nombrados y número de frutas nombradas. La VD en Letras y Números fue el número de ensayos que el sujeto consiguió reproducir correctamente. En el test Cambios la VD fue el número de soluciones correctas emitidas por el sujeto. Para el Finger Tapping se emplearon como VD la media de pulsaciones emitidas por el sujeto para la mano dominante y para la no dominante. En cuanto al STROOP, las VD consideradas fueron: número de palabras leídas, número de colores denominados, número de colores denominados correctamente en la condición de interferencia, y puntuación de interferencia. Por último, en el TAVEC se seleccionaron las siguientes VD: Número de palabras recordadas en el ensayo 1, número de palabras recordadas a lo largo de los cinco ensayos, recuerdo libre a corto plazo, recuerdo libre a largo plazo, reconocimiento y discriminabilidad.

En los casos en los que se encontraron diferencias estadísticamente significativas en el análisis anterior se procedió a realizar comparaciones a posteriori dos a dos (grupos NBC vs. BC, grupos NBC vs. AN y grupos BC vs. AN) mediante el análisis de Mann-Whitney.

En el caso de la Figura Compleja de Rey los resultados mostraron que existían diferencias estadísticamente significativas entre los tres grupos para todas las variables

estudiadas excepto para el número de falsas alarmas: copia [$\chi^2(2)= 6,15; p<0,046$], recuerdo inmediato [$\chi^2(2)= 8,83; p<0,012$], recuerdo demorado [$\chi^2(2)= 6,25; p<0,044$] y reconocimiento [$\chi^2(2)= 9,34; p<0,009$]. En cuanto a las comparaciones a posteriori, en el rendimiento de la copia y en el reconocimiento demorado se encontró una superioridad estadística del grupo NBC sobre el grupo BC. En cuanto al reconocimiento, el rendimiento fue estadísticamente superior en el grupo NBC con respecto al grupo AN. Contrario a la hipótesis prevista, el rendimiento del grupo BC en el recuerdo inmediato fue estadísticamente inferior al del grupo AN (ver Tabla 34).

Para el Test d2, los resultados mostraron que existían diferencias estadísticamente significativas entre los grupos para todas las variables estudiadas excepto para TR, errores de omisión y VAR: TA [$\chi^2(2)= 8,67; p<0,013$], errores de comisión [$\chi^2(2)= 10,97; p<0,004$], TOT [$\chi^2(2)= 10,13; p<0,006$], y CON [$\chi^2(2)= 15,16; p<0,001$]. Las comparaciones a posteriori mostraron que para las variables TA y TOT la ejecución de los AN fue estadísticamente inferior a la del grupo NBC. En cuanto a los errores de comisión, el grupo de AN fue estadísticamente superior a los otros dos. Por último, el grupo NBC obtuvo una ejecución estadísticamente superior a la de los otros dos grupos en la variable CON (ver Tabla 34).

En cuanto a las pruebas de fluidez, los resultados mostraron que existían diferencias estadísticamente significativas entre los tres grupos sólo para la variable frutas [$\chi^2(2)= 8,12; p<0,017$]. Los análisis a posteriori mostraron que la ejecución del grupo NBC era estadísticamente superior a la del grupo BC y estadísticamente inferior a la del grupo AN (ver tabla 34).

Para la variable del Letras y Números no se encontraron diferencias estadísticamente significativas entre los grupos (ver tabla 34).

Tabla 34. Medias, desviaciones típicas y análisis a posteriori de las pruebas neuropsicológicas no específicas para los grupos de prevalencia diferencial.

Prueba	Variables	Grupo NBC		Grupo BC		Grupo AN		X ²	p	Mann-Whitney
		Media	Dt	Media	Dt	Media	Dt			
FCR	Copia	33,43	4,67	29,61	8,14	32,30	4,33	6,15	0,046	(NBC>BC)=AN
	R Inmediato	17,50	6,71	15,63	3,96	20,42	4,69	8,83	0,012	NBC=(BC<AN)
	R Demorado	18,36	6,16	13,77	4,56	17,05	6,55	6,25	0,044	(NBC>BC)=AN
	Reconocimiento	8,47	3,31	7,27	2,19	6,19	2,75	9,34	0,009	BC=(NBC>AN)
Test d2	Falsas alarmas	1,21	2,52	0,77	1,00	1,19	2,65	0,27	0,870	NBC=BC=AN
	TR	380,78	77,89	312,52	105,29	336,18	101,14	5,05	0,080	NBC=BC=AN
	TA	140,53	35,00	109,73	43,81	111,00	34,88	8,67	0,013	BC=(NBC>AN)
	Errores o	18,53	15,47	21,43	23,75	29,77	32,59	1,99	0,368	NBC=BC=AN
Fluidez	Errores c	4,32	11,54	5,17	14,68	22,48	47,14	10,97	0,004	(NBC=BC)<AN
	TOT	357,92	81,52	285,91	99,08	103,04	84,41	10,13	0,006	BC=(NBC>AN)
	CON	139,50	37,98	102,73	50,50	88,14	51,77	15,16	0,001	NBC>(BC=AN)
	VAR	13,89	4,59	15,34	8,23	14,25	5,08	0,39	0,981	NBC=BC=AN
Fluidez	Fonética	26,66	11,41	23,00	8,91	26,92	9,81	1,86	0,393	NBC=BC=AN
	Sem. Animales	19,22	7,94	15,65	4,32	15,67	5,83	4,75	0,092	NBC=BC=AN
	Sem. Frutas	14,48	6,42	11,65	2,82	12,11	6,69	8,12	0,017	NBC>BC; NBC<AN; BC=AN
	Total	9,48	3,30	7,42	2,33	8,37	2,54	7,23	0,065	NBC=BC=AN
LyN	Total	9,92	5,64	7,18	4,13	12,77	5,37	11,69	0,003	(NBC=BC)<AN
	Mano Dom.	41,26	9,71	34,60	12,27	34,09	12,26	6,37	0,041	NBC>(BC=AN)
	Mano No-Dom	35,32	7,58	31,84	11,54	30,72	10,51	3,77	0,141	NBC=BC=AN
	Palabras	93,33	17,63	77,87	24,76	75,55	22,31	11,48	0,003	NBC>(BC=AN)
STROOP	Colores	65,73	14,07	55,87	16,25	58,59	15,55	5,28	0,071	(NBC>BC)=AN
	PalxCol	39,23	12,34	31,04	11,74	34,74	12,92	4,09	0,129	(NBC>BC)=AN
	Interferencia	0,74	8,36	-1,16	7,98	3,41	8,03	2,48	0,289	NBC=BC=AN
	Ensayo 1	7,03	2,10	5,79	1,79	6,62	2,04	5,17	0,075	NBC=BC=AN
TAVEC	Total ensayos	55,06	9,32	44,87	14,26	43,92	14,26	13,42	0,001	NBC>(BC=AN)
	RL-CP	11,36	2,83	7,87	3,15	7,88	3,68	17,51	0,000	NBC>(BC=AN)
	RL-LP	11,60	3,06	7,83	3,96	6,60	4,28	20,48	0,000	NBC>(BC=AN)
	Reconocimiento	15,06	0,98	12,79	3,14	11,70	3,12	24,37	0,000	NBC>(BC=AN)
TAVEC	Discriminabilidad	94,95	4,98	80,32	19,76	83,48	10,36	26,41	0,000	NBC>(BC=AN)

Nota: BC= búsqueda de compensación; NBC= no búsqueda de compensación; AN= análogos; FCR= Test de la Figura Compleja de Rey; R= recuerdo; o=omisión; c= comisión; Sem. = semántica; LyN= letras y números; Finger Tapp= Finger Tapping; PalxCol= palabras denominadas en la condición de interferencia; TAVEC= Test de Aprendizaje Verbal España-Complutense; RL-CP= recuerdo libre a corto plazo; RL-LP= recuerdo libre a largo plazo.

En la variable del Cambios los resultados mostraron diferencias estadísticamente significativas entre los grupos [$\chi^2(2)= 11,69$; $p<0,003$], descubriendo los análisis a posteriori una ejecución del grupo AN superior a la de los otros dos, contrariamente a lo planteado por nuestras hipótesis (ver Tabla 34).

En el Finger Tapping no se encontraron diferencias entre los grupos en la ejecución de la mano no dominante, pero sí de la dominante [$\chi^2(2)= 6,37$; $p<0,041$], mostrando los análisis a posteriori la superioridad del grupo NBC sobre los otros dos (ver Tabla 34).

En el caso del STROOP los resultados mostraron que existían diferencias estadísticamente significativas entre los tres grupos estudiados sólo en el número de palabras leídas [$\chi^2(2)= 11,48$; $p<0,003$], encontrándose en los análisis a posteriori la superioridad en la ejecución del grupo NBC sobre los otros dos (ver Tabla 34).

Para finalizar, en el TAVEC los resultados mostraron que existían diferencias estadísticamente significativas para todas las variables excepto para el número de palabras recordadas en el ensayo 1: total ensayos [$\chi^2(2)= 13,42$; $p<0,001$], recuerdo libre a corto plazo [$\chi^2(2)= 17,51$; $p<0,000$], recuerdo libre a largo plazo [$\chi^2(2)= 20,48$; $p<0,000$], reconocimiento [$\chi^2(2)= 24,37$; $p<0,000$] y discriminabilidad [$\chi^2(2)= 26,41$; $p<0,000$]. En cuanto a las comparaciones a posteriori, en todas las variables se encontró una ejecución del grupo NBC estadísticamente superior a la de los otros dos grupos (BC y AN), que no diferían entre sí (ver Tabla 34).

Estudio de los tamaños del efecto:

Para realizar un análisis más preciso de la relevancia estadística de los hallazgos se calcularon los tamaños del efecto de los análisis anteriores empleando el estadístico Delta de Cohen. Mediante este análisis se puso de manifiesto que la variable que mejor diferencia entre los grupos NBC y BC se corresponde con la variable discriminabilidad

del TAVEC ($\delta= 1,18$), y la que menos diferencia es la variable errores de comisión del Test d2 ($\delta= -0,06$) (ver Tabla 35).

Para diferenciar entre los grupos NBC y AN, la variable que se mostró más útil fue la variable TOT el Test d2 ($\delta= 3,07$), y la variable que menos diferenció entre los grupos NBC y AN fue la de falsas alarmas de la Figura Compleja de Rey ($\delta= 0,00$), obteniendo el resto de las variables niveles de diferenciación entre estos dos grupos intermedias (ver Tabla 35).

Por último, se puso de manifiesto que la variable que mejor diferencia entre los grupos BC y AN fue la puntuación TOT del Test d2 ($\delta= 1,99$), y la que menos diferencia es el RL-CP del TAVEC ($\delta= -0,00$) (ver Tabla 35).

Tabla 35. Delta de Cohen de las variables de los tests neuropsicológicos no específicos de simulación para los grupos de prevalencia diferencial

Prueba	Variables	NBC vs. BC	NBC vs. AN	BC vs. AN
FCR	Copia	0,596	0,251	0,431
	R Inmediato	0,350	0,512	1,107
	R Demorado	0,856	0,206	0,590
	Reconocimiento	0,436	0,752	0,437
	Falsas alarmas	0,250	0,007	0,230
Test d2	TR	0,745	0,498	0,229
	TA	0,781	0,845	0,032
	Errores o	0,147	0,467	0,296
	Errores c	0,064	0,618	0,560
	TOT	0,797	3,072	1,993
	CON	0,831	1,144	0,285
	VAR	0,226	0,074	0,163
	Fluidez	Fonética	0,360	0,024
	Sem. Animales	0,582	0,515	0,003
	Sem. Frutas	0,612	0,361	0,096
LyN	Total	0,731	0,380	0,390
Cambios	Total	0,560	0,517	1,176
Finger Tapp	Mano Dom.	0,606	0,652	0,041
	Mano No-Dom	0,364	0,508	0,101
STROOP	Palabras	0,729	0,890	0,098
	Colores	0,650	0,482	0,171
	PalxCol	0,680	0,355	0,300
	Interferencia	0,232	0,325	0,570
TAVEC	Ensayo 1	0,637	0,198	0,433
	Total ensayos	0,864	0,944	0,066
	RL-CP	1,167	1,069	0,002
	RL-LP	1,074	1,362	0,298
	Reconocimiento	1,101	1,639	0,348
	Discriminabilidad	1,182	1,495	0,209

Nota: BC= búsqueda de compensación; NBC= no búsqueda de compensación; AN= análogos; FCR= Test de la Figura Compleja de Rey; R= recuerdo; o=omisión; c= comisión; Sem.= semántica; LyN= letras y números; Finger Tapp= Finger Tapping; PalxCol= palabras denominadas en la condición de interferencia; TAVEC= Test de Aprendizaje Verbal España-Complutense; RL-CP= recuerdo libre a corto plazo; RL-LP= recuerdo libre a largo plazo.

4.2. Análisis de “grupos conocidos” en las pruebas no específicas de simulación (Hipótesis 2):

Una vez comprobado que algunas de las variables de los tests neuropsicológicos no específicos de simulación son capaces de diferenciar entre grupos de prevalencia diferencial, se procedió a comprobar si dichas pruebas eran capaces de diferenciar a un grupo sospechosos de simulación. Para ello, empleamos cuatro grupos: un grupo de no-búsqueda de compensación (NBC), un grupo con búsqueda de compensación pero no sospechoso de simulación (BCNSS), un grupo con búsqueda de compensación y sospechoso de simulación (BCSS) y por último un grupo de análogos (AN).

Para ello, siguiendo la estructura del objetivo anterior, se realizaron 29 análisis no paramétricos para 4 grupos independientes (Kruskal-Wallis): cinco para la Figura Compleja de Rey, siete para el Test d2, tres para la fluidez, uno para la prueba Letras y Números, uno para el Cambios, dos para el Finger Tapping, cuatro para el STROOP y seis para el TAVEC. En todos los casos la VI fue el grupo diferencial [grupo de no-búsqueda de compensación (NBC) vs. grupo de búsqueda de compensación no sospechoso de simulación (BCNSS) vs. grupo de búsqueda de compensación sospechoso de simulación (BCSS) vs. grupo de análogos (AN)]. Las VD seleccionadas fueron las mismas que para la hipótesis anterior.

En los casos en los que se encontraron diferencias estadísticamente significativas en el análisis anterior se procedió a realizar comparaciones dos a dos (grupos NBC vs. BCNSS, grupos NBC vs. BCSS, grupos NBC vs. AN, grupos BCNSS vs. BCSS, grupos BC-NSS vs. AN, y grupos BCSS vs. AN) mediante el análisis de Mann-Whitney.

En el caso de la Figura Compleja de Rey, se encontraron diferencias estadísticamente significativas entre los grupos para todas las variables excepto para la copia y las falsas alarmas: recuerdo inmediato [$\chi^2(2) = 10,75; p < 0,013$], recuerdo

demorado [$\chi^2(2)= 8,29; p<0,040$] y reconocimiento [$\chi^2(2)= 12,63; p<0,005$] (ver Tabla 36).

En cuanto al Test d2, también se encontraron diferencias estadísticamente significativas entre los grupos para todas las variables excepto para el número de errores de comisión y para la variable VAR: TR [$\chi^2(2)= 9,40; p<0,024$], TA [$\chi^2(2)= 20,52; p<0,000$], errores de comisión [$\chi^2(2)= 15,61; p<0,001$], TOT [$\chi^2(2)= 18,78; p<0,000$] y CON [$\chi^2(2)= 26,76; p<0,000$] (ver Tabla 36).

En las variables de fluidez sólo se encontraron diferencias estadísticamente significativas entre los grupos para la variable frutas [$\chi^2(2)= 8,81; p<0,032$] (ver Tabla 36).

En la variable de la prueba Letras y Números no se encontraron diferencias estadísticamente significativas entre los grupos (ver Tabla 36).

En cuanto al Finger Tapping, los resultados mostraron diferencias estadísticamente significativas entre los grupos para las dos variables estudiadas: ejecución de la mano dominante [$\chi^2(2)= 9,62; p<0,022$] y ejecución de la mano no dominante [$\chi^2(2)= 9,52; p<0,023$] (ver Tabla 36).

En el test STROOP los resultados mostraron diferencias estadísticamente significativas para todas las variables estudiadas excepto para la interferencia: palabras [$\chi^2(2)= 17,74; p<0,000$], colores [$\chi^2(2)= 9,16; p<0,027$] y colores denominados en la condición de interferencia [$\chi^2(2)= 11,58; p<0,009$] (ver Tabla 36).

Tabla 36. Medias, desviaciones típicas y análisis a posteriori de las pruebas neuropsicológicas no específicas para los “grupos conocidos”

Prueba	Variable	Grupo No-BC		Grupo BC-No SS		Grupo BC-SS		Grupo análogos		U	p	Mann-Whitney
		Media	Sd	Media	Sd	Media	Sd	Media	Sd			
FCR	Copia	33,43	4,67	30,15	9,35	28,93	6,89	32,30	4,33	7,41	0,060	NBC=BCNSS=AN; NBC>BCSS; BCNSS=BCSS; BCSS=AN
	R Inmediato	17,50	6,71	17,05	4,43	13,87	2,53	20,42	4,69	10,75	0,013	(BCNSS=BCSS)<AN; NBC>BCSS; NBC=BCNSS; NBC=AN
D2	R Demorado	18,36	6,16	15,15	5,23	12,06	3,05	17,05	6,55	8,29	0,040	NBC=BCNSS=AN; NBC>BCSS; BCNSS=BCSS; BCSS<AN
	Reconocimiento	8,47	3,31	8,20	1,39	6,12	2,53	6,19	2,75	12,63	0,005	NBC=BCNSS=BCSS; NBC=AN; BCNSS>AN; BCSS=AN
	Falsas alarmas	1,21	2,52	0,90	0,99	0,62	1,06	1,19	2,65	0,945	0,815	NBC=BCNSS=BCSS=AN
	TR	380,78	77,89	354,38	74,06	258,10	118,16	336,18	101,14	9,40	0,024	NBC=BCNSS=AN; NBC>BCSS; BCNSS>BCSS; BCSS<AN
Fluidez	TA	140,53	35,00	135,38	31,37	76,40	34,57	111,00	34,88	20,52	0,000	(NBC=BCNSS)>BCSS; NBC=AN; BCNSS>AN; BCSS<AN
	Errores o Errores c	18,53 4,32	15,47 11,54	13,23 1,15	12,11 2,11	32,10 10,40	30,98 21,62	29,77 22,48	32,59 47,14	5,74 15,61	0,125 0,001	NBC=BCNSS=BCSS=AN (BCSS=AN)>BCNSS; NBC=BCNSS; NBC=BCSS; NBC=AN
LyN	TOT	357,92	81,52	340,00	73,90	215,60	94,63	282,85	84,41	18,78	0,000	(NBC=BCNSS)>BCSS; NBC=AN; BCNSS>AN; BCSS<AN
	CON	139,50	37,98	133,07	32,00	63,30	42,54	88,14	51,77	26,76	0,000	(NBC=BCNSS)>BCSS; NBC=AN; BCNSS>AN; BCSS=AN
Sem. Frutas	VAR	13,89	4,59	14,61	8,36	16,30	8,42	14,25	5,08	0,70	0,873	NBC=BCNSS=BCSS=AN
	Fonética	26,66	11,41	23,85	9,88	21,66	7,53	26,92	9,81	2,14	0,543	NBC=BCNSS=BCSS=AN
	Sem. Animales	19,22	7,94	16,85	4,34	13,77	3,76	15,67	5,83	6,46	0,091	NBC=BCNSS=AN; NBC>BCSS; BCNSS=BCSS; BCSS=AN
Total	Sem. Frutas	14,48	6,42	12,14	3,23	10,88	1,96	12,11	6,69	8,81	0,032	BCNSS=BCSS=AN; NBC=BCNSS; NBC>BCSS; NBC=AN
	Total	9,48	3,30	8,15	2,03	6,25	2,43	8,37	2,54	7,23	0,065	NBC=BCNSS=AN; NBC>BCSS; BCNSS=BCSS; BCSS<AN

Cambios	Total	9,92	5,64	7,54	4,05	6,40	4,66	12,77	5,37	11,85	0,008	NBC=BCNSS=BCSS; AN<(NBC, BCNSS, BCSS)
Finger Tap	Mano Dom.	41,26	9,71	38,71	10,40	28,22	12,76	34,09	12,26	9,62	0,022	NBC=BCNSS=AN; NBC>BCSS; BCNSS=BCSS; BCSS=AN
	Mano No-Dom	35,32	7,58	36,68	9,90	24,32	10,12	30,72	10,51	9,52	0,023	NBC=BCNSS=AN; NBC>BCSS; BCNSS>BCSS; BCSS=AN
STROOP	Palabras	93,33	17,63	88,35	19,34	63,20	24,83	75,55	22,31	17,74	0,000	(NBC=BCNSS)>(BCSS=AN)
	Colores	65,73	14,07	62,00	12,52	47,30	17,56	58,59	15,55	9,16	0,027	(NBC=BCNSS)>BCSS; AN=(NBC, BCNSS, BCSS)
	PalxCol	39,23	12,34	36,92	8,93	22,80	10,37	34,74	12,92	11,58	0,009	(NBC=BCNSS)>BCSS; NBC=AN; BCNSS=AN; BCSS<AN
	Interferencia	0,74	8,36	0,93	8,70	-4,10	6,10	3,41	8,03	6,38	0,094	(AN=NBC)>BCSS; NBC=BCNSS; BCNSS=BCSS; BCNSS=AN
TAVEC	Ensayo 1	7,03	2,10	6,28	1,93	5,10	1,37	6,62	2,04	7,89	0,048	NBC=BCNSS=AN; NBC>BCSS; BCNSS=BCSS; BCSS<AN
	Total ensayos	55,06	9,32	49,14	10,50	38,90	11,52	43,92	14,26	16,87	0,001	BCNSS=BCSS=AN; NBC=BCNSS; NBC>BCSS; NBC>AN
	RL-CP	11,36	2,83	8,85	2,68	6,50	3,37	7,88	3,68	19,75	0,000	BCNSS=BCSS=AN; NBC=BCNSS; NBC>BCSS; NBC>AN
	RL-LP	11,60	3,06	9,35	3,71	5,70	3,40	6,70	4,28	24,42	0,000	BCNSS>BCSS=AN; NBC>BCNSS; NBC>BCSS; NBC>AN
	Reconocimiento	15,06	0,98	13,92	1,73	11,20	4,02	11,70	3,12	26,82	0,000	NBC>(BCSS=AN); NBC=BCNSS; BCNSS=BCSS; BCNSS>AN
	Discriminabilidad	94,95	4,98	85,55	19,87	73,00	18,04	83,48	10,36	32,33	0,000	NBC>(BCSS=AN); NBC=BCNSS; BCNSS>BCSS; BCNSS=AN

Nota: BC= búsqueda de compensación; NBC= no búsqueda de compensación; AN= análogos; FCR= Test de la Figura Compleja de Rey; R= recuerdo; o=omisión; c= comisión; Sem.= semántica; LyN= letras y números; Finger Tapp= Finger Tapping; PalxCol= palabras denominadas en la condición de interferencia; TAVEC= Test de Aprendizaje Verbal España-Complutense; RL-CP= recuerdo libre a corto plazo; RL-LP= recuerdo libre a largo plazo.

Por último, en el TAVEC los resultados mostraron diferencias estadísticamente significativas entre los grupos estudiados para todas las variables: ensayo 1 [$\chi^2(2)= 7,89$; $p<0,048$], palabras recordadas en todos los ensayos [$\chi^2(2)= 16,87$; $p<0,001$], recuerdo libre a corto plazo [$\chi^2(2)= 19,75$; $p<0,000$], recuerdo libre a largo plazo [$\chi^2(2)= 24,42$; $p<0,000$], reconocimiento [$\chi^2(2)= 26,82$; $p<0,000$] y discriminabilidad [$\chi^2(2)= 32,33$; $p<0,000$] (ver Tabla 36).

Debido a la complejidad y al número de comparaciones realizadas a posteriori, se procedió a agrupar los resultados de la Tabla 25 atendiendo a los patrones de significación clínica. En primer lugar se comprobó en cuáles de las VDs el rendimiento del grupo NBC era similar al del grupo BC-NSS. Los resultados mostraron que la condición se cumplía para las siguientes variables: copia, recuerdo inmediato, recuerdo demorado, reconocimiento y falsas alarmas en el Test de la Figura Compleja de Rey; TR, TA, errores de omisión, errores de comisión, TOT, CON y VAR en el Test d2; fonética, animales y frutas en fluidez; el total del Letras y Números; el total del Cambios, la ejecución de la mano dominante y de la no-dominante en el Finger Tapping; palabras, colores, colores denominación de colores en la condición de interferencia y puntuación de interferencia en el STROOP; ensayo 1, total de los ensayos, recuerdo libre a corto plazo, reconocimiento y discriminabilidad en el TAVEC. Así, 28 de las 29 variables cumplen este patrón, lo que constituye el 96% (ver Tabla 37).

Tabla 37. Patrones esperados en los análisis a posteriori para los “grupos conocidos” en las pruebas neuropsicológicas no específicas de simulación

Patrón esperado	Variables que cumplen el patrón esperado	% de variables que cumplen el patrón esperado
NBC=BCNSS	<p>TFCR: copia, R. inmediato, R. demorado, reconocimiento, falsas alarmas</p> <p>Test d2: TR, TA, err o, err c, TOT, CON, VAR;</p> <p>Fluidez: fonética, semántica A, semántica F</p> <p>Letras y Números: total</p> <p>Cambios: total</p> <p>Finger Tapping: mano dominante, mano no-dominante</p> <p>STROOP: palabras, colores, PalxCol, Interferencia</p> <p>TAVEC: ensayo 1, total ensayos, RL-CP, reconocimiento, discriminabilidad</p>	96% (28/29)
BCSS=AN	<p>TFCR: copia, reconocimiento, falsas alarmas</p> <p>D2: err o, err c, CON, VAR;</p> <p>Fluidez: fonética, semántica A, semántica F</p> <p>Finger Tapping: mano dominante, mano no-dominante</p> <p>STROOP: palabras, colores</p> <p>TAVEC: total ensayos, RL-CP, RL-LP, reconocimiento, discriminabilidad</p>	65% (19/29)
BCNSS≠BCSS	<p>TFCR: copia</p> <p>D2: TR, TA, err c, TOT, CON</p> <p>Finger Tapping: mano no-dominante</p> <p>STROOP: palabras, colores, palxCol</p> <p>TAVEC: RL-LP, discriminabilidad</p>	41% (12/29)
(NBC=BCNSS) ≠ (BCSS=AN)	STROOP: palabras	3% (1/29)

Nota: BC= búsqueda de compensación; NBC= No búsqueda de compensación; BCNSS= búsqueda de compensación no sospechoso de simulación; BCSS= búsqueda de compensación sospechoso de simulación; TFCR= Test de la Figura Compleja de Rey; R= recuerdo; err o= errores de omisión; err c= errores de comisión; A= animales; F= frutas; PalxCol= palabras emitidas en la condición de interferencia; TAVEC= Test de Aprendizaje Verbal España-Complutense; RL-CP= recuerdo libre a corto plazo; RL-LP= recuerdo libre a largo plazo.

En segundo lugar, se comprobó en cuáles de las VD el rendimiento del grupo BC-SS era similar al del grupo AN. Los resultados mostraron que esta condición se cumplía para las variables: copia, reconocimiento y falsas alarmas de la Figura Compleja de Rey; errores de omisión, errores de comisión, CON y VAR del Test d2; total de los ensayos, recuerdo libre a corto plazo, recuerdo libre a largo plazo, reconocimiento y discriminabilidad del TAVEC, lo que constituye el 65% del total de las variables (19/29) (ver Tabla 37).

En tercer lugar, se procedió a comprobar cuáles de las VD eran capaces de diferenciar entre el grupo BC-NSS y el grupo BC-SS. Los resultados mostraron que las variables que eran capaces de diferenciar al grupo sospechoso de simulación eran las siguientes: copia de la Figura Compleja de Rey, TR, TA, errores de comisión, TOT y CON el Test d2; ejecución de la mano no-dominante del Finger Tapping; palabras, colores y denominación de colores en la condición de interferencia en el STROOP; recuerdo libre a largo plazo y discriminabilidad en el TAVEC. En total, 12 de las 29 variables, o lo que es lo mismo, el 41%, fueron capaces de distinguir entre el grupo de búsqueda de compensación a los individuos sospechosos de simulación (ver Tabla 37).

Por último, se comprobó cuáles de las VD cumplían todos los patrones esperados, esto es, qué variables identificaban al grupo NBC igual al grupo BC-NSS, pero estadísticamente diferentes de los grupos BC-SS y AN, iguales entre sí. Los resultados mostraron que tan sólo la variable palabras del STROOP cumplía este patrón, lo que constituye el 3% del total de las variables (ver Tabla 37).

Estudio de los tamaños del efecto:

Para realizar un análisis más preciso de la relevancia estadística de los hallazgos se calcularon los tamaños del efecto de los análisis anteriores empleando el estadístico Delta de Cohen. Mediante este análisis se puso de manifiesto que la variable que mejor

diferencia entre los grupos NBC y BCNSS se corresponde con el RL-CP ($\delta= 0,91$) y la que menos diferencia el recuerdo inmediato de la Figura Compleja de Rey ($\delta= 0,8$), consiguiendo el resto de las variables valores intermedios (ver Tabla 38).

Para diferenciar los grupos NBC y BCSS la variable discriminabilidad del TAVEC es la que se muestra superior ($\delta= 1,90$), y la variable que menos diferencia es la correspondiente a las falsas alarmas del mismo test ($\delta=0,32$). La diferenciación entre estos dos grupos es se encuentra entre estos dos valores para el resto de las variables (ver Tabla 38).

La variable reconocimiento del TAVEC es la que más diferencia entre los grupos NBC y AN ($\delta= 1,54$), y la variable falsas alarmas de la Figura Compleja de Rey la que menos diferencia entre ambos ($\delta= 0,00$), siendo la diferenciación entre estos dos grupos intermedia para el resto de las variables estudiadas (ver Tabla 38).

En cuanto a la diferenciación entre los grupos BCNSS y BCSS, la variable CON del Test d2 es la que muestra mejores resultados ($\delta= 1,87$), siendo la variable que menos diferencia entre estos grupos la variable VAR de la misma prueba ($\delta= -0,20$). En el resto de las variables la diferenciación obtiene valores intermedios (ver Tabla 38).

A la hora de diferenciar entre los grupos BCNSS y AN es la variable aciertos totales del Cambios la que mejores resultados obtiene ($\delta= -1,11$), siendo la variable que menos diferencia estos grupos la variable VAR del Test d2 ($\delta= 0,05$). Los valores del resto de las variables obtienen puntuaciones intermedias (ver Tabla 38).

Por último, la variable recuerdo inmediato de la Figura Compleja de Rey es la que más diferencia entre los grupos BCSS y AN ($\delta= -1,81$), y la variable reconocimiento del mismo test es la que menos diferencia estos grupos. Los valores han sido intermedios para el resto de las variables (ver Tabla 38).

Tabla 38. Delta de Cohen de las variables de los tests neuropsicológicos no específicos de simulación para los grupos conocidos.

Prueba	Variable	NBC vs. BCNSS	NBC vs. BCSS	NBC vs. AN	BCNSS vs. BCSS	BCNSS vs. AN	BCSS vs. AN	
FCR	Copia	0,467	0,778	0,251	0,150	0,314	0,600	
	R Inmediato	0,080	0,785	0,512	0,913	0,739	1,814	
	R Demorado	0,563	1,368	0,206	0,746	0,322	1,039	
	Reconocimiento	0,114	0,804	0,752	1,061	0,971	0,026	
	Falsas alarmas	0,176	0,329	0,007	0,273	0,159	0,307	
D2	TR	0,347	1,251	0,498	1,001	0,207	0,712	
	TA	0,155	1,843	0,845	1,788	0,736	0,996	
	Errores o	0,384	0,584	0,467	0,875	0,740	0,073	
	Errores c	0,464	0,366	0,618	0,779	0,866	0,351	
	TOT	0,230	1,615	0,904	1,476	0,722	0,751	
	CON	0,183	1,892	1,144	1,872	1,072	0,526	
	VAR	0,111	0,370	0,074	0,201	0,053	0,303	
	Fluidez	Fonética	0,263	0,527	0,024	0,251	0,311	0,606
		Sem. Animales	0,385	0,931	0,515	0,760	0,232	0,396
	Sem. Frutas	0,484	0,859	0,361	0,485	0,006	0,284	
LyN	Total	0,499	1,127	0,380	0,852	0,096	0,853	
Cambios	Total	0,491	0,683	0,517	0,261	1,110	1,270	
Finger T	Mano Dom.	0,253	1,160	0,652	0,905	0,407	0,469	
	Mano No-Dom	0,155	1,242	0,508	1,234	0,584	0,620	
STROOP	Palabras	0,269	1,419	0,890	1,138	0,614	0,523	
	Colores	0,280	1,165	0,482	0,977	0,242	0,681	
	PalxCol	0,217	1,446	0,355	1,463	0,199	1,025	
	Interferencia	0,022	0,669	0,325	0,679	0,296	1,062	
TAVEC	Ensayo 1	0,372	1,112	0,198	0,715	0,171	0,891	
	Total ensayos	0,597	1,550	0,944	0,930	0,421	0,389	
	RL-CP	0,911	1,567	1,069	0,776	0,305	0,391	
	RL-LP	0,664	1,826	1,335	1,026	0,663	0,260	
	Reconocimiento	0,841	1,544	1,639	0,946	0,915	0,140	
	Discriminabilidad	0,756	1,907	1,495	0,662	0,136	0,738	

Nota: BCNSS= búsqueda de compensación no sospechoso de simulación; BCSS= búsqueda de compensación sospechoso de simulación; NBC= no búsqueda de compensación; AN= análogos; FCR= Test de la Figura Compleja de Rey; R= recuerdo; o= omisión; c= comisión; Sem.= semántica; LyN= letras y números; Finger Tapp= Finger Tapping; PalxCol= palabras denominadas en la condición de interferencia; TAVEC= Test de Aprendizaje Verbal España-Complutense; RL-CP= recuerdo libre a corto plazo; RL-LP= recuerdo libre a largo plazo.

4.3. Sensibilidad, especificidad y puntos de corte en las pruebas no específicas de simulación (Hipótesis 3):

Una vez comprobado que algunas de las variables de los tests neuropsicológicos no específicos de simulación son capaces de diferenciar tanto a los grupos con una prevalencia de simulación diferente como a los sujetos sospechosos de simulación, se procedió a obtener los posibles puntos de corte para las distintas variables con sus respectivas sensibilidades y especificidades. Dicha obtención de los puntos de corte se

realizó sólo en aquellas variables para las que se encontraron diferencias estadísticamente significativas entre los grupos en los objetivos anteriores.

Como ya se ha mencionado en el Estudio 2, se empleará como criterio de estudio de los puntos de corte una especificidad superior al 90%. Para estudiar los puntos de corte que diferenciaban entre pacientes con TCE-L (tanto no buscadores de compensación como buscadores de compensación que no obtuvieron puntuaciones indicativas de simulación en las pruebas específicas) y el grupo sospechoso de simulación se realizaron 21 análisis de curvas ROC: 3 para la Figura Compleja de Rey, 5 para el Test d2, 1 para la fluidez, 1 para el Cambios, 2 para el Finger Tapping, 3 para el STROOP y 6 para el TAVEC.

Los resultados mostraron sensibilidades moderadas (70%) en las variables TR, TOT y CON del Test d2, con puntos de corte ≤ 283 , ≤ 248 y ≤ 78 , respectivamente, igual que la variable palabras del Test STROOP, con un punto de corte ≤ 66 (ver Tabla 39).

Ligeramente inferiores (60%) fueron las sensibilidades encontradas para las variables TA del Test d2 y discriminabilidad del TAVEC, con puntos de corte respectivos ≤ 82 y ≤ 77 (ver Tabla 39).

Sensibilidades bajas obtuvieron las variables colores (con un punto de corte ≤ 47) y palabras emitidas en la condición de interferencia (con un punto de corte ≤ 18) para el Test STROOP, y las variables del TAVEC recuerdo libre a corto plazo (con un punto de corte ≤ 6) y reconocimiento (con un punto de corte ≤ 11) (ver Tabla 39).

En el resto de las variables las sensibilidades encontradas fueron inaceptablemente bajas (ver Tabla 39).

Tabla 39. Sensibilidad y especificidad asociada a los puntos de corte de las distintas variables de los tests neuropsicológicos no específicos para diferenciar entre pacientes y sospechosos de simulación

Prueba	Variable	Punto de corte	Sensibilidad	IC 95%	Especificidad	IC 95%	Área ROC
FCR	R Inmediato	<=9,5	20,0	3,1- 55,6	90,9	78,3- 97,4	0,723
	R Demorado	<=9,5	40,0	12,4- 73,6	90,9	78,3- 97,4	0,757
	Reconocimiento	<=4	37,5	9,0- 75,3	93,9	79,7- 99,1	0,725
D2	TR	<=283	70,0	34,8- 93,0	90,2	76,9- 97,2	0,776
	TA	<=82	60,0	26,4- 87,6	95,1	83,4- 99,3	0,894
	Errores c	> 5	30,0	7,0- 65,2	90,2	76,9- 97,2	0,710
	TOT	<=248	70,0	34,8 - 93,0	90,2	76,9 - 97,2	0,856
	CON	<=78	70,0	34,8 - 93,0	92,7	80,1 - 98,4	0,910
Fluidez	Sem. Frutas	<=8	11,1	1,8- 48,3	90,9	78,3- 97,4	0,718
Cambios	Total	<=3	40,0	6,5 - 84,6	92,3	79,1 - 98,3	0,654
Finger Tapp	Mano Dom.	<=18,6	44,4	14,0 - 78,6	97,7	87,9 - 99,6	0,769
	Mano No-Dom	<=20,2	44,4	14,0 - 78,6	95,5	84,5 - 99,3	0,808
STROOP	Palabras	<=66	70,0	34,8 - 93,0	90,9	78,3 - 97,4	0,828
	Colores	<=47	50,0	18,9 - 81,1	90,9	78,3 - 97,4	0,785
	PalxCol	<=18	50,0	18,9 - 81,1	100,0	91,9 - 100,0	0,848
TAVEC	Ensayo 1	<=3	10,0	1,7 - 44,5	93,2	81,3 - 98,5	0,760
	Total ensayos	<=38	50,0	18,9 - 81,1	90,9	78,3 - 97,4	0,831
	RL-CP	<=6	50,0	18,9 - 81,1	93,2	81,3 - 98,5	0,820
	RL-LP	<=5	40,0	12,4 - 73,6	93,2	81,3 - 98,5	0,857
	Reconocimiento Discriminabilidad	<=11	50,0	18,9 - 81,1	97,7	87,9 - 99,6	0,793
		<=77	60,0	26,4 - 87,6	93,2	81,3 - 98,5	0,916

Nota: IC= intervalo de confianza; FCR= Test de la Figura Compleja de Rey; R= recuerdo; Sem.= semántica; Finger Tapp= Finger Tapping; Dom.= dominante; No-Dom= no dominante; PalxCol= palabras emitidas en la condición de interferencia; TAVEC= Test de Aprendizaje Verbal España-Complutense; RL-CP= recuerdo libre a corto plazo; RL-LP= recuerdo libre a largo plazo.

Para terminar, con la finalidad de dotar a nuestro hallazgos de una mayor utilidad clínica, se procedió a obtener los valores predictivos positivos (VPP) y los valores predictivos negativos (VPN) asociados a distintas prevalencias de base (“*base rate*”) de simulación para las principales variables que pueden emplearse como índices de simulación en cada test neuropsicológico no específico de simulación empleado en este estudio.

Siguiendo la estructura del estudio 2, se han como prevalencias de base 15%, 30% y 45%, por las razones expuestas en el capítulo anterior. Para cada una de estas prevalencias se proponen varios puntos de corte, de modo que el neuropsicólogo pueda decidir entre mayor o menor certeza a la hora de emitir un juicio clínico (ver Anexo 2).

4.4. Estabilidad de los hallazgos con las curvas ROC

Igual que en el estudio 2, debido a que el número de sospechosos de simulación (grupo BCSS) era reducido ($n=10$), se procedió a realizar un nuevo análisis para estudiar la estabilidad de los hallazgos encontrados con las curvas ROC en un grupo de mayor tamaño como es el grupo de análogos ($n=54$). Para este nuevo análisis se seleccionaron aquellas variables que había demostrado en los análisis anteriores que eran estadísticamente iguales entre el grupo de BCSS y el grupo de análogos (AN) y que, al mismo tiempo, eran diferentes a los grupo de NBC y BCNSS, patrón que sólo cumplió la variable número de palabras leídas del test STROOP. De esta forma, se podría comprobar en un grupo de mayor tamaño (AN), que se comportaba clínicamente igual que el grupo de simuladores, si los puntos de corte, sensibilidad, especificidad y área bajo la curva ROC encontrados para dicho grupo eran estables.

Los resultados mostraron que los puntos de corte y la especificidad eran idénticos para los grupos BCSS y AN (ver Tabla 40). Sin embargo, han aparecido

ciertas diferencias en cuanto a la sensibilidad y el área bajo la curva ROC en ambos grupos (ver Tabla 40).

Tabla 40. Puntos de corte, sensibilidad, especificidad y área bajo la curva ROC de los grupos BCSS y AN.

Prueba	Variable	Punto de corte		Sensibilidad		Especificidad		Área ROC	
		BCSS n=10	AN N=27	BCSS n=10	AN N=27	BCSS n=10	AN N=27	BCSS n=10	AN N=27
STROOP	Palabras	<=66	<=66	70,0	40,7	90,9	90,9	0,828	0,741

Nota: Área ROC= área bajo la curva ROC; BCSS= búsqueda de compensación sospechosos de simulación; AN= análogos; Combo= puntuación combinada; Errores c= errores de comisión; VSVT= Victoria Symptom Validity Test.

4.5. La SPECT-PCR en la detección de simulación (Hipótesis 4)

Para comprobar si existían diferencias en el resultado de la SPECT-PCR de los sujetos simuladores y los no simuladores se efectuó un análisis de ANOVA univariada siendo la VI el grupo (NBC, BCSS y AN) y la VD el resultado de la cuantificación de la SPECT-PCR de los sujetos. Los resultados mostraron que existían diferencias estadísticamente significativas entre los resultados de la neuroimagen de los tres grupos [$F(2,61)= 5,07; p=0,009$], mostrando los análisis a posteriori que estas diferencias se limitan a los grupos NBC y AN, encontrándose en estos últimos un menor nivel de hallazgos en la prueba de neuroimagen (ver Tabla 41).

Tabla 41. Medias, desviaciones típicas y análisis a posteriori de la prueba de neuroimagen para los diferentes grupos

	Grupo NBC		Grupo BCSS		Grupo AN		p	Bonferroni
	Media	Dt	Media	Dt	Media	Dt		
SPECT-PCR	1,71	1,41	1,66	1,22	0,74	0,90	0,009	BCSS=NBC BCSS=AN NBC>AN

Nota: SPECT-PCT= Tomografía por Emisión de Fotón Único de Perfusión Cerebral Regional; NBC= no búsqueda de compensación; BCSS= búsqueda de compensación sospechoso de simulación; AN= análogos.

A continuación presentamos una SPECT-PCR a modo de ejemplo de los hallazgos típicos que hemos encontrado en los pacientes con síndrome post-conmocional y otro perteneciente a un análogo (ver Figura 13 y 14).

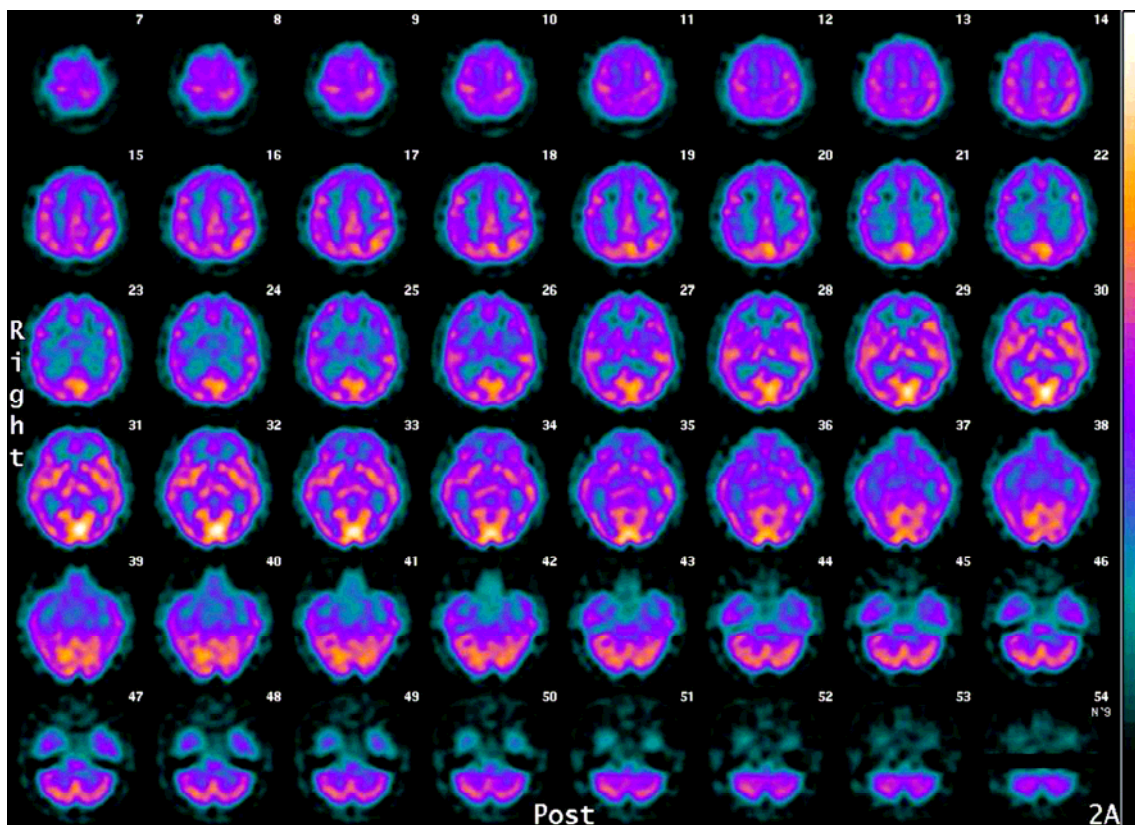


Figura 13. Ejemplo de una exploración tomográfica (SPECT) cerebral de perfusión regional de un paciente en la que se aprecia una hipocaptación generalizada del trazador.

