



Utilidad de los índices de memoria verbal para detectar simulación en población española

Raquel Vilar López^{a*}, Mónica Aparicio^a, Manuel Gómez Río^b y Miguel Pérez García^a

^aUniversidad de Granada, España

^bHospital Universitario Virgen de las Nieves, Granada, España

INFORMACIÓN ARTÍCULO

Manuscrito recibido: 30/09/2013

Revisión recibida: 01/10/2013

Aceptado: 10/10/2013

Palabras clave:

Simulación

Esfuerzo/motivación insuficiente

TAVEC

Evaluación neuropsicológica

RESUMEN

Introducción. A pesar de que la literatura internacional ha demostrado la utilidad de diversos índices de las pruebas neuropsicológicas tradicionales de memoria verbal para ayudar en la detección de la simulación de problemas cognitivos, ningún estudio similar ha sido realizado todavía en nuestro contexto.

Metodología. Este estudio ha incluido a 84 participantes divididos en 4 grupos: tres grupos de pacientes con traumatismo craneoencefálico leve (30 participantes que no buscan compensación económica, 14 participantes que buscan compensación económica pero no son sospechosos de simulación y un grupo de 10 participantes que busca una compensación económica y es sospechoso de simulación atendiendo a diversas pruebas específicas de simulación), además de un grupo de 30 análogos. Todos los participantes realizaron una extensa batería neuropsicológica en la que estaba incluido el Test de Aprendizaje Verbal España Complutense (TAVEC) además de distintas pruebas específicas de simulación (Victoria Symptom Validity Test, Test of Memory Malingering, Dot Counting Test, el Test de la b y el Test de los 15 ítems de Rey).

Resultados. Se han encontrado diferencias estadísticamente significativas entre los grupos para todos los índices del TAVEC propuestos como indicativos de simulación y tamaño del efecto elevado. Con una especificidad superior al 90%, algunos índices del TAVEC han mostrado una sensibilidad adecuada para detectar simulación: 50% el total de palabras recordadas en los cinco ensayos, el recuerdo a largo plazo y el reconocimiento, 40% el recuerdo libre a corto plazo y 60% la discriminabilidad.

Conclusiones. Algunos índices del TAVEC han mostrado su utilidad para ser empleados como un apoyo más en el complejo proceso de toma de decisiones que supone el diagnóstico de simulación, si bien nunca deberían interpretarse de modo aislado.

© 2013 Colegio Oficial de Psicólogos de Madrid. Todos los derechos reservados.

Usefulness of verbal memory indexes to detect malingering in Spanish populations

ABSTRACT

Keywords:

Malingering

Insufficient effort/motivation

TAVEC

Neuropsychological assessment

Introduction. Although international literature has demonstrated the usefulness of different traditional neuropsychological verbal memory indexes to detect malingered cognitive deficits, no study has been conducted in Spain with such a purpose yet.

Methodology. The present study included 84 patients divided into four groups: three groups of mild traumatic brain injury patients (30 patients who do not request an economic compensation, 14 participants who requested an economic compensation but are not suspected of malingering, and 10 participants suspected of malingering according to different malingering tests), and a group of 30 analogues. All participants completed an extensive neuropsychological battery including the Test de Aprendizaje Verbal España Complutense (Spanish version of the California Verbal Learning Test) and several malingering tests (Victoria Symptom Validity Test, Test of Memory Malingering, Dot Counting Test, the b Test and the Rey 15-item Test).

Results. Significant differences were found between the groups of the study for all indexes, with large effect sizes. With a specificity of at least 90%, some of the indexes from the TAVEC showed adequate sensitivity to

*La correspondencia sobre este artículo debe enviarse a Raquel Vilar López.

Departamento de Personalidad, Evaluación y Tratamiento Psicológico. Facultad de Psicología. Campus de la Cartuja s/n. 18071 Granada. Email: rvilar@ugr.es

detect malingering: total number of words recalled in the five trials, short-term memory, and recognition, which obtained a sensitivity of 50%, long-term memory (40%), and discriminability index (60%).

Conclusions. Some of the TAVEC indexes demonstrated efficacy in detecting malingering. However, they should never be employed or interpreted in isolation.

© 2013 Colegio Oficial de Psicólogos de Madrid. All rights reserved.

La validez de la ejecución de las pruebas es una parte esencial de las evaluaciones neuropsicológicas, no solo de las forenses, sino también, atendiendo a los estándares internacionales actuales, de las evaluaciones neuropsicológicas clínicas rutinarias (ver las recomendaciones de la National Academy of Neuropsychology, Bush, NAN Policy y Planning Committee, 2005 y del American Board of Clinical Neuropsychology, 2007).

En la actualidad disponemos de una amplia variedad de métodos y herramientas para detectar simulación que, en líneas generales, podemos dividir en dos bloques: las pruebas específicamente diseñadas para la detección de simulación y el empleo de pruebas neuropsicológicas tradicionales, de las que se derivan o bien puntos de corte para determinadas variables o bien puntuaciones compuestas basadas en funciones discriminantes o análisis de regresión logística para detectar a los posibles simuladores (para una revisión ver Vilar-López, 2009). Entre las ventajas de emplear las pruebas tradicionales con el doble propósito de evaluar la función neuropsicológica y la posible simulación podemos citar que no requieren un tiempo extra en la evaluación y que permiten examinar la existencia de simulación en situaciones en las que se requiere una revisión retrospectiva de un caso (Mittenberg, Patton, Canary y Condit, 2002), que proporciona información sobre la validez de pruebas concretas (Mathias, Greve, Bianchini, Houston y Crouch, 2002) y que son menos susceptibles de entrenamiento que las pruebas específicas (Ashendorf, O'Briant y McCaffrey, 2003).

