

SÉPTIMA EVALUACIÓN DE TEST EDITADOS EN ESPAÑA

Gómez Sánchez, Laura E.
SÉPTIMA EVALUACIÓN DE TEST EDITADOS EN ESPAÑA
Papeles del Psicólogo, vol. 40, núm. 3, 2019
Consejo General de Colegios Oficiales de Psicólogos, España
Disponible en: <https://www.redalyc.org/articulo.oa?id=77864998007>
DOI: <https://doi.org/10.23923/pap.psicol2019.2909>

SÉPTIMA EVALUACIÓN DE TEST EDITADOS EN ESPAÑA

Laura E. Gómez Sánchez
Facultad de Psicología. Universidad de Oviedo, España
gomezlaura@uniovi.es

DOI: <https://doi.org/10.23923/pap.psicol2019.2909>
Redalyc: <https://www.redalyc.org/articulo.oa?id=77864998007>

Recepción: 09 Abril 2019
Aprobación: 01 Agosto 2019

RESUMEN:

La Comisión de test del Consejo General de la Psicología del Colegio Oficial de Psicólogos lleva a cabo anualmente desde 2010 una evaluación de la calidad de los test que se publican en España. Como en las ediciones anteriores, el objetivo de este trabajo es describir el método y presentar los resultados de la séptima edición de evaluación de test, en la que ocho pruebas recientemente publicadas por cuatro casas editoriales (CEPE, EOS, PEARSON y TEA) fueron revisadas por un total de 16 expertos (i.e., dos revisores independientes por prueba) mediante la aplicación del Cuestionario de Evaluación de Test-Revisado (CET-R). De los resultados puede concluirse que, en términos generales, la calidad de los test evaluados es adecuada, destacando entre los puntos fuertes la calidad de las adaptaciones, las evidencias que proporcionan acerca de la consistencia interna, y la calidad de los baremos y la interpretación de sus puntuaciones. No obstante, se apuntan también algunas limitaciones y sugerencias de mejora.

PALABRAS CLAVE: Uso de los test, Evaluación de test, Psicometría, Calidad de los test, Validez, Fiabilidad..

ABSTRACT:

Every year since 2010, the Test Commission of the Spanish Psychological Association carries out a quality assessment of the tests published in Spain. As in previous editions, the aim of this paper is to describe the method and present the main results of the seventh test review, in which eight recently-published tests by four editorials (CEPE, EOS, Pearson, and TEA) were assessed by 16 experts (i.e. two independent reviewers for each test) through the application of the Test Review Questionnaire-Revised (CET-R). Results allow us to conclude that, in general terms, the quality of the assessed tests is adequate, highlighting as strong points the quality of the adaptations, the evidence on internal consistency, and the quality of standards and interpretation of scores. However, some limitations and suggestions for improvement are also noted.

KEYWORDS: Use of tests, Evaluation of tests, Psychometrics, Quality of the tests, Validity, Reliability..

Una de los deseos más reclamados por los psicólogos profesionales cuando se les pregunta acerca de sus opiniones sobre el uso de las pruebas se centra en disponer de información sobre estas que les ayude a tomar decisiones adecuadas (Evers, Muiz, Bartram et al., en prensa; Muñiz y Fernández-Hermida, 2010). Por ello, como viene siendo habitual desde 2010, la Comisión de Test del Consejo General de la Psicología del Colegio Oficial de Psicólogos (COP) ha abordado de nuevo la tarea de evaluar un pequeño número de test publicados por distintas editoriales en España.

Esta séptima edición, llevada a cabo en 2018, tiene como finalidad ayudar a los profesionales en la toma de decisiones de uso de los test proporcionándoles información acerca de su calidad mediante criterios teóricos, prácticos y psicométricos (Hernández et al., 2016; Muñiz y Fernández-Hermida, 2000; Muñiz, Hernández y Ponsoda, 2015). Tras esta edición, suman 75 los test evaluados cuyo informe puede encontrarse en la página web del Consejo General de la Psicología en España para su libre descarga (<https://www.cop.es/index.php?page=evaluacion-tests-editados-en-espana>). En la misma dirección se ubican los enlaces a los correspondientes artículos que describen el proceso de evaluación de cada edición en Papeles del Psicólogo (Elosua y Geisinger, 2016; Fonseca-Pedrero y Muñiz, 2017; Hernández, Tomás, Ferreres y Lloret, 2015; Hidalgo y Hernández, 2019; Muñiz, Fernández-Hermida, Fonseca-Pedrero, Campillo-Álvarez y Peña-Suárez, 2011; Ponsoda y Hontangas, 2013).

