



SEVENTH REVIEW OF TESTS PUBLISHED IN SPAIN

Laura E. Gómez Sánchez
Universidad de Oviedo

La Comisión de tests del Consejo General de la Psicología del Colegio Oficial de Psicólogos lleva a cabo anualmente desde 2010 una evaluación de la calidad de los tests 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 tests, 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 Tests-Revisado (CET-R). De los resultados puede concluirse que, en términos generales, la calidad de los tests 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 tests, Evaluación de tests, Psicometría, Calidad de los tests, Validez, Fiabilidad.*

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.

Key words: *Use of tests, Evaluation of tests, Psychometrics, Quality of the tests, Validity, Reliability.*

One of the wishes most requested by professional psychologists, when asked about their opinions on the use of tests, focuses on having information about them that helps the psychologists make appropriate decisions (Evers, Muiz, Bartram, et al., in press; Muñiz & Fernández-Hermida, 2010). Therefore, as has been the custom since 2010, the Test Commission of the General Council of the Spanish Psychological Association (COP in Spanish) has once again undertaken the task of evaluating a small number of tests published by different publishers in Spain.

As with the previous editions, this seventh edition of the review, carried out in 2018, is intended to help professionals in making decisions regarding the use of tests, by providing information about their quality through theoretical, practical, and psychometric criteria (Hernández et al., 2016; Muñiz & Fernández-Hermida, 2000; Muñiz, Hernández, & Ponsoda, 2015). After this edition, there are a grand total of 75 evaluated tests whose reports can be found on the website of the Spanish Psychological Association for free download (<https://www.cop.es/index.php?page=evaluacion-tests-editados-en-espana>). On the same address, there are links to

the corresponding articles that describe the review process of each edition in *Papeles del Psicólogo* (Elosua & Geisinger, 2016; Fonseca-Pedrero & Muñiz, 2017; Hernández, Tomás, Ferreres & Lloret, 2015; Hidalgo & Hernández, 2019; Muñiz, Fernández-Hermida, Fonseca-Pedrero, Campillo-Álvarez, & Peña-Suárez, 2011; Ponsoda & Hontangas, 2013).

METHOD

Participants

In order to review the tests chosen for the seventh edition, 19 reviewers were contacted, but two of them refused to participate in the review process for different reasons and we did not receive an answer from one other. Table 1 lists the 16 people who finally participated as reviewers of the tests selected for this edition (i.e., two reviewers per test). As can be seen, gender parity and geographic diversity were sought. All of the reviewers were university professors, most in the areas of psychometrics and methodology of behavioral sciences, although the aim was also for each test to be evaluated not only by an expert in these areas but also in the content or variables evaluated. Therefore, evaluators were from the areas of personality, evaluation, and psychological treatments, as well as evolutionary and educational psychology. In selecting these reviewers, as well as considering their psychometric or theoretical knowledge about the content of the test, it was established that they had no conflict of interest or direct relationship with the authors.

Received: 9 April 2019 - Accepted: 1 August 2019

Correspondence: Laura E. Gómez. Facultad de Psicología. Universidad de Oviedo. Plaza Feijoo s/n. 33003 Oviedo. España.

E-mail: gomezlaura@uniovi.es



Instrument

As in the previous editions, the instrument used was the revised version of the Test Evaluation Questionnaire, known as CET-R (Hernández, Ponsoda, Muñiz, Prieto, & Elosua, 2016; Prieto & Muñiz, 2000), which is based on the Test Review Model developed by the European Federation of Professional Psychologists Associations (Evers et al., 2013).

The CET-R is preceded by brief instructions for users aimed at facilitating its completion and providing a link to a glossary of psychometric terms. The questionnaire is composed of three basic sections: (a) general description of the test: data collected include test name, authors, editors, publication date, variables to be measured, application areas, number and format of the items, population to which it is addressed, correction procedure, and the qualification required to use the test and price; (b) evaluation of the characteristics of the test: with a scale of 1 (inadequate) to 5 (excellent), aspects assessed relate to materials and documentation, theoretical foundation, adaptation, item analysis, validity and reliability evidence, the scales, and the interpretation of the scores; and (c) overall assessment of the test: a qualitative evaluation of the test is requested, highlighting its strengths and weaknesses, as well as a quantitative assessment of its general characteristics indicating the average of the ratings issued in the corresponding sections. Thus, the CET-R combines quantitative items that involve scoring various aspects with open questions in order to collect the arguments that justify the scores provided in each section. Finally, the CET-R includes a brief list of references cited in the questionnaire.