El California Verbal Learning Test (CVLT, Delis, Kramer, Kaplan y Ober, 1987) es uno de los tests neuropsicológicos tradicionales que más investigación ha suscitado en cuanto a la evaluación del esfuerzo insuficiente y la cuarta prueba neuropsicológica más empleada por los clínicos para evaluar simulación (Sharland y Gfeller, 2007). Entre los trabajos que se han centrado en el estudio de los índices aislados como posibles indicadores de simulación en esta prueba es pionero el trabajo de Trueblood (1994), que empleó un diseño de grupos conocidos de pacientes con daño cerebral traumático breve para estudiar la utilidad del número total de palabras recordadas en los cinco ensayos (total 1-5) y del total de palabras acertadas en el reconocimiento para la detección de simulación, alcanzando una especificidad del 90% y una sensibilidad alrededor del 70%. Sin embargo, el trabajo fundamental en este área lo realizaron Millis, Putnam, Adams y Ricker (1995) con un diseño muy similar y añadiendo dos nuevas variables: el recuerdo a largo plazo con claves y la discriminabilidad. El 93% de los casos fue correctamente clasificado con la variable discriminabilidad y el 89% con los aciertos del reconocimiento. El recuerdo a largo plazo con claves y el número total de palabras recordadas obtuvieron resultados de clasificación inferiores pero también satisfactorios (87% y 83% respectivamente). Los índices propuestos en este trabajo han sido estudiados en numerosas ocasiones por diferentes autores, empleando tanto análogos (Coleman, Rappaport, Millis, Ricker y Farchione, 1998) como pacientes con traumatismos moderado-severos (Slick, Iverson y Green, 2000; Sweet et al., 2000) o grupos de envejecimiento normal (Ashendorf et al., 2003). En general, estos estudios encuentran una sensibilidad inferior a la referida por Millis y destacan la posibilidad de cometer falsos positivos en la clasificación de los individuos como simuladores empleando los índices expuestos.

En el trabajo seminal de Millis et al. (1995) también se incluyó una función lineal discriminante con la que se alcanzó un nivel de clasi-

ficación correcta del 91% y una función cuadrática que consiguió una clasificación correcta del 96% de los casos. 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, obteniendo la discriminabilidad una exactitud de clasificación similar a la de las funciones multivariantes más complejas. En trabajos posteriores estos autores propusieron diversos modelos de regresión logística (Millis y Putnam, 1997; Millis y Volinski, 2001). Estas regresiones han sido también replicadas en diversos estudios (Ashendorf et al., 2003; Demakis, 2004; Martens, Donders y Millis, 2001; Sweet et al., 2000). Entre ellos, destaca el trabajo de Curtis, Greve, Bianchini y Brennan (2006), en el que se revisan tanto las cuatro puntuaciones simples propuestas para detectar simulación (total de palabras recordadas en los cinco ensayos, recuerdo a largo plazo con claves, reconocimiento y discriminabilidad) como ocho puntuaciones compuestas, incluyendo todas las propuestas por Millis. De este trabajo, el más completo hasta la fecha sobre la utilidad del CVLT para detectar simuladores, se puede concluir que esta prueba posee una adecuada especificidad (90-95%) pero una sensibilidad moderada (40-60%).

La segunda edición del CVLT (Delis, Kramer, Kaplan y Ober, 2000) ha incluido una prueba de reconocimiento de elección forzosa opcional diseñada específicamente para evaluar el nivel de esfuerzo de los sujetos. Este índice se ha mostrado sensible a los efectos de la búsqueda de compensación económica (Moore y Donders, 2004) y ha demostrado su validez en diseños de grupos conocidos (Bauer, Yantz, Ryan, Warden y McCaffrey, 2005; Root, Robbins, Chang y Van Gorp, 2006), pudiendo suponer una mejora sobre su edición predecesora. Sin embargo, entendemos que esta medida es una prueba específica y no un índice extraído de una prueba neuropsicológica tradicional, por lo que no nos centraremos en el análisis de sus resultados. Por otro lado, se ha estudiado la utilidad de los índices empleados en el CVLT original para detectar simulación con la segunda edición de la prueba, encontrando que los índices individuales y las regresiones son igual de efectivos que en la versión original, pero que es necesario modificar los puntos de corte para no cometer falsos positivos (Greve, Curtis, Bianchini y Ord, 2009). Las puntuaciones de discriminabilidad y recuerdo a largo plazo parecen especialmente sensibles a la simulación en el CVLT-II (Wolfe et al., 2010).

Otra prueba similar al CVLT que ha recibido investigación en relación al tema de la simulación es el Rey Auditory Verbal Learning Test (RAVLT), en la que se ha encontrado que el número de palabras recordadas en el ensayo 1 (Bernard, 1990; Bernard, Houston y Natoli, 1993), el recuerdo a largo plazo (Bernard, 1990; Bernard et al., 1993; King, Gfeller y Davis, 1998), el reconocimiento (Meyers, Morrison y Miller, 2001) y el recuerdo y reconocimiento a los 60 minutos (Barrash, Suhr y Manzel, 2004) son útiles a la hora de detectar simuladores con esta prueba. Por otro lado, los patrones de ejecución de esta prueba han mostrado ser diferentes entre simuladores y no simuladores, especialmente en las variables relativas al recuerdo y al reconocimiento (Bernard, 1990; Greiffenstein, Baker y Gola, 1994; Powell, Gfeller, Hendricks y Sharland, 2004; Suhr, 2002; Suhr y Gunstad, 2000; Sullivan, Deffenti y Keane, 2002).

En definitiva, numerosas investigaciones en la literatura internacional han demostrado que distintos índices en las pruebas de memoria verbal de listas de palabras son útiles a la hora de detectar simulación. Sin embargo, esto nunca ha sido estudiado en población

española. Por ello, la finalidad de este estudio es comprobar la efectividad de un test de memoria verbal para detectar simulación en población española. Para ello emplearemos el Test de Aprendizaje Verbal España Complutense (TAVEC), versión española del CVLT, por ser el test de memoria verbal más frecuentemente empleado en nuestro contexto, y nos centraremos en los índices simples extraídos de la prueba, considerando que la literatura previa no ha encontrado diferencias importantes en los niveles de clasificación de estos índices en comparación con las funciones discriminantes y que estas últimas son más complicadas de aplicar en el ámbito clínico. Para ello empleamos un diseño mixto, planteando la hipótesis de que los pacientes no sospechosos de simulación, busquen o no una compensación económica, puntuarán de modo similar en los índices estudiados del TAVEC, mientras que los pacientes sospechosos de simulación se comportarán de modo similar a los análogos, siendo la ejecución de los dos primeros grupos diferente de la de los dos últimos.