MÉTODO

Participantes

Con el fin de que revisaran los test elegidos para la séptima edición se contactó con 19 revisores, pero dos de ellos rechazaron participar en el proceso de evaluación por distintas razones y de uno no se obtuvo respuesta. En la Tabla 1 se detallan las 16 personas que participaron finalmente revisando los test seleccionados para esta edición (i.e., dos revisores por test). Como se puede observar, se procuró la paridad por sexo y la diversidad geográfica. Todos ellos eran profesores de universidad, la mayoría de las áreas de Psicometría y Metodología de las Ciencias del Comportamiento, si bien también se procuró que cada prueba fuera evaluada no solo por un experto de estas áreas sino también en el contenido o las variables medidas. Por ello, participaron evaluadores de las áreas de Personalidad, Evaluación y Tratamientos Psicológicos, así como de Psicología Evolutiva y de la Educación. Para su selección, además de por sus conocimientos psicométricos o teóricos sobre el contenido del test, se tuvo en cuenta que no tuviesen conflicto de intereses ni relación directa con los autores.

Instrumento

El instrumento utilizado, como en las ediciones anteriores, fue la versión revisada del Cuestionario para la Evaluación de los Test, conocido como CET-R (Hernández, Ponsoda, Muñiz, Prieto y Elosua, 2016; Prieto y Muñiz, 2000), que está basado en el Modelo de Evaluación de Test elaborado por la European Federation of Professional Psychologists Associations (Evers et al., 2013).

El CET-R está precedido por unas breves instrucciones para los aplicadores que tienen como objetivo facilitar su cumplimentación y proporcionar un enlace a un glosario de términos psicométricos. El cuestionario está compuesto por tres secciones básicas: (a) descripción general del test: se recogen datos como nombre del test, autores, editores, fecha de publicación, variables que pretende medir, áreas de aplicación, número y formato de los ítems, población a la que se dirige, procedimiento de corrección, cualificación requerida para usar el test y precio, entre otros; (b) valoración de las características del test: con una escala de 1 (inadecuado) a 5 (excelente) se valoran aspectos relacionados con los materiales y la documentación, la fundamentación teórica, la adaptación, el análisis de los ítems, las evidencias de validez y fiabilidad, los baremos y la interpretación de las puntuaciones; y (c) valoración global del test: se pide realizar una evaluación cualitativa del test resaltando sus puntos fuertes y débiles, así como una valoración cuantitativa de sus características generales indicando el promedio de las calificaciones emitidas en los correspondientes apartados. De este modo, el CET-R combina ítems de carácter cuantitativo que implican puntuar diversos aspectos con preguntas abiertas con objeto de recoger los argumentos que justifican las puntuaciones proporcionadas en cada apartado. Finalmente, el CET-R incluye una breve lista de referencias citadas en el cuestionario.

La séptima edición de test tiene como novedad con respecto a las anteriores en que, por primera vez, se ha utilizado una versión electrónica tipo formulario protegido en Word que, mediante el uso de casillas de verificación, listas desplegables y cuadros de texto permite a los revisores responder a los ítems de una forma más rápida y sencilla. Además, en esta última versión, se introdujeron mejoras de formato para hacer el cuestionario más atractivo y claro, tratando de mejorar la experiencia de los usuarios mediante el uso de colores, iconos y la clarificación visual de su estructura. Finalmente, se llevaron a cabo también unas ligeras modificaciones en el CET-R que no afectaban tanto a su contenido esencial como, más bien, a aspectos formales (p. ej., adaptar a normas APA, actualización de referencias) y lingüísticos (p. ej., corrección de erratas, simplificación de expresiones, uso inclusivo del lenguaje). Esta última versión del CET-R está

disponible para su consulta y descarga en la página web del Consejo General de la Psicología de España (<http://www.cop.es/uploads/pdf/CET-R.pdf>).

