



## FIFTH REVIEW OF TESTS PUBLISHED IN SPAIN: LOOKING BACK, BUILDING THE FUTURE

Eduardo Fonseca-Pedrero<sup>1</sup> and José Muñiz<sup>2</sup>

<sup>1</sup>Universidad de La Rioja. <sup>2</sup>Universidad de Oviedo

*La psicología como ciencia necesita de una continua y rigurosa evaluación tanto de sus métodos y procedimientos como de sus técnicas e instrumentos. Los tests, una de las principales herramientas que utilizan los psicólogos en su quehacer diario, no pueden ser una excepción. Se hace necesario someter a escrutinio los instrumentos de evaluación psicológica mediante un proceso de revisión sistemático, riguroso e independiente que garantice su calidad y utilidad. Con este fin, desde el año 2010 la Comisión Nacional de Tests del Colegio Oficial de Psicólogos viene revisando anualmente algunos de los tests más utilizados por los psicólogos españoles. Tal proceso de revisión se basa en un modelo europeo de evaluación de la calidad de los tests adaptado al español y que se operativiza mediante el Cuestionario de Evaluación de Tests-Revisado. En este trabajo se presenta la quinta evaluación de tests editados en España. Se evaluaron un total de once tests, participaron un total de cuatro casas editoriales y 24 expertos en un proceso de revisión por pares. En líneas generales puede afirmarse que la calidad de los tests evaluados fue adecuada, si bien se han encontrado algunas fortalezas y debilidades. Finalmente se comentan las principales conclusiones y se analizan las posibles líneas de futuro en la evaluación de los tests en España.*

**Palabras clave.** Tests, Uso de los tests, Evaluación de tests, Psicometría, Calidad de los tests, Medición.

*Psychology, as a science, needs the continuous and rigorous evaluation of its methods and procedures as well as its techniques and instruments. The test is one of the main tools used by psychologists in their daily work, and cannot be an exception. The measuring instruments of psychological evaluation must be analyzed using a systematic, rigorous and independent process that guarantees their quality and usefulness. To this end, since 2010 the National Tests Commission of the Spanish Psychological Association has reviewed annually some of the tests most used by Spanish psychologists. This review process is based on a European test review model that has been adapted to Spanish and it is performed through the Tests Review Questionnaire-Revised. In the present work the fifth review of tests published in Spain is presented. A total of eleven tests were evaluated, with the participation of four publishing houses, and 24 experts in a peer review process. In general terms, the quality of the measurement instruments evaluated in this fifth review was adequate, although there are some strengths and weaknesses. Finally the main conclusions are discussed and the possible future lines of research in the evaluation of tests in Spain are analyzed.*

**Keywords.** Tests, Use of tests, Evaluation of tests, Psychometrics, Quality of the tests, Measurement.

**P** sychology as a science needs the continuous and rigorous evaluation of its methods, procedures, techniques and instruments. The test, one of the main tools used by psychologists in their professional practice, cannot be an exception. Based on tests, decisions are made that affect the lives of individuals and organizations, and have a clear impact on society. Therefore, tests must be subjected to scrutiny through a systematic, rigorous and independent review process that guarantees their quality and usefulness. As Muñiz et al. (2011) indicate, the correct use of tests requires, on the one hand, instruments with adequate psychometric properties (e.g., reliability of scores, validity evidence, etc.), and, on the other hand, appropriate training to ensure their correct use by the

professionals. The tests must have empirically demonstrated quality and rigor. In addition, psychologists must be competent and they must have verified information that allows them to select the most suitable tests for their purpose (Hernández, Ponsoda, Muñiz, Prieto, & Elosua, 2016). All of this is reflected in the Code of Ethics of the Psychologist and in the different guidelines developed for this purpose both nationally and internationally (Muñiz, 1997; Muñiz & Bartram, 2007). For example, Article 17 of the Code of Ethics mentions that psychologists must be professionally trained and specialized in the use of methods, instruments, techniques and procedures that they adopt in their work and they must recognize the limits of their competence and the limitations of the techniques.

The Spanish Psychological Association (COP) has reviewed the tests published in Spain annually since the year 2010 (Elosua & Geisinger, 2016; Hernández, Tomás, Ferreres, & Lloret, 2015; Muñiz et al., 2011; Ponsoda & Hontangas, 2013). Hernández et al. (2016) indicate that the ultimate objective of assessing the

Received: 10 mayo 2017 - Accepted: 30 agosto 2017  
Correspondence: Eduardo Fonseca-Pedrero. Departamento de Ciencias de la Educación. Universidad de La Rioja. C/ Luis de Ulloa, 2, Edificio VIVES. 26004 Logroño, La Rioja. España.  
E-mail: eduardo.fonseca@unirioja.es