Método

Participantes

En este estudio participaron 65 pacientes diagnosticados con traumatismo craneoencefálico leve (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). El diagnóstico de traumatismo craneoencefálico leve (TCE-L) se obtuvo de las historias clínicas de los pacientes. El SPC se evaluó con The Rivermead Post Concussion Symptoms Questionnaire (RPQ), empleando el criterio de al menos tres ítems con una puntuación de severidad igual o mayor de 3 (King, Crawford, Wenden, Moss y Wade, 1995). Ninguno de los pacientes obtuvo hallazgos en la tomografía axial computerizada (TAC) ni presentó alteraciones psiquiátricas o consumo de sustancias premórbidas. Debido a estos motivos fueron excluidos del estudio 11 participantes, por lo que el tamaño final de la muestra fue de 54 pacientes. Para asegurar que ningún paciente con psicopatología fuese incluido en el estudio se realizó una entrevista clínica de cribado, tanto con los pacientes como con algún familiar o persona cercana, y las historias clínicas se revisaron en profundidad. De los 11 pacientes excluidos, dos lo fueron por diagnósticos psicopatológicos previos al evento traumático, dos por ser consumidores de cocaína y siete por tener una historia de abuso de alcohol. Todos los pacientes fueron evaluados al menos 6 meses después de sufrir el TCE-L y su edad estaba comprendida entre los 18 y los 55 años, con la finalidad de incluir solo a pacientes mayores de edad en los que no se espera la existencia de alteraciones neuropsicológicas o neurofuncionales debidas al deterioro asociado al envejecimiento normal.

Los participantes fueron divididos en tres grupos. El primero estaba formado por pacientes que no estaban manteniendo ningún tipo de proceso del que se pudiese obtener algún tipo de beneficio económico debido a las consecuencias de su traumatismo (NC). Este grupo estaba compuesto por 16 hombres y 14 mujeres ($N = 30$), con una media de edad de 32.5 años ($DT = 13.67$) y una media de 9.3 años de escolaridad ($SD = 3.45$). El tiempo medio transcurrido desde el traumatismo hasta el momento de la evaluación fue de 263.13 días ($DT = 378.66$). El segundo grupo de pacientes estaba formado por participantes en búsqueda de compensación económica no sospechosos de simulación (NS), integrado por 8 hombres y 6 mujeres ($N = 14$), con una edad media de 35.92 años ($DT = 10.88$), una educación media de 9.14 años ($DT = 3.5$) y una media de tiempo transcurrido hasta la evaluación de 287.78 días ($DT = 448.2$). El tercer grupo estaba compuesto por pacientes en búsqueda de compensación económica sospechosos de simulación (SS). Para ello debían cumplir los criterios de simulación probable propuestos por Slick, Sherman e Iverson (1999) y obtener puntuaciones indicativas de simulación en al menos dos índices específicos (Victoria Symp-

tom Validity Test, Test of Memory Malinger, el Test de la b, Dot Counting Test o el Test de los 15 Ítems de Rey). Este grupo estaba compuesto por 8 hombres y dos mujeres ($N = 10$), con una media de edad de 35.2 años ($DT = 11.1$), una educación media de 8.5 años ($DT = 1.35$) y una media de tiempo transcurrido desde el traumatismo de 490 días ($DT = 583.88$).

Además de estos grupos de pacientes también se incluyó en el estudio a 30 estudiantes de cuarto de psicología con conocimientos de neuropsicología para constituir el grupo de análogos (AN). La media de edad de este grupo era de 20.92 años ($DT = 3.08$), con un nivel de educación medio de 13.92 años ($DT = 2.35$). Ninguno de estos participantes tenía una historia previa de daño cerebral.

Instrumentos

Todos los pacientes realizaron una batería neuropsicológica extensa en la que también estaban incluidas las pruebas mencionadas en el presente estudio:

- *Test Auditivo-Verbal España-Complutense* (TAVEC, Benedet y Alejandre 1998). Las variables seleccionadas para este estudio fueron: el número total de palabras recordadas en los cinco ensayos (R), el número total de palabras recordadas en el recuerdo a corto plazo (RL-CP), el número total de palabras recordadas en el recuerdo a largo plazo (RL-LP), el número de aciertos en el reconocimiento (REC) y la discriminabilidad (DIS).

Para clasificar a los pacientes buscadores de compensación como simuladores o no simuladores se emplearon las siguientes pruebas:

- *Victoria Symptom Validity Test* (VSVT, Slick, Hopp, Strauss y Thompson, 1997)
- *Test of Memory Malinger* (TOMM, Tombaugh, 1996)
- *The b Test* (Boone, Lu y Herzberg, 2002)
- *Dot Counting Test* (Boone, Lu, Back et al., 2002)
- *Test de los 15 Ítems de Rey* (Rey, 1964)

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 del Hospital Universitario Virgen de las Nieves de Granada a lo largo de 16 meses consecutivos. En todos estos pacientes se había realizado una exploración de TAC sin que ninguno presentara alteraciones relevantes (TAC negativo) y la evaluación se realizó al menos 6 meses después del traumatismo. Se contactó telefónicamente con los pacientes al menos dos veces (a los 3 y a los 6 meses posteriores al traumatismo) para realizar una entrevista telefónica en la que se exploraba la existencia del SPC. Si dicha exploración era positiva el paciente era citado para una evaluación neuropsicológica individual exhaustiva. De los 385 pacientes que componían la lista inicial, 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. 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. Todas las evaluaciones fueron realizadas por una licenciada en Psicología con experiencia en evaluación neuropsicológica.

El protocolo del estudio era explicado claramente a cada paciente y tras la obtención de su consentimiento informado se procedía a su evaluación. En primer lugar se realizaba una entrevista y a continuación se administraba la batería neuropsicológica, siendo la duración total de la exploración de dos horas y media aproximadamente, por lo que se realizaba un descanso de unos 20 minutos a mitad de la sesión. En ocasiones puntuales, cuando las características particula-

res del caso así lo requerían, se realizaban más períodos de descanso o se empleaban dos sesiones para realizar la evaluación.