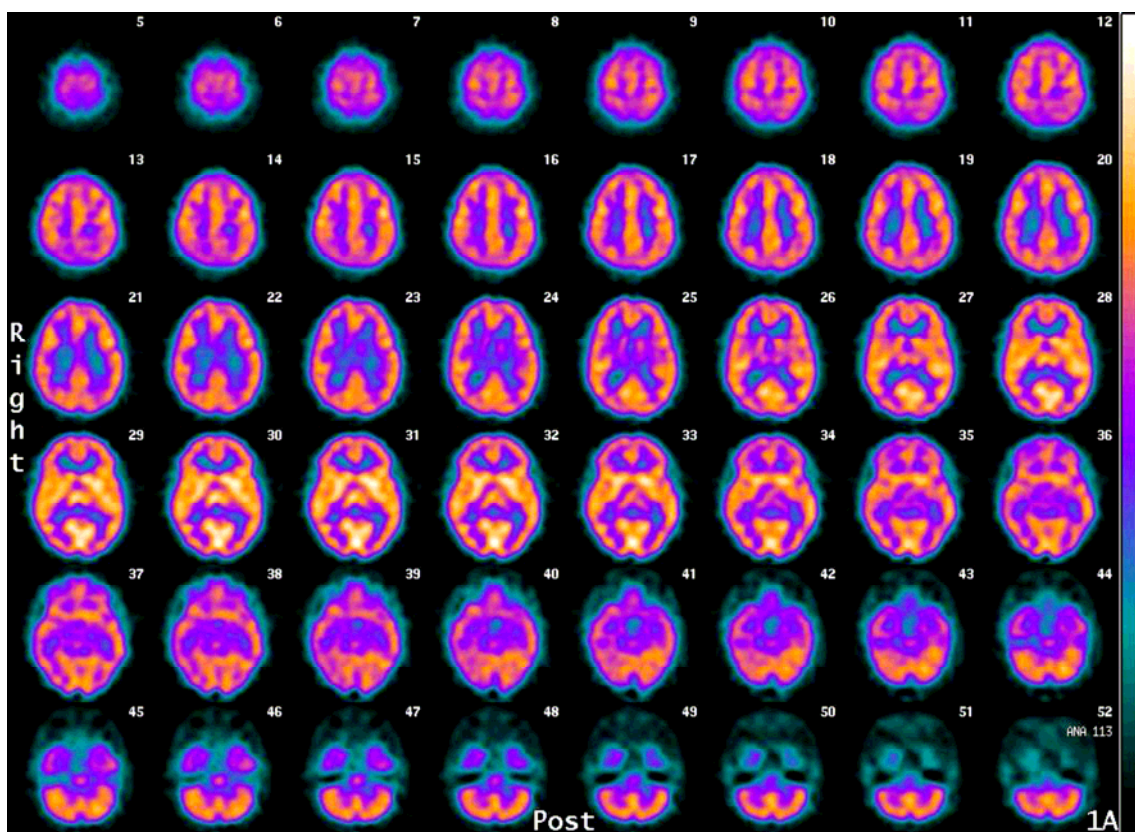


Figura 14. Ejemplo de una exploración tomográfica (SPECT) cerebral de perfusión regional de un análogo en la que no se aprecian hallazgos en la captación del trazador.

4.6. Las medidas psicofisiológicas asociadas a la ejecución de una prueba específica de simulación (Hipótesis 5).

1. Conductancia Eléctrica de la Piel.

Para comprobar si existían diferencias en la conductancia eléctrica de la piel cuando los sujetos simulaban se realizaron 4 Análisis de Varianzas de medidas repetidas, empleando en todos los casos la corrección de Greenhouse-Geisser.

El primer ANOVA, con un factor de medidas repetidas (Aciertos= cuando la respuesta del VSVT es correcta; Errores= cuando la respuesta es incorrecta; y Control= la respuesta de los participantes en los ensayos en los que no simulaban) se realizó para la selección temporal que hemos llamado Tiempo de Reacción, que era el tiempo que cada participante tardaba en responder. Las medias y desviaciones típicas las podemos ver en la tabla 42. Los resultados mostraron que existían diferencias estadísticamente significativas para las tres condiciones [$F(2, 54)=13,43; p=0,000$].

Tabla 42. Medias y desviaciones típicas de la Conductancia (microSiemens) en el espacio temporal de TR

	Media	Dt
ACIERTOS	0,1890	0,13865
ERRORES	0,2684	0,26293
CONTROL	0,0956	0,05803

Nota: Dt= desviación típica

Los análisis a posteriori mostraron que existen diferencias significativas entre los Aciertos y los Errores, entre los Aciertos y el Control y entre los Errores y el Control (ver tabla 43).

Tabla 43. Análisis post hoc para la conductancia en el espacio temporal de TR

	Diferencia entre medias	Significación (ajuste de Bonferroni)
ACIERTOS-ERRORES	-0,079(*)	0,047
ACIERTOS-CONTROL	0,093(*)	0,001
ERRORES-CONTROL	0,173(*)	0,002

* La diferencia de las medias es significativa al nivel .05

La figura 15 representa los resultados de las pruebas de contrastes intrasujetos, indicando que existen efectos significativos para el componente lineal y cuadrático, respectivamente [$F(1,27)=18,19; p=0.000$], [$F(1,27)=12,12; p=0,002$].

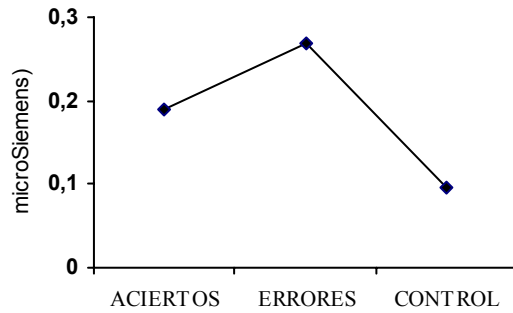


Figura 15 Conductancia máxima para Aciertos, Errores y Control en el espacio temporal de TR

Con la finalidad de facilitar la comprensión de estos datos, a continuación mostramos un ejemplo de respuesta correcta e incorrecta en la condición de simulación, y un ejemplo de respuesta en la condición de no-simulación (control) (ver Figura 16).

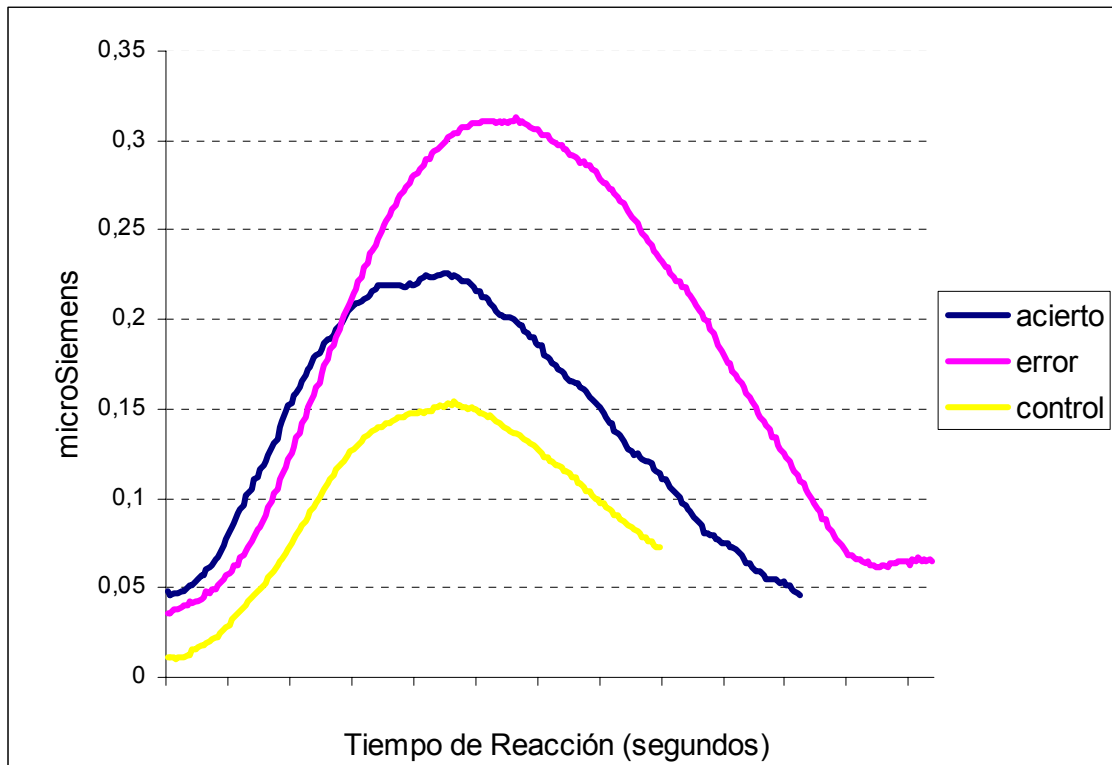


Figura 16. Ejemplo de la respuesta psicofisiológica de conductancia para un ítem acertado y uno errado en la condición de simulación y para un ítem de la condición control.

El segundo Análisis de Varianza de medidas repetidas se realizó para el espacio temporal que hemos denominado 2 segundos (que es el tiempo que transcurre desde que los participantes responden hasta que aparece un nuevo estímulo objetivo). El factor de medidas repetidas tenía los mismos valores que para el análisis anterior, Aciertos, Errores y Control. Las medias y las desviaciones típicas podemos verlas en la tabla 44. En este análisis no se encontraron diferencias estadísticamente significativas [$F(2, 54)=1,804$; $p=0,181$].

Tabla 44. Medias y desviaciones típicas de la Conductancia (microSiemens) en el espacio temporal de 2 segundos

	Media	Dt
ACIERTOS	0,1576	0,11772
ERRORES	0,2090	0,23168
CONTROL	0,1725	0,12281

Nota: Dt= desviación típica

El tercer Análisis de Varianza para la Conductancia se realizó en el espacio temporal denominado Tiempo de Reacción, pero en este caso el factor analizado tenía los siguientes niveles: Conductancia máxima para los Aciertos, conductancia máxima para los Errores y conductancia máxima en el IEE (Intervalo Entre. Estímulos o tiempo que transcurre entre la aparición del objetivo y las alternativas de respuesta). En la tabla 45 se aprecian las medias y desviaciones típicas para cada uno de los niveles, no encontrándose diferencias estadísticamente significativas entre ellos [$F(2, 54)=3,33$; $p=0,064$].

Tabla 45 Medias y desviaciones típicas de la Conductancia (microSiemens) en el espacio temporal de TR

	Media	Dt
ACIERTOS	0,1890	0,13865
ERRORES	0,2684	0,26293
IEE	0,1840	0,11689

Nota: Dt= desviación típica

La figura 17 representa los resultados de las pruebas de contrastes intrasujetos, indicando que no existen efectos significativos para el componente lineal [$F(1,27)=0,031$; $p=0,862$] y sí para el componente cuadrático [$F(1,27)=4,76$; $p=0.038$].

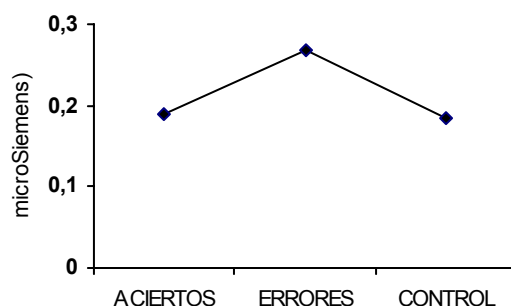


Figura 17 Conductancia máxima para Aciertos, Errores e IEE en el espacio temporal de TR

El último análisis que realizamos para la variable Conductancia fue un Análisis de Varianza de medidas repetidas en el espacio temporal de 2 segundos. En este caso, el factor tenía los mismos niveles que en el análisis anterior: Conductancia máxima para los Aciertos, conductancia máxima para los Errores y conductancia máxima en el IEE. En la tabla 46 podemos apreciar las medias y desviaciones típicas para cada uno de los niveles, entre los que no se encontraron diferencias estadísticamente significativas [$F(2, 54)=0,869$; $p=0,376$]

Tabla 46. Medias y desviaciones típicas de la Conductancia (microSiemens) en el espacio temporal de 2 segundos

	Media	Dt
ACIERTOS	0,1576	0,11772
	0,2090	0,23168
IEE	0,1840	0,11689

Nota: Dt= desviación típica

2. Pulso sanguíneo.

Para comprobar si existían diferencias en el pulso sanguíneo cuando los sujetos simulaban se realizaron ocho Análisis de Varianzas de medidas repetidas, empleando

como variables dependientes la Tasa (Pulso por minuto, en inglés, BPM) y la Altura Cíclica. En todos los casos utilizamos la corrección de Greenhouse-Geisser.

El primer análisis consistió en un ANOVA, con un factor de medidas repetidas (Aciertos= cuando la respuesta del VSVT es correcta; Errores= cuando la respuesta es incorrecta; y Control= la respuesta de los participantes en los ensayos en los que no simulaban) para la variable dependiente Tasa. Este análisis se realizó para el espacio temporal Tiempo de Reacción, que era el tiempo que cada participante tardaba en responder. Las medias y desviaciones típicas las podemos ver en la tabla 47. No se encontraron diferencias estadísticamente significativas en este análisis [$F(2, 52)=1,21$; $p=0,289$]

Tabla 47. Medias y desviaciones típicas de la Tasa (BPM) en el espacio temporal de TR

	Media	Dt
ACIERTOS	76,0775	10,22149
ERRORES	76,4496	9,99703
CONTROL	75,0858	12,37456

Nota: Dt= desviación típica

A continuación realizamos el mismo análisis que acabamos de describir con la variable dependiente Altura cíclica. El ANOVA de medidas repetidas tenía un factor con tres niveles: Aciertos, Errores y Control, encontrándose diferencias estadísticamente significativas [$F(2, 52)=15,83$; $p=0,000$]. En la tabla 48 aparecen las medias y desviaciones típicas de esta variable.

Tabla 48 Medias y desviaciones típicas de la Altura cíclica (Voltios) en el espacio temporal de TR

	Media	Dt
	3,0297	1,31393
ERRORES	3,0032	1,33505
	2,4302	1,27488

Nota: Dt= desviación típica

Los análisis a posteriori mostraron que existían diferencias estadísticamente significativas entre los Aciertos y el Control y entre los Errores y el Control (ver tabla 49).

Tabla 49. Análisis post hoc para la Altura cíclica en el TR

	m	Significación (ajuste de Bonferroni)
ACIERTOS-ERRORES	0,27	1
	0,600(*)	0,000
ERRORES-CONTROL	0,573(*)	0,003

* La diferencia de las medias es significativa al nivel .05

La figura 18 representa los resultados de las pruebas de contrastes intrasujetos, indicando que existen efectos significativos tanto para el componente lineal [$F(1,26)=20,319$; $p=0,000$] como para el componente cuadrático [$F(1,26)=8,812$; $p=0,006$].

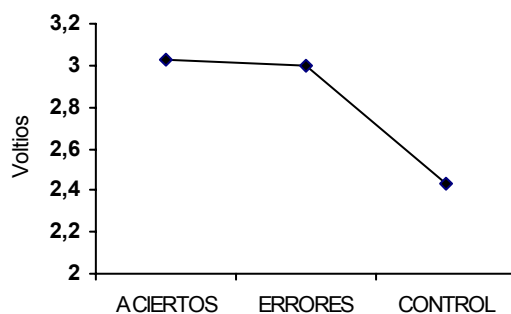


Figura 18. Altura cíclica para Aciertos, Errores y Control en el TR

El tercer ANOVA de medidas repetidas lo realizamos para la variable dependiente Tasa (BPM) en el espacio temporal de 2 segundos. Los niveles de la variable fueron Aciertos, Errores y Control, no encontrándose diferencias estadísticamente significativas [$F(2, 52)=2,403$; $p=0,132$] En la tabla 50 podemos apreciar las medias y desviaciones típicas de esta variable.

Tabla 50. Medias y desviaciones típicas de la Tasa (BPM) en el espacio temporal de 2 segundos.

	Media	Dt
	76,1431	10,13131
ERRORES	76,2132	10,42190
	71,6051	17,49257

Nota: Dt= desviación típica

El cuarto ANOVA lo realizamos para la variable dependiente Altura cíclica en el espacio temporal de 2 segundos y con los niveles de la variable: Aciertos, Errores y Control. En la tabla 51 aparecen las medias y desviaciones típicas de esta variable, para la que se encontraron diferencias estadísticamente significativas [$F(2, 52)=13,57$; $p=0,021$]

Tabla 51. Medias y desviaciones típicas de la Altura cíclica (Voltios) en el espacio temporal de 2 segundos

	Media	Dt
ACIERTOS	3,0306	1,28985
ERRORES	2,9595	1,28955
CONTROL	2,1176	2,32863

Nota: Dt= desviación típica

Los análisis a posteriori con el ajuste de Bonferroni mostraron que existían diferencias marginalmente significativas entre los Aciertos y el Control, con un grado de significatividad de 0,54.

En la figura 19 podemos ver la representación gráfica de las medias para cada una de las condiciones experimentales. Los resultados de las pruebas de contrastes intrasujetos indican que existen efectos significativos en el componente lineal [$F(1,26)=11,225$; $p=0,018$] y en el componente cuadrático [$F(1,26)=2,674$; $p=0,0452$].

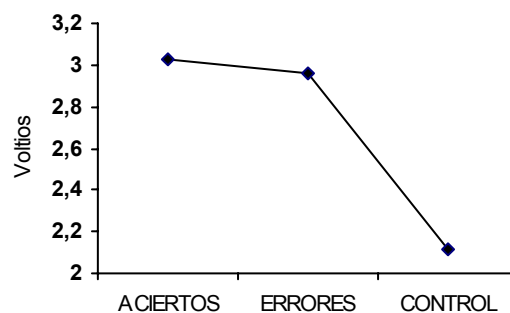


Figura 19. Altura cíclica para Aciertos, Errores y Control en 2 segundos

Para el quinto análisis realizamos un ANOVA de medidas repetidas en el espacio temporal que hemos denominado Tiempo de Reacción. La variable dependiente que analizamos fue el pulso por minuto o Tasa, con siguientes niveles de la variable dependiente: Aciertos, Errores e Intervalo Entre Estímulos (IEE). Las medias y las desviaciones típicas aparecen en la tabla 52, mostrando los resultados del ANOVA que no existían diferencias estadísticamente significativas [$F(2, 54)=0,047$; $p=0,836$].

Tabla 52. Medias y desviaciones típicas de la Tasa (BPM) para los Aciertos, Errores e IEE en el espacio temporal TR

	Media	Dt
ACIERTOS	75,4330	10,594432
ERRORES	75,7439	10,496922
IEE	76,1694	10,87083

Nota: Dt= desviación típica

El sexto ANOVA que realizamos fue para la variable Altura cíclica, en el espacio temporal de que hemos denominado Tiempo de Reacción. Las medias y las desviaciones típicas para las condición de Aciertos, Errores e IEE podemos verlas en la tabla 53, mostrando el análisis de varianza que no existen diferencias estadísticamente significativas entre estas condiciones para la Altura Cíclica [$F(2, 54)=0,356$; $p=0,562$].

Tabla 53. Media y desviaciones típicas de la Altura cíclica (Voltios) para los Aciertos, Errores e IEE en TR

	Media	Dt
ACIERTOS	3,0575	1,29776
ERRORES	3,0195	1,31292
IEE	2,8609	1,25641

Nota: Dt= desviación típica

El séptimo ANOVA de medidas repetidas se realizó para la variable Tasa (BPM) en la condición que hemos denominado 2 segundos. Las medias y desviaciones típicas podemos observarlas en la tabla 54. Los resultados mostraron que no existían diferencias estadísticamente significativas para esta variable [$F(2, 54)=0,052$; $p=0,830$].

Tabla 54. Media y desviaciones típicas en la Tasa (BPM) para los Aciertos, Errores e IEE en 2 segundos

	Media	Dt
ACIERTOS	75,4547	10,5881
ERRORES	75,5329	10,8421
IEE	76,1694	10,8708

Nota: Dt= desviación típica

El último ANOVA de medidas repetidas que realizamos fue para la variable Altura cíclica, en el espacio temporal de 2 segundos. Las medias y desviaciones típicas podemos verlas en la tabla 55. Los resultados mostraron que no existen diferencias estadísticamente significativas entre Aciertos, Errores e IEE [$F(2, 54)=0,394$; $p=0,542$].

Tabla 55. Medias y desviaciones típicas de la Altura cíclica (Voltios) para los Aciertos, Errores e IEE en 2 segundos

	Media	Dt
ACIERTOS		1,29107
ERRORES	3,0209	1,30507
IEE	2,8609	1,25641

Nota: Dt= desviación típica

5. Discusión:

En primer lugar, en este estudio 3 hemos analizado la utilidad de las pruebas neuropsicológicas no específicas en la detección de simulación, siguiendo el mismo esquema empleado en el estudio 2 para las pruebas específicas. Por tanto, empleamos un diseño de grupos de prevalencia diferencial y a continuación un diseño de grupos conocidos. Los resultados mostraron que, en general, las variables de las distintas pruebas neuropsicológicas diferenciaban entre los diversos grupos de estudio, si bien los tamaños del efecto y los análisis a posteriori pusieron de manifiesto que las variables no eran capaces de diferenciar a los sujetos simuladores de los no simuladores. El establecimiento de puntos de corte correspondientes a una especificidad superior al 90% llevó asociadas unas sensibilidades generales muy pobres, especialmente si son comparadas por las obtenidas en el estudio 2 con las pruebas específicas. No obstante,

algunas variables del Test d2 y la variable número de palabras leídas del test STROOP han obtenido resultados satisfactorios, y que además se han mostrado estables.

En segundo lugar, hemos comprobado la utilidad del empleo de distintas variables psicofisiológicas de conductancia y pulso registradas durante la ejecución de una prueba específica (el VSVT) en la detección de simulación. Los resultados han mostrado que la conductancia máxima puede ser empleada para diferenciar con éxito cuándo un sujeto está simulando o está poniendo su máximo esfuerzo en la realización de la tarea, mientras las variables de pulso no han sido tan concluyentes.

En tercer y último lugar, hemos analizado los resultados de la SPECT-PCR para ver si nos permitían identificar un patrón de simulación. La obtención de dicho patrón no ha sido posible, si bien la ausencia de diferencias entre los pacientes sospechosos de simulación y los pacientes no simuladores constituye un apoyo más a las diferencias encontradas entre los grupos en las pruebas neuropsicológicas, que no pueden ser explicadas por diferencias relativas al daño cerebral entre los grupos.

Diseño de grupos de prevalencia diferencial

A la hora de diferenciar entre el grupo de pacientes buscadores de compensación (BC), no buscadores de compensación (NBC) y análogos (AN) los tests neuropsicológicos no presentan demasiadas dificultades, ya que 17 de las 29 variables diferencian entre dichos grupos. Parece que la clave de las variables que diferencian entre los grupos de prevalencia diferencial está en si estas son o no medidas de memoria, puesto que el 90% de las variables de memoria son capaces de diferenciar entre los grupos, mientras sólo el 47% de las variables de las otras áreas lo consiguen. Esto está en consonancia con los dos únicos trabajos que estudian este aspecto en profundidad, que concluyen que los problemas de memoria son el síntoma más frecuentemente fingido (Iverson, 1995; Tan y cols, 2002). Es más, sólo en 8 de las

variables se encuentra el patrón esperado, es decir, una ejecución superior del grupo NBC con respecto a los grupos BC y AN, que se comportan de modo similar, y de estas 8 variables, 5 pertenecen al TAVEC, por lo que cabría hipotetizar que las pruebas de memoria verbal sean superiores a la hora de identificar sujetos que buscan una compensación económica y análogos, si bien esta hipótesis necesita ser probada.

En todas las variables estudiadas se encuentra una peor ejecución del grupo buscador de compensación con respecto al grupo de pacientes no buscadores de compensación, si bien estas diferencias no suelen ser estadísticamente significativas. Esta tendencia al menor rendimiento entre los sujetos que buscan una compensación económica está bien documentada en la literatura (King y cols., 2002; Langeluddecke y Lucas, 2003; Meyers y cols., 1999; Meyers y Volbrecht, 2003; Mittenberg y cols., 1995; Ross y cols., 2003), de lo que suele inferirse que las pruebas neuropsicológicas pueden ser útiles como indicadoras de simulación. Sin embargo, en nuestro estudio las diferencias entre los dos grupos de pacientes (NBC y BC) son mínimas, como lo indica el tamaño del efecto, lo cual podría ser debido a diversos motivos. En primer lugar, puede que en el grupo de pacientes buscadores de compensación no existan simuladores, lo que nos parece realmente poco probable después de lo expuesto en el estudio 2. En segundo lugar, la ausencia de diferencias podría ser debida a que nos hemos centrado en variables aisladas y no en patrones de ejecución, como es frecuente encontrarse en este tipo de investigaciones (Meyers y Volbrecht, 1999, 2003; Meyers y cols., 1999), si bien en los índices por separado también suelen encontrarse diferencias entre estos grupos (Langeluddecke y Lucas, 2003; Meyers y cols., 2001). Por último, la ausencia de diferencias entre los grupos de pacientes buscadores y no buscadores de compensación puede ser debida a que las pruebas neuropsicológicas no sean sensibles a la búsqueda de compensación económica, como se ha encontrado en algún otro estudio

(O'Bryant y cols., 2003). En cualquier caso, la ausencia de diferencias entre estos grupos implica que los sujetos que buscan compensación no están más dañados cognitivamente, lo que constituye un apoyo más a las pruebas específicas de simulación vistas en el capítulo anterior, ya que las diferencias encontradas entre los grupos no pueden ser explicadas por la existencia de un mayor deterioro cognitivo en el grupo que persigue un incentivo económico.

Por otro lado, los tamaños del efecto también demuestran que las diferencias del grupo de análogos con respecto al grupo búsqueda de compensación son mínimas, como era esperable. Sin embargo, tampoco aparecen diferencias entre el grupo de análogos y el grupo de pacientes que no buscan compensación. Esto podría ser debido a que los análogos sean tan buenos simuladores que no puedan ser detectados, si bien esta hipótesis queda directamente descartada puesto que en el estudio 1 eran descubiertos sin problemas por las pruebas específicas de simulación. Más bien, la ausencia de diferencias entre los análogos y los pacientes (sean estos buscadores de compensación o no) nos parece debida a que las pruebas neuropsicológicas no específicas no son sensibles a la simulación, sino al daño cognitivo. En este sentido, es preciso recordar que la mayor parte de la investigación sobre indicadores de simulación en los tests neuropsicológicos se sustenta en el estudio de análogos frente a controles, diferenciación fácilmente realizada por las diferentes pruebas. Sin embargo, los problemas surgen a la hora de distinguir entre análogos y pacientes (Suhr y Gunstad, 2000). Nuestra hipótesis es que, dado que las pruebas neuropsicológicas miden habilidades, que estarán reducidas en los pacientes con daño cerebral, y la estrategia de los análogos es realizar una ejecución reducida en los tests, la puntuación de ambos grupos es similar. Quizá si hubiésemos seleccionado variables en las que no es esperable un déficit dado el tipo de daño estudiado hubiesen emergido las diferencias

esperables entre el grupo de pacientes no buscadores de compensación y los análogos. Si bien en nuestros hallazgos la hipótesis se vería confirmada por los resultados del reconocimiento en el TAVEC, no sucede lo mismo con el reconocimiento en la Figura Compleja de Rey. Por todo ello, nos parece interesante una futura investigación que permita confirmar o no dicha hipótesis.

La prueba que parece ser la excepción es el TAVEC (concretamente sus variables recuerdo a corto y a largo plazo, reconocimiento y discriminabilidad) ya que es la única que obtiene un tamaño del efecto adecuado para diferenciar entre NBC y BC por un lado, y NBC y AN por otro, no encontrando diferencias entre los grupos BC y AN. Como hemos comentado con anterioridad, cabe plantearse la posibilidad de que la memoria verbal sea un dominio especialmente sensible a la simulación.

Diseño de grupos conocidos

Por otro lado, en cuanto a los “grupos conocidos”, 21 de las 29 variables muestran diferencias estadísticamente significativas entre los pacientes no buscadores de compensación (NBC), los buscadores de compensación no sospechosos de simulación (BCNSS), los sospechosos de simulación (BCSS) y los análogos (AN). De las 8 variables que no diferencian entre estos grupos la mitad son medidas frontales, por lo que quizá, en las pruebas neuropsicológicas no específicas, las variables menos útiles para detectar a posibles simuladores son aquellas que pertenecen a áreas complejas, como las medidas de la función ejecutiva. Siguiendo esta lógica, las variables más útiles serían aquellas que se basan en los dominios y habilidades más simples y por tanto más resistentes al daño cerebral. Si bien esto es puramente hipotético, y ha de ser estudiado en futuras investigaciones, está en consonancia con las teorías que rigen el desarrollo de los tests específicos de simulación, por lo que no parece una idea descabellada. Esta hipótesis no coincide con la propuesta en el apartado anterior, en la que se afirmaba que

las mejores variables para detectar simulación son aquellas que se basan en medidas de memoria, por lo que sería necesario seguir investigando para dilucidar cuál de las dos alternativas es la correcta.

A pesar de las diferencias estadísticamente significativas encontradas entre los grupos, los análisis a posteriori no corroboran los patrones previstos. Es más, sólo una de las 29 variables (es decir, el 3%) cumple todos los criterios esperados, de modo que permite diferenciar a los grupos NBC y BCNSS (sin diferencias entre ellos) de los grupos BCSS y AN (también iguales entre ellos). Esta variable es el número de palabras leídas en el Test STROOP, que al ser una medida de la velocidad de procesamiento, y por tanto básica y sencilla, estaría en consonancia y sería un apoyo a la hipótesis que acabamos de proponer. Por otro lado, hay que recordar que frente a este 3%, el porcentaje de variables que cumplía la totalidad de los patrones entre las pruebas específicas era del 39%, mostrándose por tanto claramente superiores a las no específicas.

Especialmente notable nos parece el hecho de que la variable VAR del test d2 no encuentre diferencias estadísticamente significativas entre los grupos, ya que es la única variable de todas las estudiadas de los tests neuropsicológicos no específicos propuesta en el manual como posible indicador de una falta de motivación hacia la prueba por parte de los sujetos. No obstante, en el propio manual se indica que esta es una de las variables menos fiables del test (Brickenkamp, 1962), lo que explicaría sus pobres resultados.

A pesar de lo dicho hasta el momento, un detenido análisis de las puntuaciones nos revela que las peores puntuaciones corresponden, en todas las variables, al grupo BCSS. Por tanto, sí existe una tendencia entre los sujetos sospechosos de simulación a obtener puntuaciones inferiores a los pacientes en las variables neuropsicológicas, como

ha sido señalado en otras investigaciones (Arnold y cols., 2005; Curtis y cols., 2006; Langeluddecke y Lucas, 2003; Lu y cols., 2003), si bien estas diferencias en la ejecución no son suficientes para distinguirlos, como lo demuestra también el análisis del tamaño del efecto. Mediante este análisis podemos comprobar que, en la línea de lo esperado, las variables no encuentran diferencias entre los grupos de pacientes no simuladores (NBC y BCNSS). Así, como ya ha sido comentado en el estudio anterior, parece que las variables relacionadas con los procesos de litigio no afectan a los sujetos lo suficiente como para hacer que se modifiquen sus puntuaciones en las medidas cognitivas. Por otro lado, el grupo BCSS se comporta igual que el grupo AN, y de modo diferente al grupo NBC, lo que apuntaría en la línea de que las pruebas neuropsicológicas parecen ser capaces de identificar a los simuladores. Sin embargo, el grupo AN se comporta igual que los dos grupos de pacientes no simuladores (NBC y BCNSS), y además no se encuentran diferencias entre los dos grupos de búsqueda de compensación, sean o no sospechosos de simulación (es decir, hayan puntuado como simuladores o no en las pruebas específicas). Todos estos resultados parecen estar indicando que las pruebas neuropsicológicas no específicas no son lo suficientemente sensibles a la simulación y que los tests neuropsicológicos poseen poca habilidad para diferenciar a sospechosos de simulación de pacientes no simuladores (Greiffenstein y cols., 1994).

Además, los tamaños del efecto nos muestran que este tipo de pruebas se comportan mejor a la hora de detectar sospechosos de simulación que a la hora de detectar análogos. Esto puede ser debido a diversos motivos: por un lado, podría ser que los resultados obtenidos con el grupo AN no sean generalizables, si bien después de lo encontrado en el capítulo anterior no nos parece probable. Por otro lado, podría ser que el conocimiento neuropsicológico de nuestros análogos los dotase de la capacidad para

realizar las pruebas neuropsicológicas como verdaderos pacientes, de modo que no podrían ser detectados como simuladores. Sin embargo, esto no impidió que estos sujetos fueran detectados por las pruebas específicas, por lo que nuevamente queda patente su superioridad. En este sentido, hay que recordar que, como ya hemos mencionado previamente, es necesario el apoyo convergente de los resultados de diseños con análogos y de grupos conocidos para admitir la validez de una medida de simulación (Rogers, 1997^b), convergencia que no se produce con las pruebas neuropsicológicas no específicas de simulación, por lo que podemos concluir que las variables neuropsicológicas empleadas en nuestro estudio no gozan del apoyo necesario para ser empleadas como índices de simulación.

Sensibilidad, especificidad y puntos de corte

A pesar de estos resultados globalmente negativos, algunas variables de las pruebas permiten una adecuada clasificación de los sospechosos de simulación. En concreto, algunas de las variables del Test d2 (TR, TOT y CON) y del Test STROOP (palabras) consiguen una sensibilidad del 70%, manteniendo una especificidad superior al 90%. Otra prueba que obtiene unos resultados aceptables, aunque inferiores a estos, es el TAVEC. Para todos estos casos, el área obtenida bajo la curva ROC se situó en torno a 0,8, lo que confirma la validez de los datos.

Por otro lado, los resultados que hemos obtenido con las curvas ROC parecen estables, ya que al repetir los análisis con un grupo mayor (el grupo de análogos) los puntos de corte obtenidos, así como las especificidades y el área bajo la curva ROC eran idénticos o muy similares. Por tanto, esta ausencia de diferencias entre el grupo sospechoso de simulación y el grupo de análogos nos hace concluir que nuestros hallazgos son consistentes a pesar de las diferencias encontradas en la sensibilidad, que ya han sido comentadas en el estudio anterior.

En general, nuestros resultados están en la línea de lo encontrado en el trabajo de O'Bryant y cols. (2004), que afirman que existe poca evidencia empírica que sustente el empleo de puntos de corte en los tests neuropsicológicos para su empleo en la detección de simulación.

A pesar de esto, en la literatura sobre las pruebas neuropsicológicas no específicas suelen encontrarse diferencias entre los sujetos que buscan compensación y los que no la buscan (Meyers y cols., 2001; Meyers y Volbrecht, 1999; 2003; Ross y cols., 2003), así como en los diseños que emplean grupos conocidos (Arnold y cols., 2005; Curtis y cols., 2006; Larrabee, 2003^a; Langeluddecke y Lucas, 2003; Lu y cols., 2003) o cuando se emplean diseños de análogos (Demakis, 1999; Forrest, Allen y Goldstein, 2004; Osimani y cols., 1997; Suhr, 2002; Suhr y Gunstad, 2002). Nuestros resultados no muestran de modo claro tales diferencias, lo que podría ser debido a diversos motivos. En primer lugar, puede ser que no hayamos seleccionado las variables adecuadas dentro de los diversos tests para la detección de simulación, ya que nos hemos centrado en las variables generalmente empleadas en la evaluación del daño cognitivo, y no en los índices propuestos específicamente para la detección de simulación o en el empleo de patrones de ejecución. Por ejemplo, nuestros resultados con el Test de la Figura Compleja de Rey son muy similares en las variables recuerdo inmediato y reconocimiento a los encontrados por Lu y cols. (2003), ya que los puntos de corte propuestos en ambos trabajos son prácticamente idénticos. Si bien en el estudio de Lu y cols., mediante el empleo de una puntuación combinada con una variable que no hemos tenido en cuenta ("errores atípicos"), consiguieron aumentar notablemente la sensibilidad de la prueba, por lo que quizá si nosotros hubiésemos tenido en cuenta dicha variable también hubiésemos encontrado mejores resultados para este test.

En segundo lugar, nuestros resultados podrían ser debidos al empleo de puntos de corte para las distintas variables, en lugar de haber empleado funciones discriminantes que tengan en cuenta medidas de diversas pruebas, que además ha sido propuesto como el mejor método para diferenciar a simuladores y no simuladores (Rogers, 1997^b; Surh y Gunstad, 2000). En este sentido, nuestros resultados con el TAVEC para distintas variables (total ensayos, reconocimiento y discriminabilidad) han sido prácticamente idénticos en cuanto a los puntos de corte propuestos a los de Curtis y cols. (2006) con el California Verbal Learning Test, si bien las sensibilidades encontradas han sido superiores en nuestro trabajo. Curtis y cols. aumentaron la sensibilidad de esta prueba mediante el empleo de diversos modelos y funciones discriminantes cuyo empleo quizá hubiese aumentado también la sensibilidad encontrada en nuestro trabajo. Sin embargo, es preciso destacar que sí hemos estudiado el empleo de funciones discriminantes en nuestro estudio, si bien los resultados no superaban los obtenidos por las variables aisladas, por lo que finalmente no han sido incluidas. Por tanto, nos parece que podemos descartar esta posibilidad.

Por último, es posible que hayamos sido demasiado exigentes en nuestros criterios para considerar los índices como válidos. En realidad no existe ningún criterio consensuado a la hora de aceptar una variable dentro de las pruebas no específicas de simulación como válido. Una propuesta reciente es la de Curtis y cols. (2006), que aceptan como válido un índice si su valor predictivo positivo es superior a 0,51, equivalente a una especificidad del 95% y una sensibilidad mínima del 20%. Observando nuestros resultados en el Anexo 2 puede comprobarse que prácticamente la totalidad de las variables estudiadas cumplen dicho criterio. En definitiva, cabe plantearse que nuestros resultados no son significativamente diferentes de lo que suele encontrarse en la literatura, si bien nosotros hemos sido más estrictos en su

interpretación, ya que consideramos que, debido al tema que nos ocupa, deben tomarse las mayores precauciones posibles.

Como conclusión, podemos decir que los índices derivados de las pruebas neuropsicológicas no específicas para la detección de simulación no carecen de utilidad empleadas como un apoyo más en el complejo proceso de toma de decisiones que supone el diagnóstico de simulación, si bien nunca deberían emplearse de modo aislado o como sustitutivas de las pruebas específicas.

La SPECT-PCR en la detección de simulación

No se han encontrado diferencias en la neuroimagen de los pacientes sospechosos de simulación (BCSS) con respecto a los pacientes no simuladores (NBC) o a los análogos (AN), pero sí se han encontrado diferencias entre los pacientes no simuladores y los análogos. Por tanto, no podemos decir que hayamos encontrado un perfil de simulador mediante la SPECT-PCR. Consideramos que esta ausencia de diferencias entre los pacientes simuladores y no simuladores puede ser debida a que esta prueba de neuroimagen es realizada en condiciones de reposo, en las cuales no existe ningún motivo para que la captación de las áreas cerebrales sea diferente entre los pacientes que deciden simular a la hora de realizar una evaluación neuropsicológica. Probablemente, el empleo de la neuroimagen tenga su utilidad en la detección de la simulación si se realiza de modo simultáneo a la evaluación neuropsicológica, y no en momentos separados. Quizá por este motivo sí se han detectado diferencias en otros estudios con resonancia magnética funcional cuando esta se realizó al mismo tiempo que una prueba específica de simulación (Lee y cols., 2005).

Por otro lado, la propia composición de los grupos está dificultando el hallazgo de diferencias, ya que hay que tener en cuenta que nuestro grupo sospechoso de simulación está formado por pacientes verdaderos (es decir, ha sido comprobado que

todos ellos han sufrido realmente un TCE-L) por lo que parece más probable que dichos pacientes estén magnificando sus síntomas, y no inventándolos, ya que mediante esta exploración ha quedado patente que los sospechosos de simulación realmente tienen algún tipo de daño, a diferencia del grupo de análogos. En este sentido, nuestros resultados están en consonancia con los de otros autores que han encontrado hallazgos en la SPECT-PCR de los pacientes con síndrome post-conmocional (Hofman y cols., 2001; Jacobs y cols, 1994; Kant y cols., 1997). Podría ser que si nuestra fuente de reclutamiento hubiese sido forense y no hospitalaria, dichos hallazgos hubiesen sido menores en el grupo sospechoso de simulación, y por tanto más similares a los de los análogos, si bien esta afirmación es puramente hipotética.

No obstante, la ausencia de diferencias entre los pacientes no simuladores y los sospechosos de simulación en la neuroimagen constituye un apoyo más a las medidas de simulación neuropsicológicas que hemos venido estudiando, ya que las diferencias encontradas en dichas medidas no pueden ser atribuidas a diferencias en el daño cerebral entre los grupos, como demuestran los resultados de la SPECT-PCR.

Las medidas psicofisiológicas

Los resultados de los análisis de la Conductancia Eléctrica de la Piel nos permiten concluir que existe una mayor activación cuando los participantes cometen errores que cuando aciertan. Además, también podemos afirmar que la conductancia nos permite diferenciar cuándo los participantes están en la condición de simulación y cuándo están realizando la tarea lo mejor posible.

En concreto, la mayor amplitud de la respuesta la encontramos cuando, en la condición de simulación, los participantes cometen errores, después cuando aciertan, y por último cuando no simulan (condición control). Estas diferencias son significativas cuando analizamos la conductancia en el transcurso de la decisión de elegir la respuesta

correcta o no (lo que hemos denominado espacio temporal Tiempo de Reacción). Los mismos resultados se encuentran cuando comparamos la Conductancia en los Aciertos, en los Errores y en el Intervalo Entre Estímulos (mayor amplitud en los errores, después cuando aciertan y después cuando están en el IEE), aunque aquí las diferencias encontradas son sólo marginalmente significativas.

Podemos concluir, por tanto, que la amplitud máxima de la Conductancia eléctrica de la piel nos va a permitir comprobar si los participantes han decidido simular su respuesta equivocándose (cometiendo un error), acertando (en la condición de simulación) o si no están simulando (en la condición control).

En general, los datos obtenidos en la Conductancia eléctrica son los que esperábamos, si bien los resultados encontrados con la variable pulso no son tan claros.

En cuanto a las variables del pulso podemos concluir que, en general, la Tasa era menor cuando los participantes estaban en la condición de no simulación que cuando simulaban (control versus simulación –aciertos y errores). Igualmente, su Tasa era menor cuando simulan que cuando están en el IEE (memorizando el número que tienen que reconocer). Sin embargo, la Altura cíclica era menor cuando no simulaban que cuando lo hacían, y también era menor en el IEE que cuando simulaban.

Por lo tanto, en nuestro caso el pulso no ha sido una variable determinante a la hora de diferenciar las respuestas de simulación frente a las de no simulación. Estos resultados pueden ser debidos a la propia medida en sí, ya que en la mayoría de los estudios la Tasa Cardíaca se medía a través del Electrocardiograma y no a través del pulso sanguíneo (por ejemplo Hickling, Blanchard, Mundy & Galovski, 2002).

Atendiendo a los resultados que hemos obtenido, podemos afirmar que la Conductancia eléctrica de la piel es una medida más clara de la simulación que el pulso sanguíneo. Resultados similares los encontramos en Corral, Otero, Barrenetxea, y

Landeta (1998), en los que encuentran, como nosotros, que la conductancia es una medida más clara para detectar el engaño, y no encuentran ninguna diferencia en la tasa cardíaca.

De todos modos, hay que señalar que es muy complicado encontrar estudios similares a los nuestros en los que se intenta comprobar simultáneamente evaluación psicofisiológica, con medidas periféricas, y neuropsicológica de la simulación o el engaño, y debemos asumir que estos datos sólo pueden explicarse en las condiciones experimentales descritas: un grupo de participantes que deben hacer una tarea muy fácil, el VSVT, a los que se les insta a simular durante el transcurso de la misma (simulación) y posteriormente se les motiva para que a continuación hagan la tarea lo mejor posible (control). No obstante, teniendo en cuenta los resultados encontrados, las variables psicofisiológicas parecen ser unas medidas prometedoras para su empleo conjunto con las medidas neuropsicológicas en la detección de la simulación.

CAPÍTULO 10
DISCUSIÓN GENERAL, CONCLUSIONES
Y PERSPECTIVAS FUTURAS

1. Discusión general

El objetivo general de esta tesis ha sido el estudio de diversos métodos propuestos para la detección de simulación de déficits cognitivos en nuestro contexto. Dentro de este trabajo pueden concretarse dos focos principales de interés. Por un lado, estábamos interesados en comprobar cómo se comportan algunas de las pruebas específicas de simulación más extensamente investigadas y empleadas dentro del ámbito de la Neuropsicología Forense. Por otro lado, estábamos interesados en estudiar las aportaciones del empleo de otro tipo de procedimientos en la detección de simulación, en concreto las pruebas neuropsicológicas tradicionales, la SPECT-PCR cerebral, y el empleo de medidas psicofisiológicas en combinación con una prueba específica de simulación.

La relevancia de los resultados de esta tesis radica, fundamentalmente, en que es la primera vez que el estudio de la simulación se realiza de modo sistemático con población española, habiendo sido señalada recientemente por la NAN (Bush, Ruff y cols., 2005) la necesidad de investigar el comportamiento de las pruebas de simulación en poblaciones culturalmente diferentes.

Por otro lado, nuestros resultados se sustentan en diferentes tipos de apoyo, que le dan un mayor peso a nuestros hallazgos. Así, en primer lugar hemos empleado un diseño de grupos de prevalencia diferencial, para determinar si las pruebas son sensibles a la búsqueda de compensación. Además, hemos establecido un grupo de pacientes sospechosos de simulación (diseño de grupos conocidos), lo que nos ha permitido determinar que la prevalencia de base de simulación en nuestro contexto es del 41%, lo cual está en consonancia con los trabajos más recientes y sólidos realizados con población norteamericana sobre la prevalencia de simulación en traumatismos leves (39% según Mittenberg, Patton y cols., 2002; 40% según Larrabee y cols., 2003). Por

último, nuestros resultados se ven respaldados por el empleo de análogos altamente sofisticados. En este sentido hay que resaltar que los análogos empleados en esta investigación tienen un alto nivel educativo, poseen conocimientos neuropsicológicos relacionados con el daño cerebral y su educación, y además han recibido un amplio entrenamiento tanto de los síntomas típicos del trastorno que debían simular como, en algunos casos, de estrategias encaminadas específicamente a superar las pruebas de simulación. Todas las características de este grupo están encaminadas, por tanto, a dificultar la detección, y a pesar de ello, los resultados obtenidos han sido excelentes. Esto, y la ausencia de diferencias encontradas entre este grupo de análogos y el grupo de sospechosos de simulación, hacen que nuestros datos resulten generalizables (Rogers, 1997^{a,b}). En el caso de las pruebas específicas de simulación, otra fuente de apoyo a nuestros datos es la ausencia de diferencias entre nuestros datos con respecto a los datos de las poblaciones de origen.