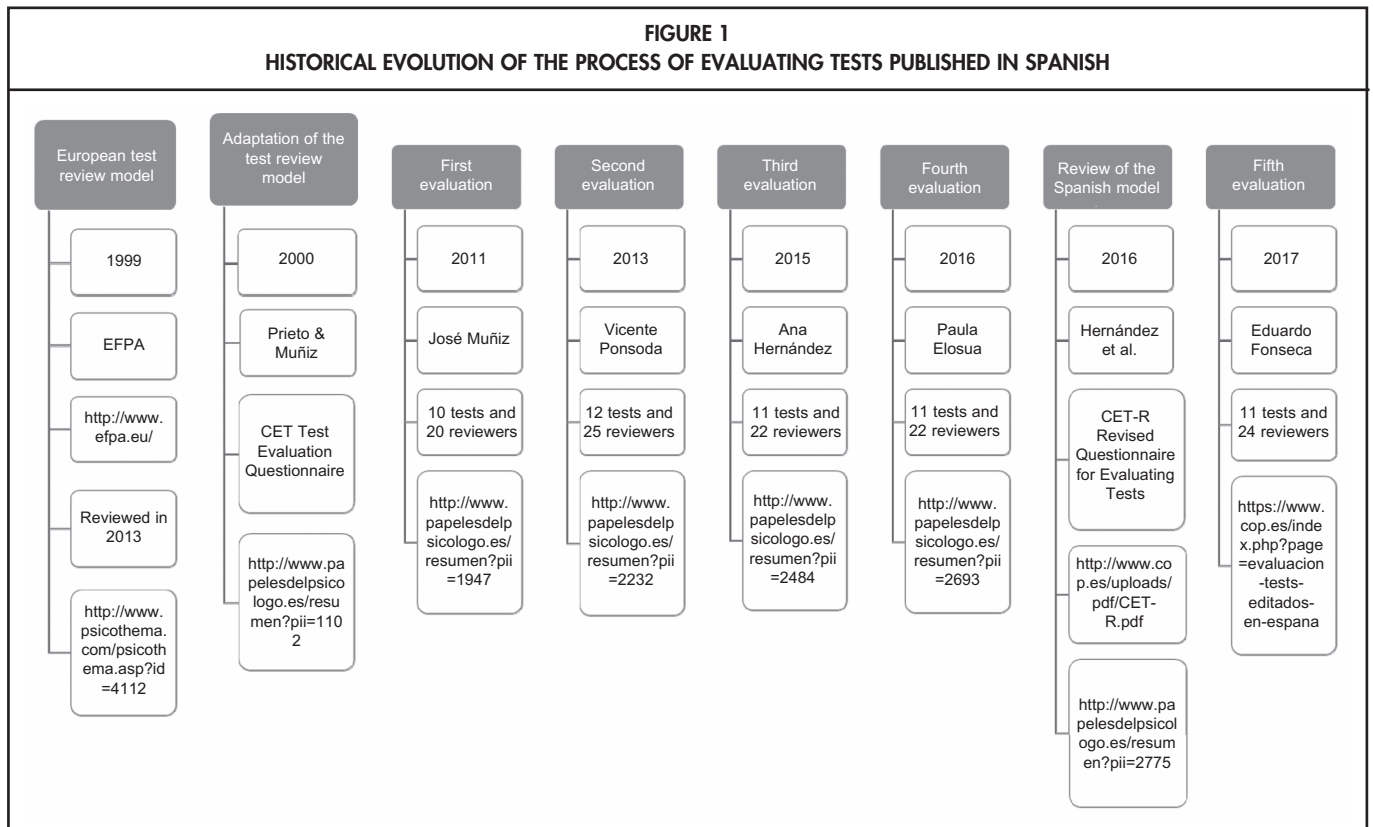
quality of tests used in Spain is to provide users of psychological measurement and evaluation instruments with accurate and accessible information on the quality of the tests available. The fifth edition has just been completed. The main objective of this article is to present the main results of this most recent review of the tests published in Spanish. Firstly, a brief historical outline is provided that examines the main milestones in the process of reviewing the tests published in our country. The aim is to offer an overview of the current state of the question. Next, the evaluation process and the main results of the fifth test review are described in detail. Thirdly, and finally, the main conclusions and future lines of investigation are discussed.

**TEST REVIEWS IN SPAIN: THE ROAD TRAVELED**

When psychology practitioners are asked about the main demands regarding the use of tests, they mention, among other things, the need to provide technical and psychometric information about the tests that will help them to make informed decisions (Muñiz & Fernández-Hermida, 2010; Evers et al., 2017). To answer these and other questions, the Standing Committee on Tests and Testing of the European Federation of Psychologists' Associations (EFPA, <http://www.efpa.eu/>) developed a test review model that was adapted into Spanish by Prieto and Muñiz (2000). In 2010, the National Test Commission of the COP decided to launch the first test review,