El grupo de análogos fue reclutado de la asignatura de Neuropsicología Aplicada, para asegurar su conocimiento sobre los síntomas y evaluación del daño cerebral. Siguiendo las recomendaciones propuestas por Rogers (1997), a los participantes se les proporcionaba unas instrucciones específicas en las que se ofrecía un escenario en el que podían identificarse (víctimas de un accidente de tráfico). Las explicaciones verbales dadas por el experimentador presentaban la evaluación como un desafío a los participantes, para ver si eran capaces de simular un daño en una situación real. Se tuvieron en cuenta los incentivos tanto positivos (puntos extra en una asignatura) como negativos (vergüenza social por la exposición pública de los nombres de los peores simuladores) y se especificaba que la simulación debía ser lo suficientemente creíble como para evitar la detección. Además, en las instrucciones se explicaban algunos de los síntomas más comunes del trastorno que los sujetos debían simular, así como información sobre los métodos concretos de las pruebas de simulación. 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. Tras la evaluación los participantes cumplimentaban un informe con la finalidad de comprobar el recuerdo de las instrucciones y su comprensión, así como la implicación de los participantes en la simulación. La duración aproximada de todo el proceso (instrucciones, evaluación y cuestionario posterior) era de una hora y cuarto.

Resultados

1. Diferencias entre los grupos

Para comprobar si las variables del TAVEC eran capaces de diferenciar los distintos grupos del estudio se realizaron 6 análisis no paramétricos para 4 grupos independientes (Kruskal-Wallis), siendo la variable independiente (VI) el grupo diferencial (grupo de no-búsqueda de compensación [NC] vs. grupo de búsqueda de compensación no sospechoso de simulación [NS] vs. grupo de búsqueda de compensación sospechoso de simulación [SS] vs. grupo de análogos [AN]) y las variables dependientes (VD) el número de palabras recordadas a lo largo de los cinco ensayos (R), el recuerdo libre a corto plazo (RL-CP), el recuerdo libre a largo plazo (RL-LP), el reconocimiento (REC) y la discriminabilidad (DIS). En los casos en los que se encontraron diferencias estadísticamente significativas en el análisis anterior se procedió a realizar comparaciones dos a dos

(grupos NC vs. NS, grupos NC vs. SS, grupos NC vs. AN, grupos NS vs. SS, grupos NS vs. AN y grupos SS vs. AN) mediante el análisis de Mann-Whitney.

Los resultados mostraron diferencias estadísticamente significativas entre los grupos estudiados para todas las variables: R, $\chi^2(2) = 16.87, p < .001$; RL-CP, $\chi^2(2) = 19.75, p < .000$; RL-LP, $\chi^2(2) = 24.42, p < .000$; REC, $\chi^2(2) = 26.82, p < .000$; DIS, $\chi^2(2) = 32.33, p < .000$. Las comparaciones a posteriori por pares mostraron, de modo global, que el NC puntúa más alto que el resto de los grupos, mientras que el grupo SS es el que obtiene puntuaciones inferiores y similares a las del grupo AN (ver tabla 1).

Estudio del tamaño del efecto

Para realizar un análisis más preciso de la relevancia estadística de los hallazgos se calculó el tamaño 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 NC y NS se corresponde con el RL-CP ($\delta = 0.91$), mientras que la variable que menos los diferencia es R ($\delta = 0.59$). En el caso de NC y SS la que mejor diferencia es la variable DIS ($\delta = 1.90$) y la que menos es el REC ($\delta = 1.54$). Para los grupos NC y AN la que mejor diferencia es el REC ($\delta = 1.63$) y la que peor es R ($\delta = 0.94$). Entre los grupos NS y SS la variable que mejor diferencia a los grupos es el RL-LP ($\delta = 1.02$) y la que menos es DIS ($\delta = 0.66$). Para los grupos NS y AN la variable que mejor diferencia es el REC ($\delta = 0.91$) y la que menos es la DIS ($\delta = 0.13$). Por último, en la comparación entre los grupos SS y AN, la variable que mejor diferencia es la DIS ($\delta = 0.73$) y la que menos es el REC ($\delta = 0.14$) (ver tabla 2).

2. Sensibilidad, especificidad y puntos de corte de las pruebas

Debido al problema que suponen los falsos positivos en el estudio de la detección de la simulación, siguiendo las recomendaciones de la literatura (Greve y Bianchini, 2004), se decidió utilizar como criterio una especificidad superior al 90%. Con el objetivo de estudiar los puntos de corte que diferenciaban entre pacientes con TCE-L (tanto NC como NS) y el grupo sospechoso de simulación se realizaron 6 análisis de curvas ROC. El valor predictivo positivo (VPP) y negativo (VPN) han sido calculados tomando la prevalencia de base de simulación del 30% (Binder, 1993; Constantinou, Bauer, Ashendorf, Fisher y McCaffrey, 2005). Los puntos de corte no se muestran en el presente artículo para salvaguardar la validez de las medidas, pero diferentes puntos de corte para cada una de las variables con su VPP y VPN asociado a diferentes prevalencias de base (15% y 45%) están dispo-

Tabla 1
Medias, desviaciones típicas, análisis de las diferencias entre los grupos del estudio y análisis a posteriori de los índices del TAVEC

	Grupo NC	Grupo NS	Grupo SS	Grupo AN	U	p	Mann-Whitney
	Media (DT)	Media (DT)	Media (DT)	Media (DT)			
Total ensayos	55.06 (9.32)	49.14 (10.50)	38.90 (11.52)	43.92 (14.26)	16.87	.001	NS = SS = AN; NC = NS; NC > SS; NC > AN
RL-CP	11.36 (2.83)	8.85 (2.68)	6.50 (3.37)	7.88 (3.68)	19.75	.000	NS = SS = AN; NC = NS; NC > SS; NC > AN
RL-LP	11.60 (3.06)	9.35 (3.71)	5.70 (3.40)	6.70 (4.28)	24.42	.000	NS > SS = AN; NC > NS; NC > SS; NC > AN
Reconocimiento	15.06 (0.98)	13.92 (1.73)	11.20 (4.02)	11.70 (3.12)	26.82	.000	NC > (SS = AN); NC = NS; NS = SS; NS > AN
Discriminabilidad	94.95 (4.98)	85.55 (19.87)	73.00 (18.04)	83.48 (10.36)	32.33	.000	NC > (SS = AN); NC = NS; NS > SS; NS = AN

Nota. NC = grupo de pacientes no buscadores de compensación; NS = grupo de pacientes buscadores de compensación no sospechosos de simulación; SS = grupo de pacientes buscadores de compensación sospechosos de simulación; AN = análogos; TAVEC = Test de Aprendizaje Verbal España-Complutense; RL-CP = recuerdo libre a corto plazo; RL-LP = recuerdo libre a largo plazo.