En relación con el primer foco de interés de esta tesis, es decir, el comportamiento de las pruebas específicas de simulación, los resultados indican que ciertas variables de algunas de las pruebas específicas de simulación estudiadas parecen muy prometedoras para su empleo en la detección de simulación en la práctica clínica, ya que presentan unas excelentes sensibilidades y especificidades. Especialmente útiles parecen las variables ítems totales del VSVT (mediante el empleo de un punto de corte ≤ 38), los ensayos 1 y 2 del TOMM (con unos puntos de corte ≤ 37 y ≤ 44 , respectivamente) y las variables e-score y errores d del Test de la b (puntos de corte > 95 y > 1). Dichas variables, además de obtener los mejores resultados, llevan asociados unos puntos de corte idénticos o muy similares a los publicados en los manuales de las pruebas o en la literatura (VSVT: Inman y cols., 1998; Macciocchi y cols., 2006; TOMM: Rees y cols., 1998, 2001; Tan y cols., 2002; Powell y cols., 2004; Etherton y

cols., 2005; Greve y cols., 2006; Haber y Fichtenberg, 2006; Test de la b: Boone, Lu y Herzberg, 2002). De los puntos de corte derivados del VSVT y del TOMM se desprende que el criterio de puntuaciones significativamente por debajo del azar es inadecuadamente estricto, ya que mediante el establecimiento de puntos de corte podemos conseguir una elevada sensibilidad sin que esto produzca un descenso importante en la especificidad (Bender y Rogers, 2004; Gervais, Rohling, Green y Ford, 2004; Guilmette, Hart y Giuliano, 1993; Guilmette, Hart y cols., 1994; Greiffenstein y cols., 1994; Hiscock, Branham y Hiscock, 1994; Holmquist y Wanlass, 2002; Martin, Bolter, Todd, Gouvier y Nicolls, 1993; Martin, Hayes y Gouvier, 1996; Rees, Tombaugh, Gansler y Moczynski, 1998; Rose, Hall y Szalda-Petree, 1995; Slick, Hopp, Strauss, Hunter y Pinch, 1994). No obstante, este criterio no carece de importancia, ya que nos proporciona información sobre la intencionalidad del sujeto (Rogers, Harrel y Liff, 1993). Es decir, si un paciente puntúa significativamente por debajo del nivel del azar, sabremos que realmente conoce la respuesta verdadera y de modo intencional decide seleccionar la opción incorrecta. La prueba que se ha mostrado superior a la hora de detectar ejecuciones inferiores al nivel del azar ha sido el VSVT (resultado también encontrado por Tan y cols., 2002), por lo que su inclusión en las evaluaciones forenses parece adecuada.

Contrapuesto a estos resultados nos encontramos con el Test de los 15-Ítems de Rey, que comete un inaceptable número de falsos negativos, así como un elevado número de falsos positivos. Esto es especialmente relevante si tenemos en cuenta que es la segunda prueba más empleada para detectar simulación (Slick, Tan, Strauss y Hultsch, 2004), y que los neuropsicólogos forenses la consideran una prueba “aceptable” (Lally, 2003). Puede comprenderse el extenso empleo de este test teniendo en cuenta su accesibilidad, así como su rápida administración e interpretación, aunque

nuestros datos indican que su utilización no es aconsejable, o que, como mínimo, sus resultados han de ser interpretados con la máxima precaución y cautela. Por último, los resultados del Dot Counting pueden describirse como moderadamente aceptables, ya que en el contexto de una magnífica especificidad su sensibilidad es moderada. Por tanto, no clasifica erróneamente a los pacientes, pero su capacidad de detección de simuladores es bastante limitada, como suele señalarse en la literatura (Rose y cols., 1998; Strauss y cols., 2002; Hilsabeck y Gouvier, 2005).

Por último, otro dato interesante es que, debido a las moderadas correlaciones encontradas entre las diferentes pruebas de simulación, parece recomendable el empleo de diversas pruebas en las evaluaciones forenses (Tan y cols., 2002; Strauss y cols., 2002; Cato y cols., 2002; Hilsabeck y Gouvier, 2005).

En cuanto al segundo foco de interés, en primer lugar algunas de las variables de los test neuropsicológicos tradicionales se han mostrado útiles para la detección de simulación. En concreto, las variables TR, TOT y CON y la variable número de palabras leídas en el STROOP han obtenido unos buenos resultados de sensibilidad y especificidad. Esta es la primera vez que dichas variables se plantean como posibles indicadores de simulación, por lo que, si bien se han mostrado prometedores, han de ser replicados para poder confirmar su utilidad. En cualquier caso, los resultados obtenidos por las pruebas neuropsicológicas no específicas de simulación han sido notablemente inferiores a los obtenidos por las pruebas específicas. En este sentido, sólo el 3% de las variables no específicas cumplieron el patrón esperado (es decir, una ejecución igual en pacientes sospechosos de simulación y análogos, y diferente de la ejecución de pacientes sin incentivos para simular y pacientes no sospechosos de simulación, a su vez iguales entre si), frente al 39% de las variables específicas. Además, a igual especificidad, las variables neuropsicológicas no específicas de simulación obtienen

sensibilidades mucho más bajas que las de las variables específicas, lo que ha llevado a algunos investigadores a catalogar a las pruebas neuropsicológicas no específicas de simulación como inefectivas para diferenciar la simulación del daño legítimo (Slick y cols., 1999). Esta superioridad de las pruebas específicas está en concordancia con lo referido por la literatura (Van Gorp y cols., 1999; Inman y Berry, 2002; Schagen y cols., 1997; Greiffenstein y cols., 1994; Iverson y Binder, 2000; O'Bryant y cols., 2003; Tan y cols., 2002).

Por último, los datos provenientes del registro psicofisiológico y el SPECT cerebral debemos considerarlos preliminares. Aunque existe una abundante literatura sobre el uso del polígrafo en la detección de la mentira, no existen trabajos que simultáneamente a las pruebas neuropsicológicas hayan estudiado su capacidad para detectar simuladores de sintomatología cognitiva. En este sentido, la medición de variables psicofisiológicas de pulso y conductancia asociadas a la ejecución de un test específico de simulación (el VSVT) se ha mostrado como un método prometedor para la detección de simulación, ya que los sujetos muestran una elevación de su activación cuando simulan. Sin embargo, los resultados obtenidos en este estudio deben corroborarse en un diseño de "grupos conocidos" para poder determinar su utilidad real (Rogers, 1997).

En relación a la SPECT-PCR, no se han encontrado diferencias entre los pacientes sospechosos de simulación y los demás, lo que nos ha permitido concluir que las diferencias encontradas en la ejecución de las pruebas neuropsicológicas no puede ser explicada por la existencia de un mayor daño cerebral en el grupo sospechoso de simulación, lo que sustenta la idea de que los pacientes sospechosos están, en realidad, simulando. Sin embargo, necesitaremos más estudios para conocer la utilidad de esta

técnica para la detección de la simulación ya que, nuestro estudio no nos ha permitido obtener un patrón típico de los sujetos simuladores.

Como recomendación clínica, debemos recordar que, a pesar de los buenos resultados obtenidos por algunos de los métodos específicos aquí estudiados, el diagnóstico de simulación NUNCA ha de recaer sobre la puntuación obtenida por un sujeto en uno o varios tests (Hilsabeck y Gouvier, 2005), sino que deberá ser el resultado de una combinación de medidas específicas de simulación, tests neuropsicológicos, datos de la entrevista, medidas fisiológicas, información colateral y autoinformes (Martin, Franzen y Orey, 1998).

Finalmente, debemos considerar algunos problemas que podrían limitar los resultados de esta Tesis. Quizás el principal problema sea su reducido n en el grupo de sospechosos de simulación. Sin embargo, hay que mencionar que han sido todos los sospechosos que hemos detectado en el transcurso de la evaluación de todos los traumatismos leves que hemos estudiado durante tres años de investigación. Además, nuestro $n=10$ del grupo sospechoso de simulación es muy similar al empleado por otros trabajos notables, como el manual del TOMM, en el que se emplea un $n=11$ en el grupo sospechoso de simulación. Por otro lado, nuestros datos han sido muy consistentes con otros estudios publicados en la literatura que han utilizado grupos de mayor tamaño.

2. Conclusiones

En función de los resultados obtenidos, de esta tesis se derivan las siguientes conclusiones:

1. El VSVT, el TOMM y el Test de la b son instrumentos útiles para su aplicación en la detección de simulación en nuestro contexto

2. El Test 15-Ítems de Rey ha obtenido unos resultados inaceptables para recomendar su uso en la detección de la simulación.
3. La detección de simulación es mayor mediante el establecimiento de puntos de corte que mediante el criterio de ejecución por debajo del azar.
4. Las pruebas específicas de simulación son claramente superiores a las no específicas en la detección de simulación.

3. Perspectivas futuras

De nuestros resultados se derivan diversas perspectivas de investigación futura que enumeramos a continuación:

1. Estudiar más en profundidad el efecto del entrenamiento y las instrucciones sobre la capacidad de simulación, mediante el empleo de diversos grupos con diferentes grados de conocimiento neuropsicológico y mediante el empleo de instrucciones que varíen en su contenido, así como el empleo de un grupo sin instrucciones.
2. Estudiar la resistencia de los tests específicos de simulación al autoentrenamiento examinando la capacidad de los sujetos para valerse de la información sobre la detección de simulación accesible en la actualidad (Internet, libros, revistas...) animando a un grupo de sujetos a que durante un tiempo determinado se entrenen mediante todos los medios que dispongan para pasar con éxito una evaluación forense.
3. Estudiar la influencia de diversos factores: edad, escolaridad, inteligencia, género, tipo de incentivo empleado... sobre la capacidad de simular de los sujetos.

4. Estudio de otras pruebas de simulación específicas en nuestro contexto (por ejemplo, Amsterdam Short-Term Memory Test, Word Memory Test), así como el diseño y desarrollo de nuevas pruebas específicas.
5. Estudiar el empleo de los “patrones de ejecución” y diversos índices propuestos como indicativos de simulación en la literatura en las pruebas no específicas de simulación.
6. Estudio del comportamiento de las pruebas de simulación en nuestro contexto en diversos grupos clínicos (depresión, esquizofrenia, traumatismos moderados y severos) así como en población normal.
7. Estudio de la simulación de otro tipo de trastornos además del TCE-L (estrés post-traumático y diversos trastornos psiquiátricos).
8. Profundizar en la utilidad del empleo del polígrafo asociado al empleo de pruebas específicas de simulación, mediante el estudio de otras variables psicofisiológicas y aplicándolas a diseños de grupos conocidos.
9. Examinar la utilidad del empleo de otras pruebas de neuroimagen funcional en la detección de simulación, especialmente de la resonancia magnética funcional mientras el sujeto se enfrenta a pruebas de simulación.

REFERENCIAS

- Abdel-Dayem, H.M., Abu-Judeh, H., Kumar, M., Atay, S., Naddaf, S., El-Zeftawy, H., & Luo, J.Q. (1998). SPECT brain perfusion abnormalities in mild or moderate traumatic brain injury. *Clinical Nuclear Medicine, 23*, 309-317.
- Ackerman, M.J. (2006). Forensic Report Writing. *Journal of Clinical Psychology, 62*, 59-72.
- Allen, L.M., Conder, R.L., Green, P., & Cox, D.R. (1997). *CARB '97 Manual for the Computerized Assessment of Response Bias*. CogniSyst, Inc., Durham, N.C.
- Allen, L.M., Iverson, G.L., & Green, P. (2002). Computerized Assessment of Response Bias in forensic neuropsychology. *Journal of Forensic Neuropsychology, 3*, 205-225.
- American Psychiatric Association (1994). *Diagnostic and statistical manual for mental disorders* (4th ed.). Washington, DC: American Psychiatric Association.
- Arnett, P.A., Hammeke, T.A., & Schwartz, L. (1995). Quantitative and Qualitative Performance on Rey's 15-Item Test in Neurological Patients and Dissimulators. *The Clinical Neuropsychologist, 9*, 17-26.
- Arnold, G., Boone, K.B., Lu, P., Dean, A., Wen, J., Nitch, S., & McPherson, S. (2005). Sensitivity and specificity of Finger Tapping Test scores for the detection of suspect effort. *The Clinical Neuropsychologist, 19*, 105-120.
- Ashendorf, L., Constantinou, M., & McCaffrey, R.J. (2004). The effect of depression and anxiety on the TOMM in community-dwelling older adults. *Archives of Clinical Neuropsychology, 19*, 125-130.
- Ashendorf, L., O'Bryant, S.E., & McCaffrey, R.J. (2003). Specificity of malingering detection strategies in older adults using the CVLT and WCST. *The Clinical Neuropsychologist, 17*, 255-262.
- Audenaert, K., Jansen, H.M.L., Otte, A., Peremans, K., Vervaeke, M., Crombez, R., et al. (2003). Imaging of mild traumatic brain injury using ⁵⁷Co and ^{99m}Tc HMPAO SPECT as compared to other diagnostic procedures. *Medical Science Monitor, 9*, MT112-117
- Bagby, R.M., Buis, T., & Nicholson, R.A. (1995). Relative effectiveness of the standard validity scales in detecting fake-bad and fake-good responding: replication and extension. *Psychological Assessment, 7*, 84-92.
- Barrash, J., Suhr, J.A., & Manzel, K. (1998). A brief, sensitive, and highly specific procedure for detecting malingered memory impairment. *Journal of the International Neuropsychological Society, 4*, 28.
- Barth, J.T., Ryan, T.V., & Hawk, G.L. (1992). Forensic neuropsychology: A reply to the method skeptics. *Neuropsychology Review, 2*, 251-266.
- Bauer, L., & McCaffrey, R.J. (2006). Coverage of the Test of Memory Malingering, Victoria Symptom Validity Test, and Word Memory Test on the Internet : Is test security threatened ? *Archives of Clinical Neuropsychology, 21*, 121-126.
- Bender, S.D., & Rogers, R. (2004). Detection of neurocognitive feigning: development of a multi-strategy assessment. *Archives of Clinical Neuropsychology, 19*, 49-60.
- Benedet, M.J., & Alejandre, M.A. (1998). *TAVEC. Test de aprendizaje verbal España-Complutense*. Madrid: TEA.
- Bernard, L.C. (1990). Prospects for faking believable memory deficits on neuropsychological tests and the use of incentives in simulation research. *Journal of Clinical and Experimental Neuropsychology, 12*, 715-728.

- Bernard, L.C., McGrath, M.J., & Houston, W. (1996). The differential effects of simulating malingering, closed head injury, and other CNS pathology on the Wisconsin Card Sorting Test: Support for the "pattern of performance" hypothesis. *Archives of Clinical Neuropsychology, 11*, 231-245.
- Berry, D.T.R., & Butcher, J.N. (1998). Detection of feigning of head injury symptoms on the MMPI-2. In C.R. Reynolds (Ed.), *Detection of Malingering during Head Injury Litigation*. New York: Plenum Press.
- Bianchini, K.J., Greve, K.W., & Love, J.M. (2003). Definite Malingered Neurocognitive Dysfunction in moderate/severe traumatic brain injury. *The Clinical Neuropsychologist, 17*, 574-580.
- Bianchini, K.J., Mathias, C.W., & Greve, K.W. (2001). Symptom Validity Testing: A critical review. *The Clinical Neuropsychologist, 15*, 19-45.
- Bianchini, K.J., Mathias, C.W., Greve, K.W., Houston, R.J., & Crouch, J.A. (2001). Classification accuracy of the Portland Digit Recognition Test in traumatic brain injury. *The Clinical Neuropsychologist, 15*, 461-470.
- Bigler, E.D. (2006). Can author bias be determined in forensic neuropsychology research published in Archives of Clinical Neuropsychology? *Archives of Clinical Neuropsychology, 21*, 503-508.
- Binder, L.M. (1986). Persisting symptoms after mild head injury: a review of the postconcussive syndrome. *Journal of Clinical and Experimental Neuropsychology, 8*, 323-346.
- Binder, L.M. (1990). Malingering following minor head trauma. *The Clinical Neuropsychologist, 4*, 25-36.
- Binder, L.M. (1993). Assessment of malingering after mild head trauma with the Portland Digit Recognition Test. *Journal of Clinical and Experimental Neuropsychology, 15*, 170-182.
- Binder, L.R. (1997). A review of mild head trauma, Part II: Clinical implications. *Journal of Clinical and Experimental Neuropsychology, 19*, 432-457.
- Binder, L.M., & Pankratz, L. (1987). Neuropsychological evidence of a factitious memory complaint. *Journal of Clinical and Experimental Neuropsychology, 15*, 167-171.
- Binder, L.M., & Rohling, M.L. (1996). Money matters: A meta-analytic review of the effects of financial incentives on recovery after closed-head injury. *The American Journal of Psychiatry, 153*, 7-10.
- Binder, L.M., & Willis, S.C. (1991). Assessment of motivation after financially compensable minor head trauma. *Psychological Assessment, 3*, 175-181.
- Binks, P.G., Gouvier, D., & Waters, W.F. (1997). Malingering detection with the Dot Counting Test. *Archives of Clinical Neuropsychology, 12*, 41-46.
- Blanchard, E.B., Hickling, E.J., Buckley, T.C., Taylor, A.E., Vollmer, A., & Loss, W.R. (1996). Psychophysiology of Posttraumatic Stress Disorder related to motor vehicle accidents: Replication and extension. *Journal of Consulting and Clinical Psychology, 64*, 742-751.
- Blanchard, E.B., Kolb, L.C., Gerardi, R.J., Ryan, P., & Pallmeyer, T.P. (1986). Cardiac response to relevant stimuli as an adjunctive tool for diagnosing post-traumatic stress disorder in Vietnam veterans. *Behavior Therapy, 17*, 592-606.
- Blanchard, E.B., Kolb, L.C., Pallmeyer, T.P., & Gerardi, R.J. (1982). A psychophysiological study of post-traumatic stress disorder in Vietnam veterans. *Psychiatric Quarterly, 54*, 220-229.
- Bleiberg, J., Garmoe, W.S., Halpern, E.L., Reeves, D.L., & Nadler, J.D. (1997). Consistency of within-day and across-day performance after mild brain injury. *Neuropsychiatry, Neuropsychology, and Behavioral Neurology, 10*, 247-253.

- Blostein, P., & Jones, S.J. (2003). Identification and evaluation of patients with mild traumatic brain injury: Results of a national survey of Level I trauma centers. *The Journal of Trauma Injury, Infection, and Critical Care*, *55*, 450-453.
- Bohnen, N., Twijnstra, A., & Jolles, J. (1993). Persistence of postconcussional symptoms in uncomplicated mildly head-injured patients: A prospective cohort study. *Neuropsychiatry, Neuropsychology and Behavioural Neurology*, *6*, 193-200.
- Bolan, B., Foster, J.K., Schmand, B., & Bolan, S. (2002). A comparison of three tests to detect feigned amnesia: The effects of feedback and the measurement of response latency. *Journal of Clinical and Experimental Neuropsychology*, *24*, 154-167.
- Boone, K., Lu, P., & Herzberg, D.S. (2002^a). *The b Test. Manual*. Los Angeles: Western Psychological Services.
- Boone, K.B., Lu, P., Back, C., King, C., Lee, A., Philpott, L., Shamieh, E., & Warner-Chacon K. (2002^b). Sensitivity and specificity of the Rey Dot Counting Test in patients with suspect effort and various clinical samples. *Archives of Clinical Neuropsychology*, *17*, 625-642.
- Boone, K.B., Lu, P., Sherman, D., Palmer, B., Back, C., Shamieh, E., Warner-Chacon, K., & Berman, N.G. (2000). Validation of a new technique to detect malingering of cognitive symptoms: The b Test. *Archives of Clinical Neuropsychology*, *15*, 227-241.
- Brickenkamp, R. (1962). *Test d2*. Zurich: Verlag für Psychologie.
- Brown, L., Sherbenou, R., & Johnsen, S. (1982). *Test of Nonverbal Intelligence*. Austin: Pro-Ed.
- Bush, S.S., & NAN Policy & Planning Committee (2005). Independent and court-ordered forensic neuropsychological examinations: Official statement of the National Academy of Neuropsychology. *Archives of Clinical Neuropsychology*, *20*, 997-1007.
- Bush, S.S., Ruff, R.M., Tröster, A.I., Barth, J.T., Koffler, S.P., Pliskin, N-H., Reynolds, C.R., & Silver, C.H. (2005). Symptom validity assessment: Practice issues and medical necessity. *Archives of Clinical Neuropsychology*, *20*, 419-426.
- Butcher, J.N., Dahlstrom, W.G., Graham, J.R., Tellegen, A., & Kaemmer, B. (1989). *Manual for administration and scoring the Minnesota Multiphasic Personality Inventory-2*. Minneapolis: University of Minnesota Press.
- Cassidy, J.D., Carroll, L.J., Cote, P., Lernstra, M., Berglund, A., & Nygren, H. (2000). Effect of eliminating compensation for pain and suffering on the outcome of insurance claims for whiplash. *New England Journal of Medicine*, *342*, 1179-1186.
- Cato, M.A., Brewster, J., Ryan, T., & Giuliano, A.J. (2002). Coaching and the ability to simulate mild traumatic brain injury symptoms. *The Clinical Neuropsychologist*, *16*, 524-535.
- Cicerone, K.D. (1996). Attention deficits and dual task demands after mild traumatic brain injury. *Brain Injury*, *10*, 79-89.
- Constantinou, M., Bauer, L., Ashendorf, L., Fisher, F.M., & McCaffrey, R.J. (2005). Is poor performance on recognition memory effort measures indicative of generalized poor performance on neuropsychological tests? *Archives of Clinical Neuropsychology*, *20*, 191-198.
- Constantinou, M., & McCaffrey, R.J., (2003). Using the TOMM to evaluate children's effort to perform optimally on neuropsychological measures. *Child Neuropsychology*, *9*, 81-90.
- Corral, S., Otero, J., Barrenetxea, A., & Landeta, O. (1998). Información y Test de Conocimiento Culpable en la Detección del Engaño. *Psicológica*, *19*, 187-199.

- Cruz, M.V., Seisdedos, N., & Cordero, A. (1997). *Rey. Test de Copia y de reproducción de memoria de figuras geométricas complejas*. Madrid: TEA.
- Curtis, K.L., Greve, K.W., Bianchini, K.J., & Brennan, A. (2006). California Verbal Learning Test indicators of malingered neurocognitive dysfunction. Sensitivity and specificity in traumatic brain injury. *Assessment, 13*, 46-61.
- Daubert, S.D., & Metzler, A.E. (2000). The detection of fake-bad and fake-good responding on the Millon Clinical Multiaxial Inventory-III. *Psychological Assessment, 12*, 418-424.
- Davalos, D.B., & Bennet, T.L. (2002). A review of the use of single-photon emission computerized tomography as a diagnostic tool in mild traumatic brain injury. *Applied Neuropsychology, 9*, 92-105.
- Delis, D.C., Kramer, J.H., Kaplan, E., Ober, B.A. (1987). *California Verbal Learning Test*. New York: Psychological Corporation.
- Demakis, G.J. (1999). Serial Malingering on Verbal and Nonverbal Fluency and Memory Measures: An Analog Investigation. *Archives of Clinical Neuropsychology, 14*, 401-410.
- Denney, R.L., & Wynkoop, T.F. (2000). Clinical Neuropsychology in the criminal forensic setting. *The Journal of Head Trauma Rehabilitation, 15*, 804-828.
- DiCarlo, M.A., Gfeller, J.D., & Oliveri, M.V. (2000). Effects of Coaching on Detecting Feigned Cognitive Impairment with the Category Test. *Archives of Clinical Neuropsychology, 15*, 399-413.
- Donders, J.D. (1999). Brief report: Sensitivity of a malingering formula for the Wisconsin Card Sorting Test. *Journal of Forensic Neuropsychology, 1*, 34-43.
- Duncan, A. (2005). The impact of cognitive and psychiatric impairment of psychotic disorders on the Test of Memory Malingering (TOMM). *Assessment, 12*, 123-129.
- Ellwagner, J., Tenhula, W.N., Rosenfeld, P., & Sweet, J.J. (1999). Identifying simulators of cognitive deficit though combined use of neuropsychological tests of performance and event-related potentials. *Journal of Clinical and Experimental Neuropsychology, 21*, 866-879.
- Erdal, K. (2004). The effects of motivation, coaching, and knowledge of neuropsychology on the simulated malingering of head injury. *Archives of Clinical Neuropsychology, 19*, 73-88.
- Esing, S.M., Mittenberg, W., Petersen, R.S., Strauman, S., & Cooper, J.T. (2001). Practices in forensic neuropsychology: Perspectives of neuropsychologists and trial attorneys. *Archives of Clinical Neuropsychology, 16*, 271-291.
- Etherton, J.L., Bianchini, K.J., Greve, K.W., & Ciota, M.A. (2005). Test of Memory Malingering performance is unaffected by laboratory-induced pain: implications for clinical use. *Archives of Clinical Neuropsychology, 20*, 375-384.
- Faust, D. (1991). Forensic neuropsychology: The art of practicing a science that not yet exist. *Neuropsychology Review, 2*, 205-231.
- Faust, D., & Ackley, M.A. (1998). Did you think it was going to be easy? Some methodological suggestions for the investigation and development of malingering detection techniques. In C.R. Reynolds (Ed.), *Detection of Malingering During Head Injury Litigation*, (pp. 1-54). New York: Plenum Press.
- Faust, D., Hart, K., & Guilmette, T.J. (1988). Pediatric malingering: The capacity of children to fake believable deficits of neuropsychological testing. *Journal of Consulting and Clinical Psychology, 56*, 578-582.

- Fee, C.R., & Rutherford, W.H. (1988). A study of the effect of legal settlement on post-concussion symptoms. *Archives of Emergency Medicine, 5*, 12-17.
- Forrest, T.J., Allen, D.N., & Goldstein, G. (2004). Malingering indexes for the Halstead Category Test. *Archives of Clinical Neuropsychology, 18*, 334-347.
- Fox, D.D., Lees-Haley, P.R., Earnest, K., & Dolezal-Wood, S. (1995). Post-Concussive Symptoms: Base rates and etiology in psychiatric patients. *The Clinical Neuropsychologist, 9*, 89-92.
- Frederick, R.I. (2002). A review of Rey's strategies for detecting malingered neuropsychological impairment. *Journal of Forensic Neuropsychology, 2*, 1-25.
- Frederick, R.I., & Foster, H.G. (1991). Multiple measures of malingering on a forced-choice test of cognitive ability. *Psychological Assessment, 3*, 596-602.
- Frederick, R.I., Sarfaty, S.D., Johnston, J.D., & Powel, J. (1994). Validation of a detector of response bias on a forced-choice test of nonverbal ability. *Neuropsychology, 8*, 118-125.
- Garzón, A. (1990). *Psicología y justicia*. Valencia: Promolibro.
- Gavett, B.E., O'Bryant, S.E., Fisher, J.M., & McCaffrey, R.J. (2005). Hit rates of adequate performance based on the Test of Memory Malingering (TOMM) Trial 1. *Applied Neuropsychology, 12*, 1-4.
- Gentilini, N., Nichelli, P., & Schoenhuber, R. (1989). *Assessment of Attention in Mild Head Injury*. New York, NY: Oxford University Press.
- Gervais, R.O., Rohling, M.L., Green, P., & Ford, W. (2004). A comparison of WMT, CARB, and TOMM failure rates in non-head injury disability claimants. *Archives of Clinical Neuropsychology, 19*, 475-487.
- Gervais, R.O., Russell, A.S., Green, P., Allen, L.M., Ferrari, R., & Pieschl, S.D. (2001). Effort testing in fibromyalgia patients with disability incentives. *Journal of Rheumatology, 28*, 1892-1899.
- Gierok, S.D., Dickson, A.L., Cole, J.A. (2005). Performance of forensic and non-forensic adult psychiatric inpatients on the Test of Memory Malingering. *Archives of Clinical Neuropsychology, 20*, 755-760.
- Golden, J.C. (1994). *Stroop. Test de Colores y Palabras. Manual*. Madrid: TEA.
- Golden, C.J., & Grier, C.A. (1998). Detecting malingering on the Luria-Nebraska Neuropsychological Battery. In *Detection of Malingering During Head Injury Litigation*, Reynolds (Ed.), Plenum Press, New York.
- Goldberg, J.O., & Miller, H.R. (1986). Performance of psychiatric inpatients and intellectually deficient individuals on a task that assesses the validity of memory complaints. *Journal of Clinical Psychology, 42*, 792-795.
- Gouvier, W.D., Lees-Haley, P.R., & Hammer, J.H. (2003). The neuropsychological examination in the detection of malingering in the forensic arena: Costs and benefits. In G.P. Prigatano, & N.H. Pliskin (Eds.), *Clinical Neuropsychology and Cost Outcome Research* (pp. 405-424). New York: Taylor and Francis Books, Inc.
- Gowda, N.K., Agrawal, D., Bal, C., Chandrashekar, N., Tripathi, M., Bandopadhyaya, G.P., Malhotra, A., & Mahapatra, A.K. (2006). Technetium Tc-99m ethyl cysteinate dimer brain single-photon emission CT in mild traumatic brain injury: A prospective study. *American Journal of Neuroradiology, 27*, 447-451.
- Green, P., Allen, L., & Astner, K. (1996). *The Word Memory Test: A user's guide to the oral and computer-administered forms. US version 1.1*. Durham, NC: CogniSyst.

- Green, P., & Iverson, G.L. (2001). Validation of the Computerized Assessment of Response Bias in litigating patients with head injuries. *The Clinical Neuropsychologist*, *15*, 492-497.
- Green, P., Iverson, G.L., & Allen, L. (1999). Detecting malingering in head injury litigation with the Word Memory Test. *Brain Injury*, *13*, 813-819.
- Green, P., Rohling, M.L., Lees-Haley, P.R., & Allen, L.M. (2001). Effort has a greater effect on neuropsychological test scores than severe brain injury. *Brain Injury*, *15*, 1045-1060.
- Greiffenstein, M.F., Baker, W.J., & Gola, T. (1994). Validation of malingered amnesia measures with a large clinical sample. *Psychological Assessment*, *6*, 218-224.
- Greiffenstein, M.F., Baker, W.J., & Gola, T. (1996). Motor dysfunction profiles in traumatic brain injury and postconcussion syndrome. *Journal of the International Neuropsychological Society*, *2*, 477-485.
- Greiffenstein, M.F., & Cohen, L. (2005). Neuropsychology and the Law: Principles of productive attorney-neuropsychologist relations. In Larrabee, G.J. (Ed.), *Forensic Neuropsychology: A scientific approach* (pp. 29-91). New York: Oxford University Press.
- Greub, B.L., & Suhr, J.A. (2006). The validity of the Letter Memory Test as a measure of memory malingering: Robustness to coaching. *Archives of Clinical Neuropsychology*, *21*, 249-254.
- Greve, K.W., & Bianchini, K.J. (2002). Using the Wisconsin Card Sorting Test to detect malingering: An analysis of the specificity of two methods in nonmalingering normal and patient samples. *Journal of Clinical and Experimental Neuropsychology*, *24*, 48-54.
- Greve, K.W., & Bianchini, K.J. (2004). Setting empirical cut-offs on psychometric indicators of negative response bias: a methodological commentary with recommendations. *Archives of Clinical Neuropsychology*, *19*, 533-541.
- Greve, K.W., & Bianchini, K.J. (2006). Should the Retention trial of the Tomm of Memory Malingering be optional? *Archives of Clinical Neuropsychology*, *21*, 117-119.
- Greve, K.W., Bianchini, K.J., & Doane, B.M. (2006). Classification accuracy of the Test of Memory Malingering in traumatic brain injury: Results of a known-groups analysis. *Journal of Clinical and Experimental Neuropsychology*, *28*, 1176-1190.
- Gronwall, D. (1987). Advances in assessment of attention and information processing after head injury. In H.S. Levin, J. Grafman & H.M. Eisenberg (Eds.), *Neurobehavioral Recovery From Head Injury* (pp. 355-371). New York, NY: Oxford University Press.
- Gronwall, D. (1991). Minor head injury. *Neuropsychology*, *4*, 253-265.
- Grosz, H., & Zimmerman, J. (1965). Experimental analysis of hysterical blindness. *Archives of General Psychiatry*, *13*, 225-260.
- Grote, C. (2005). Ethical practice of Forensic Neuropsychology. In Larrabee, G.J. (Ed.), *Forensic Neuropsychology: A scientific approach* (pp. 93-114). New York: Oxford University Press.
- Grote, C.L., Kooker, E.K., Garron, D.C., Nyenhuis, D.L., Smith, C.A., & Mattingly, M.L. (2000). Performance of compensation seeking and non-compensation seeking samples on the Victoria Symptom Validity Test: Cross-validation and extension of a standardization study. *Journal of Clinical and Experimental Neuropsychology*, *22*, 709-719.
- Guilmette, T.J., Hart, K.J., & Giuliano, A.J. (1993). Malingering detection: The use of a forced-choice method in identifying organic versus simulated memory impairment. *The Clinical Neuropsychologist*, *7*, 59-69.

- Guilmette, T.J., Hart, K.J., Giuliano, A.J., & Leininger, B.E. (1994). Detecting simulated memory impairment: Comparison of the Rey Fifteen-Item Test and the Hiscock forced-choice procedure. *The Clinical Neuropsychologist, 8*, 283-294.
- Guilmette, T.J., Whelihan, W., Sparadeo, F.R., & Buongiorno, G. (1994). Validity of neuropsychological test results in disability evaluations. *Perceptual and Motor Skills, 78*, 1179-1186.
- Guilmette, T.J., Whelihan, W.M., Hart, K.J., Sparadeo, F.R., & Buongiorno, G. (1996). Order effects in the administration of a forced-choice procedure for detection of malingering in disability claimants' evaluations. *Perceptual and Motor Skills, 83*, 1007-1016.
- Gunstad, J., & Suhr, J.A. (2001). Efficacy of the full and abbreviated forms of the Pórtland Digit Recognition Test: Vulnerability to Coaching. *The Clinical Neuropsychologist, 15*, 397-404.
- Gutiérrez, J.M., & Gur, R.C. (1998). Detection of malingering using forced-choice techniques. In C.R. Reynolds (Ed.), *Detection of Malingering during Head Injury Litigation*. New York: Plenum Press.
- Haber, A.H., & Fichtenberg, N.L. (2006). Replication of the Test of Memory Malingering (TOMM) in a traumatic brain injury and head trauma sample. *The Clinical Neuropsychologist, 20*, 524-532.
- Haines, M.E., & Norris, M.P. (1995). Detecting the malingering of cognitive deficits: An update. *Neuropsychological Review, 5*, 125-148.
- Hathaway, S.R., & McKinley, J.C. (1940). A multiphasic personality schedule:I. Construction of the schedule. *Journal of Psychology, 10*, 204-210.
- Heaton, R.K. (1981). *A manual for the Wisconsin Card Sorting Test*. Odessa: PAR.
- Heaton, R.K., Smith, H.H., Lehman, R.A.W., & Vogt, A.T. (1978). Prospects for faking believable deficits on neuropsychological testing. *Journal of Consulting and Clinical Psychology, 46*, 892-900.
- Heilbronner, R.L. (2004). A status report on the practice of forensic neuropsychology. *The Clinical Neuropsychologist, 18*, 312-326.
- Hickling, E.J., Blanchard, E.B., Mundy, E., & Galovski, T.E. (2002). Detection of malingered MVA related Posttraumatic Stress Disorder: An investigation of the ability to detect professional actors by experienced clinicians, psychological tests and psychophysiological assessment. *Journal of Forensic Psychology Practice, 2*, 33-53.
- Hilsabeck, R.C., & Gouvier, W.D. (2005). Detecting simulated memory impairment: Further validation of the Word Completion Memory Test (WCMT). *Archives of Clinical Neuropsychology, 20*, 1025-1041.
- Hilsabeck, R. C., LeCompte, D. C., Marks, A. R., & Grafman, J. (2001). The Word Completion Memory Test (WCMT): A new test to detect malingered memory deficits. *Archives of Clinical Neuropsychology, 16*, 669-677.
- Hiscock, C.K., Branham, J.D., & Hiscock, M. (1994). Detection of feigned cognitive impairment: The two-alternative forced-choice method compared with selected conventional tests. *Journal of Psychopathology & Behavioral Assessment, 16*, 95-110.
- Hiscock, M., & Hiscock, C.K. (1989). Refining the forced-choice method of detection of malingering. *Journal of Clinical and Experimental Neuropsychology, 11*, 967-974.
- Hofman, P.A.M., Stapert, S.Z., van Kroonenburg, M.J.P.G., Jolles, J., Jelle de Kruijk, & Wilmink, J.T. (2001). MR imaging, single-photon emission CT, and neurocognitive performance after mild traumatic brain injury. *American Society of Neuroradiology, 22*, 441-449.

- Hom, J. (2003). Forensic neuropsychology: are we there yet? *Archives of Clinical Neuropsychology*, *18*, 827-845.
- Holmquist, L.A., & Wanlass, R.L. (2002). A multidimensional approach towards malingering detection. *Archives of Clinical Neuropsychology*, *17*, 143-156.
- Iacono, W.G. (2000). The detection of deception. In J.C. Cacioppo, L.G. Tassinary, & G.G. Berntson (Eds.), *Handbook of Psychophysiology* (pp. 772-793). Cambridge: University Press.
- Ichise, M., Chung, D., Wang, P., Wortzman, G., Gray, B., & Franks, W. (1994). Technetium-99m-HMPAO SPECT, CT and MRI in the evaluation of patients with chronic traumatic brain injury: a correlation with neuropsychological performance. *Journal of Nuclear Medicine*, *35*, 217-225.
- Inman, T.H., & Berry, D.T.R. (2002). Cross-validation of indicators of malingering. A comparison of nine neuropsychological tests, four tests of malingering, and behavioral observations. *Archives of Clinical Neuropsychology*, *17*, 1-23.
- Inman, T.H., Vickery, C.D., Lamb, D.G., Edwards, C.L., & Smith, G.T. (1998). Development and initial validation of a new procedure for evaluating adequacy of effort given during neuropsychological testing: The Letter Memory Test. *Psychological Assessment*, *10*, 128-139.
- Iverson, G. (1995). Qualitative aspects of malingered memory deficits. *Brain Injury*, *9*, 35-40.
- Iverson, G.L., & Binder, L.M. (2000). Detecting Exaggeration and Malingering in Neuropsychological Assessment. *Journal of Head Trauma Rehabilitation*, *15*, 829-858.
- Iverson, G.L., & Franzen, M.D. (1996). Using multiple objective memory procedures to detect simulated malingering. *Journal of Clinical and Experimental Neuropsychology*, *18*, 38-51.
- Iverson, G.L., & Franzen, M.D. (1998). Detecting malingered memory deficits with the Recognition Memory Test. *Brain Injury*, *12*, 275-282.
- Iverson, G.L., Franzen, M.D., & McCracken, M. (1991). Evaluation of an objective assessment technique for the detection of malingered memory deficits. *Law and Human Behavior*, *15*, 667-676.
- Iverson, G.L., Green, P., & Gervais, R. (1999). Using the Word Memory Test to detect biased responding in head injury litigation. *The Journal of Cognitive Rehabilitation*, *March-April*, 4-8.
- Iverson, G.L., Ropacki, M.T., Allen, L.M., & Elias, J.W. (2001). Analog malingering on the Computerized Assessment of Response Bias (Abstract). *Archives of Clinical Neuropsychology*, *16*, 850-851.
- Jacobs, A., Put, E., Ingels, M., & Bossuyt, A. (1994). Prospective evaluation of technetium-99m-HMPAO SPECT in mild and moderate traumatic brain injury. *Journal of Nuclear Medicine*, *35*, 942-947.
- Kant, R., Smith-Seemiller, L., Isaac, G., & Duffy, J. (1997). Tc-HMPAO SPECT in persistent post-concussion syndrome after mild head injury comparison with MRI/CT. *Brain Injury*, *11*, 115-124.
- Kay, T., Harrington, D.E., Adams, R., Anderson, T., Berrol, S., Cicerone, K., et al. (1993). Definition of mild traumatic brain injury. *Journal of Head Trauma Rehabilitation*, *8*, 86-87.
- Keller, M., Hiltbrunner, B., Dill, C., & Kesselring, J. (2000). Reversible neuropsychological deficits after mild traumatic brain injury. *Journal of Neurology, Neurosurgery, and Psychiatry*, *68*, 761-764.
- King, N.S. (2003). Post-concussion syndrome: clarity amid the controversy? *The British Journal of Psychiatry*, *183*, 276-278.
- King, J.H., Sweet, J.J., Sherer, M., Curtiss, G., & Vanderploeg, R.D. (2002). Validity indicators within the Wisconsin Card Sorting Test: Application of new and previously researched multivariate

- procedures in multiple traumatic brain injury samples. *The Clinical Neuropsychologist*, 16, 506-523.
- Langeluddecke, P.M., & Lucas, S.K. (2003). Quantitative measures of memory malingering on the Wechsler Memory Scale-Third edition in mild head injury litigants. *Archives of Clinical Neuropsychology*, 18, 181-197.
- Larrabee, G.J. (2000). Forensic neuropsychological assessment. In R.D. Vanderploeg, (Ed.), *Clinician's guide to neuropsychological assessment*, (pp. 301-335). Mahwah, NJ: Lawrence Erlbaum Associates.
- Larrabee, G.J. (2003^a). Detection of malingering using atypical performance patterns on standard neuropsychological tests. *The Clinical Neuropsychologist*, 17, 410-425.
- Larrabee, G.J. (2003^b). Detection of symptom exaggeration with the MMPI-2 in litigants with malingered neurocognitive dysfunction. *The Clinical Neuropsychologist*, 17, 54-68.
- Larrabee, G.J. (2005^a). Mild traumatic brain injury. In Larrabee, G.J. (Ed.), *Forensic Neuropsychology: A scientific approach* (pp. 209-236). New York: Oxford University Press.
- Larrabee, G.J. (2005^b). A scientific approach to Forensic Neuropsychology. In Larrabee, G.J. (Ed.), *Forensic Neuropsychology: A scientific approach* (pp. 209-236). New York: Oxford University Press.
- Lee, A., Boone, K.B., Lesser, I., Wohl, M., Wilkins, S., & Parks, C. (2000). Performance of older depressed patients on two cognitive malingering tests: Positive rates for the Rey 15-item Memorization and Dot Counting Tests. *The Clinical Neuropsychologist*, 14, 303-308.
- Lee, G.P., Loring, D.W., & Martin, R.C. (1992). Rey's 15-Item Visual Memory Test for the Detection of Malingering: Normative Observations on Patients with Neurological Disorders. *Psychological Assessment*, 4, 43-46.
- Lee, T.M.C., Liu, H.L., Chan, C.C.H., Ng, Y.B., Fox, P.T., & Gao, J.H. (2005). Neural correlates of feigned memory impairment. *NeuroImage*, 28, 305-313.
- Lees-Haley P. R. (1997). MMPI-2 base rates for 492 personal injury plaintiffs: Implications and challenges for forensic assessment. *Journal of Clinical Psychology*, 53, 745-755.
- Lees-Haley, P.R., & Brown, R.S. (1993). Neuropsychological complaint base rates of 170 personal injury claimants. *Archives of Clinical Neuropsychology*, 8, 203-209.
- Leininger, B.E., Gramling, S.E., Farrell, A.D., Kreuzer, J.S., & Peck, E.A. (1990). Neuropsychological deficits in symptomatic minor head injury patients after concussion and mild concussion. *Journal of Neurology, Neurosurgery and Psychiatry*, 53, 293-296.
- Levin, H.S., Benton, A.L., & Grossman, R.G. (1982). *Neurobehavioral consequences of closed head injury*. New York: Oxford University Press.
- Lewis, J.L., Simcox, A.M., & Berry, D.T.R. (2002). Screening for feigned psychiatric symptoms in a forensic sample by using the MMPI-2 and the Structured Inventory of Malingered Symptomatology. *Psychological Assessment*, 14, 170-176.
- Lezak, M.D. (1995). *Neuropsychological Assessment* (3rd Edition). New York: Oxford University Press.
- Lezak, M.D., Howieson, D.B., & Loring, D.W. (2004). *Neuropsychological Assessment* (4th Edition). New York: Oxford University Press.
- Lu, P.H., Boone, K.B., Cozolino, L., & Mitchell, C. (2003). Effectiveness of the Rey-Osterrieth Complex Figure Test and the Meyers and Meyers Recognition trial in the detection of suspect effort. *The Clinical Neuropsychologist*, 17, 426-440.