TABLA 1 PARTICIPANTES COMO REVISORES EN LA SÉPTIMA EVALUACIÓN DE TEST	
Nombre y apellidos	Filiación
Albert Sesé Abad	Univesitat de les Illes Balears
Ana M ^a Hernández Baeza	Universidad de Valencia
Beatriz Lucas Molina	Universidad de Valencia
Eduardo Fonseca Pedrero	Universidad de La Rioja
Francisco José Abad García	Universidad Autónoma de Madrid
Ignacio Pedrosa García	Universidad de Oviedo
Isabel Benítez Baena	Universidad Loyola Andalucía
Joan Guàrdia Olmos	Universitat de Barcelona
M. Ángeles Alcedo Rodríguez	Universidad de Oviedo
M. Carme Viladrich Segués	Universidad Autónoma de Barcelona
Maite Garaigordobil Landazabal	Universidad del País Vasco
Patricia Navas Macho	Universidad de Salamanca
Paula Elosua Oliden	Universidad del País Vasco
Urbano Lorenzo Seva	Universitat Rovira i Virgili
Vicente Ponsoda Gil	Universidad Autónoma de Madrid
Victor B. Arias González	Universidad de Salamanca

TABLA 1
PARTICIPANTES COMO REVISORES EN LA SÉPTIMA EVALUACIÓN DE TEST

Procedimiento

En esta ocasión fueron las editoriales (CEPE, EOS, PEARSON y TEA) las que propusieron a la Comisión de Test del Consejo General de la Psicología en España las pruebas que querían someter a evaluación (i.e., ocho test). De modo similar a como se procedió en ediciones anteriores, una vez aceptada la propuesta por la Comisión, la coordinadora de esta séptima edición (autora de este artículo) seleccionó el panel de expertos para el proceso de revisión, de tal forma que cada prueba fuera revisada por dos revisores independientes: uno con un perfil técnico-psicométrico y otro experto en la(s) variable(s) medida(s).

En mayo de 2018, la coordinadora envió por correo electrónico a los evaluadores seleccionados que accedieron a colaborar la versión electrónica del CET-R. Por su parte, las editoriales pusieron a disposición del Consejo General de la Psicología, de manera gratuita, tres ejemplares completos de cada test. El Consejo General de la Psicología fue el encargado de enviar por correo postal un ejemplar de la prueba a cada par de revisores y a la coordinadora. De este modo, cada evaluador revisó un único test y la coordinadora revisó los ocho test seleccionados.

La tarea de los revisores y de la coordinadora, una vez recibidos los test a evaluar, consistió en aplicar el CET-R a la prueba asignada a cambio de una compensación económica simbólica de 50 euros (que algunos declinaron) y de quedarse de forma gratuita con el ejemplar del test. Todos los evaluadores culminaron su tarea y enviaron su versión cumplimentada del CET-R a la coordinadora entre los meses de mayo y julio del mismo año. En ese tiempo, la coordinadora estuvo siempre disponible para responder dudas, solucionar posibles problemas y recoger comentarios que los revisores quisieran hacer constar.

A finales del mes de julio, la coordinadora, con las dos versiones del CET-R cumplimentadas por los revisores independientes más su propia evaluación, elaboró un informe provisional para cada uno de los

test. En la gran mayoría de los casos, solo hubo que integrar las tres evaluaciones disponibles para cada prueba, observándose una alta coincidencia en las puntuaciones otorgadas. En los pocos casos en los que hubo discrepancias entre los dos revisores, fue la evaluación independiente de la coordinadora la tomada en cuenta para resolver las diferencias.

En el mismo mes de julio los informes provisionales fueron enviados al COP, que se encargó este de hacérselos llegar inmediatamente a las respectivas editoriales. Las editoriales tuvieron de plazo hasta mediados de septiembre para realizar las alegaciones que estimaran oportunas a los informes provisionales. Así lo hicieron la mayoría de ellas, de forma muy detallada, precisando, corrigiendo, aclarando y comentando aquellos aspectos que consideraron de los informes de evaluación, con la única excepción de la editorial EOS que no deseó expresar ninguna alegación.