Tabla 2

Delta de Cohen de las variables del TAVEC de los distintos grupos del estudio

Variable	NC vs. NS	NC vs. SS	NC vs. AN	NS vs. SS	NS vs. AN	SS vs. AN
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. NC = grupo de pacientes no buscadores de compensación; NS = grupo de pacientes buscadores de compensación no sospechosos de simulación; SS = grupo de pacientes buscadores de compensación sospechosos de simulación; AN = análogos; TAVEC = Test de Aprendizaje Verbal España-Complutense; RL-CP = recuerdo libre a corto plazo; RL-LP = recuerdo libre a largo plazo.

nibles para los lectores interesados a través del contacto con los autores de este artículo.

Los resultados mostraron una sensibilidad moderada para las distintas variables del TAVEC: 60% en la variable discriminabilidad, 50% para el total de palabras recordadas en los cinco ensayos, el recuerdo libre a corto plazo y en el reconocimiento y 40% para el recuerdo libre a largo plazo (ver tabla 3).

Discusión

En este estudio se ha analizado la utilidad de algunos de los índices del TAVEC para detectar la simulación de problemas de memoria, empleando un diseño mixto de grupos conocidos y análogos. Los resultados mostraron que, en general, las variables diferencian entre los diversos grupos del estudio, con un elevado tamaño del efecto, y que el establecimiento de puntos de corte correspondientes a una especificidad superior al 90% se asocia a una sensibilidad moderada entorno al 50%.

Las bajas puntuaciones del grupo sospechoso de simulación, si bien no siempre son estadísticamente inferiores a las obtenidas por los otros grupos, están en consonancia con la literatura, que indica una tendencia entre los simuladores a obtener puntuaciones inferiores a los pacientes en las variables neuropsicológicas, como ha sido señalado en otras investigaciones (Arnold et al., 2005; Curtis et al., 2006; Langeluddecke y Lucas, 2003; Lu, Boone, Cozolino y Mitchell, 2003). El análisis de las medias de los grupos y del tamaño del efecto pone de manifiesto que puede ser complicado diferenciar a los pacientes buscadores de compensación de los simuladores atendiendo a los índices del TAVEC, ya que ambos grupos obtienen puntuaciones similares en el número de palabras recordadas en los cinco ensayos, recuerdo libre a corto plazo y reconocimiento, si bien las variables recuerdo libre a largo plazo y discriminabilidad sí permiten diferenciar a estos grupos. En general, los resultados indican que en estas dos medidas los pacientes (sean o no buscadores de compensación) se comportan de modo diferente a los simuladores, que a su vez se comportan de modo similar a los análogos. En este sentido, hay que recordar que 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), convergencia que se produce en el caso del recuerdo a largo plazo y la discriminabilidad, variables que parecen adecuadas para ser empleadas como índices de simulación. Este hallazgo es congruente con investigaciones previas en las que se ha encontrado que el índice de discriminabilidad es la mejor medida para detectar el esfuerzo insuficiente de los pacientes en el CVLT (Curtis et al., 2006; Millis et al., 1995). En líneas generales, nuestros resultados son concordantes con la literatura sobre las pruebas neuropsicológicas de memoria verbal de listas de palabras, en la que suelen encontrarse diferencias entre los sujetos que buscan compensación y los que no la buscan (Meyers et al., 2001), así como con los estudios que encuentran diferencias empleando diseños de grupos conocidos (Curtis et al., 2006) y con los resultados encontrados empleando grupos de análogos (Demakis, 1999; Suhr, 2002; Suhr y Gunstad, 2000).

En cuanto a la sensibilidad y especificidad de las medidas, todas las variables del TAVEC han obtenido un área bajo la curva ROC similar o superior a .8, lo que confirma la validez de los datos. Nuestros resultados han sido prácticamente idénticos en cuanto a los puntos de corte propuestos a los de Curtis et al. (2006) con el CVLT, si bien la sensibilidad encontrada ha sido ligeramente superior en nuestro trabajo. Si seguimos la propuesta de estos autores, para aceptar como válido un índice de simulación debe obtener un VPP superior a 0.51 (equivalente a una especificidad del 95% y una sensibilidad mínima del 20%). Dado que todas las variables estudiadas se aproximan a dicho criterio, parece que el TAVEC puede ser una herramienta interesante a la hora de apoyar un diagnóstico de simulación. Además, nuestros resultados son muy similares a la sensibilidad y especificidad referidas por Vickery, Berry, Inman, Harris y Orey (2001) en su metaanálisis sobre las pruebas específicas de simulación y a las referidas por Larrabee (2003) para los índices derivados de pruebas no específicas, lo que apoya la generabilidad de nuestros resultados.

A pesar de todo lo anteriormente mencionado, es necesario considerar que los valores del VPP y VPN sugieren que un resultado positivo de los índices es más definitivo e indicativo de la presencia de esfuerzo suficiente que los resultados negativos indicativos de un

Tabla 3

Sensibilidad y especificidad asociada a los puntos de corte de las distintas variables del TAVEC para diferenciar entre pacientes y sospechosos de simulación

Variable	Sensibilidad	IC 95%	Especificidad	IC 95%	Área ROC	VPP	VPN
Total ensayos	50.0	18.9-81.1	90.9	78.3-97.4	0.831	70.2	80.9
RL-CP	50.0	18.9-81.1	93.2	81.3-98.5	0.820	75.9	81.3
RL-LP	40.0	12.4-73.6	93.2	81.3-98.5	0.857	71.5	78.4
Reconocimiento	50.0	18.9-81.1	97.7	87.9-99.6	0.793	90.4	82
Discriminabilidad	60.0	26.4-87.6	93.2	81.3-98.5	0.916	79	84.5