- Lynch, W.J. (2004). Determination of Effort Level, Exaggeration, and Malingering in Neurocognitive Assessment. *The Journal of Head Trauma Rehabilitation*, *19*, 277-283.
- Macciocchi, S.N., Seel, R.T., Alderson, A., & Godsall, R. (2006). Victoria Symptom Validity Test performance in acute severe traumatic brain injury: Implications for test interpretation. *Archives of Clinical Neuropsychology*, *21*, 395-404.
- Martens, M., Donders, J., & Millis, S.R. (2001). Evaluation of invalid response sets after traumatic head injury. *Journal of Forensic Neuropsychology*, *2*, 1-18.
- Martin, N.J., & Franzen, M.D. (1992). The malingering of memory deficits among sophisticated subjects. *The Clinical Neuropsychologist*, *6*, 329.
- Martin, R.C., Bolter, J.F., Todd, M.E., Gouvier, W.D., & Niccolls, R. (1993). Effects of sophistication and motivation on the detection of malingered memory performance using a computerized forced-choice task. *Journal of Clinical and Experimental Neuropsychology*, *15*, 867-880.
- Martin, R.C., Franzen, M.D., & Orey, S. (1998). Magnitude of error as a strategy to detect feigned memory impairment. *The Clinical Neuropsychologist*, *12*, 84-91.
- Martin, R.C., Hayes, J.S., & Gouvier, W.D. (1996). Differential vulnerability between postconcussion self-report and objective malingering tests in identifying simulated mild head injury. *Journal of Clinical and Experimental Neuropsychology*, *18*, 265-275.
- Martzke, J., Swan, C., & Varney, N. (1991). Post traumatic anosmia and orbital frontal damage: Neuropsychological and neuropsychiatric correlates. *Neuropsychology*, *5*, 213-225.
- Masdeu, J.C., Van Heertum, R.L., Kleiman, A., Anselmi, G., Kissane, K., Horng, J., et al. (1994). Early single-photon emission computed tomography in mild head trauma. A controlled study. *Journal of Neuroimaging*, *4*, 177-181.
- Mateer, C.A. (2000). Attention. In S.A. Raskin & C.A. Mateer (Eds.), *Neuropsychological management of mild traumatic brain injury* (pp. 73-92). New York: Oxford University Press.
- Mateer, C.A., & D'Arcy, R.C.N. (2000). Current concepts and approaches to management. In S.A. Raskin & C.A. Mateer (Eds.), *Neuropsychological management of mild traumatic brain injury* (pp. 3-22). New York: Oxford University Press.
- McFlynn, G., Montgomery, E.A., Fenton, G.W., & Rutherford, W. (1984). Measurement of reaction time following minor head injury. *Journal of Neurology, Neurosurgery and Psychiatry*, *47*, 1326-1331.
- McKinzey, R.K., Podd, M.G., Krehbiel, M.A., Mensch, A.J., & Trombka, C.C. (1997). Detection of malingering on the Luria-Nebraska Neuropsychological Battery: An initial and cross-validation. *Archives of Clinical Neuropsychology*, *12*, 505-512.
- McKinzey, R.K., & Russell, E.W. (1997). Detection of malingering on the Halstead-Reitan Battery: A cross-validation. *Archives of Clinical Neuropsychology*, *12*, 585-590.
- Mensch, A.J., & Woods, D.J. (1986). Patterns of feigning brain damage on the LNNB. *The International Journal of Clinical Neuropsychology*, *8*, 59-63.
- Merckelbach, H., & Smith, G.P. (2003). Diagnostic accuracy of the Structured Inventory of Malingering Symptomatology (SIMS) in detecting instructed malingering. *Archives of Clinical Neuropsychology*, *18*, 145-152.
- Merten, T., Green, P., Henry, M., Blaskewitz, N., & Brockhaus, R. (2005). Analog validation of German-language symptom validity tests and the influence of coaching. *Archives of Clinical Neuropsychology*, *20*, 719-726.

- Meyers, J.E., Bayless, J.D., & Meyers, K.R. (1996). Rey Complex Figure: memory error patterns and functional abilities. *Applied Neuropsychology*, 3, 89-92.
- Meyers, J.E., Galinsky, A.M., & Volbrecht, M. (1999). Malingering and mild brain injury: How low is too low. *Applied Neuropsychology*, 6, 208-216.
- Meyers, J.W., & Meyers, K.R. (1995). *Rey Complex Figure Test and Recognition Trial. Professional manual*. Odessa: PAR.
- Meyers, J.E., Morrison, A.L., & Miller, J.C. (2001). How low is too low, revisited: Sentence repetition and AVLT-Recognition in the detection of malingering. *Applied Neuropsychology*, 8, 234-241.
- Meyers, J.E., & Volbrecht, M. (1998). Validation of reliable digits for detection of malingering. *Assessment*, 5, 303-307.
- Meyers, J.E., & Volbrecht, M. (1999). Detection of malingerers using the Rey Complex Figure and Recognition Trial. *Applied Neuropsychology*, 6, 201-207.
- Meyers, J.E., & Volbrecht, M.E. (2003). A validation of multiple malingering detection methods in a large clinical sample. *Archives of Clinical Neuropsychology*, 18, 261-276.
- Miller, A., Donders, J., & Suhr, J. (2000). Evaluation of malingering with the Wisconsin Card Sorting Test : A cross-validation. *Clinical Neuropsychological Assessment*, 2, 141-149.
- Millis, H., & Home, G. (1986). Whiplash: man-made disease. *New Zealand Medical Journal*, 99, 373-374.
- Millis, S.R., & Kler, S. (1995). Limitations of the Rey Fifteen-Item Test in the Detection of Malingering. *The Clinical Neuropsychologist*, 9, 241-244.
- Millis, S.R., & Putnam, S.H. (1994). The Recognition Memory Test in the assessment of memory impairment after financially compensable mild head injury: a replication. *Perceptual and Motor Skills*, 79, 384-386.
- Millis, S.R., & Putnam, S.H. (1996). Evaluation of malingering in the neuropsychological examination of mild head injury. *NeuroRehabilitation*, 7, 55-65.
- Millis, S.R., & Putnam, S.H. (1997). *The CVLT in the assessment of financially compensable mild head injury: Further development*. Paper presented at the meeting of the International Neuropsychological Society, Bergen, Norway.
- Millis, S.R., Putnam, S.H., Adams, K.M., & Ricker, J.H. (1995). The California Verbal Learning Test in the Detection of Incomplete Effort in Neuropsychological Evaluation. *Psychological Assessment*, 7, 463-471.
- Millon, T., Davis, R.D., & Millon, C. (1997). *Manual for the Millon Clinical Multiaxial Inventory-III (MCMI-III)*. Minneapolis, MN: National Computer Systems.
- Mittenberg, W., Aguila-Puentes, G., Patton, C., Canyock, E.M., & Heilbronner, R.L. (2002). Neuropsychological profiling of symptom exaggeration and malingering. *Journal of Forensic Neuropsychology*, 3, 227-240.
- Mittenberg, W., Azrin, R., Millsaps, C. & Heilbronner, R. (1993). Identification of malingered head injury on the Wechsler Memory Scale-Revised. *Psychological Assessment*, 5, 34-40.
- Mittenberg, W., Patton, C., Canyock, E.M., & Condit, D.C. (2002). Base rates of malingering and symptom exaggeration. *Journal of Clinical and Experimental Neuropsychology*, 24, 1094-1102.

- Mittenberg, W., Rotholz, A., Russell, E., & Heilbronner, R. (1996). Identification of malingered head injury on the Halstead-Reitan Battery. *Archives of Clinical Neuropsychology, 11*, 271-281.
- Mittenberg, W., Theroux-Fichera, S., Zielinski, R.E., & Heilbronner, R.L., (1995). Identification of Malingered Head Injury on the Wechsler Adult Intelligence Scale-Revised. *Professional Psychology: Research and Practice, 26*, 491-498.
- Moore, B.A., & Donders, J. (2004). Predictors of invalid neuropsychological test performance after traumatic brain injury. *Brain Injury, 18*, 975-984.
- Muñoz-Céspedes, J.M., Pelegrín-Valero, C., Tirapu-Ustarroz, J., & Fernández-Guinea, S. (1998). Sobre la naturaleza, diagnóstico y tratamiento del síndrome posconmocional. *Revista de neurología, 27*, 844-853.
- Nelson, N.W., Boone, K., Dueck, A., Wagener, L., Lu, P., & Grills, C. (2003). Relationships between eight measures of suspected effort. *The Clinical Neuropsychologist, 17*, 263-272.
- Newcombe, F., Rabbit, P., & Briggs, M. (1994). Minor head injury: pathophysiological or iatrogenic sequelae? *Journal of Neurology, Neurosurgery, and Psychiatry, 57*, 709-716.
- Nies, K.J., & Sweet, J.L. (1994). Neuropsychological assessment and malingering: A critical review of past and present strategies. *Archives of Clinical Neuropsychology, 9*, 501-552.
- O'Bryant, S.E., Duff, K., Fisher, J., & McCaffrey, R.J. (2004). Performance profiles and cut-off scores on the Memory Assessment Scales. *Archives of Clinical Neuropsychology, 19*, 489-496.
- O'Bryant, S.E., Hilsabeck, R.C., Fisher, J.M., & McCaffrey, R.J. (2003). Utility of the Trail Making Test in the assessment of malingering in a sample of mild traumatic brain injury litigants. *The Clinical Neuropsychologist, 17*, 69-74.
- Osimani, A., Alon, A., Berger, A., & Abarbanel, J.M. (1997). Use of the Stroop phenomenon as a diagnostic tool for malingering. *Journal of Neurology, Neurosurgery, and Psychiatry, 62*, 617-621.
- Pankratz, L. (1983). A new technique of the assessment and modification of feigned memory deficit. *Perceptual and Motor Skills, 57*, 367-372.
- Pankratz, L., Fausti, S.A., & Peed, S. (1975). A forced-choice technique to evaluate deafness in a hysterical or malingering patient. *Journal of Consulting and Clinical Psychology, 43*, 421-422.
- Patrick, C. J. & Iacono, W. G. (1991) A comparison of field and laboratory polygraphs in the detection of deception. *Psychophysiology, 28*, 632 –638.
- Pelegrín-Valero, C., Fernández-Guinea, S., Tirapu-Ustarroz, J., & Muñoz-Céspedes, J.M. (2001). Diagnóstico diferencial del síndrome posconmocional. *Revista de Neurología, 32*, 867-884.
- Pelfrey, W.V. (2004). The relationship between malingerers' intelligence and MMPI-2 knowledge and their ability to avoid detection. *International Journal of Offender Therapy and Comparative Criminology, 48*, 649-663.
- Powell, M.R., Gfeller, J.D., Hendricks, B.L., & Sharland, M. (2004). Detecting symptom- and test-coached simulators with the Test of Memory Malingering. *Archives of Clinical Neuropsychology, 19*, 693-702.
- Prigatano, G.P., & Amin, K. (1993). Digit memory test: Unequivocal cerebral dysfunction and suspected malingering. *Journal of Clinical and Experimental Neuropsychology, 15*, 537-546.
- Prigatano, G.P., Smason, I., Lamb, D.G., & Bortz, J.J. (1997). Suspected malingering and the Digit Memory Test: A replication and extension. *Archives of Clinical Neuropsychology, 12*, 609-619.

- Puente, A., Pérez, M. (2000^a). Neuropsychological assessment of ethnic minorities: Clinical issues. In I. Cuellar, & F.A. Paniagua. *Handbook of multicultural mental health* (pp. 419-435).
- Puente, A., Pérez M. (2000^b). Psychological assessment of ethnic minorities. In G. Goldstein, & M. Hersen. *Handbook of psychological assessment* (3rd Ed.) (pp. 527-551).
- Ramírez, R.M., Chirivella-Garrido, J., Caballero, M.C., Ferri-Campos, J., & Noé-Sebastián, E. (2004). Inteligencia, memoria y simulación: correlaciones entre instrumentos de medida. *Revista de Neurología*, 38, 28-33.
- Rapport, L.J., Farchione, T.J., Coleman, R.D., & Axelrod, B.N. (1998). Effects of coaching on malingered motor function profiles. *Journal of Clinical and Experimental Neuropsychology*, 20, 89-97.
- Raskin, S.A. (2000). Executive functions. In S.A. Raskin & C.A. Mateer (Eds.), *Neuropsychological management of mild traumatic brain injury* (pp. 113-133). New York: Oxford University Press.
- Raskin, S., & Rearick, E. (1996). Verbal Fluency in Individuals with Mild Traumatic Brain Injury. *Neuropsychology*, 10, 416-422.
- Rees, L.M., Tombaugh, T.N., & Boulay, L. (2001). Depression and the Test of Memory Malingering. *Archives of Clinical Neuropsychology*, 16, 501-506.
- Rees, L.M., Tombaugh, T.N., Gansler, D.A., & Moczynski, N.P. (1998). Five validation experiments of the Test of Memory Malingering (TOMM). *Psychological Assessment*, 10, 10-20.
- Reitan, R.M., & Wolfson, D. (1995^a). Consistency of responses on retesting among head-injured subjects in litigation versus head-injured subjects not in litigation. *Applied Neuropsychology*, 2, 67-71.
- Reitan, R.M., & Wolfson, D. (1995^b). *The Halstead-Reitan Neuropsychological Test Battery: Theory and clinical interpretation* (2nd ed.). Tucson: Neuropsychology Press.
- Reitan, R.M., & Wolfson, D. (1996). The question of validity of neuropsychological test scores among head-injured litigants: Development of a Dissimulation Index. *Archives of Clinical Neuropsychology*, 11, 573-580.
- Reitan, R.M., & Wolfson, D. (1997). Consistency of neuropsychological test scores of head-injured subjects involved in litigation compared with head-injured subjects not involved in litigation: development of the Retest Consistency Index. *The Clinical Neuropsychologist*, 11, 69-76.
- Reitan, R.M., & Wolfson, D. (1998). Detection of Malingering and invalid test results using the Halstead-Reitan Battery. In *Detection of Malingering During Head Injury Litigation*, Reynolds (Ed.), Plenum Press, New York.
- Rey, A. (1964). *L'examen clinique en psychologie*. Paris : Presses Universitaires de France.
- Reynolds, C.R. (1998). Common sense, clinicians, and actuarialism in the detection of malingering during head injury litigation. In C.R. Reynolds (Ed.), *Detection of Malingering during Head Injury Litigation*. New York: Plenum Press.
- Reznek, L. (2005). The Rey 15-item memory test for malingering: A meta-analysis. *Brain Injury*, 19, 539-543.
- Richardson, J.T.E. (2000). *Clinical and neuropsychological aspects of closed head injury* (2nd Ed). Hove: Psychology Press.
- Rogers, R. (1984). Towards an empirical model of malingering and deception. *Behavioral Sciences and the Law*, 2, 93-112.

- Rogers, R. (1990). Models of feigned mental illness. *Professional Psychology: Research and Practice*, *21*, 182-188.
- Rogers, R. (1996). *Test of Cognitive Abilities*. Unpublished test, University of North Texas.
- Rogers, R. (1997^a). Introduction. In R. Rogers (Ed.). *Clinical Assessment of Malingering and Deception* (2nd ed.) (pp. 1-19). New York: The Guilford Press.
- Rogers, R. (1997^b). Researching dissimulation. In R. Rogers (Ed.). *Clinical Assessment of Malingering and Deception* (2nd ed.) (pp. 398-426). New York: The Guilford Press.
- Rogers, R., Bagby, R.M., & Dickens, S.E. (1992). *Structured Interview of Reported Symptoms (SIRS) and professional manual*. Odessa, FL: Psychological Assessment Resources.
- Rogers, R., & Bender, S.D. (2003). Evaluation of malingering and deception. In A.M. Goldstein (Ed.), *Handbook of Psychology, Volume 11* (pp. 109-129). New York: CUNY.
- Rogers, R., & Cavanaugh, J.L. (1983). "Nothing but the truth"... a reexamination of malingering. *The Journal of Psychiatry & Law*, *11*, 443-459.
- Rogers, R., & Cruise, K.R. (1998). Assessment of malingering with simulation designs: Threats to external validity. *Law and Human Behavior*, *22*, 273-285.
- Rogers, R., Harrell, E.H., & Liff, C.D. (1993). Feigning neuropsychological impairment: A critical review of methodological and clinical considerations. *Clinical Psychology Review*, *13*, 255-274.
- Rogers, R., Sewell, K.W., & Goldstein, A. (1994). Explanatory models of malingering: A prototypical analysis. *Law and Human Behaviour*, *18*, 543-552.
- Rohling, M.L., Green, P., Allen, L., & Iverson, G.L. (2002). Depressive symptoms and neurocognitive test scores in patients passing symptom validity tests. *Archives of Clinical Neuropsychology*, *17*, 205-222.
- Rose, F.E., Hall, S., & Szalda-Petree, A.D. (1995). Portland Digit Recognition Test-computerized: The measurement of response latency improves the detection of malingering. *The Clinical Neuropsychologist*, *9*, 124-134.
- Rose, J.E., Hall, S., Szalda-Petree, A.D., & Bach, P.J. (1998). A comparison of four tests of malingering and the effects of coaching. *Archives of Clinical Neuropsychology*, *13*, 349-363.
- Rosenfeld, J.P., Ellwanger, J.W., Nolan, K., Wu, S., Berman, R.G., & Sweet, J. (1999). P300 Scalp amplitude distribution as an index of deception in a simulated cognitive deficit model. *International Journal of Psychophysiology*, *33*, 3-19.
- Ross, S.R., Krukowski, R.A., Putnam, S.H., & Adams, K.M. (2003). The Memory Assessment Scales in the detection of incomplete effort in mild head injury. *The Clinical Neuropsychologist*, *17*, 581-591.
- Ruff, R., Levin, H., Mattis, S., High, W., Marshall, L., Eisenberg, H. & Tabbador, K. (1989). Recovery of memory after mild head injury: a three-center study. In H. Levin, H. Eisenberg, & A. Benton (Eds.), *Mild Head Injury*. New York, NY: Oxford University Press, pp 176-188.
- Sbordone, R.J., & Saul, R.E (Eds.), *Neuropsychology for health care professionals and attorneys* (2nd ed.). Boca Raton, Fla.: St. Lucie Press (2000).
- Schacter, D., & Crovitz, H. (1977). Memory function after closed head injury: a review of the quantitative research. *Cortex*, *13*, 105-176.

- Schagen, S., Schmand, B., Sterke, S., & Lindeboom, J. (1997). Amsterdam Short-Term Memory Test: A new procedure for the detection of feigned memory deficits. *Journal of Clinical and Experimental Neuropsychology, 19*, 43-51.
- Schoenberg, M.R., Dorr, D., & Morgan, C. (2003). The ability of the Millon Clinical Multiaxial Inventory-Third Edition to detect malingering. *Psychological Assessment, 15*, 198-204.
- Schretlen, D., Brandt, J., Krafft, L., & Van Gorp, W. (1991). Some Caveats in Using the Rey 15-Item Memory Test to Detect Malingered Amnesia. *Psychological Assessment, 3*, 667-672.
- Schwartz, S.M., Gramling, S.E., Kerr, K.L., & Morin, C. (1998). Evaluation of intellect and deficit specific information on the ability to fake memory deficits. *International Journal of Law and Psychiatry, 21*, 261-272.
- Seisdedos, N. (2000). *Cambios. Test de flexibilidad cognitiva*. Madrid: TEA.
- Slick, D.J., Hopp, G., Strauss, E., Hunter, M., & Pinch, D. (1994). Detecting dissimulation: Profiles of simulated malingerers, traumatic brain-injury patients, and normal controls on a revised version of Hiscock and Hiscock's Forced-Choice Memory Test. *Journal of Clinical and Experimental Neuropsychology, 16*, 472-481.
- Slick, D.J., Hopp, G., Strauss, E., & Spellacy, F.J. (1996). Victoria Symptom Validity Test: Efficiency for detecting feigned memory impairment and relationship to neuropsychological tests and MMPI-2 validity scales. *Journal of Clinical and Experimental Neuropsychology, 18*, 911-922.
- Slick, D., Hopp, G., Strauss, E., & Thompson, G. (1997). *The Victoria Symptom Validity Test*. Odessa: PAR.
- Slick, D.J., Sherman, E.M.S., & Iverson, G.L. (1999). Diagnostic Criteria for Malingered Neurocognitive Dysfunction: Proposed Standards for Clinical Practice and Research. *The Clinical Neuropsychologist, 13*, 545-561.
- Slick, D.J., Tan, J.E., Strauss, E., Mateer, C.A., Harnadek, & Sherman, E.M.S. (2003). Victoria Symptom Validity Test scores of patients with profound memory impairment: NonLitigant case studies. *The Clinical Neuropsychologist, 17*, 390-394.
- Smith, G.P., & Burger, G.K. (1997). Detection of malingering: validation of the Structured Inventory of Malingered Symptomatology (SIMS). *Journal of the American Academy of Psychiatry and the Law, 25*, 183-189.
- Spreen, O., & Strauss, E. (1998). *A compendium of neuropsychological tests and norms*. New York: Oxford University Press.
- Stablum, R., Mogentale, C., & Umilta, C. (1996). Executive functioning following mild closed head injury. *Cortex, 32*, 261-278.
- Strauss, E., Hultsch, D.F., Hunter, M., Slick, D.J., Patry, B., & Levy-Bencheton, J. (2000). Using intraindividual variability to detect malingering in cognitive performance. *The Clinical Neuropsychologist, 14*, 420-432.
- Strauss, E., Slick, D.J., Levy-Bencheton, J., Hunter, M., MacDonald, S.W.S., & Hultsch, D.F. (2002). Intraindividual variability as an indicator of malingering in head injury. *Archives of Clinical Neuropsychology, 17*, 423-444.
- Stuss, D., & Gow, C. (1992). "Frontal dysfunction" after traumatic brain injury. *Neuropsychiatry, Neuropsychology and Behavioural Neurology, 5*, 272-282.
- Surh, J.A. (2002). Malingering, coaching, and the serial position effect. *Archives of Clinical Neuropsychology, 17*, 69-77.

- Surh, J.A., & Boyer, D. (1999). Use of the Wisconsin Card Sorting Test in the detection of malingering in student simulator and patient samples. *Journal of Clinical and Experimental Neuropsychology*, *21*, 701-708.
- Surh, J.A., & Gunstad, J. (2000). The effects of coaching on the sensitivity and specificity of malingering measures. *Archives of Clinical Neuropsychology*, *15*, 415-424.
- Surh, J., Tranel, D., Wefel, J., & Barrash, J. (1997). Memory performance after head injury: Contributions of malingering, litigation status, psychological factors, and medication use. *Journal of Clinical and Experimental Neuropsychology*, *19*, 500-514.
- Sweet, J.J., King, F.H., Malina, A.C., Bergman, M.A., & Simmons, A. (2002). Documenting the prominence of forensic neuropsychology at national meetings and in relevant professional journals from 1990 to 2000. *The Clinical Neuropsychologist*, *16*, 481-494.
- Sweet, J.J., Moberg, P.J., & Suchi, Y. (2000). Ten-Year follow-up survey of clinical neuropsychologists: Part I: Practices and Beliefs. *The Clinical Neuropsychologist*, *14*, 18-37.
- Sweet, J.J., Peck, E.A., Abramowitz, C., & Etzweiler, S. (2002). National Academy of Neuropsychology/Division 40 of the American Psychological Association practice survey of clinical neuropsychology in the United States, Part I: Practitioner and practice characteristics, professional activities, and time requirements. *The Clinical Neuropsychologist*, *16*, 109-127.
- Sweet, J.J., Wolfe, P., Sattlberger, E., Numan, B., Rosenfeld, J.P., Clingerman, S., & Nies, K.J. (2000). Further Investigation of Traumatic Brain Injury Versus Insufficient Effort with the California Verbal Learning Test. *Archives of Clinical Neuropsychology*, *15*, 105-113.
- Tan, J.E., Slick, D.J., Strauss, E., & Hultsch, D.F. (2002). How'd they do it? Malingering strategies on Symptom Validity Tests. *The Clinical Neuropsychologist*, *16*, 495-505.
- Tardiff, H.P., Barry, R.J., Fox, A.M., & Johnstone, S.J. (2000). Detection of feigned recognition memory impairment using the old new effect of the event-related potential. *International Journal of Psychophysiology*, *36*, 1-9.
- Tardiff, H.P., Barry, R.J., & Johnstone, S.J. (2002). Event-related potentials reveal processing differences in honest vs. malingered memory performance. *International Journal of Psychophysiology*, *46*, 147-158.
- Taylor, L.A., Kreutzer, J.S., & West, D.D. (2003). Evaluation of malingering cut-off scores for the Rey 15-Item Test: a brain injury case study series. *Brain Injury*, *4*, 295-308.
- Teichner, G., & Wagner, M.T. (2004). The Test of Memory Malingering (TOMM): normative data from cognitively intact, cognitively impaired, and elderly patients with dementia. *Archives of Clinical Neuropsychology*, *19*, 455-464.
- Tenhula, W.N., & Sweet, J.J. (1996). Double cross-validation of the Booklet Category Test in detecting malingered traumatic brain injury. *The Clinical Neuropsychologist*, *10*, 104-116.
- Thomas, J., & Wanlass, R. (1994). Development of individual and family assessment instruments for the identification of neuropsychological malingering. *Archives of Clinical Neuropsychology*, *9*, 191.
- Thompson, G.B. (2002). The Victoria Symptom Validity Test: An enhanced test of symptom validity. *Journal of Forensic Neuropsychology*, *2*, 43-67.
- Thompson, L.L., & Cullum, C.M. (1991). Pattern of performance on neuropsychological tests in relation to effort in mild head injury patients. *Archives of Clinical Neuropsychology*, *6*, 231.
- Tombaugh, T.N. (1996). *Test of Memory Malingering, TOMM*. New York/Toronto: MHS.

- Trueblood, W. (1994). Qualitative and quantitative characteristics of malingered and other invalid WAIS-R and clinical memory data. *Journal of Clinical and Experimental Neuropsychology*, *16*, 597-607.
- Trueblood, W., & Binder, L.M. (1995, October). *Psychologists' accuracy in identifying neuropsychological test protocols of clinical malingerers*. Trabajo presentado en la National Academy of Neuropsychology, San Francisco.
- Trueblood, W., & Binder, L.M. (1997). Psychologists' accuracy in identifying neuropsychological test protocols of clinical malingerers. *Archives of Clinical Neuropsychology*, *12*, 13-27.
- Trueblood, W., & Schmidt, M. (1993). Malingering and other validity considerations in the neuropsychological evaluation of mild head injury. *Journal of Clinical and Experimental Neuropsychology*, *15*, 578-590.
- Umile, E.M., Plotkin, R.C., & Sandel, M.E. (1998). Functional assessment of mild traumatic brain injury using SPECT and neuropsychological testing. *Brain Injury*, *12*, 577-594.
- Van Gorp, W.G., Humphrey, L.A., Kalechstein, A., Brumm, V.L., McMullen, W.J., Stoddard, M., & Pachana, N.A. (1999). How well do standard clinical neuropsychological tests identify malingering? A preliminary analysis. *Journal of Clinical and Experimental Neuropsychology*, *21*, 245-250.
- Varney, N.R., Bushnell, D.L., Nathan, M., Kahn, D., Roberts, R., Rezai, K., et al. (1995). NeuroSPECT correlates of disabling mild head injury: preliminary findings. *Journal of Head Trauma Rehabilitation*, *10*, 18-28.
- Verdejo-García, A., Alcázar, M.A., Gómez-Jarabo, G., & Pérez-García, M. (2004). Pautas para el desarrollo científico y profesional de la neuropsicología forense. *Revista de neurología*, *39*, 60-73.
- Vickery, C.D., Berry, D.T.R., Dearth, C.S., Vagnini, V.L., Baser, R.E., Cragar, D.E., & Orey, S.A. (2004). Head injury and the ability to feign neuropsychological deficits. *Archives of Clinical Neuropsychology*, *19*, 37-48.
- Vickery, C.D., Berry, D.T.R., Inman, T.H., Harris, M.J., & Orey, S.A. (2001). Detection of inadequate effort on neuropsychological testing: A meta-analytic review of selected procedures. *Archives of Clinical Neuropsychology*, *16*, 45-73.
- Vila, J. (1996). Una introducción a la psicofisiología clínica. Madrid: Pirámide.
- Vilar, R., Gómez-Río, M., Verdejo-García, A., Llamas-Elvira, J.M., & Pérez-García, M. (In press). Detection of malingering in a Spanish population using three specific malingering tests. *Archives of Clinical Neuropsychology*.
- Vos, P.E., Battistin, L., Birbamer, G., Gerstenbrand, F., Potapov, A., Prevec, T. et al. (2002). EFNS guideline in mild traumatic brain injury : report of an EFNS task force. *European Journal of Neurology*, *9*, 207-219.
- Wade, D.T., King, N.S., Wenden, F.J., Crawford, S., & Caldwell, F.E. (1998). Routine follow up after head injury: a second randomised controlled trial. *Journal of Neurology, Neurosurgery and Psychiatry*, *65*, 177-183.
- Wang, Y., Chan, R.C.K., & Deng, Y. (2006). Examination of postconcussion-like symptoms in healthy university students: Relationships to subjective and objective neuropsychological function performance. *Archives of Clinical Neuropsychology*, *21*, 339-347.
- Warrington, E.K. (1984). *Recognition Memory Test*. Windsor, UK: NFER-Nelson.
- Weschler, D. (1945). A standardized memory scale for clinical use. *Journal of Psychology*, *19*, 87-95.

- Weschler, D. (1981). *Weschler Adult Intelligence Scale, revised*. New York: Psychological Corporation.
- Weschler, D. (1987). *The Wechsler Memory Scale-Revised*. New York: Psychological Corporation.
- Weschler, D. (1997). *The Wechsler Memory Scale-III*. San Antonio: Psychological Corporation.
- Weschler, D. (1997^b). *The Wechsler Adult Intelligence Scale-III*. San Antonio: Psychological Corporation.
- Wetter, M.W., & Corrigan, S.K. (1995). Providing information to clients about psychological tests: A survey of attorney's and law students' attitudes. *Professional Psychology: Research and Practice*, 26, 474-477.
- Williams, M.J. (1991). *Memory Assessment Scales (MAS): Professional Manual*. Odessa: Psychological Assessment Resources.
- Willison, J., & Tombaugh, T.N. (2006). Detecting simulation of attention deficits using reaction times tests. *Archives of Clinical Neuropsychology*, 21, 41-52.
- Willson, J.T., Wiedmann, K.D., Hadley, D.M., Condon, B., Teasdale, G., & Brooks, D.N. (1988). Early and late magnetic resonance imaging and neuropsychological outcome after head injury. *Journal of Neurology, Neurosurgery and Psychiatry*, 51, 391-396.
- World Health Organization (1992). *International Statistical Classification of Diseases and Related Health Problems, 10th revision (ICD-10)*. Geneva: World Health Organization.
- Wynkoop, T.F., & Denney, R.L. (2001). Computerized Assessment of Response Bias (CARB), Word Memory Test (WMT), and Memory Complaints Inventory (MCI). *Journal of Forensic Neuropsychology*, 2, 71-77.
- Yanez, Y.T., Fremouw, W., Tennant, J., Strunk, J., & Coker, K. (2006). Effects of severe depression on TOMM performance among disability-seeking outpatients. *Archives of Clinical Neuropsychology* 21, 161-165.
- Youngjohn, J. R. (1995). Confirmed attorney coaching prior to neuropsychological evaluation. *Assessment*, 2, 279-283.
- Youngjohn, J.R., Burrows, L., & Erdal, K. (1995). Brain damage or compensation neurosis? The controversial post-concussion syndrome. *Clinical Neuropsychologist*, 9, 112-123.
- Youngjohn, J.R., Lees-Haley, P.R., & Binder, L.M. (1999). Comment: Warning malingerers produces more sophisticated malingering. *Archives of Clinical Neuropsychology* 14, 511-515.
- Zakzanis, K.K. (2001). Statistics to tell the truth, the whole truth and nothing but the truth. *Archives of Clinical Neuropsychology*, 16, 653-667.

APÉNDICE 1
LA DOBLE TRADUCCIÓN CRUZADA

APÉNDICE 1.1. INSTRUCCIONES ORIGINALES**VSVT:**

You will be taking a test of memory that requires concentration. People with memory problems often find this test to be difficult. On this screen you will see a five-digit number. You will have 5 seconds to look at the number. Study the number for the full 5 seconds so that you can recognize it when you see it again. Five seconds after the number disappears, you will see two more five-digit numbers on the screen, one to the left and one to the right. One of the two numbers will always be the same as the number you saw just before. You must always choose one of the numbers, even if you have to guess. As soon as you have decided which number was the one you saw before, you will press one of two keys. Press the <Z> key if the number on the left is the same as the one you saw just before. Press the forward slash key (</>) if the number on the right is the same as the one you saw before. After you press the key, a new number to remember will appear. Do not worry if you have to guess or if you make a mistake; just concentrate on the next number to remember. Take as much time as you need to decide which number is the one you saw just before, but you should respond as soon as you have decided. Just try to decide as quickly as you can without making mistakes. During the test, rest the index finger of each hand on the <Z> and forward slash keys at all times. Do you have any questions about what you are to do?

You may find the next section of the test more difficult because you will have to remember the number for 10 seconds instead of 5.

You may find the last section of this test to be the most difficult because you will have to remember the number for 15 seconds instead of 10.

TOMM:

This is a test of your ability to learn and to remember pictures of common objects. First, I'll show you 50 pictures, one at a time. Then, I'll give you a chance to see how many of them you can remember.

Let's try a sample trial first that contains only two pictures. Look carefully at each picture and try to remember it. You don't have to learn the name of each picture. Just look at each one and try to remember it.

Point to the picture I showed you before. Was it this one or this one?

I'm going to show you 50 pictures, one at a time. Look at each one and try to remember it. Do you have any questions before we begin?

Point to the picture I showed you previously.

Correct.

No, that's not right. It was this one.

I'm going to show you the same 50 pictures again. Look at each one and try to remember it. Any questions?

Remember the booklet I showed you earlier that contained the 50 pictures. Let's see how many you can remember.

The b Test:

I want you to circle all the letter "b"s on these pages as quickly as you can. The letters you circle should look exactly like this.

You may find that the test becomes more difficult as you go from page to page, but try to do your best.

Please open the booklet and begin.

Rey Fifteen-Item Memory Test:

I am going to show you a card with 15 things on it to remember. When I take the card away, I want you to write down as many of the 15 things as you can remember. Be sure to look at all of them. (Goldberg y Miller, 1986).

APÉNDICE 1.2. TRADUCCIÓN DE LAS INSTRUCCIONES AL CASTELLANO

VSTV:

Usted realizará una prueba de memoria que requiere concentración. Normalmente, la gente con problemas de memoria lo encuentra difícil. En la pantalla verá un número de cinco dígitos. Dispondrá de 5 segundos para mirar el número. Estudie el número durante estos 5 segundos completos para que pueda reconocerlo cuando lo vuelva a ver. Cinco segundos después de que el número desaparezca, verá otros dos números de cinco dígitos más en la pantalla, uno a la izquierda y otro a la derecha. Uno de los dos números siempre será el mismo número vio antes. Debe escoger uno de los números, incluso si tiene que adivinar. Tan pronto haya decidido qué número era el que vio antes, pulsará una de las dos teclas. Pulse la tecla <Z> si el número a la izquierda es el mismo que el que acabó de ver antes. Pulse la tecla de barra oblicua (</>) si el número a la derecha es el mismo que el que vio antes. Después de pulsar la tecla, aparecerá un nuevo número que deberá recordar. No se preocupe si tiene que adivinar o si comete un error, tan sólo concéntrese en el siguiente número que debe recordar. Tome tanto tiempo como necesite para decidir qué número es el que acabó de ver antes, pero debe responder tan pronto lo haya decidido. Sólo intente decidir tan rápido como pueda sin cometer errores. Durante la prueba, apoye en todo momento el dedo índice de cada mano sobre la <Z> y (</>). ¿Tiene alguna pregunta sobre lo que tiene que hacer? Es posible que la siguiente sección de la prueba le resulte más difícil ya que tendrá que recordar el número durante 10 segundos en vez de 5.

Es posible que la última sección le resulte la más difícil ya que tendrá que recordar el número en 15 segundos en lugar de 10.

TOMM:

Esta es una prueba para medir su capacidad para aprender y recordar dibujos de objetos comunes. En primer lugar, le mostraré 50 dibujos, uno a uno. Posteriormente, le daré una oportunidad para ver cuántos puede recordar.

Hagamos primero una prueba que contiene sólo dos dibujos. Mire atentamente a cada dibujo e intente recordarlo. No tiene que aprender el nombre de cada dibujo. Tan sólo mire cada uno de ellos e intente recordarlo.

Señale el dibujo que le mostré anteriormente. ¿Era éste o éste?

Le voy a mostrar 50 dibujos, uno a uno. Mire cada uno de ellos e intente recordarlos.

¿Tiene alguna pregunta antes de que empecemos?

Señale el dibujo que le mostré anteriormente.

Correcto.

No, eso no es correcto. Era éste.

Le voy a mostrar de nuevo 50 dibujos. Mire a cada uno de ellos e intente recordarlos.

¿Tiene alguna pregunta?

Recuerde el cuadernillo que le mostré antes que contenía 50 dibujos. Vamos a ver cuántos puede recordar.

La Prueba b:

Quiero que haga un círculo alrededor de todas las letras “b” en estas páginas tan rápido como pueda. Las letras que rodee deben tener exactamente el mismo aspecto que ésta.

Es posible que le resulte más difícil la prueba a medida que avanza de página, pero intente hacerlo lo mejor que pueda.

Por favor, abra el cuadernillo y empiece.

Test de 15-Ítems de Rey:

Le voy a mostrar una tarjeta con 15 cosas que debe recordar. Cuando retire la tarjeta, quiero que escriba tantas de las 15 cosas como recuerde. Asegúrese de que las mira todas (Goldberg y Miller, 1986)

APÉNDICE 1.3. TRADUCCIÓN AL INGLÉS

VSTV:

You are going to take a memory test that requires concentration. People with memory problems usually find this test difficult. On the screen you will see a five digit number.

You will have 5 seconds to look at the number. Study the number for the entire 5 seconds, so that you will recognize it when you see it again. Five seconds after the number disappears, you will see two more 5 digit numbers on the screen, one on the left and one on the right. One of the two numbers will always be the same number you saw earlier. You must choose one of the two numbers, even if you have to guess. As soon as you have decided which of the two numbers you saw before, press one of the two keys. Press the <Z> key if the number on the left is the same one you saw earlier. Press the slanted bar key (</>) if the number on the right is the same one you saw before. After pressing the key, a new number will appear that you will have to remember. Do not worry if you have to guess or if you make a mistake, just concentrate on the next number you have to remember. Take as much time as you need to decide which number you saw before, but you must respond as soon as you have decided. Just try to decide as quickly as you can without making mistakes. During the test, always rest your index finger of each hand on the <Z> and (</>) keys. Do you have any questions about what you have to do?

It is possible that the next section of the test will seem more difficult, as you will have to remember the number for 10 seconds instead of 5.

It is possible that the final section of the test will seem more difficult, as you will have to remember the number for 15 seconds instead of 10.

TOMM:

This is a test for measuring your ability to learn and remember drawings of everyday objects. In the first place, I will show you 50 drawings, one by one. Then, I will give you a chance to see how many you can remember.

First we will do a practice trial that only has two drawings. Look closely at each drawing and try to remember it. You do not have to learn the name of each drawing. Just look at each of them and try to remember them.

Point to the drawing I showed you earlier. Was it this one or this one?

I am going to show you 50 drawings, one by one. Look at each of them and try to remember them. Do you have any questions before we begin?

Point to the drawing I showed you earlier.

Correct.

No, that is not correct. It was this one.

I am going to show you 50 new drawings. Look at each of them and try to remember them. Do you have any questions?

Think about the booklet I showed you earlier that contained 50 drawings. Let's see how many you can remember.

The b Test:

I want you to make a circle around all the letters "b" on these pages, as fast as you can.

The letters you circle must look exactly like this one.

It is possible that the test will seem more difficult as you go through the pages, but try to do the test as well as you can.

Please, open the booklet and begin.

15-Item Rey Test:

I am going to show you a card with 15 things you must remember. When I take the card away, I want you to describe as many of the 15 things as you can remember. Make sure you look at all of them (Goldberg and Miller, 1986)

APÉNDICE 2

LAS INSTRUCCIONES PROPORCIONADAS A LOS ANÁLOGOS

APÉNDICE 2.1. INSTRUCCIONES TIPO A

Imagina que has tenido un accidente de coche. Intenta realizar los test de modo que convenzas al examinador de que has sufrido un daño cerebral como consecuencia de dicho accidente. Para hacerlo bien, debes intentar hacer que los problemas sean lo más creíbles posible. La gente que ha tenido un golpe en la cabeza generalmente experimenta el llamado síndrome post-comnacional: dolores de cabeza, visión borrosa, fatiga, mareos, ansiedad, irritabilidad, y problemas de concentración y memoria. En la mayoría de los casos estos síntomas no duran más de 1 mes, pero alguna gente los experimenta 6 meses o 1 año después del golpe. Esta gente tiene problemas a la hora de prestar atención, recordar cosas, y aprender material nuevo. Además, su pensamiento es un poco más lento de lo que solía ser.

Tu función consiste en realizar los test de modo que convenzas al examinador de que has sufrido un daño cerebral permanente a raíz del accidente. Tienes que simular los problemas del modo más severo que puedas sin hacerlos demasiado obvios para el examinador. Esto significa que tu daño cerebral debe ser creíble. Algunos síntomas demasiado exagerados, como no recordar absolutamente nada, son fáciles de detectar. Si el examinador cree que no tienes ningún problema no ganarás tu juicio, y podrás ser encarcelado. Si por el contrario consigues engañar al experimentador, ganarás el juicio y obtendrás una suculenta suma de dinero.

Recuerda que si superas los test con éxito conseguirás el doble de puntos en esta práctica. Los nombres y apellidos de las 5 personas que mejor hayan engañado al experimentador serán presentados en la puerta 386. Además, también se expondrán públicamente los nombres y apellidos de la persona que peor haya simulado.

Nombre y apellidos: _____

APÉNDICE 2.2. INSTRUCCIONES TIPO B

Imagina que has tenido un accidente de coche. Intenta realizar los test de modo que convenzas al examinador de que has sufrido un daño cerebral como consecuencia de dicho accidente. Para hacerlo bien, debes intentar hacer que los problemas sean lo más creíbles posible. La gente que ha tenido un golpe en la cabeza generalmente experimenta el llamado síndrome post-comnacional: dolores de cabeza, visión borrosa, fatiga, mareos, ansiedad, irritabilidad, y problemas de concentración y memoria. En la mayoría de los casos estos síntomas no duran más de 1 mes, pero alguna gente los experimenta 6 meses o 1 año después del golpe. Esta gente tiene problemas a la hora de prestar atención, recordar cosas, y aprender material nuevo. Además, su pensamiento es un poco más lento de lo que solía ser.

Tu función consiste en realizar los test de modo que convenzas al examinador de que has sufrido un daño cerebral permanente a raíz del accidente. Tienes que simular los problemas del modo más severo que puedas sin hacerlos demasiado obvios para el examinador. Esto significa que tu daño cerebral debe ser creíble. Algunos síntomas demasiado exagerados, como no recordar absolutamente nada, son fáciles de detectar. Si el examinador cree que no tienes ningún problema no ganarás tu juicio, y podrás ser encarcelado. Si por el contrario consigues engañar al experimentador, ganarás el juicio y obtendrás una suculenta suma de dinero.

ADVERTENCIA: Algunos test tienen métodos para detectar a las personas que intentan simular. Algunos de estos métodos son los siguientes:

- Detectar si realizas mejor los ítems fáciles que los difíciles
- Detectar cuánto tiempo necesitas para realizar cada ítem y compararlo con la dificultad del ítem
- Compara tu habilidad en ítems semejantes
- Detectar si realizas errores particulares, especialmente en ítems muy fáciles
- Detectar si tus errores coinciden con los de la mayoría de la gente
- Detectar si cometes más errores que los esperados por el azar sólo

No te preocupes por estos métodos, sólo intenta hacerlo lo mejor que puedas.

Recuerda que si superas los test con éxito conseguirás el doble de puntos en esta práctica. Los nombres y apellidos de las 5 personas que mejor hayan engañado al experimentador serán presentados en la puerta 386. Además, también se expondrán públicamente los nombres y apellidos de la persona que peor haya simulado.

Nombre y apellidos: _____

APÉNDICE 3
EL INFORME POSTERIOR A LA EVALUACIÓN

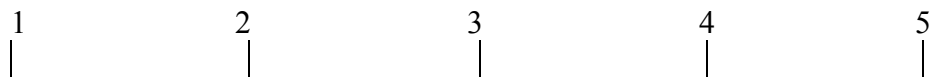
Nombre y apellidos:

Tipo de instrucciones: A / B

- ¿Qué se te ha pedido que hagas? ¿Cuáles fueron las instrucciones?

- ¿Qué significan estas instrucciones para ti?

- ¿Qué porcentaje de tiempo seguiste *cuidadosamente* las instrucciones?



1= todo el tiempo

2= casi todo el tiempo

3= algunas veces

4= casi nunca

5= nunca

- ¿Hasta qué punto crees que has logrado engañar al experimentador?



1= estoy seguro de que he conseguido engañarlo

2= estoy casi seguro de que he conseguido engañarlo

3= no sé si he conseguido engañarlo

4= creo que no he conseguido engañarlo

5= estoy seguro de que no he podido engañarlo

- ¿Qué métodos has seguido para engañarlo?