As a novelty with respect to the previous editions, the seventh edition of the test review used, for the first time, an electronic

version - a protected form in Word that, through the use of check boxes, drop-down lists, and text boxes, allows reviewers to respond to items more quickly and easily. In addition, in this latest version, format improvements were introduced to make the questionnaire more attractive and clearer, attempting to improve the user experience through the use of colors, icons, and the visual clarification of its structure. Finally, slight modifications were also made to the CET-R that did not affect either its essential content or, especially, formal aspects (e.g., adapting to APA standards, updating references) and linguistic aspects (e.g., correction of typos, simplification of expressions, inclusive use of language). This latest version of the CET-R is available for consultation and download on the website of the Spanish Psychological Association (www.cop.es/uploads/pdf/CET-R.pdf).

Process

On this occasion it was the publishers (CEPE, EOS, PEARSON, and TEA) who proposed the tests they wished to submit to the Test Commission of the General Council of the Spanish Psychological Association (i.e., eight tests). Similar to previous editions, once the proposal was accepted by the Commission, the coordinator of the seventh edition (the author of this article) selected the panel of experts for the review process, such that each test was reviewed by two independent reviewers: one with a technical-psychometric profile and the other an expert in the variable(s) measured.

In May 2018, the coordinator emailed the selected evaluators who agreed to collaborate with the electronic version of the CET-R. For their part, the publishers made available to the Spanish Psychological Association, free of charge, three complete copies of each test. The Spanish Psychological Association was responsible for mailing a hard copy of the test to each pair of reviewers and the coordinator. Thus, each evaluator reviewed a single test and the coordinator reviewed the eight selected tests.

The task of the reviewers and the coordinator, after receiving the tests to be evaluated, was to apply the CET-R to the test assigned in exchange for a symbolic monetary compensation of 50 euros (which some declined) and to keep the test copy for free. All reviewers finished the task and sent the completed version of the CET-R to the coordinator between the months of May and July of the same year. During that time, the coordinator was always available to answer questions, solve possible problems, and collect any comments that the reviewers wished to record.

At the end of July, the coordinator, using the two versions of the CET-R completed by the independent reviewers plus her own evaluation, prepared an interim report for each of the tests. In the vast majority of cases, only the three evaluations available for each test had to be integrated, and there was a high coincidence in the scores awarded. In the few cases in which there were discrepancies between the two reviewers, the independent evaluation of the coordinator was taken into account to resolve the differences.

In the same month, July, the provisional reports were sent to

TABLE 1
PARTICIPANT REVIEWERS IN THE
SEVENTH TEST REVIEW

Name and surnames	Affiliation
Albert Sesé Abad	University of the Balearic Islands
Ana M ^a Hernández Baeza	University of Valencia
Beatriz Lucas Molina	University of Valencia
Eduardo Fonseca Pedrero	University of La Rioja
Francisco José Abad García (Paco)	Autonomous University of Madrid
Ignacio Pedrosa García	University of Oviedo
Isabel Benítez Baena	Loyola University Andalucía
Joan Guàrdia Olmos	University of Barcelona
M. Ángeles Alcedo Rodríguez	University of Oviedo
M. Carme Viladrich Segué	Autonomous University of Barcelona
Maite Garaigordobil Landazabal	University of the Basque Country
Patricia Navas Macho	University of Salamanca
Paula Elosua Oliden	University of the Basque Country
Urbano Lorenzo Seva	Rovira i Virgili University
Vicente Ponsoda Gil	Autonomous University of Madrid
Víctor B. Arias González	University of Salamanca



the COP, which was responsible for sending them immediately to the respective publishers. The publishers had a deadline until mid-September to make the allegations they deemed appropriate in the provisional reports. Most of them did so in great detail, specifying, correcting, clarifying, and discussing the aspects considered in the evaluation reports, with the sole exception of the publisher EOS, who did not wish to express any allegation.