Tras una lectura detenida de las alegaciones y los comentarios realizados por los autores y las editoriales de los test, la coordinadora integró los comentarios que estimó oportunos y corrigió las puntuaciones que consideró justificadas en el informe final de cada prueba. Llegados a este punto cabe mencionar que para esta edición no solo se creó y actualizó la versión electrónica del CET-R, sino que también se elaboró la versión electrónica tipo formulario protegido en Word del informe final, tratando de mejorar su formato y adecuarlo a la estética de la imagen corporativa del COP. Así, los informes finales se realizaron en diciembre, mes en que se enviaron al COP y fueron publicados en su página web. El proceso de evaluación de la séptima edición de evaluación de test, ilustrado en la Figura 1, comenzó y fue concluido por tanto dentro del año natural 2018.

Análisis de datos

Para realizar los informes finales, la coordinadora tuvo en cuenta: (a) los comentarios de los revisores para llevar a cabo un análisis cualitativo que los integrara; y (b) las puntuaciones asignadas en los ítems para llevar a cabo un análisis descriptivo, utilizando básicamente el promedio de las puntuaciones consignadas en los ítems para calcular la puntuación final en 14 apartados. Las puntuaciones en los ítems podían ser: inadecuado=1; adecuado, pero con algunas carencias=2; adecuado=3; bueno=4 y excelente=5.

RESULTADOS

Los informes detallados correspondientes a los ocho test facilitados por las cuatro casas editoriales participantes que han sido valorados en esta séptima edición de evaluación de test publicados en España pueden consultarse y descargarse en la página web del COP, dentro del apartado correspondiente al año 2018 (<https://www.cop.es/index.php?page=evaluacion-tests-editados-en-espana>). Los ocho test valorados en esta edición se detallan en la Tabla 2, habiendo sometido a evaluación tres test tanto CEPE como TEA y uno las editoriales EOS y PEARSON. Todos los test seleccionados fueron publicados por tales editoriales entre 2017 y 2018.

En la Tabla 3 se recoge un resumen de las puntuaciones medias obtenidas por los ocho test, con los aspectos evaluados organizados por filas y los test sometidos a evaluación por columnas. En la última columna, se incluye la puntuación media obtenida por los ocho test en cada uno de los 14 aspectos evaluados. Teniendo en cuenta que las puntuaciones pueden oscilar entre 1 (inadecuado) y 5 (excelente), en términos generales, los resultados fueron entre adecuados ($M=3,3$) y excelentes ($M=4,8$) para la mayoría de los 14 aspectos estimados. Ninguno de ellos recibió una puntuación que significara “inadecuado” (1) y las puntuaciones correspondientes a una valoración de “adecuado, pero con carencias” (2) fueron excepcionales. De hecho, la calificación media en los 14 aspectos corresponde con una valoración de “bueno” ($M=4,1$) y las calificaciones medias de los ocho test oscilaron entre 3,5 (adecuado-bueno) y 4,7 (bueno-excelente).

CONCLUSIONES

En conclusión, los puntos más fuertes de los test evaluados en esta séptima edición de test son su consistencia interna, los baremos e interpretación de las puntuaciones, los materiales y la documentación. Cuando se trata de adaptaciones, el modo en que estas se llevan a cabo también es valorado como excelente. En general, el proceso de adaptación se lleva a cabo siguiendo las Directrices de la Comisión Internacional de Test (Hambleton, Merenda y Spielberg, 2005; International Test Commission, 2018; Muñiz, Elosua y Hambleton, 2013).

Por el contrario, los aspectos que presentan mayores limitaciones y son más susceptibles de mejora tienen que ver con la provisión de mejores evidencias de validez relacionas con el contenido y con la estructura interna de los test, si bien en muchas ocasiones no se trata tanto de un problema de carencia o inadecuación de este tipo de evidencias sino más bien que se obvian o se incluyen de forma superficial en los manuales. Por ello, para próximas ediciones, se recomienda a las editoriales que sometan a valoración no solo los test y los manuales de administración y corrección sino cualquier otro tipo de material adicional que recoja dicha información (por ejemplo, artículos científicos).