Nota. TAVEC = Test de Aprendizaje Verbal España-Complutense; RL-CP = recuerdo libre a corto plazo; RL-LP = recuerdo libre a largo plazo; IC = intervalo de confianza; VPP = valor predictivo positivo; VPN = valor predictivo negativo.

esfuerzo adecuado (Wolfe et al., 2010) y que estamos hablando de una sensibilidad que, en la mayoría de los casos, no supera el 50%. Esto quiere decir que con el empleo de estos índices no específicos de simulación solo seríamos capaces de detectar con el TAVEC aproximadamente a la mitad de las personas que estén simulando un daño en su evaluación neuropsicológica. En este tipo de resultados se centran algunos autores (p. ej., O'Bryant, Duff, Fisher y McCaffrey, 2004) para afirmar 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. Esto se hace más evidente si comparamos estos resultados con los obtenidos generalmente por las pruebas específicas de simulación, que obtienen una sensibilidad muy superior a la referida en este estudio y que también han sido probadas con población española (Vilar-López et al., 2007; Vilar-López, Gómez-Río, Caracuel-Romero, Llamas-Elvira y Pérez-García 2008; Vilar-López, Gómez-Río, Santiago-Ramajo et al., 2008; Vilar-López, Pérez-García, Sánchez-Barrera, Rodríguez-Fernández y Gómez-Río, 2011).

En cuanto a la aplicación clínica de los datos, es preciso recordar que ninguna medida de simulación es suficiente para determinar si un individuo está o no simulando. Dicha decisión debe estar basada siempre en un juicio clínico en el que se considere toda la información sobre el caso particular. Además, es muy importante considerar que este estudio se ha realizado con pacientes con TCE-L y que las puntuaciones por debajo de los puntos de corte propuestos en este estudio podrían ser frecuentes en individuos con daños cerebrales más severos, por lo que la aplicación de estos resultados a pacientes más severos podría llevar a un incremento importante de falsos positivos.

En lo referente a las limitaciones de este estudio, la principal es el bajo número de participantes, que hace necesaria la replicación de los resultados en estudios futuros. Otra posible limitación es que no hemos considerado las posibles diferencias demográficas entre los grupos, si bien esto es porque este tipo de medidas no influyen en los índices de simulación (i.e., Slick et al., 2000). El grupo de análogos no está equiparado al de pacientes en edad y escolaridad porque consideramos que los estudiantes de cuarto de Psicología son idóneos para simular problemas cognitivos de un modo realista ya que, debido a sus conocimientos sobre evaluación psicológica y daño cerebral, es probable que eludan la detección mejor. Además, no hemos incluido como grupo control un grupo asintomático porque la ocurrencia de un evento traumático en el grupo control elimina la posibilidad de que las diferencias encontradas entre los grupos se deban al traumatismo. Por último, solo hemos tenido en cuenta algunas variables del TAVEC y no hemos incluido funciones discriminantes que incorporen las distintas medidas de la prueba. En este sentido es necesario destacar que no hemos encontrado mejores resultados con el empleo de funciones discriminantes que con las variables aisladas, por lo que finalmente no han sido incluidas en el presente estudio. Este hallazgo está en consonancia con estudios previos que refieren la ausencia de diferencias entre el empleo de funciones discriminantes y las variables por separado en el CVLT (Coleman et al., 1998; Millis et al., 1995).

En conclusión, este es el primer estudio que explora la utilidad de los índices del TAVEC para detectar simulación. Nuestros resultados son consistentes con los referidos por otros autores para el CVLT (Curtis et al., 2006; Millis et al., 1995; Slick et al., 2000; Sweet et al., 2000; Trueblood, 1994), e indican que dichos índices pueden ser útiles a la hora de detectar problemas de memoria inventados o exagerados. No obstante, dichos índices han de ser empleados como un apoyo más en el complejo proceso de toma de decisiones que supone el diagnóstico de simulación y nunca deberían emplearse de modo aislado o como sustitutos de las pruebas específicas.

Extended summary

The validity of test execution is an essential part of neuropsychological testing. Several studies have indicated that the

California Verbal Learning Test (CVLT) is a useful tool in detecting malingering (i.e., Ashendorf, O'Briant, & McCaffrey 2003; Curtis, Greve, Bianchini, & Brennan, 2006; Demakis, 2004; Martens, Donders, & Millis, 2001; Millis, Putnam, Adams, & Ricker, 1995; Slick, Iverson, & Green 2000; Sweet et al., 2000; Wolfe et al., 2010). The aim of the present study is to test the efficacy of the Test de Aprendizaje Verbal España-Complutense (Spanish version of the CVLT) to detect malingering among Spanish individuals.

Methodology

Participants

Participants were 65 patients with normal CT scans diagnosed with MTBI (duration of the loss of consciousness less than or equal to 30 minutes, Glasgow Coma Scale score between 13 and 15, and posttraumatic amnesia for less than 24 hours), who presented Post-Concussion Syndrome (PCS) according to The Rivermead Post Concussion Symptoms Questionnaire (RPQ) (King, Crawford, Wenden, Moss, & Wade, 1995). Eleven individuals were excluded from this study because they had psychopathological diagnoses, clinical signs, or premonitory substance abuse. Therefore, the final number of participants was 54. Their age range was between 18 and 55 years.