After a careful reading of the allegations and comments made by the authors and the publishers of the tests, the coordinator included the comments that she considered appropriate and corrected the scores that she considered justified in the final report of each test. At this point it should be mentioned that, for this edition, not only was the electronic version of the CET-R created and updated, but also the electronic version of the protected form in Word of the final report was prepared, in an attempt to improve its format and adapt it to the aesthetics of the corporate image of the COP. Thus, the final reports were produced in December, when they were sent to the COP and published on its website. The evaluation process of the seventh edition of the test review, illustrated in Figure 1, began and was therefore concluded within the calendar year 2018.

Data analysis

To produce the final reports, the coordinator took into account: (a) the comments of the reviewers, in order to provide a qualitative analysis that includes them; and (b) the scores assigned in the items in order to carry out a descriptive analysis, basically using the average of the scores recorded in the items to calculate the final score in 14 sections. As mentioned earlier, the scores on the items were: inappropriate = 1; adequate, but with some deficiencies = 2; adequate = 3; good = 4; and excellent = 5.

Results

The detailed reports corresponding to the eight tests provided by the four participating publishing houses that were assessed in this seventh edition of the evaluation of tests published in Spain can be consulted and downloaded, as already mentioned, on the COP website, within the section corresponding to the year 2018 (<https://www.cop.es/index.php?page=evaluacion-tests-editados-en-espana>). The eight tests assessed in this edition are detailed in Table 2, with three CEPE and TEA tests each being submitted to evaluation, as well as one each from the publishers EOS and PEARSON. All the selected tests were published by these publishers between 2017 and 2018.

Table 3 shows a summary of the average scores obtained by the eight tests, with the 14 aspects evaluated organized by rows and the tests submitted to evaluation by columns. In the last column, the average score obtained by the eight tests in each of the 14 aspects evaluated is shown. Taking into account that the scores can range between 1 (inadequate) and 5 (excellent), in general terms, the results were between adequate ($M = 3.3$) and excellent ($M = 4.8$) for most of the 14 aspects evaluated. None of them received a score that meant “inadequate” (1) and the

scores corresponding to an assessment of “adequate, but with some deficiencies” (2) were rare. In fact, the average score in the 14 aspects corresponds to an assessment of “good” ($M = 4.1$) and the average scores of the eight tests ranged between.

CONCLUSIONS

In conclusion, the strongest points of the tests evaluated in this seventh edition of tests are their internal consistency, the scales and interpretation of the scores, the materials, and the documentation. When it comes to adaptations, the way in which these are carried out was also valued as excellent. In general, the adaptation process was carried out following the Guidelines of the International Test Commission (Hambleton, Merenda, & Spielberg, 2005; International Test Commission, 2018; Muñoz, Elosua, & Hambleton, 2013).

On the contrary, the aspects that presented the most limitations and had the most room for improvement concerned the provision of better validity evidence related to the content and the internal structure of the tests, although in many cases it was not so much a problem of this type of evidence being missing or inadequate but rather that it was overlooked or superficially included in the manuals. Therefore, for future editions, publishers are recommended to submit not only the tests and administration and correction manuals for evaluation, but also any other type of additional material that provides this information (for example, scientific articles).