Cabe subrayar que sería conveniente para futuras ediciones tratar de abordar los aspectos que han sido más obviados en esta y en previas ediciones, siempre y cuando sean estos pertinentes, procurando proporcionar evidencias de fiabilidad interjueces, de equivalencia y mediante la Teoría de Respuesta al ítem (TRI). De hecho, llama la atención que mientras en la edición anterior casi un tercio de los instrumentos analizados habían abordado la precisión de la medida desde los modelos de la TRI (Hidalgo y Hernández, 2019), no lo hizo ninguno en la presente edición. Por otro lado, solo dos de las pruebas proporcionan evidencias del funcionamiento diferencial de los ítems (DIF), número que si bien es igual al de la evaluación anterior y superior al de ediciones previas, convendría incrementar progresivamente la inclusión de este relevante criterio en la construcción y validación de los test publicados (Muñiz y Fonseca-Pedrero, 2019).

Finalmente, con respecto a la aplicación del CET-R, se ponía el acento en ediciones anteriores en la dificultad de interpretar, analizar y puntuar determinados apartados, si bien tales dificultades no se han manifestado de forma notable en esta séptima edición, quizás gracias a las últimas revisiones y mejoras realizadas en el CET-R tanto en forma como en contenido basadas en las sugerencias realizadas en las evaluaciones previas. A pesar de no observarse grandes dificultades en la comprensión y puntuación de los ítems, de la aplicación del CET-R y de los comentarios realizados por los participantes de esta edición (expertos, editores, autores y coordinadora) se desprenden algunas reflexiones y sugerencias para el futuro.

En primer lugar, sería conveniente evolucionar de una versión electrónica del CET-R a una versión online, vinculada a una base de datos, que permitiera calcular la valoración numérica (media de cada apartado) de forma automática, optimizando así el uso de las nuevas tecnologías en la evaluación de los test (Fonseca-Pedrero y Muñiz, 2017; Hidalgo y Hernández, 2019; Muñiz y Fernández-Hermida, 2010). Del mismo modo, sobre el CET-R, sería conveniente revisar el sistema de puntuación para evitar posibles sesgos en la evaluación, pues tal y como se viene realizando, para obtener la valoración global de un apartado no se tienen en cuenta aquellos aspectos sobre los que no hay información, de tal modo que resulta más ventajoso no presentar ninguna información (—) que presentar alguna información y que esta sea inadecuada (1) o adecuada, pero con carencias (2). Por ello, se apunta como recomendación que se utilicen categorías como “no aplicable” (n/a) y “no se aporta” (0).

En cuanto a los expertos, conviene subrayar la importancia de combinar expertos en psicometría con expertos en el contenido de los test por la riqueza de los comentarios que unos y otros aportan, pues más que contradecirse suelen ser complementarios. De hecho, en esta edición llama la atención el alto grado de acuerdo observado entre los revisores independientes. No obstante, cabe subrayar que en el mundo de la revisión por pares solemos tener un cierto sesgo a enfatizar las limitaciones de lo que evaluamos y a describir brevemente aquellos aspectos que consideramos adecuados o excelentes. En este sentido, se recomienda para

futuras ediciones hacer hincapié a los expertos en la necesidad de subrayar tanto las limitaciones como los aspectos fuertes de la prueba con objeto de que los comentarios cualitativos que se incluyen en los informes finales puedan estar más equilibrados y ser más precisos.

Igualmente, se recomienda a autores y editores que realicen alegaciones concretas pero detalladas, especificando exactamente las partes en las que observan desacuerdos y proponiendo formulaciones alternativas, especialmente con ánimo de contrarrestar ese posible sesgo de la revisión por pares que tiende a subrayar más las limitaciones. Finalmente, conviene recordar que los test se valoran con la información proporcionada por las editoriales que, en esta edición como en las anteriores, se limita a los tests y sus manuales de aplicación, pero existe la posibilidad de incluir entre los materiales sometidos a evaluación referencias o adjuntos donde pueda encontrarse información complementaria que por alguna razón se haya decidido no incluir en el manual.