ANEXO 1

**PUNTOS DE CORTE, SENSIBILIDADES Y ESPECIFICIDADES DE LAS
PRUEBAS ESPECÍFICAS DE SIMULACIÓN PARA DISTINTAS
PREVALENCIAS DE BASE**

Anexo 1.1.1. Sensibilidad, especificidad y valores predictivos positivos y negativos asociados a un porcentaje de base del **15%** para diversos puntos de corte de las variables del **VSVT**

Variable	Punto de corte	Sensibilidad	IC 95%	Especificidad	IC 95%	VP+	VP-
Ítems fáciles	< 15	0,0	0,0 - 37,1	100,0	87,5 - 100,0		85,0
	<=20	50,0	16,0 - 84,0	100,0	87,5 - 100,0	100,0	91,9
	<=22	50,0	16,0 - 84,0	92,9	76,5 - 98,9	55,3	91,3
	<=23	62,5	24,7 - 91,0	82,1	63,1 - 93,9	38,2	92,5
	<=24	100,0	62,9 - 100,0	0,0	0,0 - 12,5	15,0	
Ítems difíciles	< 4	0,0	0,0 - 37,1	100,0	87,5 - 100,0		85,0
	<=13	62,5	24,7 - 91,0	100,0	87,5 - 100,0	100,0	93,8
	<=16	62,5	24,7 - 91,0	92,9	76,5 - 98,9	60,7	93,3
	<=17	75,0	35,0 - 96,1	89,3	71,7 - 97,6	55,3	95,3
	<=18	87,5	47,4 - 97,9	85,7	67,3 - 95,9	51,9	97,5
	<=19	87,5	47,4 - 97,9	75,0	55,1 - 89,3	38,2	97,1
	<=21	100,0	62,9 - 100,0	67,9	47,7 - 84,1	35,4	100,0
	<=24	100,0	62,9 - 100,0	0,0	0,0 - 12,5	15,0	
Ítems totales	< 21	0,0	0,0 - 37,1	100,0	87,5 - 100,0		85,0
	<=38	75,0	35,0 - 96,1	100,0	87,5 - 100,0	100,0	95,8
	<=40	75,0	35,0 - 96,1	92,9	76,5 - 98,9	64,9	95,5
	<=41	87,5	47,4 - 97,9	89,3	71,7 - 97,6	59,0	97,6
	<=43	87,5	47,4 - 97,9	71,4	51,3 - 86,7	35,1	97,0
	<=45	100,0	62,9 - 100,0	57,1	37,2 - 75,5	29,2	100,0
	<=48	100,0	62,9 - 100,0	0,0	0,0 - 12,5	15,0	

Variable	Punto de corte	Sensibilidad	IC 95%	Especificidad	IC 95%	VP+	VP-
TR ítems fáciles	>=0,69	100,0	62,9-100,0	0,0	0,0 - 13,3	15,0	
	>1,44	100,0	62,9 - 100,0	53,8	33,4 - 73,4	27,7	100,0
	>1,5	87,5	47,4 - 97,9	53,8	33,4 - 73,4	25,1	96,1
	>1,91	87,5	47,4 - 97,9	76,9	56,3 - 91,0	40,1	97,2
	>2	75,0	35,0 - 96,1	76,9	56,3 - 91,0	36,4	94,6
	>2,55	75,0	35,0 - 96,1	88,5	69,8 - 97,4	53,4	95,2
	>3,75	25,0	3,9 - 65,0	88,5	69,8 - 97,4	27,7	87,0
	>4,47	25,0	3,9 - 65,0	96,2	80,3 - 99,4	53,4	87,9
	>6,58	12,5	2,1 - 52,6	96,2	80,3 - 99,4	36,4	86,2
	>7,2	12,5	2,1 - 52,6	100,0	86,7 - 100,0	100,0	86,6
	>7,39	0,0	0,0 - 37,1	100,0	86,7 - 100,0		85,0
TR ítems difíciles	>=0,92	100,0	62,9 - 100,0	0,0	0,0 - 13,3	15,0	
	>2,19	100,0	62,9 - 100,0	57,7	36,9 - 76,6	29,4	100,0
	>2,26	87,5	47,4 - 97,9	57,7	36,9 - 76,6	26,7	96,3
	>3,25	87,5	47,4 - 97,9	69,2	48,2 - 85,6	33,4	96,9
	>3,48	62,5	24,7 - 91,0	69,2	48,2 - 85,6	26,4	91,3
	>3,74	62,5	24,7 - 91,0	80,8	60,6 - 93,4	36,4	92,4
	>4,19	50,0	16,0 - 84,0	80,8	60,6 - 93,4	31,5	90,2
	>4,65	50,0	16,0 - 84,0	88,5	69,8 - 97,4	43,3	90,9
	>5,39	25,0	3,9 - 65,0	88,5	69,8 - 97,4	27,7	87,0
		>10,13	25,0	3,9 - 65,0	100,0	86,7 - 100,0	100,0
	>12,31	0,0	0,0 - 37,1	100,0	86,7 - 100,0		85,0

Variable	Punto de corte	Sensibilidad	IC 95%	Especificidad	IC 95%	VP+	VP-
TR ítems totales	>=0,81	100,0	62,9 - 100,0	0,0	0,0 - 13,3	15,0	
	>1,76	100,0	62,9 - 100,0	53,8	33,4 - 73,4	27,7	100,0
	>1,88	87,5	47,4 - 97,9	53,8	33,4 - 73,4	25,1	96,1
	>2,71	87,5	47,4 - 97,9	73,1	52,2 - 88,4	36,4	97,1
	>2,74	75,0	35,0 - 96,1	73,1	52,2 - 88,4	33,0	94,3
	>3,11	75,0	35,0 - 96,1	80,8	60,6 - 93,4	40,8	94,8
	>3,22	62,5	24,7 - 91,0	80,8	60,6 - 93,4	36,4	92,4
	>3,48	62,5	24,7 - 91,0	88,5	69,8 - 97,4	48,9	93,0
	>4,36	25,0	3,9 - 65,0	88,5	69,8 - 97,4	27,7	87,0
	>7,3	25,0	3,9 - 65,0	100,0	86,7 - 100,0	100,0	88,3
	>9,44	0,0	0,0 - 37,1	100,0	86,7 - 100,0		85,0

Nota: IC= intervalo de confianza; TR= tiempo de reacción; VP+= valor predictivo positivo; VP-= valor predictivo negativo; VSVT= Victoria Symptom Validity Test.

Anexo 1.1.2. Sensibilidad, especificidad y valores predictivos positivos y negativos asociados a un porcentaje de base del **30%** para diversos puntos de corte de las variables del **VSVT**

Variable	Punto de corte	Sensibilidad	IC 95%	Especificidad	IC 95%	VP+	VP-
Ítems fáciles	< 15	0,0	0,0 - 37,1	100,0	87,5 - 100,0		70,0
	<=20	50,0	16,0 - 84,0	100,0	87,5 - 100,0	100,0	82,4
	<=22	50,0	16,0 - 84,0	92,9	76,5 - 98,9	75,0	81,2
	<=23	62,5	24,7 - 91,0	82,1	63,1 - 93,9	60,0	83,6
	<=24	100,0	62,9 - 100,0	0,0	0,0 - 12,5	30,0	
Ítems difíciles	< 4	0,0	0,0 - 37,1	100,0	87,5 - 100,0		70,0
	<=13	62,5	24,7 - 91,0	100,0	87,5 - 100,0	100,0	86,2
	<=16	62,5	24,7 - 91,0	92,9	76,5 - 98,9	78,9	85,2
	<=17	75,0	35,0 - 96,1	89,3	71,7 - 97,6	75,0	89,3
	<=18	87,5	47,4 - 97,9	85,7	67,3 - 95,9	72,4	94,1
	<=19	87,5	47,4 - 97,9	75,0	55,1 - 89,3	60,0	93,3
	<=21	100,0	62,9 - 100,0	67,9	47,7 - 84,1	57,1	100,0
	<=24	100,0	62,9 - 100,0	0,0	0,0 - 12,5	30,0	
	Ítems totales	< 21	0,0	0,0 - 37,1	100,0	87,5 - 100,0	
<=38		75,0	35,0 - 96,1	100,0	87,5 - 100,0	100,0	90,3
<=40		75,0	35,0 - 96,1	92,9	76,5 - 98,9	81,8	89,7
<=41		87,5	47,4 - 97,9	89,3	71,7 - 97,6	77,8	94,3
<=43		87,5	47,4 - 97,9	71,4	51,3 - 86,7	56,8	93,0
<=45		100,0	62,9 - 100,0	57,1	37,2 - 75,5	50,0	100,0
<=48		100,0	62,9 - 100,0	0,0	0,0 - 12,5	30,0	

Variable	Punto de corte	Sensibilidad	IC 95%	Especificidad	IC 95%	VP+	VP-
TR ítems fáciles	$\geq 0,69$	100,0	62,9 - 100,0	0,0	0,0 - 13,3	30,0	
	$> 1,44$	100,0	62,9 - 100,0	53,8	33,4 - 73,4	48,1	100,0
	$> 1,5$	87,5	47,4 - 97,9	53,8	33,4 - 73,4	44,8	91,0
	$> 1,91$	87,5	47,4 - 97,9	76,9	56,3 - 91,0	61,9	93,5
	> 2	75,0	35,0 - 96,1	76,9	56,3 - 91,0	58,2	87,8
	$> 2,55$	75,0	35,0 - 96,1	88,5	69,8 - 97,4	73,6	89,2
	$> 3,75$	25,0	3,9 - 65,0	88,5	69,8 - 97,4	48,1	73,3
	$> 4,47$	25,0	3,9 - 65,0	96,2	80,3 - 99,4	73,6	74,9
	$> 6,58$	12,5	2,1 - 52,6	96,2	80,3 - 99,4	58,2	71,9
	$> 7,2$	12,5	2,1 - 52,6	100,0	86,7 - 100,0	100,0	72,7
	$> 7,39$	0,0	0,0 - 37,1	100,0	86,7 - 100,0		70,0
TR ítems difíciles	$\geq 0,92$	100,0	62,9 - 100,0	0,0	0,0 - 13,3	30,0	
	$> 2,19$	100,0	62,9 - 100,0	57,7	36,9 - 76,6	50,3	100,0
	$> 2,26$	87,5	47,4 - 97,9	57,7	36,9 - 76,6	47,0	91,5
	$> 3,25$	87,5	47,4 - 97,9	69,2	48,2 - 85,6	54,9	92,8
	$> 3,48$	62,5	24,7 - 91,0	69,2	48,2 - 85,6	46,5	81,2
	$> 3,74$	62,5	24,7 - 91,0	80,8	60,6 - 93,4	58,2	83,4
	$> 4,19$	50,0	16,0 - 84,0	80,8	60,6 - 93,4	52,7	79,0
	$> 4,65$	50,0	16,0 - 84,0	88,5	69,8 - 97,4	65,0	80,5
	$> 5,39$	25,0	3,9 - 65,0	88,5	69,8 - 97,4	48,1	73,3
	$> 10,13$	25,0	3,9 - 65,0	100,0	86,7 - 100,0	100,0	75,7
	$> 12,31$	0,0	0,0 - 37,1	100,0	86,7 - 100,0		70,0

Variable	Punto de corte	Sensibilidad	IC 95%	Especificidad	IC 95%	VP+	VP-
TR ítems totales	$\geq 0,81$	100,0	62,9 - 100,0	0,0	0,0 - 13,3	30,0	
	$> 1,76$	100,0	62,9 - 100,0	53,8	33,4 - 73,4	48,1	100,0
	$> 1,88$	87,5	47,4 - 97,9	53,8	33,4 - 73,4	44,8	91,0
	$> 2,71$	87,5	47,4 - 97,9	73,1	52,2 - 88,4	58,2	93,2
	$> 2,74$	75,0	35,0 - 96,1	73,1	52,2 - 88,4	54,4	87,2
	$> 3,11$	75,0	35,0 - 96,1	80,8	60,6 - 93,4	62,6	88,3
	$> 3,22$	62,5	24,7 - 91,0	80,8	60,6 - 93,4	58,2	83,4
	$> 3,48$	62,5	24,7 - 91,0	88,5	69,8 - 97,4	69,9	84,6
	$> 4,36$	25,0	3,9 - 65,0	88,5	69,8 - 97,4	48,1	73,3
	$> 7,3$	25,0	3,9 - 65,0	100,0	86,7 - 100,0	100,0	75,7
	$> 9,44$	0,0	0,0 - 37,1	100,0	86,7 - 100,0		70,0

Nota: IC= intervalo de confianza; TR= tiempo de reacción; VP+= valor predictivo positivo; VP-= valor predictivo negativo; VSVT= Victoria Symptom Validity Test.

Anexo 1.1.3. Sensibilidad, especificidad y valores predictivos positivos y negativos asociados a un porcentaje de base del **45%** para diversos puntos de corte de las variables del **VSVT**

Variable	Punto de corte	Sensibilidad	IC 95%	Especificidad	IC 95%	VP+	VP-
Ítems fáciles	< 15	0,0	0,0 - 37,1	100,0	87,5 - 100,0		55,0
	<=20	50,0	16,0 - 84,0	100,0	87,5 - 100,0	100,0	71,0
	<=22	50,0	16,0 - 84,0	92,9	76,5 - 98,9	85,1	69,4
	<=23	62,5	24,7 - 91,0	82,1	63,1 - 93,9	74,1	72,8
	<=24	100,0	62,9 - 100,0	0,0	0,0 - 12,5	45,0	
Ítems difíciles	< 4	0,0	0,0 - 37,1	100,0	87,5 - 100,0		55,0
	<=13	62,5	24,7 - 91,0	100,0	87,5 - 100,0	100,0	76,5
	<=16	62,5	24,7 - 91,0	92,9	76,5 - 98,9	87,7	75,2
	<=17	75,0	35,0 - 96,1	89,3	71,7 - 97,6	85,1	81,4
	<=18	87,5	47,4 - 97,9	85,7	67,3 - 95,9	83,4	89,3
	<=19	87,5	47,4 - 97,9	75,0	55,1 - 89,3	74,1	88,0
	<=21	100,0	62,9 - 100,0	67,9	47,7 - 84,1	71,8	100,0
	<=24	100,0	62,9 - 100,0	0,0	0,0 - 12,5	45,0	
Ítems totales	< 21	0,0	0,0 - 37,1	100,0	87,5 - 100,0		55,0
	<=38	75,0	35,0 - 96,1	100,0	87,5 - 100,0	100,0	83,0
	<=40	75,0	35,0 - 96,1	92,9	76,5 - 98,9	89,6	81,9
	<=41	87,5	47,4 - 97,9	89,3	71,7 - 97,6	87,0	89,7
	<=43	87,5	47,4 - 97,9	71,4	51,3 - 86,7	71,5	87,5
	<=45	100,0	62,9 - 100,0	57,1	37,2 - 75,5	65,6	100,0
	<=48	100,0	62,9 - 100,0	0,0	0,0 - 12,5	45,0	

Variable	Punto de corte	Sensibilidad	IC 95%	Especificidad	IC 95%	VP+	VP-
TR ítems fáciles	$\geq 0,69$	100,0	62,9 - 100,0	0,0	0,0 - 13,3	45,0	
	$> 1,44$	100,0	62,9 - 100,0	53,8	33,4 - 73,4	63,9	100,0
	$> 1,5$	87,5	47,4 - 97,9	53,8	33,4 - 73,4	60,8	84,0
	$> 1,91$	87,5	47,4 - 97,9	76,9	56,3 - 91,0	75,6	88,3
	> 2	75,0	35,0 - 96,1	76,9	56,3 - 91,0	72,7	79,0
	$> 2,55$	75,0	35,0 - 96,1	88,5	69,8 - 97,4	84,2	81,2
	$> 3,75$	25,0	3,9 - 65,0	88,5	69,8 - 97,4	63,9	59,0
	$> 4,47$	25,0	3,9 - 65,0	96,2	80,3 - 99,4	84,2	61,0
	$> 6,58$	12,5	2,1 - 52,6	96,2	80,3 - 99,4	72,7	57,3
	$> 7,2$	12,5	2,1 - 52,6	100,0	86,7 - 100,0	100,0	58,3
	$> 7,39$	0,0	0,0 - 37,1	100,0	86,7 - 100,0		55,0
TR ítems difíciles	$\geq 0,92$	100,0	62,9 - 100,0	0,0	0,0 - 13,3	45,0	
	$> 2,19$	100,0	62,9 - 100,0	57,7	36,9 - 76,6	65,9	100,0
	$> 2,26$	87,5	47,4 - 97,9	57,7	36,9 - 76,6	62,9	84,9
	$> 3,25$	87,5	47,4 - 97,9	69,2	48,2 - 85,6	69,9	87,1
	$> 3,48$	62,5	24,7 - 91,0	69,2	48,2 - 85,6	62,4	69,3
	$> 3,74$	62,5	24,7 - 91,0	80,8	60,6 - 93,4	72,7	72,5
	$> 4,19$	50,0	16,0 - 84,0	80,8	60,6 - 93,4	68,0	66,4
	$> 4,65$	50,0	16,0 - 84,0	88,5	69,8 - 97,4	78,0	68,4
	$> 5,39$	25,0	3,9 - 65,0	88,5	69,8 - 97,4	63,9	59,0
	$> 10,13$	25,0	3,9 - 65,0	100,0	86,7 - 100,0	100,0	62,0
	$> 12,31$	0,0	0,0 - 37,1	100,0	86,7 - 100,0		55,0

Variable	Punto de corte	Sensibilidad	IC 95%	Especificidad	IC 95%	VP+	VP-
TR ítems totales	$\geq 0,81$	100,0	62,9 - 100,0	0,0	0,0 - 13,3	45,0	
	$> 1,76$	100,0	62,9 - 100,0	53,8	33,4 - 73,4	63,9	100,0
	$> 1,88$	87,5	47,4 - 97,9	53,8	33,4 - 73,4	60,8	84,0
	$> 2,71$	87,5	47,4 - 97,9	73,1	52,2 - 88,4	72,7	87,7
	$> 2,74$	75,0	35,0 - 96,1	73,1	52,2 - 88,4	69,5	78,1
	$> 3,11$	75,0	35,0 - 96,1	80,8	60,6 - 93,4	76,1	79,8
	$> 3,22$	62,5	24,7 - 91,0	80,8	60,6 - 93,4	72,7	72,5
	$> 3,48$	62,5	24,7 - 91,0	88,5	69,8 - 97,4	81,6	74,2
	$> 4,36$	25,0	3,9 - 65,0	88,5	69,8 - 97,4	63,9	59,0
	$> 7,3$	25,0	3,9 - 65,0	100,0	86,7 - 100,0	100,0	62,0
	$> 9,44$	0,0	0,0 - 37,1	100,0	86,7 - 100,0		55,0

Nota: IC= intervalo de confianza; TR= tiempo de reacción; VP+= valor predictivo positivo; VP-= valor predictivo negativo; VSVT= Victoria Symptom Validity Test.

Anexo 1.2.1. Sensibilidad, especificidad y valores predictivos positivos y negativos asociados a un porcentaje de base del **15%** para diversos puntos de corte de las variables del **Test of Memory Malinger**

Variable	Punto de corte	Sensibilidad	IC 95%	Especificidad	IC 95%	VP+	VP-
Ensayo 1	< 23	0,0	0,0 - 31,0	100,0	91,9 - 100,0		85,0
	<=28	30,0	7,0 - 65,2	100,0	91,9 - 100,0	100,0	89,0
	<=32	40,0	12,4 - 73,6	97,7	87,9 - 99,6	75,6	90,2
	<=37	100,0	69,0 - 100,0	97,7	87,9 - 99,6	88,6	100,0
	<=50	100,0	69,0 - 100,0	0,0	0,0 - 8,1	15,0	
Ensayo 2	< 20	0,0	0,0 - 31,0	100,0	91,9 - 100,0		85,0
	<=37	50,0	18,9 - 81,1	100,0	91,9 - 100,0	100,0	91,9
	<=39	50,0	18,9 - 81,1	97,7	87,9 - 99,6	79,5	91,7
	<=44	100,0	69,0 - 100,0	97,7	87,9 - 99,6	88,6	100,0
	<=50	100,0	69,0 - 100,0	0,0	0,0 - 8,1	15,0	

Nota: IC= intervalo de confianza; VP+= valor predictivo positivo; VP-= valor predictivo negativo.

Anexo 1.2.2. Sensibilidad, especificidad y valores predictivos positivos y negativos asociados a un porcentaje de base del **30%** para diversos puntos de corte de las variables del **Test of Memory Malinger**

Variable	Punto de corte	Sensibilidad	IC 95%	Especificidad	IC 95%	VP+	VP-
Ensayo 1	< 23	0,0	0,0 - 31,0	100,0	91,9 - 100,0		70,0
	<=28	30,0	7,0 - 65,2	100,0	91,9 - 100,0	100,0	76,9
	<=32	40,0	12,4 - 73,6	97,7	87,9 - 99,6	88,3	79,2
	<=37	100,0	69,0 - 100,0	97,7	87,9 - 99,6	95,0	100,0
	<=50	100,0	69,0 - 100,0	0,0	0,0 - 8,1	30,0	
Ensayo 2	< 20	0,0	0,0 - 31,0	100,0	91,9 - 100,0		70,0
	<=37	50,0	18,9 - 81,1	100,0	91,9 - 100,0	100,0	82,4
	<=39	50,0	18,9 - 81,1	97,7	87,9 - 99,6	90,4	82,0
	<=44	100,0	69,0 - 100,0	97,7	87,9 - 99,6	95,0	100,0
	<=50	100,0	69,0 - 100,0	0,0	0,0 - 8,1	30,0	

Nota: IC= intervalo de confianza; VP+= valor predictivo positivo; VP-= valor predictivo negativo.

Anexo 1.2.3. Sensibilidad, especificidad y valores predictivos positivos y negativos asociados a un porcentaje de base del **45%** para diversos puntos de corte de las variables del **Test of Memory Malinger**

Variable	Punto de corte	Sensibilidad	IC 95%	Especificidad	IC 95%	VP+	VP-
Ensayo 1	< 23	0,0	0,0 - 31,0	100,0	91,9 - 100,0		55,0
	<=28	30,0	7,0 - 65,2	100,0	91,9 - 100,0	100,0	63,6
	<=32	40,0	12,4 - 73,6	97,7	87,9 - 99,6	93,5	66,6
	<=37	100,0	69,0 - 100,0	97,7	87,9 - 99,6	97,3	100,0
	<=50	100,0	69,0 - 100,0	0,0	0,0 - 8,1	45,0	
Ensayo 2	< 20	0,0	0,0 - 31,0	100,0	91,9 - 100,0		55,0
	<=37	50,0	18,9 - 81,1	100,0	91,9 - 100,0	100,0	71,0
	<=39	50,0	18,9 - 81,1	97,7	87,9 - 99,6	94,7	70,5
	<=44	100,0	69,0 - 100,0	97,7	87,9 - 99,6	97,3	100,0
	<=50	100,0	69,0 - 100,0	0,0	0,0 - 8,1	45,0	

Nota: IC= intervalo de confianza; VP+= valor predictivo positivo; VP-= valor predictivo negativo.

Anexo 1.3.1. Sensibilidad, especificidad y valores predictivos positivos y negativos asociados a un porcentaje de base del **15%** para diversos puntos de corte de las variables del **Test de la b**

Variable	Punto de corte	Sensibilidad	IC 95%	Especificidad	IC 95%	VP+	VP-
e-score	≥ 23	100,0	69,0 - 100,0	0,0	0,0 - 8,1	15,0	
	$> 95,2$	100,0	69,0 - 100,0	97,7	87,9 - 99,6	88,6	100,0
	> 157	60,0	26,4 - 87,6	97,7	87,9 - 99,6	82,3	93,3
	$> 163,86$	60,0	26,4 - 87,6	100,0	91,9 - 100,0	100,0	93,4
	$> 635,73$	0,0	0,0 - 31,0	100,0	91,9 - 100,0		85,0
Errores d	≥ 0	100,0	69,0 - 100,0	0,0	0,0 - 8,1	15,0	
	> 1	100,0	69,0 - 100,0	95,5	84,5 - 99,3	79,5	100,0
	> 2	90,0	55,5 - 98,3	95,5	84,5 - 99,3	77,7	98,2
	> 3	80,0	44,4 - 96,9	100,0	91,9 - 100,0	100,0	96,6
	> 33	0,0	0,0 - 31,0	100,0	91,9 - 100,0		85,0
Errores c	≥ 0	100,0	69,0 - 100,0	0,0	0,0 - 8,1	15,0	
	> 0	60,0	26,4 - 87,6	86,4	72,6 - 94,8	43,7	92,4
	> 1	30,0	7,0 - 65,2	95,5	84,5 - 99,3	53,8	88,5
	> 2	30,0	7,0 - 65,2	100,0	91,9 - 100,0	100,0	89,0
	> 39	0,0	0,0 - 31,0	100,0	91,9 - 100,0		85,0
Errores o	≥ 0	100,0	69,0 - 100,0	0,0	0,0 - 8,1	15,0	
	> 0	90,0	55,5 - 98,3	0,0	0,0 - 8,1	13,7	0,0
	> 23	90,0	55,5 - 98,3	59,1	43,3 - 73,7	28,0	97,1
	> 24	80,0	44,4 - 96,9	59,1	43,3 - 73,7	25,7	94,4
	> 25	80,0	44,4 - 96,9	61,4	45,5 - 75,6	26,8	94,6
	> 26	70,0	34,8 - 93,0	63,6	47,8 - 77,6	25,4	92,3
	> 29	70,0	34,8 - 93,0	70,5	54,8 - 83,2	29,5	93,0
	> 30	60,0	26,4 - 87,6	72,7	57,2 - 85,0	28,0	91,2
	> 36	60,0	26,4 - 87,6	86,4	72,6 - 94,8	43,7	92,4
	> 37	50,0	18,9 - 81,1	88,6	75,4 - 96,2	43,7	90,9
	> 49	50,0	18,9 - 81,1	93,2	81,3 - 98,5	56,4	91,3
	> 50	40,0	12,4 - 73,6	93,2	81,3 - 98,5	50,9	89,8

Variable	Punto de corte	Sensibilidad	IC 95%	Especificidad	IC 95%	VP+	VP-
Errores o	>57	40,0	12,4 - 73,6	95,5	84,5 - 99,3	60,8	90,0
	>60	30,0	7,0 - 65,2	95,5	84,5 - 99,3	53,8	88,5
	>64	30,0	7,0 - 65,2	97,7	87,9 - 99,6	70,0	88,8
	>66	20,0	3,1 - 55,6	97,7	87,9 - 99,6	60,8	87,4
	>76	20,0	3,1 - 55,6	100,0	91,9 - 100,0	100,0	87,6
	>96	0,0	0,0 - 31,0	100,0	91,9 - 100,0		85,0
Tiempo	>=18,86	100,0	69,0 - 100,0	0,0	0,0 - 8,1	15,0	
	>30,26	100,0	69,0 - 100,0	43,2	28,4 - 59,0	23,7	100,0
	>30,33	90,0	55,5 - 98,3	43,2	28,4 - 59,0	21,8	96,1
	>37,83	90,0	55,5 - 98,3	70,5	54,8 - 83,2	35,0	97,6
	>38,04	80,0	44,4 - 96,9	70,5	54,8 - 83,2	32,3	95,2
	>39,4	80,0	44,4 - 96,9	79,5	64,7 - 90,2	40,8	95,8
	>40,53	70,0	34,8 - 93,0	79,5	64,7 - 90,2	37,7	93,8
	>43	70,0	34,8 - 93,0	86,4	72,6 - 94,8	47,5	94,2
	>43,46	60,0	26,4 - 87,6	86,4	72,6 - 94,8	43,7	92,4
	>44	60,0	26,4 - 87,6	88,6	75,4 - 96,2	48,2	92,6
	>45,73	50,0	18,9 - 81,1	88,6	75,4 - 96,2	43,7	90,9
	>47,2	50,0	18,9 - 81,1	93,2	81,3 - 98,5	56,4	91,3
	>59,46	20,0	3,1 - 55,6	93,2	81,3 - 98,5	34,1	86,8
	>60	20,0	3,1 - 55,6	97,7	87,9 - 99,6	60,8	87,4
	>63,23	10,0	1,7 - 44,5	97,7	87,9 - 99,6	43,7	86,0
>68,01	10,0	1,7 - 44,5	100,0	91,9 - 100,0	100,0	86,3	
>154	0,0	0,0 - 31,0	100,0	91,9 - 100,0		85,0	

Nota: Errores c= errores de comisión; Errores o= errores de omisión; IC= intervalo de confianza; VP+= valor predictivo positivo; VP-= valor predictivo negativo.

Anexo 1.3.2. Sensibilidad, especificidad y valores predictivos positivos y negativos asociados a un porcentaje de base del **30%** para diversos puntos de corte de las variables del **Test de la b**

Variable	Punto de corte	Sensibilidad	IC 95%	Especificidad	IC 95%	VP+	VP-
e-score	≥ 23	100,0	69,0 - 100,0	0,0	0,0 - 8,1	30,0	
	$> 95,2$	100,0	69,0 - 100,0	97,7	87,9 - 99,6	95,0	100,0
	> 157	60,0	26,4 - 87,6	97,7	87,9 - 99,6	91,9	85,1
	$> 163,86$	60,0	26,4 - 87,6	100,0	91,9 - 100,0	100,0	85,4
	$> 635,73$	0,0	0,0 - 31,0	100,0	91,9 - 100,0		70,0
Errores d	≥ 0	100,0	69,0 - 100,0	0,0	0,0 - 8,1	30,0	
	> 1	100,0	69,0 - 100,0	95,5	84,5 - 99,3	90,4	100,0
	> 2	90,0	55,5 - 98,3	95,5	84,5 - 99,3	89,5	95,7
	> 3	80,0	44,4 - 96,9	100,0	91,9 - 100,0	100,0	92,1
	> 33	0,0	0,0 - 31,0	100,0	91,9 - 100,0		70,0
Errores c	≥ 0	100,0	69,0 - 100,0	0,0	0,0 - 8,1	30,0	
	> 0	60,0	26,4 - 87,6	86,4	72,6 - 94,8	65,3	83,4
	> 1	30,0	7,0 - 65,2	95,5	84,5 - 99,3	73,9	76,1
	> 2	30,0	7,0 - 65,2	100,0	91,9 - 100,0	100,0	76,9
	> 39	0,0	0,0 - 31,0	100,0	91,9 - 100,0		70,0
Errores o	≥ 0	100,0	69,0 - 100,0	0,0	0,0 - 8,1	30,0	
	> 0	90,0	55,5 - 98,3	0,0	0,0 - 8,1	27,8	0,0
	> 23	90,0	55,5 - 98,3	59,1	43,3 - 73,7	48,5	93,2
	> 24	80,0	44,4 - 96,9	59,1	43,3 - 73,7	45,6	87,3
	> 25	80,0	44,4 - 96,9	61,4	45,5 - 75,6	47,0	87,7
	> 26	70,0	34,8 - 93,0	63,6	47,8 - 77,6	45,2	83,2
	> 29	70,0	34,8 - 93,0	70,5	54,8 - 83,2	50,4	84,6
	> 30	60,0	26,4 - 87,6	72,7	57,2 - 85,0	48,5	80,9
	> 36	60,0	26,4 - 87,6	86,4	72,6 - 94,8	65,3	83,4
	> 37	50,0	18,9 - 81,1	88,6	75,4 - 96,2	65,3	80,5
	> 49	50,0	18,9 - 81,1	93,2	81,3 - 98,5	75,9	81,3
	> 50	40,0	12,4 - 73,6	93,2	81,3 - 98,5	71,5	78,4
	> 57	40,0	12,4 - 73,6	95,5	84,5 - 99,3	79,0	78,8

Variable	Punto de corte	Sensibilidad	IC 95%	Especificidad	IC 95%	VP+	VP-
Errores o	>60	30,0	7,0 - 65,2	95,5	84,5 - 99,3	73,9	76,1
	>64	30,0	7,0 - 65,2	97,7	87,9 - 99,6	85,0	76,5
	>66	20,0	3,1 - 55,6	97,7	87,9 - 99,6	79,0	74,0
	>76	20,0	3,1 - 55,6	100,0	91,9 - 100,0	100,0	74,5
	>96	0,0	0,0 - 31,0	100,0	91,9 - 100,0		70,0
Tiempo	>=18,86	100,0	69,0 - 100,0	0,0	0,0 - 8,1	30,0	
	>30,26	100,0	69,0 - 100,0	43,2	28,4 - 59,0	43,0	100,0
	>30,33	90,0	55,5 - 98,3	43,2	28,4 - 59,0	40,4	91,0
	>37,83	90,0	55,5 - 98,3	70,5	54,8 - 83,2	56,6	94,3
	>38,04	80,0	44,4 - 96,9	70,5	54,8 - 83,2	53,7	89,2
	>39,4	80,0	44,4 - 96,9	79,5	64,7 - 90,2	62,6	90,3
	>40,53	70,0	34,8 - 93,0	79,5	64,7 - 90,2	59,5	86,1
	>43	70,0	34,8 - 93,0	86,4	72,6 - 94,8	68,8	87,0
	>43,46	60,0	26,4 - 87,6	86,4	72,6 - 94,8	65,3	83,4
	>44	60,0	26,4 - 87,6	88,6	75,4 - 96,2	69,4	83,8
	>45,73	50,0	18,9 - 81,1	88,6	75,4 - 96,2	65,3	80,5
	>47,2	50,0	18,9 - 81,1	93,2	81,3 - 98,5	75,9	81,3
	>59,46	20,0	3,1 - 55,6	93,2	81,3 - 98,5	55,7	73,1
	>60	20,0	3,1 - 55,6	97,7	87,9 - 99,6	79,0	74,0
	>63,23	10,0	1,7 - 44,5	97,7	87,9 - 99,6	65,3	71,7
	>68,01	10,0	1,7 - 44,5	100,0	91,9 - 100,0	100,0	72,2
>154	0,0	0,0 - 31,0	100,0	91,9 - 100,0		70,0	

Nota: Errores c= errores de comisión; Errores o= errores de omisión; IC= intervalo de confianza; VP+= valor predictivo positivo; VP-= valor predictivo negativo.

Anexo 1.3.3. Sensibilidad, especificidad y valores predictivos positivos y negativos asociados a un porcentaje de base del **45%** para diversos puntos de corte de las variables del **Test de la b**

Variable	Punto de corte	Sensibilidad	IC 95%	Especificidad	IC 95%	VP+	VP-
e-score	≥ 23	100,0	69,0 - 100,0	0,0	0,0 - 8,1	45,0	
	$> 95,2$	100,0	69,0 - 100,0	97,7	87,9 - 99,6	97,3	100,0
	> 157	60,0	26,4 - 87,6	97,7	87,9 - 99,6	95,6	74,9
	$> 163,86$	60,0	26,4 - 87,6	100,0	91,9 - 100,0	100,0	75,3
	$> 635,73$	0,0	0,0 - 31,0	100,0	91,9 - 100,0		55,0
Errores d	≥ 0	100,0	69,0 - 100,0	0,0	0,0 - 8,1	45,0	
	> 1	100,0	69,0 - 100,0	95,5	84,5 - 99,3	94,7	100,0
	> 2	90,0	55,5 - 98,3	95,5	84,5 - 99,3	94,2	92,1
	> 3	80,0	44,4 - 96,9	100,0	91,9 - 100,0	100,0	85,9
	> 33	0,0	0,0 - 31,0	100,0	91,9 - 100,0		55,0
Errores c	≥ 0	100,0	69,0 - 100,0	0,0	0,0 - 8,1	45,0	
	> 0	60,0	26,4 - 87,6	86,4	72,6 - 94,8	78,3	72,5
	> 1	30,0	7,0 - 65,2	95,5	84,5 - 99,3	84,4	62,5
	> 2	30,0	7,0 - 65,2	100,0	91,9 - 100,0	100,0	63,6
	> 39	0,0	0,0 - 31,0	100,0	91,9 - 100,0		55,0
Errores o	≥ 0	100,0	69,0 - 100,0	0,0	0,0 - 8,1	45,0	
	> 0	90,0	55,5 - 98,3	0,0	0,0 - 8,1	42,4	0,0
	> 23	90,0	55,5 - 98,3	59,1	43,3 - 73,7	64,3	87,8
	> 24	80,0	44,4 - 96,9	59,1	43,3 - 73,7	61,5	78,3
	> 25	80,0	44,4 - 96,9	61,4	45,5 - 75,6	62,9	78,9
	> 26	70,0	34,8 - 93,0	63,6	47,8 - 77,6	61,2	72,2
Errores o	> 29	70,0	34,8 - 93,0	70,5	54,8 - 83,2	66,0	74,2
	> 30	60,0	26,4 - 87,6	72,7	57,2 - 85,0	64,3	69,0
	> 36	60,0	26,4 - 87,6	86,4	72,6 - 94,8	78,3	72,5
	> 37	50,0	18,9 - 81,1	88,6	75,4 - 96,2	78,3	68,4
	> 49	50,0	18,9 - 81,1	93,2	81,3 - 98,5	85,7	69,5
	> 50	40,0	12,4 - 73,6	93,2	81,3 - 98,5	82,8	65,5
	> 57	40,0	12,4 - 73,6	95,5	84,5 - 99,3	87,8	66,0

Variable	Punto de corte	Sensibilidad	IC 95%	Especificidad	IC 95%	VP+	VP-
Errores o	>60	30,0	7,0 - 65,2	95,5	84,5 - 99,3	84,4	62,5
	>64	30,0	7,0 - 65,2	97,7	87,9 - 99,6	91,5	63,0
	>66	20,0	3,1 - 55,6	97,7	87,9 - 99,6	87,8	59,9
	>76	20,0	3,1 - 55,6	100,0	91,9 - 100,0	100,0	60,4
	>96	0,0	0,0 - 31,0	100,0	91,9 - 100,0		55,0
Tiempo	>=18,86	100,0	69,0 - 100,0	0,0	0,0 - 8,1	45,0	
	>30,26	100,0	69,0 - 100,0	43,2	28,4 - 59,0	59,0	100,0
	>30,33	90,0	55,5 - 98,3	43,2	28,4 - 59,0	56,4	84,1
	>37,83	90,0	55,5 - 98,3	70,5	54,8 - 83,2	71,4	89,6
	>38,04	80,0	44,4 - 96,9	70,5	54,8 - 83,2	68,9	81,2
	>39,4	80,0	44,4 - 96,9	79,5	64,7 - 90,2	76,2	82,9
	>40,53	70,0	34,8 - 93,0	79,5	64,7 - 90,2	73,7	76,4
	>43	70,0	34,8 - 93,0	86,4	72,6 - 94,8	80,8	77,9
	>43,46	60,0	26,4 - 87,6	86,4	72,6 - 94,8	78,3	72,5
	>44	60,0	26,4 - 87,6	88,6	75,4 - 96,2	81,2	73,0
	>45,73	50,0	18,9 - 81,1	88,6	75,4 - 96,2	78,3	68,4
	>47,2	50,0	18,9 - 81,1	93,2	81,3 - 98,5	85,7	69,5
	>59,46	20,0	3,1 - 55,6	93,2	81,3 - 98,5	70,6	58,7
	>60	20,0	3,1 - 55,6	97,7	87,9 - 99,6	87,8	59,9
	>63,23	10,0	1,7 - 44,5	97,7	87,9 - 99,6	78,3	57,0
	>68,01	10,0	1,7 - 44,5	100,0	91,9 - 100,0	100,0	57,6
>154	0,0	0,0 - 31,0	100,0	91,9 - 100,0		55,0	

Nota: Errores c= errores de comisión; Errores o= errores de omisión; IC= intervalo de confianza; VP+= valor predictivo positivo; VP-= valor predictivo negativo.

Anexo 1.4.1. Sensibilidad, especificidad y valores predictivos positivos y negativos asociados a un porcentaje de base del **15%** para diversos puntos de corte de las variables del **Dot Counting**

Variable	Punto de corte	Sensibilidad	IC 95%	Especificidad	IC 95%	VP+	VP-
Combo	>=5,1	100,0	69,0 - 100,0	0,0	0,0 - 8,9	15,0	
	>6,95	100,0	69,0 - 100,0	15,0	5,7 - 29,8	17,2	100,0
	>7,66	90,0	55,5 - 98,3	15,0	5,7 - 29,8	15,7	89,5
	>10,83	90,0	55,5 - 98,3	52,5	36,1 - 68,5	25,1	96,7
	>10,96	80,0	44,4 - 96,9	55,0	38,5 - 70,7	23,9	94,0
	>11,66	80,0	44,4 - 96,9	70,0	53,5 - 83,4	32,0	95,2
	>11,7	70,0	34,8 - 93,0	70,0	53,5 - 83,4	29,2	93,0
	>13,43	70,0	34,8 - 93,0	87,5	73,2 - 95,8	49,7	94,3
	>13,62	60,0	26,4 - 87,6	87,5	73,2 - 95,8	45,9	92,5
	>15,16	60,0	26,4 - 87,6	97,5	86,8 - 99,6	80,9	93,2
	>26	20,0	3,1 - 55,6	97,5	86,8 - 99,6	58,5	87,4
	>29,49	20,0	3,1 - 55,6	100,0	91,1 - 100,0	100,0	87,6
	>34,88	0,0	0,0 - 31,0	100,0	91,1 - 100,0		85,0
Tiempo agrupados	>=1,16	100,0	69,0 - 100,0	0,0	0,0 - 8,9	15,0	
	>3,16	100,0	69,0 - 100,0	70,0	53,5 - 83,4	37,0	100,0
	>3,26	90,0	55,5 - 98,3	70,0	53,5 - 83,4	34,6	97,5
	>3,45	90,0	55,5 - 98,3	75,0	58,8 - 87,3	38,8	97,7
	>3,51	80,0	44,4 - 96,9	75,0	58,8 - 87,3	36,1	95,5
	>5,12	80,0	44,4 - 96,9	92,5	79,6 - 98,3	65,3	96,3
	>5,65	60,0	26,4 - 87,6	92,5	79,6 - 98,3	58,5	92,9

Variable	Punto de corte	Sensibilidad	IC 95%	Especificidad	IC 95%	VP+	VP-
Tiempo agrupados	>5,83	60,0	26,4 - 87,6	95,0	83,0 - 99,2	67,9	93,1
	>5,9	50,0	18,9 - 81,1	97,5	86,8 - 99,6	77,9	91,7
	>6,83	30,0	7,0 - 65,2	97,5	86,8 - 99,6	67,9	88,8
	>9,83	30,0	7,0 - 65,2	100,0	91,1 - 100,0	100,0	89,0
	>15,39	0,0	0,0 - 31,0	100,0	91,1 - 100,0		85,0
Errores	>=0	100,0	69,0 - 100,0	0,0	0,0 - 8,9	15,0	
	>0	90,0	55,5 - 98,3	20,0	9,1 - 35,7	16,6	91,9
	>1	90,0	55,5 - 98,3	35,0	20,6 - 51,7	19,6	95,2
	>2	70,0	34,8 - 93,0	62,5	45,8 - 77,3	24,8	92,2
	>3	40,0	12,4 - 73,6	82,5	67,2 - 92,6	28,7	88,6
	>4	30,0	7,0 - 65,2	95,0	83,0 - 99,2	51,4	88,5
	>5	20,0	3,1 - 55,6	97,5	86,8 - 99,6	58,5	87,4
	>7	0,0	0,0 - 31,0	97,5	86,8 - 99,6	0,0	84,7
	>8	0,0	0,0 - 31,0	100,0	91,1 - 100,0		85,0
	Ratio	< 0,81	0,0	0,0 - 31,0	100,0	91,1 - 100,0	
<=0,82		20,0	3,1 - 55,6	100,0	91,1 - 100,0	100,0	87,6
<=0,91		20,0	3,1 - 55,6	97,5	86,8 - 99,6	58,5	87,4
<=1,07		40,0	12,4 - 73,6	97,5	86,8 - 99,6	73,8	90,2
<=1,11		40,0	12,4 - 73,6	95,0	83,0 - 99,2	58,5	90,0
<=1,19		60,0	26,4 - 87,6	95,0	83,0 - 99,2	67,9	93,1
<=1,24		60,0	26,4 - 87,6	87,5	73,2 - 95,8	45,9	92,5

Variable	Punto de corte	Sensibilidad	IC 95%	Especificidad	IC 95%	VP+	VP-
Ratio	<=1,26	70,0	34,8 - 93,0	87,5	73,2 - 95,8	49,7	94,3
	<=1,41	70,0	34,8 - 93,0	75,0	58,8 - 87,3	33,1	93,4
	<=1,42	80,0	44,4 - 96,9	72,5	56,1 - 85,4	33,9	95,4
	<=1,49	80,0	44,4 - 96,9	67,5	50,9 - 81,4	30,3	95,0
	<=1,5	90,0	55,5 - 98,3	65,0	48,3 - 79,4	31,2	97,4
	<=1,52	90,0	55,5 - 98,3	62,5	45,8 - 77,3	29,8	97,3
	<=1,55	100,0	69,0 - 100,0	60,0	43,3 - 75,1	30,6	100,0
	<=4,13	100,0	69,0 - 100,0	0,0	0,0 - 8,9	15,0	

Nota: IC= intervalo de confianza; VP+= valor predictivo positivo; VP-= valor predictivo negativo.

Anexo 1.4.2. Sensibilidad, especificidad y valores predictivos positivos y negativos asociados a un porcentaje de base del **30%** para diversos puntos de corte de las variables del **Dot Counting**

Variable	Punto de corte	Sensibilidad	IC 95%	Especificidad	IC 95%	VP+	VP-
Combo	>=5,1	100,0	69,0 - 100,0	0,0	0,0 - 8,9	30,0	
	>6,95	100,0	69,0 - 100,0	15,0	5,7 - 29,8	33,5	100,0
	>7,66	90,0	55,5 - 98,3	15,0	5,7 - 29,8	31,2	77,8
	>10,83	90,0	55,5 - 98,3	52,5	36,1 - 68,5	44,8	92,5
	>10,96	80,0	44,4 - 96,9	55,0	38,5 - 70,7	43,2	86,5
	>11,66	80,0	44,4 - 96,9	70,0	53,5 - 83,4	53,3	89,1
	>11,7	70,0	34,8 - 93,0	70,0	53,5 - 83,4	50,0	84,5
	>13,43	70,0	34,8 - 93,0	87,5	73,2 - 95,8	70,6	87,2
	>13,62	60,0	26,4 - 87,6	87,5	73,2 - 95,8	67,3	83,6
	>15,16	60,0	26,4 - 87,6	97,5	86,8 - 99,6	91,1	85,0
	>26	20,0	3,1 - 55,6	97,5	86,8 - 99,6	77,4	74,0
	>29,49	20,0	3,1 - 55,6	100,0	91,1 - 100,0	100,0	74,5
	>34,88	0,0	0,0 - 31,0	100,0	91,1 - 100,0		70,0
Tiempo agrupados	>=1,16	100,0	69,0 - 100,0	0,0	0,0 - 8,9	30,0	
	>3,16	100,0	69,0 - 100,0	70,0	53,5 - 83,4	58,8	100,0
	>3,26	90,0	55,5 - 98,3	70,0	53,5 - 83,4	56,3	94,2
	>3,45	90,0	55,5 - 98,3	75,0	58,8 - 87,3	60,7	94,6
	>3,51	80,0	44,4 - 96,9	75,0	58,8 - 87,3	57,8	89,7
	>5,12	80,0	44,4 - 96,9	92,5	79,6 - 98,3	82,1	91,5
	>5,65	60,0	26,4 - 87,6	92,5	79,6 - 98,3	77,4	84,4
	>5,83	60,0	26,4 - 87,6	95,0	83,0 - 99,2	83,7	84,7
	>5,9	50,0	18,9 - 81,1	97,5	86,8 - 99,6	89,6	82,0
	>6,83	30,0	7,0 - 65,2	97,5	86,8 - 99,6	83,7	76,5
	>9,83	30,0	7,0 - 65,2	100,0	91,1 - 100,0	100,0	76,9
	>15,39	0,0	0,0 - 31,0	100,0	91,1 - 100,0		70,0
	Errores	>=0	100,0	69,0 - 100,0	0,0	0,0 - 8,9	30,0
>0		90,0	55,5 - 98,3	20,0	9,1 - 35,7	32,5	82,4
>1		90,0	55,5 - 98,3	35,0	20,6 - 51,7	37,2	89,1
>2		70,0	34,8 - 93,0	62,5	45,8 - 77,3	44,4	82,9
>3		40,0	12,4 - 73,6	82,5	67,2 - 92,6	49,5	76,2
>4		30,0	7,0 - 65,2	95,0	83,0 - 99,2	72,0	76,0
>5		20,0	3,1 - 55,6	97,5	86,8 - 99,6	77,4	74,0
>7		0,0	0,0 - 31,0	97,5	86,8 - 99,6	0,0	69,5
>8		0,0	0,0 - 31,0	100,0	91,1 - 100,0		70,0
Ratio	< 0,81	0,0	0,0 - 31,0	100,0	91,1 - 100,0		70,0
	<=0,82	20,0	3,1 - 55,6	100,0	91,1 - 100,0	100,0	74,5
	<=0,91	20,0	3,1 - 55,6	97,5	86,8 - 99,6	77,4	74,0

Variable	Punto de corte	Sensibilidad	IC 95%	Especificidad	IC 95%	VP+	VP-
Ratio	<=1,07	40,0	12,4 - 73,6	97,5	86,8 - 99,6	87,3	79,1
	<=1,11	40,0	12,4 - 73,6	95,0	83,0 - 99,2	77,4	78,7
	<=1,19	60,0	26,4 - 87,6	95,0	83,0 - 99,2	83,7	84,7
	<=1,24	60,0	26,4 - 87,6	87,5	73,2 - 95,8	67,3	83,6
	<=1,26	70,0	34,8 - 93,0	87,5	73,2 - 95,8	70,6	87,2
	<=1,41	70,0	34,8 - 93,0	75,0	58,8 - 87,3	54,5	85,4
	<=1,42	80,0	44,4 - 96,9	72,5	56,1 - 85,4	55,5	89,4
	<=1,49	80,0	44,4 - 96,9	67,5	50,9 - 81,4	51,3	88,7
	<=1,5	90,0	55,5 - 98,3	65,0	48,3 - 79,4	52,4	93,8
	<=1,52	90,0	55,5 - 98,3	62,5	45,8 - 77,3	50,7	93,6
	<=1,55	100,0	69,0 - 100,0	60,0	43,3 - 75,1	51,7	100,0
	<=4,13	100,0	69,0 - 100,0	0,0	0,0 - 8,9	30,0	

Nota: IC= intervalo de confianza; VP+= valor predictivo positivo; VP-= valor predictivo negativo.