Participants were divided into four groups. First, a group of patients not involved in compensation seeking processes (NC) that could not obtain any type of economic gain due to the consequences of their trauma. This group was composed of 16 men and 14 women ($N = 30$), with a mean age of 32.5 years ($SD = 13.67$) and a mean of 9.3 years of education ($SD = 3.45$). The mean time that elapsed from the TBI until the evaluation was 263.13 days ($SD = 378.66$). The second group was a compensation-seeking group that was not suspected of malingering (NS). This group was composed of 8 men and 6 women ($N = 14$) with a mean age of 35.92 years ($SD = 10.88$), a mean number of years of education of 9.14 ($SD = 3.50$). The mean time that elapsed from the traumatism until the evaluation was 287.78 days ($SD = 448.2$). The third group consisted of patients who sought compensation and were suspected of malingering (SS). They met the criteria for probable malingering proposed by Slick, Sherman, and Iverson (1999). To be considered malingering suspects, the participants had to obtain scores indicative of malingering on at least two specific tests (Victoria Symptom Validity Test, Test of Memory Malingering, the b Test, Dot Counting Test or Rey 15-Item Test). This group was composed of 8 men and 2 women ($N = 10$), with a mean age of 35.20 years ($SD = 11.10$), and a mean education of 8.5 years ($SD = 1.35$). The mean time that elapsed between the traumatism and the evaluation was 490 days ($SD = 583.88$). Finally, in this study, 30 4th-year Psychology students who were knowledgeable about neuropsychology made up the group of analogues (AN), following all of the recommendations proposed by Rogers (1997). The mean age of the group was 20.92 years ($SD = 3.08$), and the mean education level was 13.92 years ($SD = 2.35$). None of these participants reported a history of brain injury.

Results

Results showed statistical differences between the groups of the study for all of the TAVEC variables: total number of words recalled 1-5, $\chi^2(2) = 16.87, p < .001$; short term memory, $\chi^2(2) = 19.75, p < .000$; long-term memory, $\chi^2(2) = 24.42, p < .000$; recognition, $\chi^2(2) = 26.82, p < .000$; and discriminability, $\chi^2(2) = 32.33, p < .000$. Post-hoc comparisons showed that the NC group obtained the highest scores, while SS obtained the lowest, which was similar to the AN group. Considering a specificity > 90% (Greve & Bianchini, 2004), the results showed moderate sensitivity for the different indexes: 60% for discriminability, 50% for the total words recalled and long-term memory, and 40% for short-term memory.

Discussion

Our results replicate other studies that also utilize the CVLT, demonstrating that discriminability is the best measure for detecting insufficient effort (Curtis et al., 2006; Millis et al., 1995) and that differences exist in verbal measures between groups of differential prevalence (Meyers, Morrison, & Miller, 2001), known groups (Curtis et al., 2006), or analogues (Demakis, 1999; Suhr, 2002; Suhr & Gunstad, 2000). Moreover, our cut-off points were very similar to those of Curtis et al. (2006) with the CVLT, with sensitivity being slightly higher in our study.

With regard to the clinical setting, it is important to consider that malingering can never be diagnosed with just a single measure, a complex clinical judgement being necessary. TAVEC indexes may be useful in achieving this goal with Spanish patients, but should never be used in isolation.

Conflicto de intereses

Los autores de este artículo declaran que no tienen ningún conflicto de intereses.

Referencias

American Board of Clinical Neuropsychology. (2007). American Academy of Clinical Neuropsychology (AACN) practice guidelines for neuropsychological assessment and consultation. *The Clinical Neuropsychologist*, 21, 209-231.

Arnold, G., Boone, K. B., Lu, P., Dean, A., Wen, J., Nitch, S. y 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., O'Bryant, S. E. y McCaffrey, R. J. (2003). Specificity of malingering detection strategies in older adults using the CVLT and WCST. *The Clinical Neuropsychologist*, 17, 255-262.

Barrash, J., Suhr, J. y Manzel, K. (2004). Detecting poor effort and malingering with an expanded version of the Auditory Verbal Learning Test (AVLTX): validation with clinical samples. *Journal of the Clinical and Experimental Neuropsychology*, 26, 125-140.

Bauer, L., Yantz, C. L., Ryan, L. M., Warden, D. L. y McCaffrey, R. J. (2005). An examination of the California Verbal Learning Test II to detect incomplete effort in a traumatic brain-injury sample. *Applied Neuropsychology*, 12, 202-207.

Benedet, M. J. y Alejandro, M. A. (1998). *TAVEC. Test de Aprendizaje Verbal Española Complutense*. Madrid: TEA ediciones.

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., Houston, W. y Natoli, L. (1993). Malingering on neuropsychological memory tests: potential objective indicators. *Journal of Clinical Psychology*, 49, 45-53.

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.

Boone, K., Lu, P. y Herzberg, D. S. (2002). *The b Test. Manual*. Los Angeles: Western Psychological Services.

Boone, K. B., Lu, P., Back, C., King, C., Lee, A., Philpott, L., ... Warner-Chacon K. (2002). 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.

Bush, S. S. y 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.

Coleman, R. D., Rapport, L. J., Millis, S. R., Ricker, J. H. y Farchione, T. J. (1998). Effects of coaching on detection of malingering on the California Verbal Learning Test. *Journal of Clinical and Experimental Neuropsychology*, 20, 201-210.

Constantinou, M., Bauer, L., Ashendorf, L., Fisher, F. M. y 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.

Curtis, K. L., Greve, K. W., Bianchini, K. J. y Brennan, A. (2006). California Verbal Learning Test indicators of malingered neurocognitive dysfunction. Sensitivity and specificity in traumatic brain injury. *Assessment*, 13, 46-61.

Delis, D. C., Kramer, J. H., Kaplan, E. y Ober, B. A. (1987). *California Verbal Learning Test*. New York: Psychological Corporation.

Delis, D. C., Kramer, J. H., Kaplan, E. y Ober, B. A. (2000). *California Verbal Learning Test* (Second Edition). San Antonio, TX: 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.

Demakis, G. J. (2004). Application of clinically-derived malingering cutoffs on the California Verbal Learning Test and the Wechsler Adult Intelligence Test-Revised to an analog malingering study. *Applied Neuropsychology*, 11, 222-228.

Greiffenstein, M. F., Baker, W. J. y Gola, T. (1994). Validation of malingered amnesia measures with a large clinical sample. *Psychological Assessment*, 6, 218-224.

Greve, K. W. y 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., Curtis, K. L., Bianchini, K. J. y Ord, J. S. (2009). Are the original and second edition of the California Verbal Learning Test equally accurate in detecting malingering? *Assessment*, 16, 237-248.

King, J. H., Gfeller, J. D. y Davis, H. P. (1998). Detecting simulated memory impairment with the Rey Auditory Verbal Learning Test: implications of base rates and study generalizability. *Journal of the Clinical and Experimental Neuropsychology*, 20, 603-612.