Para terminar, conviene hacer hincapié en el impacto que la evaluación de la calidad de los test y, más concretamente, el uso del CET-R está teniendo en el ámbito de la Psicología (Hidalgo y Hernández, 2019), pues la herramienta se está revelando no solo como esencial en la toma de decisiones de los psicólogos profesionales, sino también para la docencia universitaria (el 64% de los profesores que lo conocen lo utilizan en sus clases), la investigación (muchos investigadores lo utilizan como una herramienta de autoevaluación y lo tienen en cuenta en los procesos de construcción, adaptación y validación de las pruebas) y las propias editoriales (que lo utilizan no solo como guía sino como sistema de referencia de mejora continua)

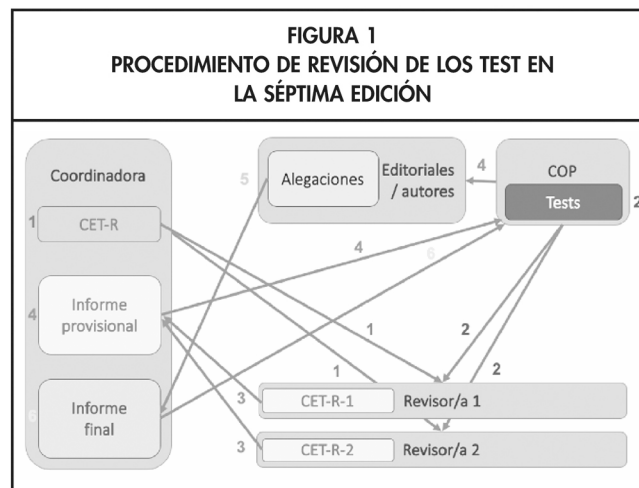


FIGURA 1
PROCEDIMIENTO DE REVISIÓN DE LOS TEST EN LA SÉPTIMA EDICIÓN

TABLA 2 TEST EVALUADOS EN LA SÉPTIMA EDICIÓN (2018)				
Acronimo	Nombre de la prueba	Autor/es originales (año publicación)	Autores adaptación española (año publicación)	Editorial
PAIB-1 Renovado	Prueba de Aspectos Instrumentales Básicos en Lenguaje y Matemáticas: E. Infantil (5 años) 1º y 2º E. Primaria	---	Ramos, Galve, Trallero y Martínez (2017)	CEPE. Ciencias de la Educación Preescolar y Especial.
PAIB-2 Renovado	Prueba de Aspectos Instrumentales Básicos en Lenguaje y Matemáticas: 3º y 4º E. Primaria	---	Ramos, Galve y Martínez (2017)	
PAIB-3 Renovado	Prueba de Aspectos Instrumentales Básicos en Lenguaje y Matemáticas: 5º y 6º E. Primaria y 1º ESO	---	Ramos, Martínez y Galve (2017)	
IAES-A	Inventario de Ansiedad Escolar - Abreviado	---	García e Inglés (2017)	EOS. Instituto de Orientación Psicológica
BYI-2	Inventarios de Beck para niños y adolescentes - 2	Beck, Beck, Jolly y Steer (2005)	Hernández, Aguilar, Paradell y Vallar (2017)	PEARSON Educación
BRIEF-2	Evaluación Conductual de la Función Ejecutiva	Gioia et al. (2015)	Maldonado et al. (2017)	TEA Ediciones
DP-3	Perfil de Desarrollo - 3	Alpern (2007)	Sánchez (2018)	
Factor g-R	Test de Inteligencia No Verbal - Revisado	Cattell, Cattell y Weiss (2006)	Departamento de I+D+i de TEA Ediciones (2017)	

TABLA 2
TEST EVALUADOS EN LA SÉPTIMA EDICIÓN (2018)

TABLA 3 RESUMEN DE LAS CALIFICACIONES DE LOS TEST EVALUADOS EN LA SÉPTIMA EDICIÓN									
	IAES	BYI -2	Factor g R	BRIEF -2	DP -3	PAIB -1	PAIB 2	PAIB -3	M
Materiales y documentación	4	4	4,5	5	5	3,5	3,5	3,5	4,1
Fundamentación teórica	4,5	3,5	5	5	4,5	3	3	3	3,9
Adaptación	---	4	4	5	5	---	---	---	4,5
Análisis de ítems	3,5	---	4,5	3	5	3,5	3,5	3,5	3,8
Validez: contenido	4	3	3	4	4,5	2,5	2,5	2,5	3,3
Validez: relación con otras variables	3	4	4	4,5	4,5	3	3	3,5	3,7
Validez: estructura interna	3,5	2	4,5	3,5	4,5	---	---	---	3,6
Validez: análisis del DIF	3,5	---	---	---	4,5	---	---	---	4,0
Fiabilidad: equivalencia	---	---	---	4,5	---	---	---	---	4,5
Fiabilidad: consistencia interna	4	4,5	4,5	5	5	5	5	5	4,8
Fiabilidad: estabilidad	4	4,5	3	4,5	4,5	---	---	---	4,1
Fiabilidad: TRI	---	---	---	---	---	---	---	---	---
Fiabilidad: interjueces	---	---	---	---	---	---	---	---	---
Baremos e interpretación de las puntuaciones	4	5	5	4	5	4	4	4	4,4