Anexo 1.4.3. Sensibilidad, especificidad y valores predictivos positivos y negativos asociados a un porcentaje de base del **45%** para diversos puntos de corte de las variables del **Dot Counting**

Variable	Punto de corte	Sensibilidad	IC 95%	Especificidad	IC 95%	VP+	VP-
Combo	>=5,1	100,0	69,0 - 100,0	0,0	0,0 - 8,9	45,0	
	>6,95	100,0	69,0 - 100,0	15,0	5,7 - 29,8	49,0	100,0
	>7,66	90,0	55,5 - 98,3	15,0	5,7 - 29,8	46,4	64,7
	>10,83	90,0	55,5 - 98,3	52,5	36,1 - 68,5	60,8	86,5
	>10,96	80,0	44,4 - 96,9	55,0	38,5 - 70,7	59,3	77,1
	>11,66	80,0	44,4 - 96,9	70,0	53,5 - 83,4	68,6	81,1
	>11,7	70,0	34,8 - 93,0	70,0	53,5 - 83,4	65,6	74,0
	>13,43	70,0	34,8 - 93,0	87,5	73,2 - 95,8	82,1	78,1
	>13,62	60,0	26,4 - 87,6	87,5	73,2 - 95,8	79,7	72,8
	>15,16	60,0	26,4 - 87,6	97,5	86,8 - 99,6	95,2	74,9
	>26	20,0	3,1 - 55,6	97,5	86,8 - 99,6	86,7	59,8
	>29,49	20,0	3,1 - 55,6	100,0	91,1 - 100,0	100,0	60,4
	>34,88	0,0	0,0 - 31,0	100,0	91,1 - 100,0		55,0
Tiempo agrupados	>=1,16	100,0	69,0 - 100,0	0,0	0,0 - 8,9	45,0	
	>3,16	100,0	69,0 - 100,0	70,0	53,5 - 83,4	73,2	100,0
	>3,26	90,0	55,5 - 98,3	70,0	53,5 - 83,4	71,1	89,5
	>3,45	90,0	55,5 - 98,3	75,0	58,8 - 87,3	74,7	90,2
	>3,51	80,0	44,4 - 96,9	75,0	58,8 - 87,3	72,4	82,1
	>5,12	80,0	44,4 - 96,9	92,5	79,6 - 98,3	89,7	85,0
	>5,65	60,0	26,4 - 87,6	92,5	79,6 - 98,3	86,7	73,9

Variable	Punto de corte	Sensibilidad	IC 95%	Especificidad	IC 95%	VP+	VP-
Tiempo agrupados	>5,83	60,0	26,4 - 87,6	95,0	83,0 - 99,2	90,8	74,4
	>5,9	50,0	18,9 - 81,1	97,5	86,8 - 99,6	94,2	70,4
	>6,83	30,0	7,0 - 65,2	97,5	86,8 - 99,6	90,8	63,0
	>9,83	30,0	7,0 - 65,2	100,0	91,1 - 100,0	100,0	63,6
	>15,39	0,0	0,0 - 31,0	100,0	91,1 - 100,0		55,0
Errores	>=0	100,0	69,0 - 100,0	0,0	0,0 - 8,9	45,0	
	>0	90,0	55,5 - 98,3	20,0	9,1 - 35,7	47,9	71,0
	>1	90,0	55,5 - 98,3	35,0	20,6 - 51,7	53,1	81,1
	>2	70,0	34,8 - 93,0	62,5	45,8 - 77,3	60,4	71,8
	>3	40,0	12,4 - 73,6	82,5	67,2 - 92,6	65,2	62,7
	>4	30,0	7,0 - 65,2	95,0	83,0 - 99,2	83,1	62,4
	>5	20,0	3,1 - 55,6	97,5	86,8 - 99,6	86,7	59,8
	>7	0,0	0,0 - 31,0	97,5	86,8 - 99,6	0,0	54,4
	>8	0,0	0,0 - 31,0	100,0	91,1 - 100,0		55,0
	Ratio	< 0,81	0,0	0,0 - 31,0	100,0	91,1 - 100,0	
<=0,82		20,0	3,1 - 55,6	100,0	91,1 - 100,0	100,0	60,4
<=0,91		20,0	3,1 - 55,6	97,5	86,8 - 99,6	86,7	59,8
<=1,07		40,0	12,4 - 73,6	97,5	86,8 - 99,6	92,9	66,5
<=1,11		40,0	12,4 - 73,6	95,0	83,0 - 99,2	86,7	65,9
<=1,19		60,0	26,4 - 87,6	95,0	83,0 - 99,2	90,8	74,4
<=1,24		60,0	26,4 - 87,6	87,5	73,2 - 95,8	79,7	72,8

Variable	Punto de corte	Sensibilidad	IC 95%	Especificidad	IC 95%	VP+	VP-
Ratio	<=1,26	70,0	34,8 - 93,0	87,5	73,2 - 95,8	82,1	78,1
	<=1,41	70,0	34,8 - 93,0	75,0	58,8 - 87,3	69,6	75,3
	<=1,42	80,0	44,4 - 96,9	72,5	56,1 - 85,4	70,4	81,6
	<=1,49	80,0	44,4 - 96,9	67,5	50,9 - 81,4	66,8	80,5
	<=1,5	90,0	55,5 - 98,3	65,0	48,3 - 79,4	67,8	88,8
	<=1,52	90,0	55,5 - 98,3	62,5	45,8 - 77,3	66,3	88,4
	<=1,55	100,0	69,0 - 100,0	60,0	43,3 - 75,1	67,2	100,0
	<=4,13	100,0	69,0 - 100,0	0,0	0,0 - 8,9	45,0	

Nota: IC= intervalo de confianza; VP+= valor predictivo positivo; VP-= valor predictivo negativo.

Anexo 1.5.1. Sensibilidad, especificidad y valores predictivos positivos y negativos asociados a un porcentaje de base del **15%** para diversos puntos de corte de las variables del **Test 15-Ítems de Rey**

Variable	Punto de corte	Sensibilidad	IC 95%	Especificidad	IC 95%	VP+	VP-
Ítems totales	< 1	0,0	0,0 - 33,8	100,0	90,7 - 100,0		85,0
	<=1	0,0	0,0 - 33,8	97,4	86,1 - 99,6	0,0	84,7
	<=2	11,1	1,8 - 48,3	97,4	86,1 - 99,6	42,7	86,1
	<=3	22,2	3,5 - 59,9	94,7	82,2 - 99,2	42,7	87,3
	<=5	22,2	3,5 - 59,9	92,1	78,6 - 98,2	33,2	87,0
	<=6	33,3	7,9 - 69,9	86,8	71,9 - 95,5	30,9	88,1
	<=7	44,4	14,0 - 78,6	84,2	68,7 - 93,9	33,2	89,6
	<=8	55,6	21,4 - 86,0	84,2	68,7 - 93,9	38,3	91,5
	<=10	55,6	21,4 - 86,0	73,7	56,9 - 86,6	27,1	90,4
	<=11	66,7	30,1 - 92,1	73,7	56,9 - 86,6	30,9	92,6
	<=12	88,9	51,7 - 98,2	50,0	33,4 - 66,6	23,9	96,2
	<=14	88,9	51,7 - 98,2	39,5	24,1 - 56,6	20,6	95,3
	<=15	100,0	66,2 - 100,0	0,0	0,0 - 9,3	15,0	

Nota: IC= intervalo de confianza; VP+= valor predictivo positivo; VP-= valor predictivo negativo.

Anexo 1.5.2. Sensibilidad, especificidad y valores predictivos positivos y negativos asociados a un porcentaje de base del **30%** para diversos puntos de corte de las variables del **Test 15-Ítems de Rey**

Variable	Punto de corte	Sensibilidad	IC 95%	Especificidad	IC 95%	VP+	VP-
Ítems totales	< 1	0,0	0,0 - 33,8	100,0	90,7 - 100,0		70,0
	<=1	0,0	0,0 - 33,8	97,4	86,1 - 99,6	0,0	69,4
	<=2	11,1	1,8 - 48,3	97,4	86,1 - 99,6	64,4	71,9
	<=3	22,2	3,5 - 59,9	94,7	82,2 - 99,2	64,4	74,0
	<=5	22,2	3,5 - 59,9	92,1	78,6 - 98,2	54,7	73,4
	<=6	33,3	7,9 - 69,9	86,8	71,9 - 95,5	52,1	75,2
	<=7	44,4	14,0 - 78,6	84,2	68,7 - 93,9	54,7	78,0
	<=8	55,6	21,4 - 86,0	84,2	68,7 - 93,9	60,1	81,6
	<=10	55,6	21,4 - 86,0	73,7	56,9 - 86,6	47,5	79,5
	<=11	66,7	30,1 - 92,1	73,7	56,9 - 86,6	52,1	83,8
	<=12	88,9	51,7 - 98,2	50,0	33,4 - 66,6	43,2	91,3
	<=14	88,9	51,7 - 98,2	39,5	24,1 - 56,6	38,6	89,2
	<=15	100,0	66,2 - 100,0	0,0	0,0 - 9,3	30,0	

Nota: IC= intervalo de confianza; RP+= ratio de probabilidad positiva; RP-= ratio de probabilidad negativa; VP+= valor predictivo positivo; VP-= valor predictivo negativo.

Anexo 1.5.3. Sensibilidad, especificidad y valores predictivos positivos y negativos asociados a un porcentaje de base del **45%** para diversos puntos de corte de las variables del **Test 15-Ítems de Rey**

Variable	Punto de corte	Sensibilidad	IC 95%	Especificidad	IC 95%	VP+	VP-
Ítems totales	< 1	0,0	0,0 - 33,8	100,0	90,7 - 100,0		55,0
	<=1	0,0	0,0 - 33,8	97,4	86,1 - 99,6	0,0	54,3
	<=2	11,1	1,8 - 48,3	97,4	86,1 - 99,6	77,6	57,2
	<=3	22,2	3,5 - 59,9	94,7	82,2 - 99,2	77,6	59,8
	<=5	22,2	3,5 - 59,9	92,1	78,6 - 98,2	69,7	59,1
	<=6	33,3	7,9 - 69,9	86,8	71,9 - 95,5	67,5	61,4
	<=7	44,4	14,0 - 78,6	84,2	68,7 - 93,9	69,7	64,9
	<=8	55,6	21,4 - 86,0	84,2	68,7 - 93,9	74,2	69,8
	<=10	55,6	21,4 - 86,0	73,7	56,9 - 86,6	63,3	67,0
	<=11	66,7	30,1 - 92,1	73,7	56,9 - 86,6	67,5	73,0
	<=12	88,9	51,7 - 98,2	50,0	33,4 - 66,6	59,3	84,6
	<=14	88,9	51,7 - 98,2	39,5	24,1 - 56,6	54,6	81,3
	<=15	100,0	66,2 - 100,0	0,0	0,0 - 9,3	45,0	

Nota: IC= intervalo de confianza; RP+= ratio de probabilidad positiva; RP-= ratio de probabilidad negativa; VP+= valor predictivo positivo; VP-= valor predictivo negativo.

ANEXO 2

**PUNTOS DE CORTE, SENSIBILIDADES Y ESPECIFICIDADES DE LAS
PRUEBAS NEUROPSICOLÓGICAS NO ESPECÍFICAS DE SIMULACIÓN
PARA DISTINTAS PREVALENCIAS DE BASE**

Anexo 2.1.1. Sensibilidad, especificidad y valores predictivos positivos y negativos asociados a un porcentaje de base del **15%** para diversos puntos de corte de las variables de la **Figura Compleja de Rey**

Variable	Punto de corte	Sensibilidad	IC 95%	Especificidad	IC 95%	VP+	VP-
Copia	< 5	0,0	0,0 - 31,0	100,0	91,9 - 100,0		85,0
	<=5	0,0	0,0 - 31,0	97,7	87,9 - 99,6	0,0	84,7
	<=13,5	10,0	1,7 - 44,5	97,7	87,9 - 99,6	43,7	86,0
	<=14,5	10,0	1,7 - 44,5	95,5	84,5 - 99,3	28,0	85,7
	<=25	20,0	3,1 - 55,6	95,5	84,5 - 99,3	43,7	87,1
	<=26,5	20,0	3,1 - 55,6	93,2	81,3 - 98,5	34,1	86,8
	<=27	30,0	7,0 - 65,2	93,2	81,3 - 98,5	43,7	88,3
	<=27,5	30,0	7,0 - 65,2	90,9	78,3 - 97,4	36,8	88,0
	<=28	40,0	12,4 - 73,6	88,6	75,4 - 96,2	38,3	89,3
	<=28,5	40,0	12,4 - 73,6	86,4	72,6 - 94,8	34,1	89,1
	<=30	50,0	18,9 - 81,1	84,1	69,9 - 93,3	35,7	90,5
	<=31	60,0	26,4 - 87,6	81,8	67,3 - 91,8	36,8	92,1
	<=32	70,0	34,8 - 93,0	72,7	57,2 - 85,0	31,2	93,2
	<=33	70,0	34,8 - 93,0	68,2	52,4 - 81,4	28,0	92,8
	<=34	80,0	44,4 - 96,9	54,5	38,9 - 69,6	23,7	93,9
	<=35	80,0	44,4 - 96,9	36,4	22,4 - 52,2	18,2	91,2
<=36	100,0	69,0 - 100,0	0,0	0,0 - 8,1	15,0		
R Inmediato	< 3	0,0	0,0 - 31,0	100,0	91,9 - 100,0		85,0
	<=3	0,0	0,0 - 31,0	97,7	87,9 - 99,6	0,0	84,7
	<=4	10,0	1,7 - 44,5	97,7	87,9 - 99,6	43,7	86,0
	<=7	10,0	1,7 - 44,5	93,2	81,3 - 98,5	20,6	85,4

Variable	Punto de corte	Sensibilidad	IC 95%	Especificidad	IC 95%	VP+	VP-
R Inmediato	<=9,5	20,0	3,1 - 55,6	90,9	78,3 - 97,4	28,0	86,6
	<=10	20,0	3,1 - 55,6	88,6	75,4 - 96,2	23,7	86,3
	<=11	30,0	7,0 - 65,2	88,6	75,4 - 96,2	31,8	87,8
	<=11,5	30,0	7,0 - 65,2	86,4	72,6 - 94,8	28,0	87,5
	<=14	50,0	18,9 - 81,1	81,8	67,3 - 91,8	32,7	90,3
	<=14,5	70,0	34,8 - 93,0	79,5	64,7 - 90,2	37,7	93,8
	<=16	70,0	34,8 - 93,0	61,4	45,5 - 75,6	24,2	92,1
	<=16,5	80,0	44,4 - 96,9	56,8	41,0 - 71,6	24,6	94,2
	<=17	90,0	55,5 - 98,3	50,0	34,6 - 65,4	24,1	96,6
	<=28	90,0	55,5 - 98,3	2,3	0,4 - 12,1	14,0	56,3
	<=29	100,0	69,0 - 100,0	2,3	0,4 - 12,1	15,3	100,0
	<=31	100,0	69,0 - 100,0	0,0	0,0 - 8,1	15,0	
	R Demorado	< 2	0,0	0,0 - 31,0	100,0	91,9 - 100,0	
<=2		10,0	1,7 - 44,5	100,0	91,9 - 100,0	100,0	86,3
<=7		10,0	1,7 - 44,5	95,5	84,5 - 99,3	28,0	85,7
<=8,5		30,0	7,0 - 65,2	93,2	81,3 - 98,5	43,7	88,3
<=9,5		40,0	12,4 - 73,6	90,9	78,3 - 97,4	43,7	89,6
<=10		40,0	12,4 - 73,6	88,6	75,4 - 96,2	38,3	89,3
<=11		50,0	18,9 - 81,1	88,6	75,4 - 96,2	43,7	90,9
<=13		50,0	18,9 - 81,1	79,5	64,7 - 90,2	30,1	90,0
<=13,5		60,0	26,4 - 87,6	79,5	64,7 - 90,2	34,1	91,8
<=14		60,0	26,4 - 87,6	75,0	59,7 - 86,8	29,8	91,4
<=14,5		70,0	34,8 - 93	70,5	54,8 - 83	29,5	93,0

Variable	Punto de corte	Sensibilidad	IC 95%	Especificidad	IC 95%	VP+	VP-
R Demorado	<=15	80,0	44,4 - 96,9	65,9	50,1 - 79,5	29,3	94,9
	<=16	90,0	55,5 - 98,3	59,1	43,3 - 73,7	28,0	97,1
	<=28	90,0	55,5 - 98,3	4,5	0,7 - 15,5	14,3	72,0
	<=29,5	100,0	69,0 - 100,0	4,5	0,7 - 15,5	15,6	100,0
	<=31	100,0	69,0 - 100,0	0,0	0,0 - 8,1	15,0	
Reconocimiento	< 2	0,0	0,0 - 37,1	100,0	89,3 - 100,0		85,0
	<=2	12,5	2,1 - 52,6	100,0	89,3 - 100,0	100,0	86,6
	<=3	12,5	2,1 - 52,6	97,0	84,2 - 99,5	42,1	86,3
	<=4	37,5	9,0 - 75,3	93,9	79,7 - 99,1	52,2	89,5
	<=5	37,5	9,0 - 75,3	87,9	71,8 - 96,5	35,3	88,8
	<=6	50,0	16,0 - 84,0	84,8	68,1 - 94,8	36,8	90,6
	<=7	50,0	16,0 - 84,0	66,7	48,2 - 82,0	20,9	88,3
	<=8	87,5	47,4 - 97,9	45,5	28,1 - 63,6	22,1	95,4
	<=9	100,0	62,9 - 100,0	24,2	11,1 - 42,3	18,9	100,0
	<=20	100,0	62,9 - 100,0	0,0	0,0 - 10,7	15,0	
Falsas alarmas	< 0	0,0	0,0 - 37,1	100,0	89,3 - 100,0		85,0
	<=0	62,5	24,7 - 91,0	51,5	33,6 - 69,2	18,5	88,6
	<=1	87,5	47,4 - 97,9	24,2	11,1 - 42,3	16,9	91,7
	<=2	87,5	47,4 - 97,9	9,1	2,0 - 24,4	14,5	80,5
	<=3	100,0	62,9 - 100,0	3,0	0,5 - 15,8	15,4	100,0
	<=12	100,0	62,9 - 100,0	0,0	0,0 - 10,7	15,0	

Nota: IC= intervalo de confianza; R= recuerdo; VP+= valor predictivo positivo; VP-= valor predictivo negativo.

Anexo 2.1.2. Sensibilidad, especificidad y valores predictivos positivos y negativos asociados a un porcentaje de base del **30%** para diversos puntos de corte de las variables de la **Figura Compleja de Rey**

Variable	Punto de corte	Sensibilidad	IC 95%	Especificidad	IC 95%	VP+	VP-
Copia	< 5	0,0	0,0 - 31,0	100,0	91,9 - 100,0		70,0
	<=5	0,0	0,0 - 31,0	97,7	87,9 - 99,6	0,0	69,5
	<=13,5	10,0	1,7 - 44,5	97,7	87,9 - 99,6	65,3	71,7
	<=14,5	10,0	1,7 - 44,5	95,5	84,5 - 99,3	48,5	71,2
	<=25	20,0	3,1 - 55,6	95,5	84,5 - 99,3	65,3	73,6
	<=26,5	20,0	3,1 - 55,6	93,2	81,3 - 98,5	55,7	73,1
	<=27	30,0	7,0 - 65,2	93,2	81,3 - 98,5	65,3	75,6
	<=27,5	30,0	7,0 - 65,2	90,9	78,3 - 97,4	58,6	75,2
	<=28	40,0	12,4 - 73,6	88,6	75,4 - 96,2	60,1	77,5
	<=28,5	40,0	12,4 - 73,6	86,4	72,6 - 94,8	55,7	77,1
	<=30	50,0	18,9 - 81,1	84,1	69,9 - 93,3	57,4	79,7
	<=31	60,0	26,4 - 87,6	81,8	67,3 - 91,8	58,6	82,7
	<=32	70,0	34,8 - 93,0	72,7	57,2 - 85,0	52,4	85,0
	<=33	70,0	34,8 - 93,0	68,2	52,4 - 81,4	48,5	84,1
	<=34	80,0	44,4 - 96,9	54,5	38,9 - 69,6	43,0	86,4
	<=35	80,0	44,4 - 96,9	36,4	22,4 - 52,2	35,0	80,9
<=36	100,0	69,0 - 100,0	0,0	0,0 - 8,1	30,0		
R Inmediato	< 3	0,0	0,0 - 31,0	100,0	91,9 - 100,0		70,0
	<=3	0,0	0,0 - 31,0	97,7	87,9 - 99,6	0,0	69,5
	<=4	10,0	1,7 - 44,5	97,7	87,9 - 99,6	65,3	71,7
	<=7	10,0	1,7 - 44,5	93,2	81,3 - 98,5	38,6	70,7

Variable	Punto de corte	Sensibilidad	IC 95%	Especificidad	IC 95%	VP+	VP-
R Inmediato	<=9,5	20,0	3,1 - 55,6	90,9	78,3 - 97,4	48,5	72,6
	<=10	20,0	3,1 - 55,6	88,6	75,4 - 96,2	43,0	72,1
	<=11	30,0	7,0 - 65,2	88,6	75,4 - 96,2	53,1	74,7
	<=11,5	30,0	7,0 - 65,2	86,4	72,6 - 94,8	48,5	74,2
	<=14	50,0	18,9 - 81,1	81,8	67,3 - 91,8	54,1	79,2
	<=14,5	70,0	34,8 - 93,0	79,5	64,7 - 90,2	59,5	86,1
	<=16	70,0	34,8 - 93,0	61,4	45,5 - 75,6	43,7	82,7
	<=16,5	80,0	44,4 - 96,9	56,8	41,0 - 71,6	44,3	86,9
	<=17	90,0	55,5 - 98,3	50,0	34,6 - 65,4	43,5	92,1
	<=28	90,0	55,5 - 98,3	2,3	0,4 - 12,1	28,3	34,7
	<=29	100,0	69,0 - 100,0	2,3	0,4 - 12,1	30,5	100,0
	<=31	100,0	69,0 - 100,0	0,0	0,0 - 8,1	30,0	
	R Demorado	< 2	0,0	0,0 - 31,0	100,0	91,9 - 100,0	
<=2		10,0	1,7 - 44,5	100,0	91,9 - 100,0	100,0	72,2
<=7		10,0	1,7 - 44,5	95,5	84,5 - 99,3	48,5	71,2
<=8,5		30,0	7,0 - 65,2	93,2	81,3 - 98,5	65,3	75,6
<=9,5		40,0	12,4 - 73,6	90,9	78,3 - 97,4	65,3	78,0
<=10		40,0	12,4 - 73,6	88,6	75,4 - 96,2	60,1	77,5
<=11		50,0	18,9 - 81,1	88,6	75,4 - 96,2	65,3	80,5
<=13		50,0	18,9 - 81,1	79,5	64,7 - 90,2	51,2	78,8
<=13,5		60,0	26,4 - 87,6	79,5	64,7 - 90,2	55,7	82,3
<=14		60,0	26,4 - 87,6	75,0	59,7 - 86,8	50,7	81,4
<=14,5		70,0	34,8 - 93	70,5	54,8 - 83	50,4	84,6

Variable	Punto de corte	Sensibilidad	IC 95%	Especificidad	IC 95%	VP+	VP-
R Demorado	<=15	80,0	44,4 - 96,9	65,9	50,1 - 79,5	50,1	88,5
	<=16	90,0	55,5 - 98,3	59,1	43,3 - 73,7	48,5	93,2
	<=28	90,0	55,5 - 98,3	4,5	0,7 - 15,5	28,8	51,5
	<=29,5	100,0	69,0 - 100,0	4,5	0,7 - 15,5	31,0	100,0
	<=31	100,0	69,0 - 100,0	0,0	0,0 - 8,1	30,0	
Reconocimiento	< 2	0,0	0,0 - 37,1	100,0	89,3 - 100,0		70,0
	<=2	12,5	2,1 - 52,6	100,0	89,3 - 100,0	100,0	72,7
	<=3	12,5	2,1 - 52,6	97,0	84,2 - 99,5	63,9	72,1
	<=4	37,5	9,0 - 75,3	93,9	79,7 - 99,1	72,6	77,8
	<=5	37,5	9,0 - 75,3	87,9	71,8 - 96,5	57,0	76,6
	<=6	50,0	16,0 - 84,0	84,8	68,1 - 94,8	58,6	79,8
	<=7	50,0	16,0 - 84,0	66,7	48,2 - 82,0	39,1	75,7
	<=8	87,5	47,4 - 97,9	45,5	28,1 - 63,6	40,7	89,5
	<=9	100,0	62,9 - 100,0	24,2	11,1 - 42,3	36,1	100,0
	<=20	100,0	62,9 - 100,0	0,0	0,0 - 10,7	30,0	
Falsas alarmas	< 0	0,0	0,0 - 37,1	100,0	89,3 - 100,0		70,0
	<=0	62,5	24,7 - 91,0	51,5	33,6 - 69,2	35,6	76,2
	<=1	87,5	47,4 - 97,9	24,2	11,1 - 42,3	33,1	81,9
	<=2	87,5	47,4 - 97,9	9,1	2,0 - 24,4	29,2	62,9
	<=3	100,0	62,9 - 100,0	3,0	0,5 - 15,8	30,7	100,0
	<=12	100,0	62,9 - 100,0	0,0	0,0 - 10,7	30,0	

Nota: IC= intervalo de confianza; R= recuerdo; VP+= valor predictivo positivo; VP-= valor predictivo negativo.

Anexo 2.1.3. Sensibilidad, especificidad y valores predictivos positivos y negativos asociada a un porcentaje de base del **45%** para diversos puntos de corte de las variables de la **Figura Compleja de Rey**

Variable	Punto de corte	Sensibilidad	IC 95%	Especificidad	IC 95%	VP+	VP-
Copia	< 5	0,0	0,0 - 31,0	100,0	91,9 - 100,0		55,0
	<=5	0,0	0,0 - 31,0	97,7	87,9 - 99,6	0,0	54,4
	<=13,5	10,0	1,7 - 44,5	97,7	87,9 - 99,6	78,3	57,0
	<=14,5	10,0	1,7 - 44,5	95,5	84,5 - 99,3	64,3	56,5
	<=25	20,0	3,1 - 55,6	95,5	84,5 - 99,3	78,3	59,3
	<=26,5	20,0	3,1 - 55,6	93,2	81,3 - 98,5	70,6	58,7
	<=27	30,0	7,0 - 65,2	93,2	81,3 - 98,5	78,3	61,9
	<=27,5	30,0	7,0 - 65,2	90,9	78,3 - 97,4	73,0	61,3
	<=28	40,0	12,4 - 73,6	88,6	75,4 - 96,2	74,2	64,4
	<=28,5	40,0	12,4 - 73,6	86,4	72,6 - 94,8	70,6	63,8
	<=30	50,0	18,9 - 81,1	84,1	69,9 - 93,3	72,0	67,3
	<=31	60,0	26,4 - 87,6	81,8	67,3 - 91,8	73,0	71,4
	<=32	70,0	34,8 - 93,0	72,7	57,2 - 85,0	67,7	74,8
	<=33	70,0	34,8 - 93,0	68,2	52,4 - 81,4	64,3	73,5
	<=34	80,0	44,4 - 96,9	54,5	38,9 - 69,6	59,0	76,9
<=35	80,0	44,4 - 96,9	36,4	22,4 - 52,2	50,7	69,0	
R Inmediato	< 3	0,0	0,0 - 31,0	100,0	91,9 - 100,0		55,0
	<=3	0,0	0,0 - 31,0	97,7	87,9 - 99,6	0,0	54,4
	<=4	10,0	1,7 - 44,5	97,7	87,9 - 99,6	78,3	57,0
	<=7	10,0	1,7 - 44,5	93,2	81,3 - 98,5	54,5	55,9
	<=9,5	20,0	3,1 - 55,6	90,9	78,3 - 97,4		

Variable	Punto de corte	Sensibilidad	IC 95%	Especificidad	IC 95%	VP+	VP-
R Inmediato	<=10	20,0	3,1 - 55,6	88,6	75,4 - 96,2	59,0	57,5
	<=11	30,0	7,0 - 65,2	88,6	75,4 - 96,2	68,4	60,7
	<=11,5	30,0	7,0 - 65,2	86,4	72,6 - 94,8	64,3	60,1
	<=14	50,0	18,9 - 81,1	81,8	67,3 - 91,8	69,2	66,7
	<=14,5	70,0	34,8 - 93,0	79,5	64,7 - 90,2	73,7	76,4
	<=16	70,0	34,8 - 93,0	61,4	45,5 - 75,6	59,7	71,4
	<=16,5	80,0	44,4 - 96,9	56,8	41,0 - 71,6	60,3	77,6
	<=17	90,0	55,5 - 98,3	50,0	34,6 - 65,4	59,6	85,9
	<=28	90,0	55,5 - 98,3	2,3	0,4 - 12,1	43,0	21,7
	<=29	100,0	69,0 - 100,0	2,3	0,4 - 12,1	45,6	100,0
<=31	100,0	69,0 - 100,0	0,0	0,0 - 8,1	45,0		
R Demorado	< 2	0,0	0,0 - 31,0	100,0	91,9 - 100,0		55,0
	<=2	10,0	1,7 - 44,5	100,0	91,9 - 100,0	100,0	57,6
	<=7	10,0	1,7 - 44,5	95,5	84,5 - 99,3	64,3	56,5
	<=8,5	30,0	7,0 - 65,2	93,2	81,3 - 98,5	78,3	61,9
	<=9,5	40,0	12,4 - 73,6	90,9	78,3 - 97,4	78,3	64,9
	<=10	40,0	12,4 - 73,6	88,6	75,4 - 96,2	74,2	64,4
	<=11	50,0	18,9 - 81,1	88,6	75,4 - 96,2	78,3	68,4
	<=13	50,0	18,9 - 81,1	79,5	64,7 - 90,2	66,7	66,0
	<=13,5	60,0	26,4 - 87,6	79,5	64,7 - 90,2	70,6	70,9
	<=14	60,0	26,4 - 87,6	75,0	59,7 - 86,8	66,3	69,6
<=14,5	70,0	34,8 - 93,0	70,5	54,8 - 83,2	66,0	74,2	

Variable	Punto de corte	Sensibilidad	IC 95%	Especificidad	IC 95%	VP+	VP-
R Demorado	<=15	80,0	44,4 - 96,9	65,9	50,1 - 79,5	65,8	80,1
	<=16	90,0	55,5 - 98,3	59,1	43,3 - 73,7	64,3	87,8
	<=28	90,0	55,5 - 98,3	4,5	0,7 - 15,5	43,5	35,7
	<=29,5	100,0	69,0 - 100,0	4,5	0,7 - 15,5	46,2	100,0
	<=31	100,0	69,0 - 100,0	0,0	0,0 - 8,1	45,0	
Reconocimiento	< 2	0,0	0,0 - 37,1	100,0	89,3 - 100,0		55,0
	<=2	12,5	2,1 - 52,6	100,0	89,3 - 100,0	100,0	58,3
	<=3	12,5	2,1 - 52,6	97,0	84,2 - 99,5	77,1	57,5
	<=4	37,5	9,0 - 75,3	93,9	79,7 - 99,1	83,5	64,8
	<=5	37,5	9,0 - 75,3	87,9	71,8 - 96,5	71,7	63,2
	<=6	50,0	16,0 - 84,0	84,8	68,1 - 94,8	73,0	67,5
	<=7	50,0	16,0 - 84,0	66,7	48,2 - 82,0	55,1	62,0
	<=8	87,5	47,4 - 97,9	45,5	28,1 - 63,6	56,8	81,6
	<=9	100,0	62,9 - 100,0	24,2	11,1 - 42,3	51,9	100,0
	<=20	100,0	62,9 - 100,0	0,0	0,0 - 10,7	45,0	
Falsas alarmas	< 0	0,0	0,0 - 37,1	100,0	89,3 - 100,0		55,0
	<=0	62,5	24,7 - 91,0	51,5	33,6 - 69,2	51,3	62,7
	<=1	87,5	47,4 - 97,9	24,2	11,1 - 42,3	48,6	70,3
	<=2	87,5	47,4 - 97,9	9,1	2,0 - 24,4	44,1	47,1
	<=3	100,0	62,9 - 100,0	3,0	0,5 - 15,8	45,8	100,0
	<=12	100,0	62,9 - 100,0	0,0	0,0 - 10,7	45,0	

Nota: IC= intervalo de confianza; R= recuerdo; VP+= valor predictivo positivo; VP-= valor predictivo negativo.

Anexo 2.2.1. Sensibilidad, especificidad y valores predictivos positivos y negativos asociados a un porcentaje de base del **15%** para diversos puntos de corte de las variables del **Test d2**

Variable	Punto de corte	Sensibilidad	IC 95%	Especificidad	IC 95%	VP+	VP-	
TR	< 129	0,0	0,0 - 31,0	100,0	91,3 - 100,0		85,0	
	<=177	40,0	12,4 - 73,6	100,0	91,3 - 100,0	100,0	90,4	
	<=188	40,0	12,4 - 73,6	97,6	87,1 - 99,6	74,3	90,2	
	<=198	50,0	18,9 - 81,1	97,6	87,1 - 99,6	78,3	91,7	
	<=233	50,0	18,9 - 81,1	95,1	83,4 - 99,3	64,4	91,5	
	<=235	60,0	26,4 - 87,6	95,1	83,4 - 99,3	68,5	93,1	
	<=271	60,0	26,4 - 87,6	90,2	76,9 - 97,2	52,0	92,7	
	<=283	70,0	34,8 - 93,0	90,2	76,9 - 97,2	55,9	94,5	
	<=383	70,0	34,8 - 93,0	41,5	26,3 - 57,9	17,4	88,7	
	<=393	90,0	55,5 - 98,3	41,5	26,3 - 57,9	21,3	95,9	
	<=452	90,0	55,5 - 98,3	9,8	2,8 - 23,1	15,0	84,7	
	<=465	100,0	69,0 - 100,0	9,8	2,8 - 23,1	16,4	100,0	
	<=545	100,0	69,0 - 100,0	0,0	0,0 - 8,7	15,0		
	TA	< 30	0,0	0,0 - 31,0	100,0	91,3 - 100,0		85,0
<=60		50,0	18,9 - 81,1	100,0	91,3 - 100,0	100,0	91,9	
<=77		50,0	18,9 - 81,1	95,1	83,4 - 99,3	64,4	91,5	
<=82		60,0	26,4 - 87,6	95,1	83,4 - 99,3	68,5	93,1	
<=86		60,0	26,4 - 87,6	90,2	76,9 - 97,2	52,0	92,7	
<=91		70,0	34,8 - 93,0	87,8	73,8 - 95,9	50,3	94,3	
<=92		80,0	44,4 - 96,9	87,8	73,8 - 95,9	53,7	96,1	
<=107		80,0	44,4 - 96,9	85,4	70,8 - 94,4	49,1	96,0	
<=110		90,0	55,5 - 98,3	85,4	70,8 - 94,4	52,0	98,0	
<=144		90,0	55,5 - 98,3	36,6	22,1 - 53,1	20,0	95,4	
<=145		100,0	69,0 - 100,0	36,6	22,1 - 53,1	21,8	100,0	
<=223		100,0	69,0 - 100,0	0,0	0,0 - 8,7	15,0		
Err o		>=0	100,0	69,0 - 100,0	0,0	0,0 - 8,7	15,0	
		>4	100,0	69,0 - 100,0	14,6	5,6 - 29,2	17,1	100,0
	>5	90,0	55,5 - 98,3	22,0	10,6 - 37,6	16,9	92,6	
	>9	90,0	55,5 - 98,3	41,5	26,3 - 57,9	21,3	95,9	
	>10	80,0	44,4 - 96,9	41,5	26,3 - 57,9	19,4	92,2	
	>11	80,0	44,4 - 96,9	48,8	32,9 - 64,9	21,6	93,3	
	>12	70,0	34,8 - 93,0	58,5	42,1 - 73,7	23,0	91,7	

Variable	Punto de corte	Sensibilidad	IC 95%	Especificidad	IC 95%	VP+	VP-
Err o	>17	70,0	34,8 - 93,0	65,9	49,4 - 79,9	26,6	92,6
	>18	60,0	26,4 - 87,6	68,3	51,9 - 81,9	25,0	90,6
	>20	50,0	18,9 - 81,1	70,7	54,5 - 83,9	23,2	88,9
	>24	50,0	18,9 - 81,1	73,2	57,1 - 85,8	24,7	89,2
	>27	30,0	7,0 - 65,2	75,6	59,7 - 87,6	17,8	86,0
	>28	30,0	7,0 - 65,2	80,5	65,1 - 91,2	21,3	86,7
	>30	20,0	3,1 - 55,6	82,9	67,9 - 92,8	17,1	85,5
	>54	20,0	3,1 - 55,6	100,0	91,3 - 100,0	100,0	87,6
	>106	0,0	0,0 - 31,0	100,0	91,3 - 100,0		85,0
	Err c	>=0	100,0	69,0 - 100,0	0,0	0,0 - 8,7	15,0
>0		80,0	44,4 - 96,9	53,7	37,4 - 69,3	23,4	93,8
>1		70,0	34,8 - 93,0	68,3	51,9 - 81,9	28,0	92,8
>2		50,0	18,9 - 81,1	75,6	59,7 - 87,6	26,6	89,5
>3		50,0	18,9 - 81,1	78,0	62,4 - 89,4	28,7	89,8
>4		30,0	7,0 - 65,2	82,9	67,9 - 92,8	23,7	87,0
>6		30,0	7,0 - 65,2	95,1	83,4 - 99,3	52,0	88,5
>12		10,0	1,7 - 44,5	95,1	83,4 - 99,3	26,6	85,7
>55		10,0	1,7 - 44,5	100,0	91,3 - 100,0	100,0	86,3
>71		0,0	0,0 - 31,0	100,0	91,3 - 100,0		85,0
TOT	>=0	100,0	69,0 - 100,0	0,0	0,0 - 8,7	15,0	
	>4	100,0	69,0 - 100,0	14,6	5,6 - 29,2	17,1	100,0
	>5	90,0	55,5 - 98,3	17,1	7,2 - 32,1	16,1	90,6
	>10	90,0	55,5 - 98,3	36,6	22,1 - 53,1	20,0	95,4
	>11	80,0	44,4 - 96,9	41,5	26,3 - 57,9	19,4	92,2
	>12	80,0	44,4 - 96,9	51,2	35,1 - 67,1	22,4	93,6
	>14	70,0	34,8 - 93,0	51,2	35,1 - 67,1	20,2	90,6
	>19	70,0	34,8 - 93,0	63,4	46,9 - 77,9	25,2	92,3
	>20	60,0	26,4 - 87,6	65,9	49,4 - 79,9	23,7	90,3
	>27	60,0	26,4 - 87,6	70,7	54,5 - 83,9	26,6	90,9
	>28	50,0	18,9 - 81,1	73,2	57,1 - 85,8	24,7	89,2
	>29	50,0	18,9 - 81,1	75,6	59,7 - 87,6	26,6	89,5
	>30	40,0	12,4 - 73,6	78,0	62,4 - 89,4	24,3	88,1
	>31	30,0	7,0 - 65,2	78,0	62,4 - 89,4	19,4	86,3
	>38	30,0	7,0 - 65,2	82,9	67,9 - 92,8	23,7	87,0
>39	20,0	3,1 - 55,6	82,9	67,9 - 92,8	17,1	85,5	
>79	20,0	3,1 - 55,6	100,0	91,3 - 100	100,0	87,6	

Variable	Punto de corte	Sensibilidad	IC 95%	Especificidad	IC 95%	VP+	VP-
TOT	>137	0,0	0,0 - 31,0	100,0	91,3 - 100,0		85,0
CON	< 3	0,0	0,0 - 31,0	100,0	91,3 - 100,0		85,0
	<=56	60,0	26,4 - 87,6	100,0	91,3 - 100,0	100,0	93,4
	<=77	60,0	26,4 - 87,6	92,7	80,1 - 98,4	59,1	92,9
	<=78	70,0	34,8 - 93,0	92,7	80,1 - 98,4	62,8	94,6
	<=89	70,0	34,8 - 93,0	87,8	73,8 - 95,9	50,3	94,3
	<=91	80,0	44,4 - 96,9	87,8	73,8 - 95,9	53,7	96,1
	<=106	80,0	44,4 - 96,9	85,4	70,8 - 94,4	49,1	96,0
	<=108	90,0	55,5 - 98,3	85,4	70,8 - 94,4	52,0	98,0
	<=137	90,0	55,5 - 98,3	43,9	28,5 - 60,2	22,1	96,1
	<=143	100,0	69,0 - 100,0	43,9	28,5 - 60,2	23,9	100,0
	<=223	100,0	69,0 - 100,0	0,0	0,0 - 8,7	15,0	
VAR	>=5	100,0	69,0 - 100,0	0,0	0,0 - 8,7	15,0	
	>5	100,0	69,0 - 100,0	2,4	0,4 - 12,9	15,3	100,0
	>7	90,0	55,5 - 98,3	7,3	1,6 - 19,9	14,6	80,6
	>8	80,0	44,4 - 96,9	9,8	2,8 - 23,1	13,5	73,4
	>9	80,0	44,4 - 96,9	14,6	5,6 - 29,2	14,2	80,6
	>10	70,0	34,8 - 93,0	29,3	16,1 - 45,5	14,9	84,7
	>11	70,0	34,8 - 93,0	41,5	26,3 - 57,9	17,4	88,7
	>12	60,0	26,4 - 87,6	51,2	35,1 - 67,1	17,8	87,9
	>13	60,0	26,4 - 87,6	58,5	42,1 - 73,7	20,3	89,2
	>14	50,0	18,9 - 81,1	70,7	54,5 - 83,9	23,2	88,9
	>15	40,0	12,4 - 73,6	73,2	57,1 - 85,8	20,8	87,4
	>17	40,0	12,4 - 73,6	75,6	59,7 - 87,6	22,4	87,7
	>18	30,0	7,0 - 65,2	80,5	65,1 - 91,2	21,3	86,7
	>19	20,0	3,1 - 55,6	82,9	67,9 - 92,8	17,1	85,5
	>25	20,0	3,1 - 55,6	95,1	83,4 - 99,3	42,0	87,1
	>26	10,0	1,7 - 44,5	95,1	83,4 - 99,3	26,6	85,7
	>31	10,0	1,7 - 44,5	100,0	91,3 - 100,0	100,0	86,3
	>34	0,0	0,0 - 31,0	100,0	91,3 - 100,0		85,0

Nota: Err c= errores de comisión; Err o= errores omisión; IC= intervalo de confianza; RP+= ratio de probabilidad positiva; RP-= ratio de probabilidad negativa; VP+= valor predictivo positivo; VP-= valor predictivo negativo.