the results of which were published in 2011 (Muñiz et al., 2011). Subsequently, almost annually, a new test review has been carried out. The historical evolution of the evaluation process of tests published in Spain as well as the relevant information from each annual review is shown in Figure 1. All of the tests evaluated in the five reviews are presented in Figure 2. Likewise, the results of each and every review carried out to date are available on the COP website: <https://www.cop.es/index.php?page=evaluacion-tests-editados-en-espana>. Practitioners are invited to visit and analyze the tests evaluated. As can be seen in Figures 1 and 2, a total of 55 tests from different publishing houses have been reviewed to date, with the participation of more than 100 reviewers and five coordinators. It is worth mentioning that the number of experts reviewing the tests is not the arithmetic sum of the five evaluations, since in a number of exceptions some of them have participated in several editions. It should also be noted that practitioners use these reports, which are available on the COP website. The number of downloads of these reports may be an indicator of this. An analysis of the volume of downloads from the year 2012 to August 2017 has been carried out. (The information is not available for the previous years.) The results of the number of downloads are as follows: Adaptive Behavior Assessment System n = 105, Brief Symptoms Inventory-18 n = 101, Wechsler Memory Scale-IV n = 65,

**FIGURE 1**  
**HISTORICAL EVOLUTION OF THE PROCESS OF EVALUATING TESTS PUBLISHED IN SPANISH**





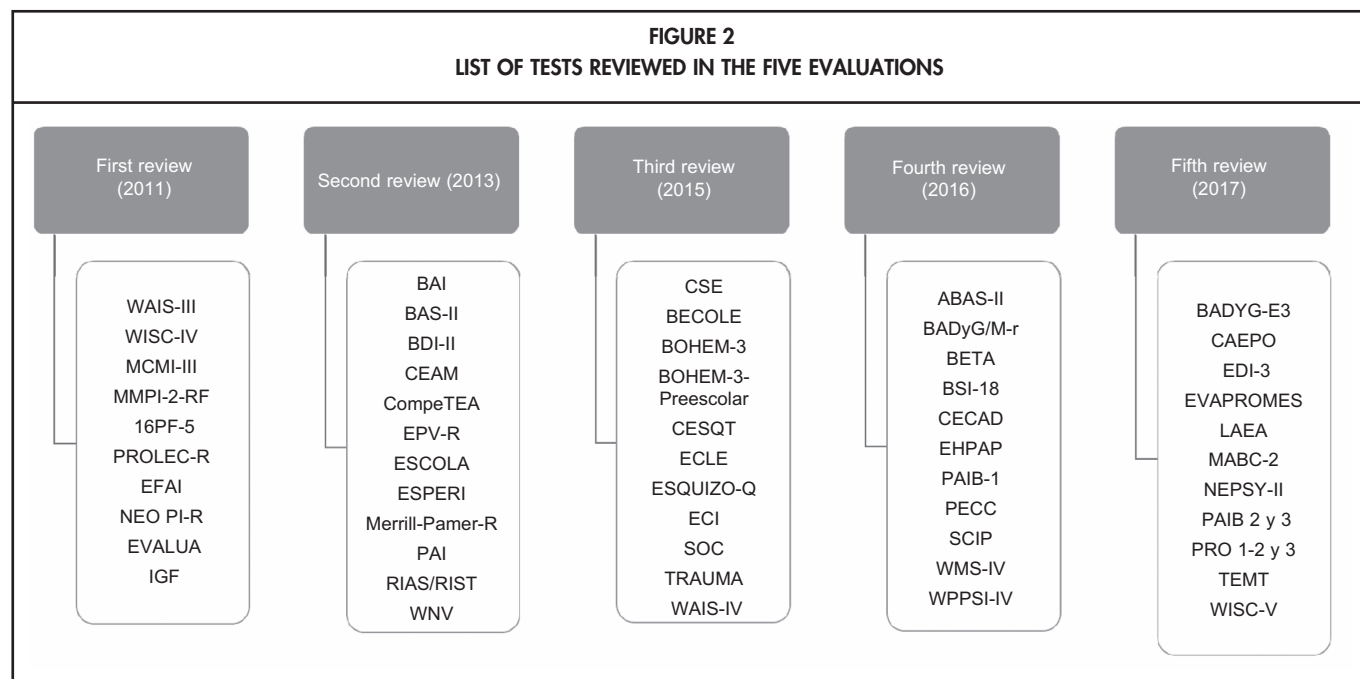
Boehm Test of Basic Concepts-3 Preschool n = 101, *Batería de Evaluación Cognitiva de la Lectura y Escritura* [Cognitive Evaluation of Reading and Writing Battery] n = 97, Oviedo Questionnaire for Schizotypy Assessment n = 94, Beck Anxiety Inventory n = 111, Beck-II Depression Inventory n = 104, *Cuestionario para la Detección de los Trastornos del Comportamiento en Niños y Adolescentes* [Questionnaire for the Detection of Behavioral Disorders in Children and Adolescents] n = 105, Minnesota Multiphasic Personality Inventory-2 Restructured n = 164, Wechsler Intelligence Scale for Children-IV n = 134 and 16 Personality Factors, fifth edition n = 119.