It should be stressed that it would be beneficial for future editions to attempt to address the aspects that have been most obviated in this and previous editions, as long as they are pertinent, aiming to provide evidence of interrater reliability, equivalence, and using item response theory (IRT). In fact, it is striking that while in the previous edition almost one third of the instruments analyzed had addressed the accuracy of the measurement using IRT models (Hidalgo & Hernández, 2019), none did so in the present edition. Furthermore, only two of the tests provided evidence of the differential functioning of the items (DIF). Although this is equal to that of the previous evaluation and superior to that of previous editions, it would be

**FIGURE
TEST REVIEW PROCEDURE,
SEVENTH EDITION**

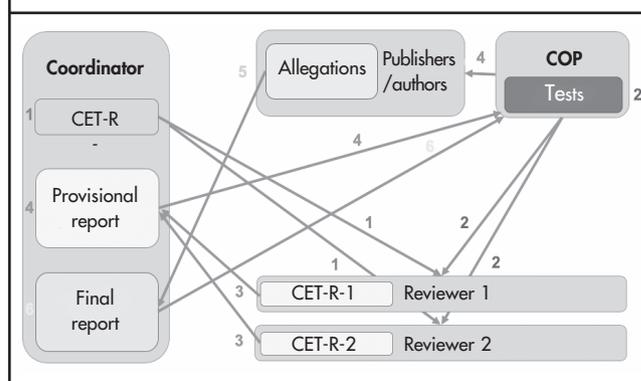


TABLE 2
TESTS EVALUATED IN THE SEVENTH EDITION (2018)

Acronym	Test name	Original author/s (year of publication)	Authors of Spanish adaptation (year of publication)	Publisher
PAIB-1 Renovado [Renewed]	Prueba de Aspectos Instrumentales Básicos en Lenguaje y Matemáticas: E. Infantil (5 años) 1º y 2º E. Primaria [Test of Basic Instrumental Aspects in Language and Mathematics: Infants (5 years old) and 1st and 2nd grade primary]	—	Ramos, Galve, Trallero, & Martínez (2017)	CEPE. Ciencias de la Educación Preescolar y Especial [Educational Science for Preschool and Special Education].
PAIB-2 Renovado [Renewed]	Prueba de Aspectos Instrumentales Básicos en Lenguaje y Matemáticas: 3º y 4º E. Primaria [Test of Basic Instrumental Aspects in Language and Mathematics: 3rd and 4th grade primary]	—	Ramos, Galve, & Martínez (2017)	
PAIB-3 Renovado [Renewed]	Prueba de Aspectos Instrumentales Básicos en Lenguaje y Matemáticas: 5º y 6º E. Primaria y 1º ESO [Test of Basic Instrumental Aspects in Language and Mathematics: 5th and 6th grade. Primary and 1st grade secondary]	—	Ramos, Martínez, & Galve (2017)	
IAES-A	Inventario de Ansiedad Escolar – Abreviado [School-related anxiety Inventory - Brief]	—	García & Inglés (2017)	EOS. Instituto de Orientación Psicológica [Institute of Psychological Orientation]
BYI-2	Inventarios de Beck para niños y adolescentes – 2 [Beck inventories for children and adolescents - 2]	Beck, Beck, Jolly, & Steer (2005)	Hernández, Aguilar, Paradell, & Vallar (2017)	PEARSON Educación
BRIEF-2	Evaluación Conductual de la Función Ejecutiva [Behavioral Evaluation of the Executive Function]	Gioia et al. (2015)	Maldonado et al. (2017)	TEA Ediciones
DP-3	Perfil de Desarrollo – 3 [Development profile]	Alpern (2007)	Sánchez (2018)	
Factor g-R	Test de Inteligencia No Verbal – Revisado [Non verbal intelligence test - Revised]	Cattell, Cattell, & Weiss (2006)	Departamento de I+D+i de TEA Ediciones [Department of R+D+i of TEA Ediciones] (2017)	