Anexo 2.2.2. Sensibilidad, especificidad y valores predictivos positivos y negativos asociados a un porcentaje de base del 30% para diversos puntos de corte de las variables del **Test d2**

Variable	Punto de corte	Sensibilidad	IC 95%	Especificidad	IC 95%	VP+	VP-
TR	< 129	0,0	0,0 - 31,0	100,0	91,3 - 100,0		70,0
	<=177	40,0	12,4 - 73,6	100,0	91,3 - 100,0	100,0	79,5
	<=188	40,0	12,4 - 73,6	97,6	87,1 - 99,6	87,5	79,1
	<=198	50,0	18,9 - 81,1	97,6	87,1 - 99,6	89,8	82,0
	<=233	50,0	18,9 - 81,1	95,1	83,4 - 99,3	81,5	81,6
	<=235	60,0	26,4 - 87,6	95,1	83,4 - 99,3	84,1	84,7
	<=271	60,0	26,4 - 87,6	90,2	76,9 - 97,2	72,5	84,0
	<=283	70,0	34,8 - 93,0	90,2	76,9 - 97,2	75,5	87,5
	<=383	70,0	34,8 - 93,0	41,5	26,3 - 57,9	33,9	76,3
	<=393	90,0	55,5 - 98,3	41,5	26,3 - 57,9	39,7	90,6
	<=452	90,0	55,5 - 98,3	9,8	2,8 - 23,1	29,9	69,5
	<=465	100,0	69,0 - 100,0	9,8	2,8 - 23,1	32,2	100,0
	<=545	100,0	69,0 - 100,0	0,0	0,0 - 8,7	30,0	
TA	< 30	0,0	0,0 - 31,0	100,0	91,3 - 100,0		70,0
	<=60	50,0	18,9 - 81,1	100,0	91,3 - 100,0	100,0	82,4
	<=77	50,0	18,9 - 81,1	95,1	83,4 - 99,3	81,5	81,6
	<=82	60,0	26,4 - 87,6	95,1	83,4 - 99,3	84,1	84,7
	<=86	60,0	26,4 - 87,6	90,2	76,9 - 97,2	72,5	84,0
	<=91	70,0	34,8 - 93,0	87,8	73,8 - 95,9	71,1	87,2
	<=92	80,0	44,4 - 96,9	87,8	73,8 - 95,9	73,8	91,1
	<=107	80,0	44,4 - 96,9	85,4	70,8 - 94,4	70,1	90,9

Variable	Punto de corte	Sensibilidad	IC 95%	Especificidad	IC 95%	VP+	VP-
TA	<=110	90,0	55,5 - 98,3	85,4	70,8 - 94,4	72,5	95,2
	<=144	90,0	55,5 - 98,3	36,6	22,1 - 53,1	37,8	89,5
	<=145	100,0	69,0 - 100,0	36,6	22,1 - 53,1	40,3	100,0
	<=223	100,0	69,0 - 100,0	0,0	0,0 - 8,7	30,0	
Err o	>=0	100,0	69,0 - 100,0	0,0	0,0 - 8,7	30,0	
	>4	100,0	69,0 - 100,0	14,6	5,6 - 29,2	33,4	100,0
	>5	90,0	55,5 - 98,3	22,0	10,6 - 37,6	33,1	83,7
	>9	90,0	55,5 - 98,3	41,5	26,3 - 57,9	39,7	90,6
	>10	80,0	44,4 - 96,9	41,5	26,3 - 57,9	36,9	82,9
	>11	80,0	44,4 - 96,9	48,8	32,9 - 64,9	40,1	85,1
	>12	70,0	34,8 - 93,0	58,5	42,1 - 73,7	42,0	82,0
	>17	70,0	34,8 - 93,0	65,9	49,4 - 79,9	46,8	83,7
	>18	60,0	26,4 - 87,6	68,3	51,9 - 81,9	44,8	79,9
	>20	50,0	18,9 - 81,1	70,7	54,5 - 83,9	42,3	76,7
	>24	50,0	18,9 - 81,1	73,2	57,1 - 85,8	44,4	77,3
	>27	30,0	7,0 - 65,2	75,6	59,7 - 87,6	34,5	71,6
	>28	30,0	7,0 - 65,2	80,5	65,1 - 91,2	39,7	72,8
	>30	20,0	3,1 - 55,6	82,9	67,9 - 92,8	33,4	70,7
	>54	20,0	3,1 - 55,6	100,0	91,3 - 100,0	100,0	74,5
	>106	0,0	0,0 - 31,0	100,0	91,3 - 100,0		70,0
	Err c	>=0	100,0	69,0 - 100,0	0,0	0,0 - 8,7	30,0
>0		80,0	44,4 - 96,9	53,7	37,4 - 69,3	42,5	86,2

Variable	Punto de corte	Sensibilidad	IC 95%	Especificidad	IC 95%	VP+	VP-
Err c	>1	70,0	34,8 - 93,0	68,3	51,9 - 81,9	48,6	84,2
	>2	50,0	18,9 - 81,1	75,6	59,7 - 87,6	46,8	77,9
	>3	50,0	18,9 - 81,1	78,0	62,4 - 89,4	49,4	78,5
	>4	30,0	7,0 - 65,2	82,9	67,9 - 92,8	43,0	73,4
	>6	30,0	7,0 - 65,2	95,1	83,4 - 99,3	72,5	76,0
	>12	10,0	1,7 - 44,5	95,1	83,4 - 99,3	46,8	71,1
	>55	10,0	1,7 - 44,5	100,0	91,3 - 100,0	100,0	72,2
	>71	0,0	0,0 - 31,0	100,0	91,3 - 100,0		70,0
	TOT	< 112	0,0	0,0 - 31,0	100,0	91,3 - 100,0	
	<=167	50,0	18,9 - 81,1	100,0	91,3 - 100,0	100,0	82,4
	<=216	50,0	18,9 - 81,1	92,7	80,1 - 98,4	74,5	81,2
	<=224	60,0	26,4 - 87,6	92,7	80,1 - 98,4	77,8	84,4
	<=241	60,0	26,4 - 87,6	90,2	76,9 - 97,2	72,5	84,0
	<=248	70,0	34,8 - 93,0	90,2	76,9 - 97,2	75,5	87,5
	<=267	70,0	34,8 - 93,0	87,8	73,8 - 95,9	71,1	87,2
	<=269	80,0	44,4 - 96,9	87,8	73,8 - 95,9	73,8	91,1
	<=351	80,0	44,4 - 96,9	43,9	28,5 - 60,2	37,9	83,7
	<=355	90,0	55,5 - 98,3	43,9	28,5 - 60,2	40,7	91,1
	<=357	90,0	55,5 - 98,3	41,5	26,3 - 57,9	39,7	90,6
	<=373	100,0	69,0 - 100,0	41,5	26,3 - 57,9	42,3	100,0
	<=534	100,0	69,0 - 100,0	0,0	0,0 - 8,7	30,0	
CON	< 3	0,0	0,0 - 31,0	100,0	91,3 - 100,0		70,0

Variable	Punto de corte	Sensibilidad	IC 95%	Especificidad	IC 95%	VP+	VP-
CON	<=56	60,0	26,4 - 87,6	100,0	91,3 - 100,0	100,0	85,4
	<=77	60,0	26,4 - 87,6	92,7	80,1 - 98,4	77,8	84,4
	<=78	70,0	34,8 - 93,0	92,7	80,1 - 98,4	80,4	87,8
	<=89	70,0	34,8 - 93,0	87,8	73,8 - 95,9	71,1	87,2
	<=91	80,0	44,4 - 96,9	87,8	73,8 - 95,9	73,8	91,1
	<=106	80,0	44,4 - 96,9	85,4	70,8 - 94,4	70,1	90,9
	<=108	90,0	55,5 - 98,3	85,4	70,8 - 94,4	72,5	95,2
	<=137	90,0	55,5 - 98,3	43,9	28,5 - 60,2	40,7	91,1
	<=143	100,0	69,0 - 100,0	43,9	28,5 - 60,2	43,3	100,0
	<=223	100,0	69,0 - 100,0	0,0	0,0 - 8,7	30,0	
VAR	>=5	100,0	69,0 - 100,0	0,0	0,0 - 8,7	30,0	
	>5	100,0	69,0 - 100,0	2,4	0,4 - 12,9	30,5	100,0
	>7	90,0	55,5 - 98,3	7,3	1,6 - 19,9	29,4	63,1
	>8	80,0	44,4 - 96,9	9,8	2,8 - 23,1	27,5	53,2
	>9	80,0	44,4 - 96,9	14,6	5,6 - 29,2	28,7	63,1
	>10	70,0	34,8 - 93,0	29,3	16,1 - 45,5	29,8	69,5
	>11	70,0	34,8 - 93,0	41,5	26,3 - 57,9	33,9	76,3
	>12	60,0	26,4 - 87,6	51,2	35,1 - 67,1	34,5	74,9
	>13	60,0	26,4 - 87,6	58,5	42,1 - 73,7	38,3	77,3
	>14	50,0	18,9 - 81,1	70,7	54,5 - 83,9	42,3	76,7
	>15	40,0	12,4 - 73,6	73,2	57,1 - 85,8	39,0	74,0
	>17	40,0	12,4 - 73,6	75,6	59,7 - 87,6	41,3	74,6

Variable	Punto de corte	Sensibilidad	IC 95%	Especificidad	IC 95%	VP+	VP-
VAR	>18	30,0	7,0 - 65,2	80,5	65,1 - 91,2	39,7	72,8
	>19	20,0	3,1 - 55,6	82,9	67,9 - 92,8	33,4	70,7
	>25	20,0	3,1 - 55,6	95,1	83,4 - 99,3	63,7	73,5
	>26	10,0	1,7 - 44,5	95,1	83,4 - 99,3	46,8	71,1
	>31	10,0	1,7 - 44,5	100,0	91,3 - 100,0	100,0	72,2
	>34	0,0	0,0 - 31,0	100,0	91,3 - 100,0		70,0

Nota: Err c= errores de comisión; Err o= errores omisión; IC= intervalo de confianza; VP+= valor predictivo positivo; VP-= valor predictivo negativo.

Anexo 2.2.3. Sensibilidad, especificidad y valores predictivos positivos y negativos asociados a un porcentaje de base del **45%** para diversos puntos de corte de las variables del **Test d2**

Variable	Punto de corte	Sensibilidad	IC 95%	Especificidad	IC 95%	VP+	VP-
TR	< 129	0,0	0,0 - 31,0	100,0	91,3 - 100,0		55,0
	<=177	40,0	12,4 - 73,6	100,0	91,3 - 100,0	100,0	67,1
	<=188	40,0	12,4 - 73,6	97,6	87,1 - 99,6	93,1	66,5
	<=198	50,0	18,9 - 81,1	97,6	87,1 - 99,6	94,4	70,5
	<=233	50,0	18,9 - 81,1	95,1	83,4 - 99,3	89,3	69,9
	<=235	60,0	26,4 - 87,6	95,1	83,4 - 99,3	91,0	74,4
	<=271	60,0	26,4 - 87,6	90,2	76,9 - 97,2	83,4	73,4
	<=283	70,0	34,8 - 93,0	90,2	76,9 - 97,2	85,4	78,6
	<=383	70,0	34,8 - 93,0	41,5	26,3 - 57,9	49,5	62,8
	<=393	90,0	55,5 - 98,3	41,5	26,3 - 57,9	55,7	83,5
	<=452	90,0	55,5 - 98,3	9,8	2,8 - 23,1	44,9	54,4
	<=465	100,0	69,0 - 100,0	9,8	2,8 - 23,1	47,6	100,0
	<=545	100,0	69,0 - 100,0	0,0	0,0 - 8,7	45,0	
TA	< 30	0,0	0,0 - 31,0	100,0	91,3 - 100,0		55,0
	<=60	50,0	18,9 - 81,1	100,0	91,3 - 100,0	100,0	71,0
	<=77	50,0	18,9 - 81,1	95,1	83,4 - 99,3	89,3	69,9
	<=82	60,0	26,4 - 87,6	95,1	83,4 - 99,3	91,0	74,4
	<=86	60,0	26,4 - 87,6	90,2	76,9 - 97,2	83,4	73,4
	<=91	70,0	34,8 - 93,0	87,8	73,8 - 95,9	82,4	78,2
	<=92	80,0	44,4 - 96,9	87,8	73,8 - 95,9	84,3	84,3
	<=107	80,0	44,4 - 96,9	85,4	70,8 - 94,4	81,7	83,9

Variable	Punto de corte	Sensibilidad	IC 95%	Especificidad	IC 95%	VP+	VP-
TA	<=110	90,0	55,5 - 98,3	85,4	70,8 - 94,4	83,4	91,3
	<=144	90,0	55,5 - 98,3	36,6	22,1 - 53,1	53,7	81,7
	<=145	100,0	69,0 - 100,0	36,6	22,1 - 53,1	56,3	100,0
	<=223	100,0	69,0 - 100,0	0,0	0,0 - 8,7	45,0	
Err o	>=0	100,0	69,0 - 100,0	0,0	0,0 - 8,7	45,0	
	>4	100,0	69,0 - 100,0	14,6	5,6 - 29,2	48,9	100,0
	>5	90,0	55,5 - 98,3	22,0	10,6 - 37,6	48,5	72,8
	>9	90,0	55,5 - 98,3	41,5	26,3 - 57,9	55,7	83,5
	>10	80,0	44,4 - 96,9	41,5	26,3 - 57,9	52,8	71,7
	>11	80,0	44,4 - 96,9	48,8	32,9 - 64,9	56,1	74,9
	>12	70,0	34,8 - 93,0	58,5	42,1 - 73,7	58,0	70,5
	>17	70,0	34,8 - 93,0	65,9	49,4 - 79,9	62,6	72,8
	>18	60,0	26,4 - 87,6	68,3	51,9 - 81,9	60,8	67,6
	>20	50,0	18,9 - 81,1	70,7	54,5 - 83,9	58,3	63,4
	>24	50,0	18,9 - 81,1	73,2	57,1 - 85,8	60,4	64,1
	>27	30,0	7,0 - 65,2	75,6	59,7 - 87,6	50,2	56,9
	>28	30,0	7,0 - 65,2	80,5	65,1 - 91,2	55,7	58,4
	>30	20,0	3,1 - 55,6	82,9	67,9 - 92,8	48,9	55,9
	>54	20,0	3,1 - 55,6	100,0	91,3 - 100,0	100,0	60,4
	>106	0,0	0,0 - 31,0	100,0	91,3 - 100,0		55,0
	Err c	>=0	100,0	69,0 - 100,0	0,0	0,0 - 8,7	45,0
>0		80,0	44,4 - 96,9	53,7	37,4 - 69,3	58,5	76,6

Variable	Punto de corte	Sensibilidad	IC 95%	Especificidad	IC 95%	VP+	VP-
Err c	>1	70,0	34,8 - 93,0	68,3	51,9 - 81,9	64,4	73,6
	>2	50,0	18,9 - 81,1	75,6	59,7 - 87,6	62,6	64,9
	>3	50,0	18,9 - 81,1	78,0	62,4 - 89,4	65,1	65,6
	>4	30,0	7,0 - 65,2	82,9	67,9 - 92,8	59,0	59,1
	>6	30,0	7,0 - 65,2	95,1	83,4 - 99,3	83,4	62,4
	>12	10,0	1,7 - 44,5	95,1	83,4 - 99,3	62,6	56,4
	>55	10,0	1,7 - 44,5	100,0	91,3 - 100,0	100,0	57,6
	>71	0,0	0,0 - 31,0	100,0	91,3 - 100,0		55,0
	TOT	< 112	0,0	0,0 - 31,0	100,0	91,3 - 100,0	
	<=167	50,0	18,9 - 81,1	100,0	91,3 - 100,0	100,0	71,0
	<=216	50,0	18,9 - 81,1	92,7	80,1 - 98,4	84,8	69,4
	<=224	60,0	26,4 - 87,6	92,7	80,1 - 98,4	87,0	73,9
	<=241	60,0	26,4 - 87,6	90,2	76,9 - 97,2	83,4	73,4
	<=248	70,0	34,8 - 93,0	90,2	76,9 - 97,2	85,4	78,6
	<=267	70,0	34,8 - 93,0	87,8	73,8 - 95,9	82,4	78,2
	<=269	80,0	44,4 - 96,9	87,8	73,8 - 95,9	84,3	84,3
	<=351	80,0	44,4 - 96,9	43,9	28,5 - 60,2	53,8	72,8
	<=355	90,0	55,5 - 98,3	43,9	28,5 - 60,2	56,8	84,3
	<=357	90,0	55,5 - 98,3	41,5	26,3 - 57,9	55,7	83,5
	<=373	100,0	69,0 - 100,0	41,5	26,3 - 57,9	58,3	100,0
	<=534	100,0	69,0 - 100,0	0,0	0,0 - 8,7	45,0	
CON	< 3	0,0	0,0 - 31,0	100,0	91,3 - 100,0		55,0

Variable	Punto de corte	Sensibilidad	IC 95%	Especificidad	IC 95%	VP+	VP-
CON	<=56	60,0	26,4 - 87,6	100,0	91,3 - 100,0	100,0	75,3
	<=77	60,0	26,4 - 87,6	92,7	80,1 - 98,4	87,0	73,9
	<=78	70,0	34,8 - 93,0	92,7	80,1 - 98,4	88,7	79,1
	<=89	70,0	34,8 - 93,0	87,8	73,8 - 95,9	82,4	78,2
	<=91	80,0	44,4 - 96,9	87,8	73,8 - 95,9	84,3	84,3
	<=106	80,0	44,4 - 96,9	85,4	70,8 - 94,4	81,7	83,9
	<=108	90,0	55,5 - 98,3	85,4	70,8 - 94,4	83,4	91,3
	<=137	90,0	55,5 - 98,3	43,9	28,5 - 60,2	56,8	84,3
	<=143	100,0	69,0 - 100,0	43,9	28,5 - 60,2	59,3	100,0
	<=223	100,0	69,0 - 100,0	0,0	0,0 - 8,7	45,0	
	VAR	>=5	100,0	69,0 - 100,0	0,0	0,0 - 8,7	45,0
>5		100,0	69,0 - 100,0	2,4	0,4 - 12,9	45,6	100,0
>7		90,0	55,5 - 98,3	7,3	1,6 - 19,9	44,3	47,2
>8		80,0	44,4 - 96,9	9,8	2,8 - 23,1	42,0	37,4
>9		80,0	44,4 - 96,9	14,6	5,6 - 29,2	43,4	47,2
>10		70,0	34,8 - 93,0	29,3	16,1 - 45,5	44,7	54,4
>11		70,0	34,8 - 93,0	41,5	26,3 - 57,9	49,5	62,8
>12		60,0	26,4 - 87,6	51,2	35,1 - 67,1	50,2	61,0
>13		60,0	26,4 - 87,6	58,5	42,1 - 73,7	54,2	64,1
>14		50,0	18,9 - 81,1	70,7	54,5 - 83,9	58,3	63,4
>15		40,0	12,4 - 73,6	73,2	57,1 - 85,8	55,0	59,8
>17	40,0	12,4 - 73,6	75,6	59,7 - 87,6	57,3	60,6	

Variable	Punto de corte	Sensibilidad	IC 95%	Especificidad	IC 95%	VP+	VP-
VAR	>18	30,0	7,0 - 65,2	80,5	65,1 - 91,2	55,7	58,4
	>19	20,0	3,1 - 55,6	82,9	67,9 - 92,8	48,9	55,9
	>25	20,0	3,1 - 55,6	95,1	83,4 - 99,3	77,0	59,2
	>26	10,0	1,7 - 44,5	95,1	83,4 - 99,3	62,6	56,4
	>31	10,0	1,7 - 44,5	100,0	91,3 - 100,0	100,0	57,6
	>34	0,0	0,0 - 31,0	100,0	91,3 - 100,0		55,0

Nota: Err c= errores de comisión; Err o= errores omisión; IC= intervalo de confianza; VP+= valor predictivo positivo; VP-= valor predictivo negativo.

Anexo 2.3.1. Sensibilidad, especificidad y valores predictivos positivos y negativos asociados a un porcentaje de base del **15%** para diversos puntos de corte de las variables de las pruebas de **fluidez**

Variable	Punto de corte	Sensibilidad	IC 95%	Especificidad	IC 95%	VP+	VP-
Fluidez fonética	< 6	0,0	0,0 - 33,8	100,0	91,7 - 100,0		85,0
	<=9	0,0	0,0 - 33,8	93,0	80,9 - 98,5	0,0	84,1
	<=10	11,1	1,8 - 48,3	90,7	77,8 - 97,3	17,4	85,3
	<=11	22,2	3,5 - 59,9	90,7	77,8 - 97,3	29,7	86,9
	<=19	22,2	3,5 - 59,9	72,1	56,3 - 84,7	12,3	84,0
	<=20	44,4	14,0 - 78,6	67,4	51,5 - 80,9	19,4	87,3
	<=21	44,4	14,0 - 78,6	65,1	49,1 - 79,0	18,4	86,9
	<=22	55,6	21,4 - 86,0	62,8	46,7 - 77,0	20,9	88,9
	<=23	55,6	21,4 - 86,0	60,5	44,4 - 75,0	19,9	88,5
	<=24	66,7	30,1 - 92,1	51,2	35,5 - 66,7	19,4	89,7
	<=25	66,7	30,1 - 92,1	46,5	31,2 - 62,3	18,0	88,8
	<=26	77,8	40,1 - 96,5	39,5	25,0 - 55,6	18,5	91,0
	<=29	77,8	40,1 - 96,5	32,6	19,1 - 48,5	16,9	89,2
	<=31	100,0	66,2 - 100,0	27,9	15,3 - 43,7	19,7	100,0
<=57	100,0	66,2 - 100,0	0,0	0,0 - 8,3	15,0		
Semántica Animales	< 7	0,0	0,0 - 33,8	100,0	91,7 - 100,0		85,0
	<=7	11,1	1,8 - 48,3	100,0	91,7 - 100,0	100,0	86,4
	<=10	22,2	3,5 - 59,9	93,0	80,9 - 98,5	36,0	87,1
	<=11	33,3	7,9 - 69,9	90,7	77,8 - 97,3	38,7	88,5
	<=13	33,3	7,9 - 69,9	79,1	64,0 - 89,9	21,9	87,0
	<=14	44,4	14,0 - 78,6	72,1	56,3 - 84,7	21,9	88,0

Variable	Punto de corte	Sensibilidad	IC 95%	Especificidad	IC 95%	VP+	VP-
	<=15	66,7	30,1 - 92,1	60,5	44,4 - 75,0	22,9	91,1
	<=16	77,8	40,1 - 96,5	55,8	39,9 - 70,9	23,7	93,4
	<=17	88,9	51,7 - 98,2	46,5	31,2 - 62,3	22,7	96,0
	<=18	88,9	51,7 - 98,2	44,2	29,1 - 60,1	21,9	95,8
	<=19	100,0	66,2 - 100,0	34,9	21,0 - 50,9	21,3	100,0
	<=32	100,0	66,2 - 100,0	0,0	0,0 - 8,3	15,0	
Semántica frutas	< 7	0,0	0,0 - 33,8	100,0	91,7 - 100,0		85,0
	<=7	11,1	1,8 - 48,3	100,0	91,7 - 100,0	100,0	86,4
	<=9	11,1	1,8 - 48,3	83,7	69,3 - 93,2	10,8	84,2
	<=10	44,4	14,0 - 78,6	79,1	64,0 - 89,9	27,3	89,0
	<=11	55,6	21,4 - 86,0	69,8	53,9 - 82,8	24,5	89,9
	<=12	88,9	51,7 - 98,2	53,5	37,7 - 68,8	25,2	96,5
	<=13	88,9	51,7 - 98,2	41,9	27,0 - 57,9	21,2	95,5
	<=14	100,0	66,2 - 100,0	32,6	19,1 - 48,5	20,7	100,0
	<=21	100,0	66,2 - 100,0	0,0	0,0 - 8,3	15,0	

Nota: IC= intervalo de confianza; VP+= valor predictivo positivo; VP-= valor predictivo negativo.

Anexo 2.3.2. Sensibilidad, especificidad y valores predictivos positivos y negativos asociados a un porcentaje de base del 30% para diversos puntos de corte de las variables de las pruebas de fluidez

Variable	Punto de corte	Sensibilidad	IC 95%	Especificidad	IC 95%	VP+	VP-
Fluidez fonética	< 6	0,0	0,0 - 33,8	100,0	91,7 - 100,0		70,0
	<=9	0,0	0,0 - 33,8	93,0	80,9 - 98,5	0,0	68,5
	<=10	11,1	1,8 - 48,3	90,7	77,8 - 97,3	33,9	70,4
	<=11	22,2	3,5 - 59,9	90,7	77,8 - 97,3	50,6	73,1
	<=19	22,2	3,5 - 59,9	72,1	56,3 - 84,7	25,4	68,4
	<=20	44,4	14,0 - 78,6	67,4	51,5 - 80,9	36,9	73,9
	<=21	44,4	14,0 - 78,6	65,1	49,1 - 79,0	35,3	73,2
	<=22	55,6	21,4 - 86,0	62,8	46,7 - 77,0	39,0	76,7
	<=23	55,6	21,4 - 86,0	60,5	44,4 - 75,0	37,6	76,0
	<=24	66,7	30,1 - 92,1	51,2	35,5 - 66,7	36,9	78,2
	<=25	66,7	30,1 - 92,1	46,5	31,2 - 62,3	34,8	76,5
	<=26	77,8	40,1 - 96,5	39,5	25,0 - 55,6	35,5	80,6
	<=29	77,8	40,1 - 96,5	32,6	19,1 - 48,5	33,1	77,4
	<=31	100,0	66,2 - 100,0	27,9	15,3 - 43,7	37,3	100,0
<=57	100,0	66,2 - 100,0	0,0	0,0 - 8,3	30,0		
Semántica Animales	< 7	0,0	0,0 - 33,8	100,0	91,7 - 100,0		70,0
	<=7	11,1	1,8 - 48,3	100,0	91,7 - 100,0	100,0	72,4
	<=10	22,2	3,5 - 59,9	93,0	80,9 - 98,5	57,7	73,6
	<=11	33,3	7,9 - 69,9	90,7	77,8 - 97,3	60,6	76,0
	<=13	33,3	7,9 - 69,9	79,1	64,0 - 89,9	40,6	73,5
	<=14	44,4	14,0 - 78,6	72,1	56,3 - 84,7	40,6	75,2

Variable	Punto de corte	Sensibilidad	IC 95%	Especificidad	IC 95%	VP+	VP-
Semántica Animales	<=15	66,7	30,1 - 92,1	60,5	44,4 - 75,0	42,0	80,9
	<=16	77,8	40,1 - 96,5	55,8	39,9 - 70,9	43,0	85,4
	<=17	88,9	51,7 - 98,2	46,5	31,2 - 62,3	41,6	90,7
	<=18	88,9	51,7 - 98,2	44,2	29,1 - 60,1	40,6	90,3
	<=19	100,0	66,2 - 100,0	34,9	21,0 - 50,9	39,7	100,0
	<=32	100,0	66,2 - 100,0	0,0	0,0 - 8,3	30,0	
Semántica frutas	< 7	0,0	0,0 - 33,8	100,0	91,7 - 100,0		70,0
	<=7	11,1	1,8 - 48,3	100,0	91,7 - 100,0	100,0	72,4
	<=9	11,1	1,8 - 48,3	83,7	69,3 - 93,2	22,6	68,7
	<=10	44,4	14,0 - 78,6	79,1	64,0 - 89,9	47,6	76,9
	<=11	55,6	21,4 - 86,0	69,8	53,9 - 82,8	44,1	78,6
	<=12	88,9	51,7 - 98,2	53,5	37,7 - 68,8	45,0	91,8
	<=13	88,9	51,7 - 98,2	41,9	27,0 - 57,9	39,6	89,8
	<=14	100,0	66,2 - 100,0	32,6	19,1 - 48,5	38,9	100,0
	<=21	100,0	66,2 - 100,0	0,0	0,0 - 8,3	30,0	

Nota: IC= intervalo de confianza; VP+= valor predictivo positivo; VP-= valor predictivo negativo.

Anexo 2.3.3. Sensibilidad, especificidad y valores predictivos positivos y negativos asociados a un porcentaje de base del 45% para diversos puntos de corte de las variables de las pruebas de **fluidez**

Variable	Punto de corte	Sensibilidad	IC 95%	Especificidad	IC 95%	VP+	VP-
Fluidez fonética	< 6	0,0	0,0 - 33,8	100,0	91,7 - 100,0		55,0
	<=9	0,0	0,0 - 33,8	93,0	80,9 - 98,5	0,0	53,2
	<=10	11,1	1,8 - 48,3	90,7	77,8 - 97,3	49,4	55,5
	<=11	22,2	3,5 - 59,9	90,7	77,8 - 97,3	66,2	58,8
	<=19	22,2	3,5 - 59,9	72,1	56,3 - 84,7	39,4	53,1
	<=20	44,4	14,0 - 78,6	67,4	51,5 - 80,9	52,8	59,7
	<=21	44,4	14,0 - 78,6	65,1	49,1 - 79,0	51,0	58,9
	<=22	55,6	21,4 - 86,0	62,8	46,7 - 77,0	55,0	63,3
	<=23	55,6	21,4 - 86,0	60,5	44,4 - 75,0	53,5	62,4
	<=24	66,7	30,1 - 92,1	51,2	35,5 - 66,7	52,8	65,2
	<=25	66,7	30,1 - 92,1	46,5	31,2 - 62,3	50,5	63,0
	<=26	77,8	40,1 - 96,5	39,5	25,0 - 55,6	51,3	68,5
	<=29	77,8	40,1 - 96,5	32,6	19,1 - 48,5	48,5	64,2
	<=31	100,0	66,2 - 100,0	27,9	15,3 - 43,7	53,2	100,0
<=57	100,0	66,2 - 100,0	0,0	0,0 - 8,3	45,0		
Semántica Animales	< 7	0,0	0,0 - 33,8	100,0	91,7 - 100,0		55,0
	<=7	11,1	1,8 - 48,3	100,0	91,7 - 100,0	100,0	57,9
	<=10	22,2	3,5 - 59,9	93,0	80,9 - 98,5	72,3	59,4
	<=11	33,3	7,9 - 69,9	90,7	77,8 - 97,3	74,6	62,4
	<=13	33,3	7,9 - 69,9	79,1	64,0 - 89,9	56,6	59,2
	<=14	44,4	14,0 - 78,6	72,1	56,3 - 84,7	56,6	61,3

Variable	Punto de corte	Sensibilidad	IC 95%	Especificidad	IC 95%	VP+	VP-
Semántica Animales	<=15	66,7	30,1 - 92,1	60,5	44,4 - 75,0	58,0	68,9
	<=16	77,8	40,1 - 96,5	55,8	39,9 - 70,9	59,0	75,4
	<=17	88,9	51,7 - 98,2	46,5	31,2 - 62,3	57,6	83,7
	<=18	88,9	51,7 - 98,2	44,2	29,1 - 60,1	56,6	82,9
	<=19	100,0	66,2 - 100,0	34,9	21,0 - 50,9	55,7	100,0
	<=32	100,0	66,2 - 100,0	0,0	0,0 - 8,3	45,0	
Semántica frutas	< 7	0,0	0,0 - 33,8	100,0	91,7 - 100,0		55,0
	<=7	11,1	1,8 - 48,3	100,0	91,7 - 100,0	100,0	57,9
	<=9	11,1	1,8 - 48,3	83,7	69,3 - 93,2	35,8	53,5
	<=10	44,4	14,0 - 78,6	79,1	64,0 - 89,9	63,5	63,5
	<=11	55,6	21,4 - 86,0	69,8	53,9 - 82,8	60,1	65,7
	<=12	88,9	51,7 - 98,2	53,5	37,7 - 68,8	61,0	85,5
	<=13	88,9	51,7 - 98,2	41,9	27,0 - 57,9	55,6	82,2
	<=14	100,0	66,2 - 100,0	32,6	19,1 - 48,5	54,8	100,0
	<=21	100,0	66,2 - 100,0	0,0	0,0 - 8,3	45,0	

Nota: IC= intervalo de confianza; VP+= valor predictivo positivo; VP-= valor predictivo negativo.

Anexo 2.4.1. Sensibilidad, especificidad y valores predictivos positivos y negativos asociados a un porcentaje de base del **15%** para diversos puntos de corte de la variable **del Letras y Números**

Variable	Punto de corte	Sensibilidad	IC 95%	Especificidad	IC 95%	VP+	VP-
Total	< 2	0,0	0,0 - 37,1	100,0	91,5 - 100,0		85,0
	<=2	12,5	2,1 - 52,6	100,0	91,5 - 100,0	100,0	86,6
	<=4	25,0	3,9 - 65,0	95,2	83,8 - 99,3	48,1	87,8
	<=5	25,0	3,9 - 65,0	88,1	74,4 - 96,0	27,0	86,9
	<=6	50,0	16,0 - 84,0	83,3	68,6 - 93,0	34,6	90,4
	<=7	75,0	35,0 - 96,1	69,0	52,9 - 82,4	30,0	94,0
	<=8	87,5	47,4 - 97,9	59,5	43,3 - 74,4	27,6	96,4
	<=9	87,5	47,4 - 97,9	38,1	23,6 - 54,4	20,0	94,5
	<=10	100,0	62,9 - 100,0	23,8	12,1 - 39,5	18,8	100,0
	<=18	100,0	62,9 - 100,0	0,0	0,0 - 8,5	15,0	

Nota: IC= intervalo de confianza; VP+= valor predictivo positivo; VP-= valor predictivo negativo.

Anexo 2.4.2. Sensibilidad, especificidad y valores predictivos positivos y negativos asociados a un porcentaje de base del **30%** para diversos puntos de corte de la variable del **Letras y Números**

Variable	Punto de corte	Sensibilidad	IC 95%	Especificidad	IC 95%	VP+	VP-
Total	< 2	0,0	0,0 - 37,1	100,0	91,5 - 100,0		70,0
	<=2	12,5	2,1 - 52,6	100,0	91,5 - 100,0	100,0	72,7
	<=4	25,0	3,9 - 65,0	95,2	83,8 - 99,3	69,2	74,8
	<=5	25,0	3,9 - 65,0	88,1	74,4 - 96,0	47,4	73,3
	<=6	50,0	16,0 - 84,0	83,3	68,6 - 93,0	56,3	79,5
	<=7	75,0	35,0 - 96,1	69,0	52,9 - 82,4	50,9	86,6
	<=8	87,5	47,4 - 97,9	59,5	43,3 - 74,4	48,1	91,7
	<=9	87,5	47,4 - 97,9	38,1	23,6 - 54,4	37,7	87,7
	<=10	100,0	62,9 - 100,0	23,8	12,1 - 39,5	36,0	100,0
	<=18	100,0	62,9 - 100,0	0,0	0,0 - 8,5	30,0	

Nota: IC= intervalo de confianza; VP+= valor predictivo positivo; VP-= valor predictivo negativo.

Anexo 2.4.3. Sensibilidad, especificidad y valores predictivos positivos y negativos asociados a un porcentaje de base del **45%** para diversos puntos de corte de la variable del **Letras y Números**

Variable	Punto de corte	Sensibilidad	IC 95%	Especificidad	IC 95%	VP+	VP-
Total	< 2	0,0	0,0 - 37,1	100,0	91,5 - 100,0		55,0
	<=2	12,5	2,1 - 52,6	100,0	91,5 - 100,0	100,0	58,3
	<=4	25,0	3,9 - 65,0	95,2	83,8 - 99,3	81,1	60,8
	<=5	25,0	3,9 - 65,0	88,1	74,4 - 96,0	63,2	58,9
	<=6	50,0	16,0 - 84,0	83,3	68,6 - 93,0	71,1	67,1
	<=7	75,0	35,0 - 96,1	69,0	52,9 - 82,4	66,5	77,1
	<=8	87,5	47,4 - 97,9	59,5	43,3 - 74,4	63,9	85,3
	<=9	87,5	47,4 - 97,9	38,1	23,6 - 54,4	53,6	78,8
	<=10	100,0	62,9 - 100,0	23,8	12,1 - 39,5	51,8	100,0
	<=18	100,0	62,9 - 100,0	0,0	0,0 - 8,5	45,0	

Nota: IC= intervalo de confianza; VP+= valor predictivo positivo; VP-= valor predictivo negativo.

Anexo 2.5.1. Sensibilidad, especificidad y valores predictivos positivos y negativos asociados a un porcentaje de base del **15%** para diversos puntos de corte de la variable del **Cambios**

Variable	Punto de corte	Sensibilidad	IC 95%	Especificidad	IC 95%	VP+	VP-
Total	< 1	0,0	0,0 - 52,0	100,0	90,9 - 100,0		85,0
	<=1	20,0	3,3 - 71,2	97,4	86,5 - 99,6	57,9	87,3
	<=2	20,0	3,3 - 71,2	92,3	79,1 - 98,3	31,5	86,7
	<=3	40,0	6,5 - 84,6	92,3	79,1 - 98,3	47,9	89,7
	<=6	40,0	6,5 - 84,6	64,1	47,2 - 78,8	16,4	85,8
	<=7	60,0	15,4 - 93,5	56,4	39,6 - 72,2	19,5	88,9
	<=8	80,0	28,8 - 96,7	53,8	37,2 - 69,9	23,4	93,8
	<=12	80,0	28,8 - 96,7	20,5	9,3 - 36,5	15,1	85,3
	<=13	100,0	48,0 - 100,0	20,5	9,3 - 36,5	18,2	100,0
	<=24	100,0	48,0 - 100,0	0,0	0,0 - 9,1	15,0	

Nota: IC= intervalo de confianza; VP+= valor predictivo positivo; VP-= valor predictivo negativo.

Anexo 2.5.2. Sensibilidad, especificidad y valores predictivos positivos y negativos asociados a un porcentaje de base del **30%** para diversos puntos de corte de la variable del **Cambios**

Variable	Punto de corte	Sensibilidad	IC 95%	Especificidad	IC 95%	VP+	VP-
Total	< 1	0,0	0,0 - 52,0	100,0	90,9 - 100,0		70,0
	<=1	20,0	3,3 - 71,2	97,4	86,5 - 99,6	77,0	74,0
	<=2	20,0	3,3 - 71,2	92,3	79,1 - 98,3	52,7	72,9
	<=3	40,0	6,5 - 84,6	92,3	79,1 - 98,3	69,0	78,2
	<=6	40,0	6,5 - 84,6	64,1	47,2 - 78,8	32,3	71,4
	<=7	60,0	15,4 - 93,5	56,4	39,6 - 72,2	37,1	76,7
	<=8	80,0	28,8 - 96,7	53,8	37,2 - 69,9	42,6	86,3
	<=12	80,0	28,8 - 96,7	20,5	9,3 - 36,5	30,1	70,5
	<=13	100,0	48,0 - 100,0	20,5	9,3 - 36,5	35,0	100,0
	<=24	100,0	48,0 - 100,0	0,0	0,0 - 9,1	30,0	

Nota: IC= intervalo de confianza; VP+= valor predictivo positivo; VP-= valor predictivo negativo.

Anexo 2.5.3. Sensibilidad, especificidad y valores predictivos positivos y negativos asociados a un porcentaje de base del **45%** para diversos puntos de corte de la variable del **Cambios**

Variable	Punto de corte	Sensibilidad	IC 95%	Especificidad	IC 95%	VP+	VP-
Total	< 1	0,0	0,0 - 52,0	100,0	90,9 - 100,0		55,0
	<=1	20,0	3,3 - 71,2	97,4	86,5 - 99,6	86,5	59,8
	<=2	20,0	3,3 - 71,2	92,3	79,1 - 98,3	68,0	58,5
	<=3	40,0	6,5 - 84,6	92,3	79,1 - 98,3	81,0	65,3
	<=6	40,0	6,5 - 84,6	64,1	47,2 - 78,8	47,7	56,6
	<=7	60,0	15,4 - 93,5	56,4	39,6 - 72,2	53,0	63,3
	<=8	80,0	28,8 - 96,7	53,8	37,2 - 69,9	58,6	76,7
	<=12	80,0	28,8 - 96,7	20,5	9,3 - 36,5	45,2	55,6
	<=13	100,0	48,0 - 100,0	20,5	9,3 - 36,5	50,7	100,0
	<=24	100,0	48,0 - 100,0	0,0	0,0 - 9,1	45,0	

Nota: IC= intervalo de confianza; VP+= valor predictivo positivo; VP-= valor predictivo negativo.

Anexo 2.6.1. Sensibilidad, especificidad y valores predictivos positivos y negativos asociados a un porcentaje de base del **15%** para diversos puntos de corte de las variables del **Finger Tapping Test**

Variable	Punto de corte	Sensibilidad	IC 95%	Especificidad	IC 95%	VP+	VP-
Mano Dominante	< 10,2	0,0	0,0 - 33,8	100,0	91,9 - 100,0		85,0
	<=10,2	11,1	1,8 - 48,3	100,0	91,9 - 100,0	100,0	86,4
	<=15,8	11,1	1,8 - 48,3	97,7	87,9 - 99,6	46,3	86,2
	<=18,6	44,4	14,0 - 78,6	97,7	87,9 - 99,6	77,5	90,9
	<=31,16	44,4	14,0 - 78,6	81,8	67,3 - 91,8	30,1	89,3
	<=32,6	55,6	21,4 - 86,0	81,8	67,3 - 91,8	35,0	91,3
	<=34	55,6	21,4 - 86,0	70,5	54,8 - 83,2	24,9	90,0
	<=35,4	77,8	40,1 - 96,5	70,5	54,8 - 83,2	31,7	94,7
	<=39,4	77,8	40,1 - 96,5	50,0	34,6 - 65,4	21,5	92,7
	<=42	88,9	51,7 - 98,2	47,7	32,5 - 63,3	23,1	96,1
	<=45,4	88,9	51,7 - 98,2	27,3	15,0 - 42,8	17,7	93,3
	<=46,4	100,0	66,2 - 100,0	27,3	15,0 - 42,8	19,5	100,0
	<=62,8	100,0	66,2 - 100,0	0,0	0,0 - 8,1	15,0	
	Mano No-dominante	< 8	0,0	0,0 - 33,8	100,0	91,9 - 100,0	
<=14,6		22,2	3,5 - 59,9	100,0	91,9 - 100,0	100,0	87,9
<=18,33		22,2	3,5 - 59,9	95,5	84,5 - 99,3	46,3	87,4
<=20,2		44,4	14,0 - 78,6	95,5	84,5 - 99,3	63,3	90,7
<=26,4		44,4	14,0 - 78,6	86,4	72,6 - 94,8	36,5	89,8
<=26,6		55,6	21,4 - 86,0	86,4	72,6 - 94,8	41,8	91,7
<=27		55,6	21,4 - 86,0	84,1	69,9 - 93,3	38,1	91,5
<=27,6		66,7	30,1 - 92,1	84,1	69,9 - 93,3	42,5	93,5
<=29,2		66,7	30,1 - 92,1	77,3	62,2 - 88,5	34,1	92,9
<=29,8		77,8	40,1 - 96,5	77,3	62,2 - 88,5	37,7	95,2
<=31,8		77,8	40,1 - 96,5	68,2	52,4 - 81,4	30,1	94,6
<=32		88,9	51,7 - 98,2	68,2	52,4 - 81,4	33,0	97,2
<=41,4		88,9	51,7 - 98,2	20,5	9,8 - 35,3	16,5	91,3
<=41,7		100,0	66,2 - 100,0	20,5	9,8 - 35,3	18,2	100,0
<=55,6	100,0	66,2 - 100,0	0,0	0,0 - 8,1	15,0		

Nota: IC= intervalo de confianza; VP+= valor predictivo positivo; VP-= valor predictivo negativo.

Anexo 2.6.2. Sensibilidad, especificidad y valores predictivos positivos y negativos asociados a un porcentaje de base del **30%** para diversos puntos de corte de las variables del **Finger Tapping Test**

Variable	Punto de corte	Sensibilidad	IC 95%	Especificidad	IC 95%	VP+	VP-
Mano Dominante	< 10,2	0,0	0,0 - 33,8	100,0	91,9 - 100,0		70,0
	<=10,2	11,1	1,8 - 48,3	100,0	91,9 - 100,0	100,0	72,4
	<=15,8	11,1	1,8 - 48,3	97,7	87,9 - 99,6	67,7	72,0
	<=18,6	44,4	14,0 - 78,6	97,7	87,9 - 99,6	89,3	80,4
	<=31,16	44,4	14,0 - 78,6	81,8	67,3 - 91,8	51,2	77,5
	<=32,6	55,6	21,4 - 86,0	81,8	67,3 - 91,8	56,7	81,1
	<=34	55,6	21,4 - 86,0	70,5	54,8 - 83,2	44,6	78,7
	<=35,4	77,8	40,1 - 96,5	70,5	54,8 - 83,2	53,0	88,1
	<=39,4	77,8	40,1 - 96,5	50,0	34,6 - 65,4	40,0	84,0
	<=42	88,9	51,7 - 98,2	47,7	32,5 - 63,3	42,2	90,9
	<=45,4	88,9	51,7 - 98,2	27,3	15,0 - 42,8	34,4	85,1
	<=46,4	100,0	66,2 - 100,0	27,3	15,0 - 42,8	37,1	100,0
	<=62,8	100,0	66,2 - 100,0	0,0	0,0 - 8,1	30,0	
Mano No-dominante	< 8	0,0	0,0 - 33,8	100,0	91,9 - 100,0		70,0
	<=14,6	22,2	3,5 - 59,9	100,0	91,9 - 100,0	100,0	75,0
	<=18,33	22,2	3,5 - 59,9	95,5	84,5 - 99,3	67,7	74,1
	<=20,2	44,4	14,0 - 78,6	95,5	84,5 - 99,3	80,7	80,0
	<=26,4	44,4	14,0 - 78,6	86,4	72,6 - 94,8	58,3	78,4
	<=26,6	55,6	21,4 - 86,0	86,4	72,6 - 94,8	63,6	81,9
	<=27	55,6	21,4 - 86,0	84,1	69,9 - 93,3	59,9	81,5
	<=27,6	66,7	30,1 - 92,1	84,1	69,9 - 93,3	64,2	85,5

Variable	Punto de corte	Sensibilidad	IC 95%	Especificidad	IC 95%	VP+	VP-
Mano No-dominante	<=29,2	66,7	30,1 - 92,1	77,3	62,2 - 88,5	55,7	84,4
	<=29,8	77,8	40,1 - 96,5	77,3	62,2 - 88,5	59,5	89,0
	<=31,8	77,8	40,1 - 96,5	68,2	52,4 - 81,4	51,2	87,7
	<=32	88,9	51,7 - 98,2	68,2	52,4 - 81,4	54,5	93,5
	<=41,4	88,9	51,7 - 98,2	20,5	9,8 - 35,3	32,4	81,1
	<=41,7	100,0	66,2 - 100,0	20,5	9,8 - 35,3	35,0	100,0
	<=55,6	100,0	66,2 - 100,0	0,0	0,0 - 8,1	30,0	

Nota: IC= intervalo de confianza; VP+= valor predictivo positivo; VP-= valor predictivo negativo.