The model developed by the EFPA for the evaluation of the quality of the tests is operationalized using a measurement instrument called the Test Review Questionnaire (CET, in Spanish) (Muñiz et al., 2011). The CET allows a qualitative and quantitative evaluation of the test that is being reviewed, that is, it aims to examine its quality and usefulness. The CET can be consulted in the work of Muñiz et al. (2011). Subsequently, in the year 2013, the EFPA reviewed this evaluation model (Evers et al., 2013). Changes in the model coupled with changes suggested in the various annual evaluations by the coordinators and reviewers as well as the new guidelines of *Standards* (American Educational Research Association, American Psychological Association, & National Council on Measurement in Education, 2014) (for example, those related to obtaining validity evidence and taxonomy) have led to the modification of the CET, resulting in the CET-Revised (CET-R). These changes have been included in the present fifth review of tests. For a more detailed analysis of the changes introduced, see the

excellent work of Hernández et al. (2016). In addition, the CET-R can be freely downloaded from the following link: <http://www.cop.es/uploads/pdf/CET-R.pdf>. The CET-R, like its previous version, consists of three extensive sections: a) a general description of the test; b) an evaluation of the characteristics of the test (e.g., reliability, validity evidence, etc.); and c) a global assessment of the test. It is an interesting tool to be used by all psychology professionals not only for practical or professional purposes but also for training purposes. Moreover, the CET-R can be used as a guideline or standard, contributing to improve the process of constructing and editing the tests published in Spain, an aspect that in turn affects the decision-making capacity of the psychology practitioners.

The test review process begins with the COP's National Test Commission designating a coordinator for that year's review. The commission in coordination with the editors also nominates the tests that are to be reviewed in that year. Once the tests have been selected (between 10 and 12, depending on the year), the coordinator selects a group of experts in the subject who review the tests through a process of peer evaluation (similar to that used in reviewing scientific articles). The selection of these reviewers is always carried out according to scientific criteria and standards, which, with occasional modifications, have remained unchanged throughout the successive evaluations. Two reviewers are selected, when circumstances permit it, one of them with a more technical-psychometric profile, and another with a profile based on more substantive issues. Their anonymity is guaranteed in relation to the test they are to evaluate, their participation is requested, and the absence of a conflict of interest confirmed. In the great majority of cases, and except for

**FIGURE 2**  
**LIST OF TESTS REVIEWED IN THE FIVE EVALUATIONS**



reasons of force majeure, the reviewers happily accept the task with which they have been entrusted. All the reviewers are given a copy of the test they are to evaluate (provided free of charge by the editors) as well as a symbolic honorarium of 50 euros. Each reviewer produces an independent report on the test and sends it to the coordinator within three months. Obviously during the whole process the coordinator is open to answer any questions or suggestions from both the reviewers and the editors. The coordinator then analyzes the comments and scores from the two reviewers and draws up a first final report. This report is sent to the respective publishing houses who, within a period of approximately one month, respond with reasoned explanations, where applicable, to the comments and/or scores they deem appropriate. In this way all of the parties involved are part of the process. Lastly, with these comments, the coordinator produces a final report for each test that is sent to the National Test Commission and is subsequently uploaded to the website for consultation (see link above). Two important aspects must be emphasized. Firstly, the process of evaluating the tests is completely independent of the editorials and authors of the tests, as well as of the COP; it only depends on the reviewers and the review coordinator. As with any process of peer evaluation, it is not perfect; nothing human is, but we believe it is as good as possible. Secondly, the evaluation of the tests does not constitute a certification of the tests by the COP's Test Commission; it is a review that highlights the strengths and weaknesses of the tests as deemed by the evaluators at a given time, following the CET-R test review model. The philosophy guiding the test reviews is not one of test censorship. It is a question of helping editors and authors to improve their tests, providing the validity evidence that guarantees the inferences made from the scores. As the great physicist Lord Kelvin points out, if you cannot measure something, you cannot improve it.

## FIFTH EVALUATION OF TESTS

This section discusses the evaluation process, the reviewed tests, the participating reviewers and the main results of the fifth review of tests published in Spain.

### *Review process*

The general process that has been followed in this review has been mentioned above, although we would like to briefly comment here on the specific aspects concerning the fifth test review. After the tests to be evaluated had been nominated, the coordinator (the first author of this work) selected the experts to review the tests and sent them a formal invitation by email (January 2016). Only two of the experts selected were unable to participate, due to completely understandable reasons, so two new reviewers were selected. Subsequently, the materials were sent. These included the review model (CET-R), instructions for its completion and the test to be evaluated (documentation, CD, correction sheets, etc.) (March/April 2016). Throughout this process, all of the difficulties and issues that came up were

solved. (For example, anecdotally, one of the tests was sent to the US, but when the reviewer received it, the test had been opened because it had been examined at customs.) Between the months of July and September 2016, the reviewers sent the evaluations to the coordinator. The preliminary report was then sent to publishers who responded efficaciously between November and December 2016. Finally, between December 2016 and January 2017, the final reports were prepared and then sent to the COP and posted on the above-mentioned website.