TABLE 3
SUMMARY OF THE TEST SCORES EVALUATED IN THE SEVENTH EDITION

	IAES	BYI -2	Factor g-R	BRIEF -2	DP -3	PAIB -1	PAIB -2	PAIB -3	M
Materials and documentation	4	4	4.5	5	5	3.5	3.5	3.5	4.1
Theoretical foundation	4.5	3.5	5	5	4.5	3	3	3	3.9
Adaptation	—	4	4	5	5	—	—	—	4.5
Item analysis	3.5	—	4.5	3	5	3.5	3.5	3.5	3.8
Validity: content	4	3	3	4	4.5	2.5	2.5	2.5	3.3
Validity: relationship with other variables	3	4	4	4.5	4.5	3	3	3.5	3.7
Validity: internal structure	3.5	2	4.5	3.5	4.5	—	—	—	3.6
Validity: DIF analysis	3.5	—	—	—	4.5	—	—	—	4.0
Reliability: equivalence	—	—	—	4.5	—	—	—	—	4.5
Reliability: internal consistency	4	4.5	4.5	5	5	5	5	5	4.8
Reliability: stability	4	4.5	3	4.5	4.5	—	—	—	4.1
Reliability: IRT	—	—	—	—	—	—	—	—	—
Reliability: interrater	—	—	—	—	—	—	—	—	—
Scales and interpretation of scores	4	5	5	4	5	4	4	4	4.4

Note. — = aspect not relevant or no data provided. Nota. — = aspecto no pertinente o sobre el que no se proporcionaron datos.



appropriate to increase this number progressively to convert the inclusion of this relevant criterion in the construction and validation of published tests (Muñiz & Fonseca-Pedrero, 2019).

Finally, with respect to the application of the CET-R, in previous editions the emphasis was placed on the difficulty of interpreting, analyzing, and scoring certain sections, whereas these difficulties have not manifested markedly in this seventh edition, perhaps thanks to the latest revisions and improvements made to the CET-R both in form and content based on the suggestions made in the previous reviews. In spite of not having observed great difficulties in the understanding and scoring of the items, from the application of the CET-R and the comments made by the participants of this edition (experts, editors, authors, and the coordinator), a number of thoughts and suggestions can be noted for the future.

Firstly, it would be useful to evolve from an electronic version of the CET-R to an online version, linked to a database, that would allow us to calculate the numerical valuation (the average of each section) automatically, thus optimizing the use of the new technologies in the evaluation of the tests (Fonseca-Pedrero & Muñiz, 2017; Hidalgo & Hernández, 2019; Muñiz & Fernández-Hermida, 2010). In the same way, regarding the CET-R, it would be useful to review the scoring system to avoid possible biases in the evaluation. As it currently stands, in order to obtain the overall assessment of a section, aspects on which we do not have information are not taken into account, so it is more advantageous not to present any information (—) than to present some information and for it to be inadequate (1) or adequate, but with deficiencies (2). Therefore, it is suggested as a recommendation that categories such as “not applicable” (n/a) and “not provided” (0) be used.

As for the experts, it is worth emphasizing the importance of combining experts in psychometrics with experts in the content of the tests, due to the value of the comments they provide, which, rather than contradicting each other, are usually complementary. In fact, in this edition the high degree of agreement observed among independent reviewers has been remarkable. However, it should be stressed that in the world of peer review we tend to have a certain bias to emphasize the limitations of what we are evaluating and to describe briefly the aspects that we consider appropriate or excellent. In this regard, it is recommended for future editions to emphasize to the experts the need to underline both the limitations and the strengths of the test so that the qualitative comments included in the final reports can be more balanced and more accurate.

Likewise, authors, and publishers are recommended to make specific but detailed allegations, specifying exactly the parts in which they observe disagreements and proposing alternative formulations, especially with the aim of counteracting the potential peer review bias that tends to underline the limitations more. Finally, it should be remembered that the tests are evaluated with the information provided by the publishers that, in this edition as in the previous ones, is limited to the tests and their application manuals. There is, however, the possibility of including, in the materials under evaluation, references or

materials where complementary information can be found that for some reason it has been decided not to include in the manual.

Finally, it is worth emphasizing the impact that the evaluation of the quality of the tests and, more specifically, the use of CET-R is having in the field of psychology (Hidalgo & Hernández, 2019). The tool is proving not only to be essential in the decision making of professional psychologists, but also for university teaching (64% of lecturers who know it use it in their classes), research (many researchers use it as a self-assessment tool and they take it into account in the processes of construction, adaptation, and validation of tests), and for the publishers themselves (who use it not only as a guide but also as a continuous improvement benchmark system).

CONFLICT OF INTERESTS

There is no conflict of interests.