Nota. --- = aspecto no pertinente o sobre el que no se proporcionaron datos.

TABLA 3
RESUMEN DE LAS CALIFICACIONES DE LOS TEST EVALUADOS EN LA SÉPTIMA EDICIÓN

CONFLICTO DE INTERESES

No existe conflicto de intereses.

REFERENCIAS

- Elosua, P., y Geisinger, K. (2016). Cuarta evaluación de test editados en España: forma y fondo. *Papeles del Psicólogo*, 37, 82-88.
- Evers, A., Muñoz, J., Hagemester, C., Høstmælingen, A., Lindley, P., Sjöberg, A., y Bartram, D. (2013). Assessing the quality of test: Revision of the EFPA review model. *Psicothema*, 25, 283-291
- Fonseca-Pedrero, E., y Muñoz, J. (2017). Quinta evaluación de test editados en España: Mirando hacia atrás, construyendo el futuro. *Papeles del Psicólogo*, 38, 161-168.
- Hambleton, R.K., Merenda, P.F., y Spielberger, C. (Eds.). (2005). *Adapting educational and psychological test for cross-cultural assessment*. Mahwah, NJ: Lawrence Erlbaum Publishers.
- Hernández, A., Ponsoda, V., Muñoz, J., Prieto, G., y Elosua, P. (2016). Revisión del modelo para evaluar la calidad de los test utilizados en España. *Papeles del Psicólogo*, 37, 192- 197.

- Hernández, A., Tomás, I., Ferreres, A., y Lloret, S. (2015). Tercera evaluación de test editados en España. *Papeles del Psicólogo*, 36, 1-8.
- Hidalgo, M. D., y Hernández, A. (2019). Sexta evaluación de test editados en España: resultados e impacto del modelo en docentes y editoriales. *Papeles del Psicólogo*, 40, 21-30.
- International Test Commission (2018). ITC Guidelines for Translating and Adapting Test (Second Edition). *International Journal of Testing*, 18, 101-134.
- Muñiz, J., Elosua, P., y Hambleton, R.K. (2013). Directrices para la traducción y adaptación de los test: segunda edición. *Psicothema*, 25, 151-157.
- Muñiz, J., y Fernández-Hermida, J.R. (2000). La utilización de los test en España. *Papeles del Psicólogo*, 76, 41-49.
- Muñiz, J., y Fernández-Hermida, J.R. (2010). La opinión de los psicólogos españoles sobre el uso de los test. *Papeles del Psicólogo*, 31, 108-121.
- Muñiz, J., Fernández-Hermida, J.R., Fonseca-Pedrero, E., Campillo-Álvarez, A., y Peña-Suárez, E. (2011). Evaluación de test editados en España. *Papeles del Psicólogo*, 32, 113-128.
- Muñiz, J., y Fonseca-Pedrero, E. (2019). Diez pasos para la construcción de un test. *Psicothema*, 31(1), 7-16. doi: 10.7334/psicothema2018.291
- Muñiz, J., Hernández, A., y Ponsoda, V. (2015). Nuevas directrices sobre el uso de los test: investigación, control de calidad y seguridad. *Papeles del Psicólogo*, 36, 161-173.
- Ponsoda, V., y Hontangas, P. (2013). Second evaluation of test published in Spain. *Papeles del Psicólogo*, 34, 82-90.
- Prieto, G., y Muñiz, J. (2000). Un modelo para evaluar la calidad de los test utilizados en España. *Papeles del Psicólogo*, 77, 65-72.