### *Selected tests and reviewers*

The tests selected in the fifth review are listed in Table 1. As can be seen, there were a total of 11 tests from four publishing houses (CEPE, EOS, PEARSON, and TEA). The original list of tests selected by the Test Commission for this fifth review underwent two slight modifications that are discussed below. First, 12 tests were originally chosen. The Personality and Preference Inventory (PAPI), marketed in Spain by FACTHUM and internationally owned by CUBIKS Limited, was initially included, although the director of the publishing house CUBIKS Limited explained in a well-reasoned letter to the COP that PAPI was in the middle of an updating process, so it would be necessary to postpone its evaluation for future editions. Second, the Verbal Intelligence Test-E2 (INVE-E2) of CEPE was also selected, but as it was also in the process of being updated, a unanimous decision was made to replace it with the PAIB 2 and 3 battery. All of these changes were agreed both by the COP's National Test Commission and by the respective editors.

Table 2 lists the 24 reviewers who participated in this fifth test review. As mentioned previously, two reviewers were selected for each test, although in two specific cases where clear divergences were found between the reviewers, a third party was selected who helped the coordinator to prepare the final report on the respective test. The selection of reviewers was carried out considering, as far as possible and as in previous evaluations, the following criteria: a) the absence of conflict of interest or other ethical issues; b) the expansion of the number of institutions participating in the national test review (e.g. health and education departments, clinical psychologists, foreign assessment and measurement bodies, etc.); c) not having participated in previous evaluations (whenever possible); and d) young researchers or practitioners of psychology. The aim was to open the range of reviewers and experts outside of academia and beyond the Spanish context, which sought, in essence, to publicize and extend the use of the Spanish test review model among psychology practitioners. We wish to express our sincere thanks for the good work carried out by all of the reviewers and publishers who participated in this fifth evaluation. This process would not have come to fruition without their valuable help and collaboration.



**Results**

Table 3 shows the main results of the fifth test review in Spain. As can be seen, the scores were, in most cases, higher than 3.5 (good to excellent). In only four cases out of all of the scores, were they less than 3. As strengths, it should be mentioned that the tests reviewed had, in general terms, excellent quality materials and documentation, adequate levels of reliability and excellent scales, in addition to sufficient evidence of validity. As for the limitations or weaknesses, these could include the lack of studies on differential item functioning in some of the tests analyzed or the total absence of psychometric analysis under the prism of Item Response Theory (IRT).

There are two aspects that must be mentioned: a) the difficulty in interpreting, analyzing and scoring certain items of the CET-R, such as, for example, those related to item analysis or criterion validity; and (b) in two specific cases, substantial differences were found between the scores and comments sent to the coordinator by the two reviewers, so a third opinion had to be sought. Regarding the first point, similar results have been found in previous reviews (Elosua & Geisinger, 2016; Hernández Baeza et al., 2015; Muñoz et al., 2011; Ponsoda & Hontangas, 2013). In this sense some of the items in the CET-R seem to be difficult for the experts to evaluate. In previous evaluations, with the purpose of correcting this limitation, it was decided to incorporate general guidelines for the use of CET/CET-R as well as a longer explanation for some of the items. However, it appears that this difficulty continues in the fifth review. Regarding the second point (partially related to the previous one), there is no doubt that the peer review process is one of many possible ways of evaluating the quality of tests, and it has its pros and cons. In the world of peer reviewing, whether of scientific articles, projects, contracts, or in our case tests, it is frequent to find discrepancies between the reviewers, so we have to resort to a third evaluator and even to several rounds of evaluation. This is intrinsic to the review process and should not necessarily be interpreted as a negative thing. In future editions, it may be necessary to further improve the understanding of some of the items in the CET-R, which could be solved, for example, by developing video tutorials, specific examples or incorporating more technical-psychometric information into the CET-R.

**SOME CONCLUSIONS**

After a lustrum of reviews of the tests published in Spain, it can be said that the process is being consolidated and it is in good health. Nevertheless, we are still in the initial stages, even more so if we compare ourselves with other nations, for example the Dutch, who have reviewed all of the tests published in the country (Evers, 2012), or the United States, where the specialized organization BUROS has already evaluated more than 10,500 tests (Elosua & Geisinger, 2016). So far in Spain, a total of 55 tests have been reviewed from different publishing houses, with the participation of more than 100 reviewers and

**TABLE 1  
LIST OF MEASURING INSTRUMENTS ANALYZED IN  
THE FIFTH TEST REVIEW**

Acronym	Test	Publisher
NEPSY-II	NEPSY-II Neuropsychological Battery for Children	Pearson Education
WISC-V	Wechsler Intelligence Scale for Children-V	Pearson Education
MABC-2	Movement Assessment Battery for Children-2	Pearson Education
TEMT	Utrecht Early Numeracy Test	EOS
EVAPROMES	Evaluation of the Metacognitive Processes in Writing	EOS
LAEA	List of Adjectives for the Evaluation of Self-concept	TEA ediciones
EDI-3	Eating Disorders Inventory -3	TEA ediciones
CAEPO	Questionnaire on Coping with Stress for Oncology Patients	TEA ediciones
BADYG-E3	Battery of Differential and General Skills Updated E3	CEPE, S.L.
PRO 1-2 & 3	Spelling Ability Primary (1-2) and Secondary (3)	CEPE, S.L.
PAIB 2 & 3	Test of Basic Instrumental Aspects in Language and Mathematics (2 & 3)	CEPE, S.L.