King, N. S., Crawford, S., Wenden, F. J., Moss, N. E. y Wade, D. T. (1995). The Rivermead Post Concussion Symptoms Questionnaire: a measure of symptoms commonly experienced after head injury and its reliability. *Journal of Neurology*, 242, 587-592.

Langeluddecke, P. M. y 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. (2003). Detection of malingering using atypical performance patterns on standard neuropsychological tests. *The Clinical Neuropsychologist*, 17, 410-425.

Lu, P. H., Boone, K. B., Cozolino, L. y 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.

Martens, M., Donders, J. y Millis, S. R. (2001). Evaluation of invalid response sets after traumatic head injury. *Journal of Forensic Neuropsychology*, 2, 1-18.

Mathias, C. W., Greve, K. W., Bianchini, K. J., Houston, R. J. y Crouch, J. A. (2002). Detecting malingered neurocognitive dysfunction using the reliable digit span in traumatic brain injury. *Assessment*, 9, 301-308.

Meyers, J. E., Morrison, A. L. y Miller, J. C. (2001). How low is too low, revisited: Sentence repetition and AVL-Recognition in the detection of malingering. *Applied Neuropsychology*, 8, 234-241.

Millis, S. R. y Volinsky, C. T. (2001). Assessment of response bias in mild head injury: Beyond malingering tests. *Journal of Clinical and Experimental Neuropsychology*, 23, 809-828.

Millis, S. R. y 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. y Ricker, J. H. (1995). The California Verbal Learning Test in the Detection of Incomplete Effort in Neuropsychological Evaluation. *Psychological Assessment*, 7, 463-471.

Mittenberg, W., Patton, C., Canyock, E. M. y Condit, D. C. (2002). Base rates of malingering and symptom exaggeration. *Journal of Clinical and Experimental Neuropsychology*, 24, 1094-1102.

Moore, B. A. y Donders, J. (2004). Predictors of invalid neuropsychological test performance after traumatic brain injury. *Brain Injury*, 18, 975-984.

O'Bryant, S. E., Duff, K., Fisher, J. y McCaffrey, R. J. (2004). Performance profiles and cut-off scores on the Memory Assessment Scales. *Archives of Clinical Neuropsychology*, 19, 489-496.

Powell, M. R., Gfeller, J. D., Hendricks, B. L. y Sharland, M. (2004). Detecting symptom- and test-coached simulators with the Test of Memory Malingering. *Archives of Clinical Neuropsychology*, 19, 693-702.

Rey, A. (1964). *L'examen clinique en psychologie*. Paris: Presses Universitaires de France.

Rogers, R. (1997). Researching dissimulation. In R. Rogers (Ed.), *Clinical Assessment of Malingering and Deception* (2nd ed.) (pp. 398-426). New York: The Guilford Press.

Root, J. C., Robbins, R. N., Chang, L. y Van Gorp, W. G. (2006). Detection of inadequate effort on the California Verbal Learning Test-Second edition: forced choice recognition and critical item analysis. *Journal of the International Neuropsychological Society*, 12, 688-696.

Sharland, M. J. y Gfeller, J. D. (2007). A survey of neuropsychologists' beliefs and practices with respect to the assessment of effort. *Archives of Clinical Neuropsychology*, 22, 213-223.

Slick, D., Hopp, G., Strauss, E. y Thompson, G. (1997). *The Victoria Symptom Validity Test*. Odessa: PAR.

Slick, D. J., Iverson, G. L. y Green, P. (2000). California Verbal Learning Test indicators of suboptimal performance in a sample of head-injury litigants. *Journal of Clinical and Experimental Neuropsychology*, 22, 569-579.

Slick, D. J., Sherman, E. M. S. e Iverson, G. L. (1999). Diagnostic Criteria for Malingered Neurocognitive Dysfunction: Proposed Standards for Clinical Practice and Research. *The Clinical Neuropsychologist*, 13, 545-561.

Sullivan, K., Deffenti, C. y Keane, B. (2002). Malingering on the RAVLT: Part II. Detection strategies. *Archives of Clinical Neuropsychology*, 17, 223-233.

Suhr, J. A. (2002). Malingering, coaching, and the serial position effect. *Archives of Clinical Neuropsychology*, 17, 69-77.

Suhr, J. A. y Gunstad, J. (2000). The effects of coaching on the sensitivity and specificity of malingering measures. *Archives of Clinical Neuropsychology*, 15, 415-424.

Sweet, J. J., Wolfe, P., Sattlberger, E., Numan, B., Rosenfeld, J. P., Clingerman, S. y 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.

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.

- Vickery, C. D., Berry, D. T. R., Inman, T. H., Harris, M. J. y 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.
- Vilar López, R. (2009). La detección de simulación. In M. Pérez-García (Ed.), *Manual de Neuropsicología Clínica*, (pp. 283-302). Madrid: Pirámide.
- Vilar-López, R., Gómez-Río, M., Caracuel-Romero, A., Llamas-Elvira, J. y Pérez-García, M. (2008). Use of specific malingering measures in a Spanish sample. *Journal of Clinical and Experimental Neuropsychology*, 30, 710-722.
- Vilar-López, R., Gómez-Río, M., Santiago-Ramajo, S., Rodríguez-Fernández, A., Puente, A. E. y Pérez-García, M. (2008). Malingering detection in a Spanish population with a known-groups design. *Archives of Clinical Neuropsychology*, 23, 365-377.
- Vilar-López, R., Pérez-García, M., Sánchez-Barrera, M. B., Rodríguez-Fernández, A. y Gómez-Río, M. (2011). Symptom validity testing and its underlying psychophysiological response pattern: a preliminary study. *Archives of Clinical Neuropsychology*, 26, 133-143.
- Vilar-López, R., Santiago-Ramajo, S., Gómez-Río, M., Verdejo-García, A., Llamas, J. M. y Pérez-García, P. (2007). Detection of malingering in a Spanish population using three specific malingering tests. *Archives of Clinical Neuropsychology*, 22, 379-388.
- Wolfe, P. L., Millis, S. R., Hanks, R., Fichtenberg, N., Larrabee, G. J. y Sweet, J. J. (2010). Effort indicators within the California Verbal Learning Test-II (CVLT-II). *Clinical Neuropsychology*, 24, 153-168.