Anexo 2.6.3. Sensibilidad, especificidad y valores predictivos positivos y negativos asociados a un porcentaje de base del **45%** para diversos puntos de corte de las variables del **Finger Tapping Test**

Variable	Punto de corte	Sensibilidad	IC 95%	Especificidad	IC 95%	VP+	VP-
Mano Dominante	< 10,2	0,0	0,0 - 33,8	100,0	91,9 - 100,0		55,0
	<=10,2	11,1	1,8 - 48,3	100,0	91,9 - 100,0	100,0	57,9
	<=15,8	11,1	1,8 - 48,3	97,7	87,9 - 99,6	80,0	57,3
	<=18,6	44,4	14,0 - 78,6	97,7	87,9 - 99,6	94,1	68,3
	<=31,16	44,4	14,0 - 78,6	81,8	67,3 - 91,8	66,7	64,3
	<=32,6	55,6	21,4 - 86,0	81,8	67,3 - 91,8	71,4	69,2
	<=34	55,6	21,4 - 86,0	70,5	54,8 - 83,2	60,6	66,0
	<=35,4	77,8	40,1 - 96,5	70,5	54,8 - 83,2	68,3	79,5
	<=39,4	77,8	40,1 - 96,5	50,0	34,6 - 65,4	56,0	73,3
	<=42	88,9	51,7 - 98,2	47,7	32,5 - 63,3	58,2	84,0
	<=45,4	88,9	51,7 - 98,2	27,3	15,0 - 42,8	50,0	75,0
	<=46,4	100,0	66,2 - 100,0	27,3	15,0 - 42,8	52,9	100,0
	<=62,8	100,0	66,2 - 100,0	0,0	0,0 - 8,1	45,0	
Mano No-dominante	< 8	0,0	0,0 - 33,8	100,0	91,9 - 100,0		55,0
	<=14,6	22,2	3,5 - 59,9	100,0	91,9 - 100,0	100,0	61,1
	<=18,33	22,2	3,5 - 59,9	95,5	84,5 - 99,3	80,0	60,0
	<=20,2	44,4	14,0 - 78,6	95,5	84,5 - 99,3	88,9	67,7
	<=26,4	44,4	14,0 - 78,6	86,4	72,6 - 94,8	72,7	65,5
	<=26,6	55,6	21,4 - 86,0	86,4	72,6 - 94,8	76,9	70,4
	<=27	55,6	21,4 - 86,0	84,1	69,9 - 93,3	74,1	69,8
	<=27,6	66,7	30,1 - 92,1	84,1	69,9 - 93,3	77,4	75,5

Variable	Punto de corte	Sensibilidad	IC 95%	Especificidad	IC 95%	VP+	VP-
	<=29,2	66,7	30,1 - 92,1	77,3	62,2 - 88,5	70,6	73,9
	<=29,8	77,8	40,1 - 96,5	77,3	62,2 - 88,5	73,7	81,0
	<=31,8	77,8	40,1 - 96,5	68,2	52,4 - 81,4	66,7	78,9
	<=32	88,9	51,7 - 98,2	68,2	52,4 - 81,4	69,6	88,2
	<=41,4	88,9	51,7 - 98,2	20,5	9,8 - 35,3	47,8	69,2
	<=41,7	100,0	66,2 - 100,0	20,5	9,8 - 35,3	50,7	100,0
	<=55,6	100,0	66,2 - 100,0	0,0	0,0 - 8,1	45,0	

Nota: IC= intervalo de confianza; VP+= valor predictivo positivo; VP-= valor predictivo negativo.

Anexo 2.7.1. Sensibilidad, especificidad y valores predictivos positivos y negativos asociados a un porcentaje de base del **15%** para diversos puntos de corte de las variables del **Stroop**

Variable	Punto de corte	Sensibilidad	IC 95%	Especificidad	IC 95%	VP+	VP-
Palabras	< 21	0,0	0,0 - 31,0	100,0	91,9 - 100,0		85,0
	<=50	40,0	12,4 - 73,6	100,0	91,9 - 100,0	100,0	90,4
	<=53	40,0	12,4 - 73,6	95,5	84,5 - 99,3	60,8	90,0
	<=60	50,0	18,9 - 81,1	95,5	84,5 - 99,3	66,0	91,5
	<=63	50,0	18,9 - 81,1	93,2	81,3 - 98,5	56,4	91,3
	<=64	60,0	26,4 - 87,6	93,2	81,3 - 98,5	60,8	93,0
	<=65	60,0	26,4 - 87,6	90,9	78,3 - 97,4	53,8	92,8
	<=66	70,0	34,8 - 93,0	90,9	78,3 - 97,4	57,6	94,5
	<=85	70,0	34,8 - 93,0	65,9	50,1 - 79,5	26,6	92,6
	<=86	80,0	44,4 - 96,9	61,4	45,5 - 75,6	26,8	94,6
	<=90	80,0	44,4 - 96,9	52,3	36,7 - 67,5	22,8	93,7
	<=93	90,0	55,5 - 98,3	52,3	36,7 - 67,5	25,0	96,7
	<=101	90,0	55,5 - 98,3	36,4	22,4 - 52,2	20,0	95,4
	<=102	100,0	69,0 - 100,0	29,5	16,8 - 45,2	20,0	100,0
<=127	100,0	69,0 - 100,0	0,0	0,0 - 8,1	15,0		
Colores	< 12	0,0	0,0 - 31,0	100,0	91,9 - 100,0		85,0
	<=12	10,0	1,7 - 44,5	100,0	91,9 - 100,0	100,0	86,3
	<=31	10,0	1,7 - 44,5	95,5	84,5 - 99,3	28,0	85,7
	<=42	40,0	12,4 - 73,6	95,5	84,5 - 99,3	60,8	90,0
	<=43	50,0	18,9 - 81,1	93,2	81,3 - 98,5	56,4	91,3
	<=47	50,0	18,9 - 81,1	90,9	78,3 - 97,4	49,3	91,2
	<=49	60,0	26,4 - 87,6	88,6	75,4 - 96,2	48,2	92,6

Variable	Punto de corte	Sensibilidad	IC 95%	Especificidad	IC 95%	VP+	VP-
Colores	<=57	60,0	26,4 - 87,6	70,5	54,8 - 83,2	26,4	90,9
	<=59	70,0	34,8 - 93,0	70,5	54,8 - 83,2	29,5	93,0
	<=60	70,0	34,8 - 93,0	61,4	45,5 - 75,6	24,2	92,1
	<=61	80,0	44,4 - 96,9	61,4	45,5 - 75,6	26,8	94,6
	<=64	80,0	44,4 - 96,9	54,5	38,9 - 69,6	23,7	93,9
	<=65	90,0	55,5 - 98,3	50,0	34,6 - 65,4	24,1	96,6
	<=70	90,0	55,5 - 98,3	34,1	20,5 - 49,9	19,4	95,1
	<=71	100,0	69,0 - 100,0	27,3	15,0 - 42,8	19,5	100,0
	<=98	100,0	69,0 - 100,0	0,0	0,0 - 8,1	15,0	
	PalxCol	< 9	0,0	0,0 - 31,0	100,0	91,9 - 100,0	
<=18		50,0	18,9 - 81,1	100,0	91,9 - 100,0	100,0	91,9
<=23		50,0	18,9 - 81,1	88,6	75,4 - 96,2	43,7	90,9
<=25		60,0	26,4 - 87,6	88,6	75,4 - 96,2	48,2	92,6
<=26		70,0	34,8 - 93,0	86,4	72,6 - 94,8	47,5	94,2
<=29		70,0	34,8 - 93,0	79,5	64,7 - 90,2	37,7	93,8
<=30		80,0	44,4 - 96,9	72,7	57,2 - 85,0	34,1	95,4
<=31		80,0	44,4 - 96,9	68,2	52,4 - 81,4	30,7	95,1
<=32		90,0	55,5 - 98,3	65,9	50,1 - 79,5	31,8	97,4
<=42		90,0	55,5 - 98,3	29,5	16,8 - 45,2	18,4	94,4
<=43		100,0	69,0 - 100,0	27,3	15,0 - 42,8	19,5	100,0
<=72		100,0	69,0 - 100,0	0,0	0,0 - 8,1	15,0	
Interferencia		< - 19,39	0,0	0,0 - 31,0	100,0	91,9 - 100,0	
	<= - 14,07	0,0	0,0 - 31,0	93,2	81,3 - 98,5	0,0	84,1

Variable	Punto de corte	Sensibilidad	IC 95%	Especificidad	IC 95%	VP+	VP-
Interferencia	<=-13,25	10,0	1,7 - 44,5	93,2	81,3 - 98,5	20,6	85,4
	<=-12,18	10,0	1,7 - 44,5	88,6	75,4 - 96,2	13,4	84,8
	<=-7,71	40,0	12,4 - 73,6	88,6	75,4 - 96,2	38,3	89,3
	<=-6,6	40,0	12,4 - 73,6	81,8	67,3 - 91,8	28,0	88,5
	<=-5,95	50,0	18,9 - 81,1	81,8	67,3 - 91,8	32,7	90,3
	<=-5,93	50,0	18,9 - 81,1	79,5	64,7 - 90,2	30,1	90,0
	<=-5,15	60,0	26,4 - 87,6	79,5	64,7 - 90,2	34,1	91,8
	<=-4,6	60,0	26,4 - 87,6	75,0	59,7 - 86,8	29,8	91,4
	<=-3,09	70,0	34,8 - 93,0	75,0	59,7 - 86,8	33,1	93,4
	<=1,33	70,0	34,8 - 93,0	52,3	36,7 - 67,5	20,6	90,8
	<=1,37	80,0	44,4 - 96,9	52,3	36,7 - 67,5	22,8	93,7
	<=4,23	80,0	44,4 - 96,9	31,8	18,6 - 47,6	17,2	90,0
	<=5,03	100,0	69,0 - 100,0	31,8	18,6 - 47,6	20,6	100,0
	<=28,11	100,0	69,0 - 100,0	0,0	0,0 - 8,1	15,0	

Nota: IC= intervalo de confianza; PalxCol= palabras emitidas en la condición de interferencia; VP+= valor predictivo positivo; VP-= valor predictivo negativo.

Anexo 2.7.2. Sensibilidad, especificidad y valores predictivos positivos y negativos asociados a un porcentaje de base del **30%** para diversos puntos de corte de las variables del **Stroop**

Variable	Punto de corte	Sensibilidad	IC 95%	Especificidad	IC 95%	VP+	VP-
Palabras	< 21	0,0	0,0 - 31,0	100,0	91,9 - 100,0		70,0
	<=50	40,0	12,4 - 73,6	100,0	91,9 - 100,0	100,0	79,5
	<=53	40,0	12,4 - 73,6	95,5	84,5 - 99,3	79,0	78,8
	<=60	50,0	18,9 - 81,1	95,5	84,5 - 99,3	82,5	81,7
	<=63	50,0	18,9 - 81,1	93,2	81,3 - 98,5	75,9	81,3
	<=64	60,0	26,4 - 87,6	93,2	81,3 - 98,5	79,0	84,5
	<=65	60,0	26,4 - 87,6	90,9	78,3 - 97,4	73,9	84,1
	<=66	70,0	34,8 - 93,0	90,9	78,3 - 97,4	76,7	87,6
	<=85	70,0	34,8 - 93,0	65,9	50,1 - 79,5	46,8	83,7
	<=86	80,0	44,4 - 96,9	61,4	45,5 - 75,6	47,0	87,7
	<=90	80,0	44,4 - 96,9	52,3	36,7 - 67,5	41,8	85,9
	<=93	90,0	55,5 - 98,3	52,3	36,7 - 67,5	44,7	92,4
	<=101	90,0	55,5 - 98,3	36,4	22,4 - 52,2	37,7	89,5
<=102	100,0	69,0 - 100,0	29,5	16,8 - 45,2	37,8	100,0	
<=127	100,0	69,0 - 100,0	0,0	0,0 - 8,1	30,0		
Colores	< 12	0,0	0,0 - 31,0	100,0	91,9 - 100,0		70,0
	<=12	10,0	1,7 - 44,5	100,0	91,9 - 100,0	100,0	72,2
	<=31	10,0	1,7 - 44,5	95,5	84,5 - 99,3	48,5	71,2
	<=42	40,0	12,4 - 73,6	95,5	84,5 - 99,3	79,0	78,8
	<=43	50,0	18,9 - 81,1	93,2	81,3 - 98,5	75,9	81,3
	<=47	50,0	18,9 - 81,1	90,9	78,3 - 97,4	70,2	80,9
	<=49	60,0	26,4 - 87,6	88,6	75,4 - 96,2	69,4	83,8

Variable	Punto de corte	Sensibilidad	IC 95%	Especificidad	IC 95%	VP+	VP-
Colores	<=57	60,0	26,4 - 87,6	70,5	54,8 - 83,2	46,5	80,4
	<=59	70,0	34,8 - 93,0	70,5	54,8 - 83,2	50,4	84,6
	<=60	70,0	34,8 - 93,0	61,4	45,5 - 75,6	43,7	82,7
	<=61	80,0	44,4 - 96,9	61,4	45,5 - 75,6	47,0	87,7
	<=64	80,0	44,4 - 96,9	54,5	38,9 - 69,6	43,0	86,4
	<=65	90,0	55,5 - 98,3	50,0	34,6 - 65,4	43,5	92,1
	<=70	90,0	55,5 - 98,3	34,1	20,5 - 49,9	36,9	88,8
	<=71	100,0	69,0 - 100,0	27,3	15,0 - 42,8	37,1	100,0
	<=98	100,0	69,0 - 100,0	0,0	0,0 - 8,1	30,0	
	PalxCol	< 9	0,0	0,0 - 31,0	100,0	91,9 - 100,0	
<=18		50,0	18,9 - 81,1	100,0	91,9 - 100,0	100,0	82,4
<=23		50,0	18,9 - 81,1	88,6	75,4 - 96,2	65,3	80,5
<=25		60,0	26,4 - 87,6	88,6	75,4 - 96,2	69,4	83,8
<=26		70,0	34,8 - 93,0	86,4	72,6 - 94,8	68,8	87,0
<=29		70,0	34,8 - 93,0	79,5	64,7 - 90,2	59,5	86,1
<=30		80,0	44,4 - 96,9	72,7	57,2 - 85,0	55,7	89,5
<=31		80,0	44,4 - 96,9	68,2	52,4 - 81,4	51,9	88,8
<=32		90,0	55,5 - 98,3	65,9	50,1 - 79,5	53,1	93,9
<=42		90,0	55,5 - 98,3	29,5	16,8 - 45,2	35,4	87,3
Interferencia	< - 19,39	0,0	0,0 - 31,0	100,0	91,9 - 100,0		70,0
	<=- 14,07	0,0	0,0 - 31,0	93,2	81,3 - 98,5	0,0	68,5

Variable	Punto de corte	Sensibilidad	IC 95%	Especificidad	IC 95%	VP+	VP-
Interferencia	<=-13,25	10,0	1,7 - 44,5	93,2	81,3 - 98,5	38,6	70,7
	<=-12,18	10,0	1,7 - 44,5	88,6	75,4 - 96,2	27,4	69,7
	<=-7,71	40,0	12,4 - 73,6	88,6	75,4 - 96,2	60,1	77,5
	<=-6,6	40,0	12,4 - 73,6	81,8	67,3 - 91,8	48,5	76,1
	<=-5,95	50,0	18,9 - 81,1	81,8	67,3 - 91,8	54,1	79,2
	<=-5,93	50,0	18,9 - 81,1	79,5	64,7 - 90,2	51,2	78,8
	<=-5,15	60,0	26,4 - 87,6	79,5	64,7 - 90,2	55,7	82,3
	<=-4,6	60,0	26,4 - 87,6	75,0	59,7 - 86,8	50,7	81,4
	<=-3,09	70,0	34,8 - 93,0	75,0	59,7 - 86,8	54,5	85,4
	<=1,33	70,0	34,8 - 93,0	52,3	36,7 - 67,5	38,6	80,3
	<=1,37	80,0	44,4 - 96,9	52,3	36,7 - 67,5	41,8	85,9
	<=4,23	80,0	44,4 - 96,9	31,8	18,6 - 47,6	33,5	78,8
	<=5,03	100,0	69,0 - 100,0	31,8	18,6 - 47,6	38,6	100,0
	<=28,11	100,0	69,0 - 100,0	0,0	0,0 - 8,1	30,0	

Nota: IC= intervalo de confianza; PalxCol= palabras emitidas en la condición de interferencia; VP+= valor predictivo positivo; VP-= valor predictivo negativo.

Anexo 2.7.3. Sensibilidad, especificidad y valores predictivos positivos y negativos asociados a un porcentaje de base del **45%** para diversos puntos de corte de las variables del **Stroop**

Variable	Punto de corte	Sensibilidad	IC 95%	Especificidad	IC 95%	VP+	VP-
Palabras	< 21	0,0	0,0 - 31,0	100,0	91,9 - 100,0		55,0
	<=50	40,0	12,4 - 73,6	100,0	91,9 - 100,0	100,0	67,1
	<=53	40,0	12,4 - 73,6	95,5	84,5 - 99,3	87,8	66,0
	<=60	50,0	18,9 - 81,1	95,5	84,5 - 99,3	90,0	70,0
	<=63	50,0	18,9 - 81,1	93,2	81,3 - 98,5	85,7	69,5
	<=64	60,0	26,4 - 87,6	93,2	81,3 - 98,5	87,8	74,0
	<=65	60,0	26,4 - 87,6	90,9	78,3 - 97,4	84,4	73,5
	<=66	70,0	34,8 - 93,0	90,9	78,3 - 97,4	86,3	78,7
	<=85	70,0	34,8 - 93,0	65,9	50,1 - 79,5	62,7	72,9
	<=86	80,0	44,4 - 96,9	61,4	45,5 - 75,6	62,9	78,9
	<=90	80,0	44,4 - 96,9	52,3	36,7 - 67,5	57,8	76,2
	<=93	90,0	55,5 - 98,3	52,3	36,7 - 67,5	60,7	86,5
	<=101	90,0	55,5 - 98,3	36,4	22,4 - 52,2	53,6	81,6
	<=102	100,0	69,0 - 100,0	29,5	16,8 - 45,2	53,7	100,0
<=127	100,0	69,0 - 100,0	0,0	0,0 - 8,1	45,0		
Colores	< 12	0,0	0,0 - 31,0	100,0	91,9 - 100,0		55,0
	<=12	10,0	1,7 - 44,5	100,0	91,9 - 100,0	100,0	57,6
	<=31	10,0	1,7 - 44,5	95,5	84,5 - 99,3	64,3	56,5
	<=42	40,0	12,4 - 73,6	95,5	84,5 - 99,3	87,8	66,0
	<=43	50,0	18,9 - 81,1	93,2	81,3 - 98,5	85,7	69,5
	<=47	50,0	18,9 - 81,1	90,9	78,3 - 97,4	81,8	69,0
	<=49	60,0	26,4 - 87,6	88,6	75,4 - 96,2	81,2	73,0

Variable	Punto de corte	Sensibilidad	IC 95%	Especificidad	IC 95%	VP+	VP-
Colores	<=57	60,0	26,4 - 87,6	70,5	54,8 - 83,2	62,4	68,3
	<=59	70,0	34,8 - 93,0	70,5	54,8 - 83,2	66,0	74,2
	<=60	70,0	34,8 - 93,0	61,4	45,5 - 75,6	59,7	71,4
	<=61	80,0	44,4 - 96,9	61,4	45,5 - 75,6	62,9	78,9
	<=64	80,0	44,4 - 96,9	54,5	38,9 - 69,6	59,0	76,9
	<=65	90,0	55,5 - 98,3	50,0	34,6 - 65,4	59,6	85,9
	<=70	90,0	55,5 - 98,3	34,1	20,5 - 49,9	52,8	80,6
	<=71	100,0	69,0 - 100,0	27,3	15,0 - 42,8	52,9	100,0
	<=98	100,0	69,0 - 100,0	0,0	0,0 - 8,1	45,0	
	PalxCol	< 9	0,0	0,0 - 31,0	100,0	91,9 - 100,0	
<=18		50,0	18,9 - 81,1	100,0	91,9 - 100,0	100,0	71,0
<=23		50,0	18,9 - 81,1	88,6	75,4 - 96,2	78,3	68,4
<=25		60,0	26,4 - 87,6	88,6	75,4 - 96,2	81,2	73,0
<=26		70,0	34,8 - 93,0	86,4	72,6 - 94,8	80,8	77,9
<=29		70,0	34,8 - 93,0	79,5	64,7 - 90,2	73,7	76,4
<=30		80,0	44,4 - 96,9	72,7	57,2 - 85,0	70,6	81,6
<=31		80,0	44,4 - 96,9	68,2	52,4 - 81,4	67,3	80,6
<=32		90,0	55,5 - 98,3	65,9	50,1 - 79,5	68,4	89,0
<=42		90,0	55,5 - 98,3	29,5	16,8 - 45,2	51,1	78,3
<=43		100,0	69,0 - 100,0	27,3	15,0 - 42,8	52,9	100,0
<=72		100,0	69,0 - 100,0	0,0	0,0 - 8,1	45,0	
Interferencia		< - 19,39	0,0	0,0 - 31,0	100,0	91,9 - 100,0	
	<= - 14,07	0,0	0,0 - 31,0	93,2	81,3 - 98,5	0,0	53,2

Variable	Punto de corte	Sensibilidad	IC 95%	Especificidad	IC 95%	VP+	VP-
Interferencia	<=-13,25	10,0	1,7 - 44,5	93,2	81,3 - 98,5	54,5	55,9
	<=-12,18	10,0	1,7 - 44,5	88,6	75,4 - 96,2	41,9	54,6
	<=-7,71	40,0	12,4 - 73,6	88,6	75,4 - 96,2	74,2	64,4
	<=-6,6	40,0	12,4 - 73,6	81,8	67,3 - 91,8	64,3	62,5
	<=-5,95	50,0	18,9 - 81,1	81,8	67,3 - 91,8	69,2	66,7
	<=-5,93	50,0	18,9 - 81,1	79,5	64,7 - 90,2	66,7	66,0
	<=-5,15	60,0	26,4 - 87,6	79,5	64,7 - 90,2	70,6	70,9
	<=-4,6	60,0	26,4 - 87,6	75,0	59,7 - 86,8	66,3	69,6
	<=-3,09	70,0	34,8 - 93,0	75,0	59,7 - 86,8	69,6	75,3
	<=1,33	70,0	34,8 - 93,0	52,3	36,7 - 67,5	54,5	68,0
	<=1,37	80,0	44,4 - 96,9	52,3	36,7 - 67,5	57,8	76,2
	<=4,23	80,0	44,4 - 96,9	31,8	18,6 - 47,6	49,0	66,0
	<=5,03	100,0	69,0 - 100,0	31,8	18,6 - 47,6	54,5	100,0
	<=28,11	100,0	69,0 - 100,0	0,0	0,0 - 8,1	45,0	

Nota: IC= intervalo de confianza; PalxCol= palabras emitidas en la condición de interferencia; VP+= valor predictivo positivo; VP-= valor predictivo negativo.

Anexo 2.8.1. Sensibilidad, especificidad y valores predictivos positivos y negativos asociados a un porcentaje de base del **15%** para diversos puntos de corte de las variables del **TAVEC**

Variable	Punto de corte	Sensibilidad	IC 95%	Especificidad	IC 95%	VP+	VP-
Ensayo 1	< 3	0,0	0,0 - 31,0	100,0	91,9 - 100,0		85,0
	<=3	10,0	1,7 - 44,5	93,2	81,3 - 98,5	20,6	85,4
	<=4	30,0	7,0 - 65,2	88,6	75,4 - 96,2	31,8	87,8
	<=5	70,0	34,8 - 93,0	75,0	59,7 - 86,8	33,1	93,4
	<=6	90,0	55,5 - 98,3	52,3	36,7 - 67,5	25,0	96,7
	<=7	90,0	55,5 - 98,3	31,8	18,6 - 47,6	18,9	94,7
	<=8	100,0	69,0 - 100,0	22,7	11,5 - 37,8	18,6	100,0
	<=11	100,0	69,0 - 100,0	0,0	0,0 - 8,1	15,0	
Total ensayos	< 18	0,0	0,0 - 31,0	100,0	91,9 - 100,0		85,0
	<=18	10,0	1,7 - 44,5	100,0	91,9 - 100,0	100,0	86,3
	<=25	10,0	1,7 - 44,5	97,7	87,9 - 99,6	43,7	86,0
	<=26	20,0	3,1 - 55,6	97,7	87,9 - 99,6	60,8	87,4
	<=32	20,0	3,1 - 55,6	95,5	84,5 - 99,3	43,7	87,1
	<=34	30,0	7,0 - 65,2	95,5	84,5 - 99,3	53,8	88,5
	<=35	30,0	7,0 - 65,2	93,2	81,3 - 98,5	43,7	88,3
	<=36	40,0	12,4 - 73,6	90,9	78,3 - 97,4	43,7	89,6
	<=38	50,0	18,9 - 81,1	90,9	78,3 - 97,4	49,3	91,2
	<=39	60,0	26,4 - 87,6	88,6	75,4 - 96,2	48,2	92,6
	<=40	60,0	26,4 - 87,6	86,4	72,6 - 94,8	43,7	92,4
	<=41	70,0	34,8 - 93,0	86,4	72,6 - 94,8	47,5	94,2
<=49	70,0	34,8 - 93,0	72,7	57,2 - 85,0	31,2	93,2	

Variable	Punto de corte	Sensibilidad	IC 95%	Especificidad	IC 95%	VP+	VP-
Total ensayos	<=50	80,0	44,4 - 96,9	68,2	52,4 - 81,4	30,7	95,1
	<=51	80,0	44,4 - 96,9	63,6	47,8 - 77,6	28,0	94,7
	<=52	90,0	55,5 - 98,3	56,8	41,0 - 71,6	26,9	97,0
	<=53	90,0	55,5 - 98,3	50,0	34,6 - 65,4	24,1	96,6
	<=55	100,0	69,0 - 100,0	45,5	30,4 - 61,1	24,4	100,0
	<=74	100,0	69,0 - 100,0	0,0	0,0 - 8,1	15,0	
RL-CP	< 0	0,0	0,0 - 31,0	100,0	91,9 - 100,0		85,0
	<=0	10,0	1,7 - 44,5	100,0	91,9 - 100,0	100,0	86,3
	<=3	10,0	1,7 - 44,5	97,7	87,9 - 99,6	43,7	86,0
	<=4	20,0	3,1 - 55,6	95,5	84,5 - 99,3	43,7	87,1
	<=5	40,0	12,4 - 73,6	93,2	81,3 - 98,5	50,9	89,8
	<=6	50,0	18,9 - 81,1	93,2	81,3 - 98,5	56,4	91,3
	<=7	70,0	34,8 - 93,0	86,4	72,6 - 94,8	47,5	94,2
	<=8	70,0	34,8 - 93,0	77,3	62,2 - 88,5	35,2	93,6
	<=9	80,0	44,4 - 96,9	65,9	50,1 - 79,5	29,3	94,9
	<=10	90,0	55,5 - 98,3	52,3	36,7 - 67,5	25,0	96,7
	<=11	90,0	55,5 - 98,3	38,6	24,4 - 54,5	20,6	95,6
	<=12	100,0	69,0 - 100,0	25,0	13,2 - 40,3	19,0	100,0
	<=16	100,0	69,0 - 100,0	0,0	0,0 - 8,1	15,0	
RL-LP	< 0	0,0	0,0 - 31,0	100,0	91,9 - 100,0		85,0
	<=0	10,0	1,7 - 44,5	100,0	91,9 - 100,0	100,0	86,3
	<=2	10,0	1,7 - 44,5	97,7	87,9 - 99,6	43,7	86,0

Variable	Punto de corte	Sensibilidad	IC 95%	Especificidad	IC 95%	VP+	VP-
	<=4	40,0	12,4 - 73,6	97,7	87,9 - 99,6	75,6	90,2
	<=5	40,0	12,4 - 73,6	93,2	81,3 - 98,5	50,9	89,8
	<=6	60,0	26,4 - 87,6	88,6	75,4 - 96,2	48,2	92,6
	<=7	80,0	44,4 - 96,9	79,5	64,7 - 90,2	40,8	95,8
	<=8	80,0	44,4 - 96,9	72,7	57,2 - 85,0	34,1	95,4
	<=9	90,0	55,5 - 98,3	65,9	50,1 - 79,5	31,8	97,4
	<=11	90,0	55,5 - 98,3	45,5	30,4 - 61,1	22,6	96,3
	<=12	100,0	69,0 - 100,0	43,2	28,4 - 59,0	23,7	100,0
	<=16	100,0	69,0 - 100,0	0,0	0,0 - 8,1	15,0	
Reconocimiento	< 4	0,0	0,0 - 31,0	100,0	91,9 - 100,0		85,0
	<=8	30,0	7,0 - 65,2	100,0	91,9 - 100,0	100,0	89,0
	<=10	30,0	7,0 - 65,2	97,7	87,9 - 99,6	70,0	88,8
	<=11	50,0	18,9 - 81,1	97,7	87,9 - 99,6	79,5	91,7
	<=12	50,0	18,9 - 81,1	95,5	84,5 - 99,3	66,0	91,5
	<=13	60,0	26,4 - 87,6	84,1	69,9 - 93,3	40,0	92,3
	<=14	80,0	44,4 - 96,9	56,8	41,0 - 71,6	24,6	94,2
	<=15	90,0	55,5 - 98,3	38,6	24,4 - 54,5	20,6	95,6
	<=16	100,0	69,0 - 100,0	0,0	0,0 - 8,1	15,0	
Discriminabilidad	< 22,7	0,0	0,0 - 31,0	100,0	91,9 - 100,0		85,0
	<=22,7	0,0	0,0 - 31,0	97,7	87,9 - 99,6	0,0	84,7
	<=52	20,0	3,1 - 55,6	97,7	87,9 - 99,6	60,8	87,4
	<=66	20,0	3,1 - 55,6	95,5	84,5 - 99,3	43,7	87,1

Variable	Punto de corte	Sensibilidad	IC 95%	Especificidad	IC 95%	VP+	VP-
	<=72	50,0	18,9 - 81,1	95,5	84,5 - 99,3	66,0	91,5
	<=75	50,0	18,9 - 81,1	93,2	81,3 - 98,5	56,4	91,3
	<=77	60,0	26,4 - 87,6	93,2	81,3 - 98,5	60,8	93,0
	<=85	60,0	26,4 - 87,6	90,9	78,3 - 97,4	53,8	92,8
	<=86	70,0	34,8 - 93,0	88,6	75,4 - 96,2	52,1	94,4
	<=88	90,0	55,5 - 98,3	86,4	72,6 - 94,8	53,8	98,0
	<=90	90,0	55,5 - 98,3	79,5	64,7 - 90,2	43,7	97,8
	<=91	100,0	69,0 - 100,0	72,7	57,2 - 85,0	39,3	100,0
	<=100	100,0	69,0 - 100,0	0,0	0,0 - 8,1	15,0	

Nota: IC= intervalo de confianza; RL-CP= recuerdo libre a corto plazo; RL-LP= recuerdo libre a largo plazo; VP+= valor predictivo positivo; VP-= valor predictivo negativo.

Anexo 2.8.2. Sensibilidad, especificidad y valores predictivos positivos y negativos asociados a un porcentaje de base del **30%** para diversos puntos de corte de las variables del **TAVEC**

Variable	Punto de corte	Sensibilidad	IC 95%	Especificidad	IC 95%	VP+	-VP
Ensayo 1	< 3	0,0	0,0 - 31,0	100,0	91,9 - 100,0		70,0
	<=3	10,0	1,7 - 44,5	93,2	81,3 - 98,5	38,6	70,7
	<=4	30,0	7,0 - 65,2	88,6	75,4 - 96,2	53,1	74,7
	<=5	70,0	34,8 - 93,0	75,0	59,7 - 86,8	54,5	85,4
	<=6	90,0	55,5 - 98,3	52,3	36,7 - 67,5	44,7	92,4
	<=7	90,0	55,5 - 98,3	31,8	18,6 - 47,6	36,1	88,1
	<=8	100,0	69,0 - 100,0	22,7	11,5 - 37,8	35,7	100,0
	<=11	100,0	69,0 - 100,0	0,0	0,0 - 8,1	30,0	
Total ensayos	< 18	0,0	0,0 - 31,0	100,0	91,9 - 100,0		70,0
	<=18	10,0	1,7 - 44,5	100,0	91,9 - 100,0	100,0	72,2
	<=25	10,0	1,7 - 44,5	97,7	87,9 - 99,6	65,3	71,7
	<=26	20,0	3,1 - 55,6	97,7	87,9 - 99,6	79,0	74,0
	<=32	20,0	3,1 - 55,6	95,5	84,5 - 99,3	65,3	73,6
	<=34	30,0	7,0 - 65,2	95,5	84,5 - 99,3	73,9	76,1
	<=35	30,0	7,0 - 65,2	93,2	81,3 - 98,5	65,3	75,6
	<=36	40,0	12,4 - 73,6	90,9	78,3 - 97,4	65,3	78,0
	<=38	50,0	18,9 - 81,1	90,9	78,3 - 97,4	70,2	80,9
	<=39	60,0	26,4 - 87,6	88,6	75,4 - 96,2	69,4	83,8
	<=40	60,0	26,4 - 87,6	86,4	72,6 - 94,8	65,3	83,4
	<=41	70,0	34,8 - 93,0	86,4	72,6 - 94,8	68,8	87,0
<=49	70,0	34,8 - 93,0	72,7	57,2 - 85,0	52,4	85,0	

Variable	Punto de corte	Sensibilidad	IC 95%	Especificidad	IC 95%	VP+	-VP
Total ensayos	<=50	80,0	44,4 - 96,9	68,2	52,4 - 81,4	51,9	88,8
	<=51	80,0	44,4 - 96,9	63,6	47,8 - 77,6	48,5	88,1
	<=52	90,0	55,5 - 98,3	56,8	41,0 - 71,6	47,2	93,0
	<=53	90,0	55,5 - 98,3	50,0	34,6 - 65,4	43,5	92,1
	<=55	100,0	69,0 - 100,0	45,5	30,4 - 61,1	44,0	100,0
	<=74	100,0	69,0 - 100,0	0,0	0,0 - 8,1	30,0	
RL-CP	< 0	0,0	0,0 - 31,0	100,0	91,9 - 100,0		70,0
	<=0	10,0	1,7 - 44,5	100,0	91,9 - 100,0	100,0	72,2
	<=3	10,0	1,7 - 44,5	97,7	87,9 - 99,6	65,3	71,7
	<=4	20,0	3,1 - 55,6	95,5	84,5 - 99,3	65,3	73,6
	<=5	40,0	12,4 - 73,6	93,2	81,3 - 98,5	71,5	78,4
	<=6	50,0	18,9 - 81,1	93,2	81,3 - 98,5	75,9	81,3
	<=7	70,0	34,8 - 93,0	86,4	72,6 - 94,8	68,8	87,0
	<=8	70,0	34,8 - 93,0	77,3	62,2 - 88,5	56,9	85,7
	<=9	80,0	44,4 - 96,9	65,9	50,1 - 79,5	50,1	88,5
	<=10	90,0	55,5 - 98,3	52,3	36,7 - 67,5	44,7	92,4
	<=11	90,0	55,5 - 98,3	38,6	24,4 - 54,5	38,6	90,0
	<=12	100,0	69,0 - 100,0	25,0	13,2 - 40,3	36,4	100,0
	<=16	100,0	69,0 - 100,0	0,0	0,0 - 8,1	30,0	
	RL-LP	< 0	0,0	0,0 - 31,0	100,0	91,9 - 100,0	
<=0		10,0	1,7 - 44,5	100,0	91,9 - 100,0	100,0	72,2
<=2		10,0	1,7 - 44,5	97,7	87,9 - 99,6	65,3	71,7

Variable	Punto de corte	Sensibilidad	IC 95%	Especificidad	IC 95%	VP+	-VP
RL-LP	<=4	40,0	12,4 - 73,6	97,7	87,9 - 99,6	88,3	79,2
	<=5	40,0	12,4 - 73,6	93,2	81,3 - 98,5	71,5	78,4
	<=6	60,0	26,4 - 87,6	88,6	75,4 - 96,2	69,4	83,8
	<=7	80,0	44,4 - 96,9	79,5	64,7 - 90,2	62,6	90,3
	<=8	80,0	44,4 - 96,9	72,7	57,2 - 85,0	55,7	89,5
	<=9	90,0	55,5 - 98,3	65,9	50,1 - 79,5	53,1	93,9
	<=11	90,0	55,5 - 98,3	45,5	30,4 - 61,1	41,4	91,4
	<=12	100,0	69,0 - 100,0	43,2	28,4 - 59,0	43,0	100,0
	<=16	100,0	69,0 - 100,0	0,0	0,0 - 8,1	30,0	
Reconocimiento	< 4	0,0	0,0 - 31,0	100,0	91,9 - 100,0		70,0
	<=8	30,0	7,0 - 65,2	100,0	91,9 - 100,0	100,0	76,9
	<=10	30,0	7,0 - 65,2	97,7	87,9 - 99,6	85,0	76,5
	<=11	50,0	18,9 - 81,1	97,7	87,9 - 99,6	90,4	82,0
	<=12	50,0	18,9 - 81,1	95,5	84,5 - 99,3	82,5	81,7
	<=13	60,0	26,4 - 87,6	84,1	69,9 - 93,3	61,8	83,1
	<=14	80,0	44,4 - 96,9	56,8	41,0 - 71,6	44,3	86,9
	<=15	90,0	55,5 - 98,3	38,6	24,4 - 54,5	38,6	90,0
	<=16	100,0	69,0 - 100,0	0,0	0,0 - 8,1	30,0	
Discriminabilidad	< 22,7	0,0	0,0 - 31,0	100,0	91,9 - 100,0		70,0
	<=22,7	0,0	0,0 - 31,0	97,7	87,9 - 99,6	0,0	69,5
	<=52	20,0	3,1 - 55,6	97,7	87,9 - 99,6	79,0	74,0
	<=66	20,0	3,1 - 55,6	95,5	84,5 - 99,3	65,3	73,6

Variable	Punto de corte	Sensibilidad	IC 95%	Especificidad	IC 95%	VP+	-VP
Discriminabilidad	<=72	50,0	18,9 - 81,1	95,5	84,5 - 99,3	82,5	81,7
	<=75	50,0	18,9 - 81,1	93,2	81,3 - 98,5	75,9	81,3
	<=77	60,0	26,4 - 87,6	93,2	81,3 - 98,5	79,0	84,5
	<=85	60,0	26,4 - 87,6	90,9	78,3 - 97,4	73,9	84,1
	<=86	70,0	34,8 - 93,0	88,6	75,4 - 96,2	72,5	87,3
	<=88	90,0	55,5 - 98,3	86,4	72,6 - 94,8	73,9	95,3
	<=90	90,0	55,5 - 98,3	79,5	64,7 - 90,2	65,3	94,9
	<=91	100,0	69,0 - 100,0	72,7	57,2 - 85,0	61,1	100,0
	<=100	100,0	69,0 - 100,0	0,0	0,0 - 8,1	30,0	

Nota: IC= intervalo de confianza; RL-CP= recuerdo libre a corto plazo; RL-LP= recuerdo libre a largo plazo; VP+= valor predictivo positivo; VP-= valor predictivo negativo.

Anexo 2.8.3. Sensibilidad, especificidad y valores predictivos positivos y negativos asociados a un porcentaje de base del **45%** para diversos puntos de corte de las variables del **TAVEC**

Variable	Punto de corte	Sensibilidad	IC 95%	Especificidad	IC 95%	VP+	VP-
Ensayo 1	< 3	0,0	0,0 - 31,0	100,0	91,9 - 100,0		55,0
	<=3	10,0	1,7 - 44,5	93,2	81,3 - 98,5	54,5	55,9
	<=4	30,0	7,0 - 65,2	88,6	75,4 - 96,2	68,4	60,7
	<=5	70,0	34,8 - 93,0	75,0	59,7 - 86,8	69,6	75,3
	<=6	90,0	55,5 - 98,3	52,3	36,7 - 67,5	60,7	86,5
	<=7	90,0	55,5 - 98,3	31,8	18,6 - 47,6	51,9	79,5
	<=8	100,0	69,0 - 100,0	22,7	11,5 - 37,8	51,4	100,0
	<=11	100,0	69,0 - 100,0	0,0	0,0 - 8,1	45,0	
Total ensayos	< 18	0,0	0,0 - 31,0	100,0	91,9 - 100,0		55,0
	<=18	10,0	1,7 - 44,5	100,0	91,9 - 100,0	100,0	57,6
	<=25	10,0	1,7 - 44,5	97,7	87,9 - 99,6	78,3	57,0
	<=26	20,0	3,1 - 55,6	97,7	87,9 - 99,6	87,8	59,9
	<=32	20,0	3,1 - 55,6	95,5	84,5 - 99,3	78,3	59,3
	<=34	30,0	7,0 - 65,2	95,5	84,5 - 99,3	84,4	62,5
	<=35	30,0	7,0 - 65,2	93,2	81,3 - 98,5	78,3	61,9
	<=36	40,0	12,4 - 73,6	90,9	78,3 - 97,4	78,3	64,9
	<=38	50,0	18,9 - 81,1	90,9	78,3 - 97,4	81,8	69,0
	<=39	60,0	26,4 - 87,6	88,6	75,4 - 96,2	81,2	73,0
	<=40	60,0	26,4 - 87,6	86,4	72,6 - 94,8	78,3	72,5
	<=41	70,0	34,8 - 93,0	86,4	72,6 - 94,8	80,8	77,9
	<=49	70,0	34,8 - 93,0	72,7	57,2 - 85,0	67,7	74,8

Variable	Punto de corte	Sensibilidad	IC 95%	Especificidad	IC 95%	VP+	VP-
Total ensayos	<=50	80,0	44,4 - 96,9	68,2	52,4 - 81,4	67,3	80,6
	<=51	80,0	44,4 - 96,9	63,6	47,8 - 77,6	64,3	79,5
	<=52	90,0	55,5 - 98,3	56,8	41,0 - 71,6	63,0	87,4
	<=53	90,0	55,5 - 98,3	50,0	34,6 - 65,4	59,6	85,9
	<=55	100,0	69,0 - 100,0	45,5	30,4 - 61,1	60,0	100,0
	<=74	100,0	69,0 - 100,0	0,0	0,0 - 8,1	45,0	
RL-CP	< 0	0,0	0,0 - 31,0	100,0	91,9 - 100,0		55,0
	<=0	10,0	1,7 - 44,5	100,0	91,9 - 100,0	100,0	57,6
	<=3	10,0	1,7 - 44,5	97,7	87,9 - 99,6	78,3	57,0
	<=4	20,0	3,1 - 55,6	95,5	84,5 - 99,3	78,3	59,3
	<=5	40,0	12,4 - 73,6	93,2	81,3 - 98,5	82,8	65,5
	<=6	50,0	18,9 - 81,1	93,2	81,3 - 98,5	85,7	69,5
	<=7	70,0	34,8 - 93,0	86,4	72,6 - 94,8	80,8	77,9
	<=8	70,0	34,8 - 93,0	77,3	62,2 - 88,5	71,6	75,9
	<=9	80,0	44,4 - 96,9	65,9	50,1 - 79,5	65,8	80,1
	<=10	90,0	55,5 - 98,3	52,3	36,7 - 67,5	60,7	86,5
	<=11	90,0	55,5 - 98,3	38,6	24,4 - 54,5	54,5	82,5
	<=12	100,0	69,0 - 100,0	25,0	13,2 - 40,3	52,2	100,0
	<=16	100,0	69,0 - 100,0	0,0	0,0 - 8,1	45,0	
RL-LP	< 0	0,0	0,0 - 31,0	100,0	91,9 - 100,0		55,0
	<=0	10,0	1,7 - 44,5	100,0	91,9 - 100,0	100,0	57,6
	<=2	10,0	1,7 - 44,5	97,7	87,9 - 99,6	78,3	57,0

Variable	Punto de corte	Sensibilidad	IC 95%	Especificidad	IC 95%	VP+	VP-
RL-LP	<=4	40,0	12,4 - 73,6	97,7	87,9 - 99,6	93,5	66,6
	<=5	40,0	12,4 - 73,6	93,2	81,3 - 98,5	82,8	65,5
	<=6	60,0	26,4 - 87,6	88,6	75,4 - 96,2	81,2	73,0
	<=7	80,0	44,4 - 96,9	79,5	64,7 - 90,2	76,2	82,9
	<=8	80,0	44,4 - 96,9	72,7	57,2 - 85,0	70,6	81,6
	<=9	90,0	55,5 - 98,3	65,9	50,1 - 79,5	68,4	89,0
	<=11	90,0	55,5 - 98,3	45,5	30,4 - 61,1	57,4	84,7
	<=12	100,0	69,0 - 100,0	43,2	28,4 - 59,0	59,0	100,0
	<=16	100,0	69,0 - 100,0	0,0	0,0 - 8,1	45,0	
	Reconocimiento	< 4	0,0	0,0 - 31,0	100,0	91,9 - 100,0	
<=8		30,0	7,0 - 65,2	100,0	91,9 - 100,0	100,0	63,6
<=10		30,0	7,0 - 65,2	97,7	87,9 - 99,6	91,5	63,0
<=11		50,0	18,9 - 81,1	97,7	87,9 - 99,6	94,7	70,5
<=12		50,0	18,9 - 81,1	95,5	84,5 - 99,3	90,0	70,0
<=13		60,0	26,4 - 87,6	84,1	69,9 - 93,3	75,5	72,0
<=14		80,0	44,4 - 96,9	56,8	41,0 - 71,6	60,3	77,6
<=15		90,0	55,5 - 98,3	38,6	24,4 - 54,5	54,5	82,5
<=16		100,0	69,0 - 100,0	0,0	0,0 - 8,1	45,0	
Discriminabilidad		< 22,7	0,0	0,0 - 31,0	100,0	91,9 - 100,0	
	<=22,7	0,0	0,0 - 31,0	97,7	87,9 - 99,6	0,0	54,4
	<=52	20,0	3,1 - 55,6	97,7	87,9 - 99,6	87,8	59,9
	<=66	20,0	3,1 - 55,6	95,5	84,5 - 99,3	78,3	59,3

Variable	Punto de corte	Sensibilidad	IC 95%	Especificidad	IC 95%	VP+	VP-
Discriminabilidad	<=72	50,0	18,9 - 81,1	95,5	84,5 - 99,3	90,0	70,0
	<=75	50,0	18,9 - 81,1	93,2	81,3 - 98,5	85,7	69,5
	<=77	60,0	26,4 - 87,6	93,2	81,3 - 98,5	87,8	74,0
	<=85	60,0	26,4 - 87,6	90,9	78,3 - 97,4	84,4	73,5
	<=86	70,0	34,8 - 93,0	88,6	75,4 - 96,2	83,4	78,3
	<=88	90,0	55,5 - 98,3	86,4	72,6 - 94,8	84,4	91,3
	<=90	90,0	55,5 - 98,3	79,5	64,7 - 90,2	78,3	90,7
	<=91	100,0	69,0 - 100,0	72,7	57,2 - 85,0	75,0	100,0
	<=100	100,0	69,0 - 100,0	0,0	0,0 - 8,1	45,0	

Nota: IC= intervalo de confianza; RL-CP= recuerdo libre a corto plazo; RL-LP= recuerdo libre a largo plazo; RP+= ratio de probabilidad positiva; RP-= ratio de probabilidad negativa; VP+= valor predictivo positivo; VP-= valor predictivo negativo.