**TABLE 2  
REVIEWERS WHO PARTICIPATED IN THE  
FIFTH TEST REVIEW**

Name	Affiliation
Carmen García García	Autonomous University of Madrid
Javier Suárez Álvarez	University of Oviedo
Beatriz Lucas Molina	University of Valencia
Alicia Pérez de Albéniz	University of La Rioja
Javier Ortuño-Sierra	University Loyola Seville
Félix Inchausti Gómez	Ministry of Health of Navarre
Jorge López Puga	Catholic University of Murcia
Ana Vanesa Valero	University of La Rioja
Irene Solbes Canales	Complutense University of Madrid
Joan Guàrdia Olmos	University of Barcelona
David Álvarez García	University of Oviedo
Ricardo Olmos Albacete	Autonomous University of Madrid
Isabel Benítez Baena	University Loyola Seville
Susana Al-Halabi Díaz	University of Oviedo
Gloria García Fernández	Complutense University of Madrid
Susana Sierra Baigrie	University of Oviedo
Cristina Anguiano Carrasco	Educational Testing Service (ETS)
Mercedes Paino Piñero	University of Oviedo
Roberto Lozano Herce	Ministry of Education of La Rioja
Silvia Sastre i Riba	University of La Rioja
Juana Gómez Benito	University of Barcelona
Joan Pere Ferrando	University of Rovira i Virgili
Nekane Balluerka Lasa	University of the Basque Country
Maite Garaigordobil Landazabal	University of the Basque Country



five coordinators. The sixth review has already begun, coordinated by Professor María Dolores Hidalgo Montesinos of the University of Murcia. In general terms, the evaluation of the quality of the tests through CET/CET-R in the last five reviews has provided results that demonstrate that our tests are of reasonably good quality, with clear strengths but also aspects to improve upon.

The incorporation of the CET-R in this fifth review has been entirely adequate. In addition, the results found both quantitatively and qualitatively have been fully satisfactory, although it is true that new data on the study of the differential item functioning or measurement invariance must be incorporated, as well as new studies from the perspective of IRT that add information, among other aspects, on the characteristic curves of the items or the information functions of the tests. It should be remembered that the study of differential item functioning, ensures, to a certain extent, equity in the measurement process. The absence of differential functioning in an item assumes that the probability of correct response depends only on the level of the participant in the variable being measured, and is not conditioned by membership of a group or characteristic (e.g. gender, culture, ethnicity, etc.). In short, we must ensure that the evaluation is carried out with equity and fairly. The authors of this paper are fully aware that researchers, psychology professionals and publishers are all working intensively on these and other lines of interest.

The CET-R appears to be a useful tool for improving the use of tests and the professional practice of psychologists. As Elosua and Geisinger (2016) point out, the publication of independent reviews subject to standardized procedures and scientific

criteria are a boost to the continuous process of improving the constructing/editing/use of tests. Furthermore, the CET-R is a tool that could be used as a gold standard, obviously whilst also recognizing its limitations, to help in the training of future psychology professionals as well as in retraining courses, and the constructing/publishing of tests. The training and awareness of professionals and different organizations is necessary since the (in)correct use of tests has a clear impact on the lives of the people or organizations that are the object of evaluation and measurement. The correct use of tests, in the numerous areas in which they are used, will undoubtedly improve professional practice, and of course, the vision of psychology as a science and a profession.

The test review process is laborious and complex, as well as enriching. This process of evaluating the tests published in Spain is framed within a broader set of organizations and international guidelines for the good use of tests. For example, as Elosua (2017) points out, the International Test Commission (ITC) has developed six sets of guidelines: (a) the translation and adaptation of tests; b) the use of tests; c) computerized tests and tests administered over the internet; d) the security of tests, examinations and other evaluations; e) quality control of test scores, their analysis and score reports; and (f) the professional use of test reviews, obsolete tests and test withdrawals. Many of these have been translated into Spanish by members of the COP National Test Commission (Hernández et al., 2016; Muñiz, Elosua, & Hambleton, 2013). For more detailed information, we invite you to visit the guidelines section of the ITC website (<https://www.intestcom.org/>) as well as the excellent work carried out previously (Leong, Bartram, Cheung, Geisinger, & Iliescu, 2016; Wells & Faulkner-Bond, 2016).