ACKNOWLEDGMENTS

We would like to thank the Test Commission of the Spanish Psychological Association, the publishers, the authors, and the reviewers of the tests evaluated for their collaboration, dedication, and professionalism.

REFERENCES

- Elosua, P., & Geisinger, K. (2016). Cuarta evaluación de tests editados en España: forma y fondo [Fourth review of tests published in Spain]. *Papeles del Psicólogo*, *37*, 82-88.
- Evers, A., Muniz, J., Hagemester, C., Høstmælingen, A., Lindley, P., Sjöberg, A., & Bartram, D. (2013). Assessing the quality of tests: Revision of the EFPA review model. *Psicothema*, *25*, 283-291.
- Fonseca-Pedrero, E., & Muniz, J. (2017). Quinta evaluación de tests editados en España: Mirando hacia atrás, construyendo el futuro [Fifth review of tests published in Spain: Looking back, building the future]. *Papeles del Psicólogo*, *38*, 161-168.
- Hambleton, R.K., Merenda, P.F., & Spielberger, C. (Eds.). (2005). *Adapting educational and psychological tests for cross-cultural assessment*. Mahwah, NJ: Lawrence Erlbaum Publishers.
- Hernandez, A., Ponsoda, V., Muniz, J., Prieto, G., & Elosua, P. (2016). Revisión del modelo para evaluar la calidad de los tests utilizados en España [Assessing the quality of tests in Spain: Revision of the Spanish test review model]. *Papeles del Psicólogo*, *37*, 192-197.
- Hernandez, A., Tomas, I., Ferreres, A., & Lloret, S. (2015). Tercera evaluación de test editados en España [Third evaluation of tests published in Spain]. *Papeles del Psicólogo*, *36*, 1-8.
- Hidalgo, M. D., & Hernández, A. (2019). Sexta evaluación de test editados en España: resultados e impacto del modelo en docentes y editoriales [Sixth review of tests published in Spain: Results and impact of the model on lecturers and publishers]. *Papeles del Psicólogo*, *40*, 21-30.



- International Test Commission (2018). ITC Guidelines for Translating and Adapting Tests (Second Edition). *International Journal of Testing, 18*, 101-134.
- Muniz, J., Elosua, P., & Hambleton, R.K. (2013). Directrices para la traducción y adaptación de los tests: segunda edición [Guidelines for the translation and adaptation of tests: Second edition]. *Psicothema, 25*, 151-157.
- Muniz, J., & Fernandez-Hermida, J.R. (2000). La utilización de los tests en España [The use of tests in Spain]. *Papeles del Psicólogo, 76*, 41-49.
- Muniz, J., & Fernandez-Hermida, J.R. (2010). La opinión de los psicólogos españoles sobre el uso de los tests [The opinion of Spanish psychologists on the use of tests]. *Papeles del Psicólogo, 31*, 108-121.
- Muniz, J., Fernandez-Hermida, J.R., Fonseca-Pedrero, E., Campillo-Alvarez, A., & Pena-Suarez, E. (2011). Evaluación de tests editados en España [Review of tests published in Spain]. *Papeles del Psicólogo, 32*, 113-128.
- Muniz, J., & Fonseca-Pedrero, E. (2019). Diez pasos para la construcción de un test [Ten steps for constructing a test]. *Psicothema, 31*(1), 7-16. doi: 10.7334/psicothema2018.291
- Muniz, J., Hernandez, A., & Ponsoda, V. (2015). Nuevas directrices sobre el uso de los tests: investigación, control de calidad y seguridad [New guidelines for test use: Research, quality control and security of tests]. *Papeles del Psicólogo, 36*, 161-173.
- Ponsoda, V., & Hontangas, P. (2013). Second evaluation of tests published in Spain. *Papeles del Psicólogo, 34*, 82-90.
- Prieto, G., & Muniz, J. (2000). Un modelo para evaluar la calidad de los tests utilizados en España [A model to evaluate the quality of the tests used in Spain]. *Papeles del Psicólogo, 77*, 65-72.