**TABLE 3**  
**SUMMARY OF THE QUALIFICATIONS OF THE TESTS ANALYZED IN THE FIFTH REVIEW**

Characteristics	Tests										
	NEPSY-II	WISC-V	MABC-2	EVAPROMES	TEMT	LAEA	EDI-3	CAEPO	BADYG-E3	PRO 1-2 & 3	PAIB 2 & 3
Materials and documentation	5	5	5	4	5	4	5	3.5	4.5	4	4
Theoretical foundation	5	5	5	4.5	5	5	5	3.5	5	5	4
Adaptation	4	4.5	5	—	5	—	5	—	—	—	—
Analysis of items	—	4.5	4	4	4.5	4	4	—	5	4	4
Validity: content	4	5	4	4	4	4	4.5	—	5	3.5	4
Validity: relationship with other variables	3.5	4.5	4	4.5	3.5	4	4	2.5	3.5	2.5	3
Validity: internal structure	—	4.5	4.5	4	3.5	4	5	3	4	3	2
Validity: DIF analysis	—	—	3	5	—	—	4	—	—	—	—
Reliability: equivalence	—	—	—	—	3	—	—	—	—	—	—
Reliability: internal consistency	4	5	4	5	4.5	4.5	4	4.5	5	4	4
Reliability: stability	3.5	4	3	—	3	5	3.5	—	—	—	—
Reliability: IRT	—	—	—	—	—	—	—	—	—	—	—
Reliability: inter-rater	5	5	5	—	—	—	—	—	—	—	—
Scales and interpretation of scores	4	5	4.5	4	4.5	5	4.5	2.5	5	3.5	3.5

Note: The scores in the table are on a scale of 1 to 5, and correspond to the following ratings: 1 = inadequate; 2 = adequate but with deficiencies; from 2.5 = adequate; from 3.5 = good; from 4.5 = excellent. When the symbol (—) appears, it means that no information is given or it is not applicable.



## BUILDING THE FUTURE

In recent years, there have been very clear advances in psychological assessment and measurement (Fonseca-Pedrero & Muñiz, 2016), and everything suggests that the future will bring us many more. Some future lines of research concerning the evaluation of the quality of the tests are mentioned below. These are inserted, obligatorily, within the more general framework of psychological and educational evaluation.

Firstly, the reviewer will have to be evaluated, that is, on the one hand we will have to estimate the psychometric properties of the CET-R evaluation model, and on the other, the inter-rater reliability of the test reviewers. These are tasks that are still pending. In addition, it would be very convenient to develop a web version of the CET-R that could be used by professionals.

Secondly, and given that psychology as a science and profession is constantly evolving, it is extremely important that the test quality review model (CET-R) is quickly adapted to the most remarkable events in the area of psychological and educational assessment and measurement. For example, the new forms of evaluation that are brought to us by new technologies are of great interest. The major change in the field of technology used in evaluation is the gradual shift from traditional paper and pencil tests to the use of computer technology. Authors such as Drasgow (2016) or Sireci and Faulkner-Bond (2016) indicate that new technologies are influencing all aspects of psychological assessment, such as test design, item construction, item presentation, the scoring of the tests and distance evaluation, to name a few. We will mention only two emerging forms of evaluation: computerized adaptive tests (CATs) and outpatient evaluation.

Although CATs have been used for many years, they have not yet become widespread, so they remain an interesting future line of research (Olea, Abad, & Barrada, 2010, Zenisky & Luecht, 2016). In Spain, we are currently working on several projects, for example, the evaluation of intelligence from the RAVEN matrix model, the evaluation of the organizational climate in work contexts and the evaluation of the entrepreneurial personality in young people.

Outpatient evaluation is a classic subject of study in psychology, although it is resurging with renewed energy today (Trull & Ebner-Priemer, 2009, 2013; van Os, Delespaul, Wigman, Myin-Germeys, & Wichers, 2013). Outpatient assessment includes a wide range of evaluation methods that seek to study the experiences of people in their natural environment and in daily life. Due to certain limitations of self-report tests (e.g., retrospective bias, artificial context of evaluation, etc.), recent research proposes to replace, or at least to incorporate, new forms of evaluation, that allow us to understand certain psychological variables and constructs from a more dynamic, personalized, contextual and ecological perspective. The aim is to analyze, using structured records of questions, the behaviors, feelings, cognitions, emotions and/or symptoms of people through mobile devices in their real daily

context. For this purpose, assessments are performed several times a day over a period of time (typically one week) to sufficiently capture the variability of the phenomena. The questions are activated by a beep within a time frame set by the researcher, for example, between 10 am and 10 pm. In addition, these beeps may occur randomly and/or at predetermined time intervals, for example, every 90 minutes. Different behavior samples are collected throughout each day, approximately 6-8 per day for 7 days. All of these data are uploaded to a platform for further analysis. It is, therefore, a complementary approach to the traditional methods of psychometric evaluation based on pencil and paper in more or less artificial contexts and it is more cross-cutting and retrospective.

Thirdly, there is the possibility of incorporating into the CET-R new psychometric models such as network analysis (Borsboom & Cramer, 2013; Fonseca-Pedrero, 2017), multidimensional models within IRT (Reckase, 2009) or other mathematical models from dynamic systems theory or chaos theory (Nelson, McGorry, Wichers, Wigman, & Hartmann, 2017) to name a few.

Finally, the continuous training of psychology professionals is the cornerstone for the coming years, as it will become more frequent and unavoidable to have knowledge of assessment techniques and procedures that go beyond the mere administration of tests or interviews. The development of neuroscience and robotics as well as translational and multidisciplinary approaches in science are clear examples.

An atmosphere of change is perceived among psychology professionals. Many of the new lines of research discussed here will undoubtedly be consolidated in the years to come. The road ahead is undeniably fascinating.

## ACKNOWLEDGMENTS

We would like to express our sincere thanks to the members of the National Test Commission and to the administrative staff of the COP, as well as to the reviewers and the different publishing houses, without whose help and collaboration this work could not have been carried out.

## REFERENCES

- American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (2014). *Standards for Educational and Psychological Testing*. Washington, D.C.: Author.
- Borsboom, D., & Cramer, A. O. (2013). Network analysis: an integrative approach to the structure of psychopathology. *Annual Review of Clinical Psychology, 9*, 91–121.
- Drasgow, F. (Ed.). (2016). *Technology and testing*. New York: Routledge.
- Elosua, P. (2017). Avances, proyectos y retos internacionales ligados al uso de tests en Psicología [International advances,



- projects and challenges related to the use of tests in Psychology]. *Estudios de Psicología*, 34, 201–210.
- Elosua, P., & Geisinger, K. F. (2016). Cuarta evaluación de tests editados en España: Forma y fondo [Fourth review of tests published in Spain: Form and content]. *Papeles del Psicólogo*, 37, 82–88.
- Evers, A. (2012). The internationalization of test reviewing: Trends, differences, and results. *International Journal of Testing*, 12, 136–156.
- Evers, A., McCormick, C., Hawley, L., Muñiz, J. et al. (2017). Testing practices and attitudes toward tests and testing: An international survey. *International Journal of Testing*, 17, 158–190.
- Evers, A., Muñiz, J., Hagemeister, C., Høtmælingen, A., Lindley, P., Sjöberg, A., & Bartram, D. (2013). Assessing the quality of tests: revision of the EFPA review model. *Psicothema*, 25, 283–91.
- Fonseca-Pedrero, E. (2017). Análisis de redes: ¿una nueva forma de comprender la psicopatología? [Network analysis: a new way of understanding psychopathology?] *Revista de Psiquiatría y Salud Mental*, <https://doi.org/10.1016/j.rpsm.2017.06.004>.
- Fonseca-Pedrero, E., & Muñiz, J. (2016). Avances en evaluación psicológica [Advances in psychological assessment]. *Papeles del Psicólogo*, 37, 2–2.
- Hernández, A., Ponsoda, V., Muñiz, 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, 161–168.
- Hernández Baeza, A., Tomás, I., Ferreres, A., & Lloret, S. (2015). Tercera evaluación de tests editados en España [Third evaluation of tests published in Spain]. *Papeles del Psicólogo*, 36, 1–8.
- Leong, F. T. L., Bartram, D., Cheung, F. M., Geisinger, K., & Iliescu, C. (2016). *The ITC International Handbook of Testing and Assessment*. New York: Oxford University Press.
- Muñiz, J. (1997). Aspectos éticos y deontológicos de la evaluación psicológica [Ethical and deontological aspects of psychological assessment]. In A. Cordero (Coord.), *La Evaluación psicológica en el año 2000 [Psychological assessment in the year 2000]*. Madrid: TEA Ediciones. (pp. 307–345).
- Muñiz, J. & Bartram, D. (2007). Improving international tests and testing. *European Psychologist*, 12(3), 206–219.
- Muñiz, 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.
- Muñiz, J., & Fernández-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.
- Muñiz, J., Fernández-Hermida, J. R., Fonseca-Pedrero, E., Campillo-Álvarez, Á., & Peña-Suárez, E. (2011). Evaluación de los tests editados en España [An evaluation of the tests published in Spain]. *Papeles del Psicólogo*, 32, 113–128.
- Nelson, B., McGorry, P. D., Wichers, M., Wigman, J. T. W., & Hartmann, J. A. (2017). Moving from static to dynamic models of the onset of mental disorder. *JAMA Psychiatry*, 74, 528–534.
- Olea, J., Abad, F., & Barrada, J. R. (2010). Tests informatizados y otros nuevos tipos de tests [Computerized tests and other new types of test]. *Papeles del Psicólogo*, 31, 94–107.
- Ponsoda, V., & Hontangas, P. (2013). Segunda evaluación de tests editados en España [Second evaluation of tests published in Spain]. *Papeles del Psicólogo*, 34, 82–90.
- Prieto, G., & Muñiz, J. (2000). Un modelo para evaluar la calidad de los tests utilizados en España [A model for evaluating the quality of the tests used in Spain]. *Papeles del Psicólogo*, 77, 65–77.
- Reckase, M. D. (2009). *Multidimensional Item Response Theory. Multidimensional Item Response Theory*. New York, NY: Springer-Verlag New York.
- Trull, T. J., & Ebner-Priemer, U. W. (2009). Using experience sampling methods/ecological momentary assessment (ESM/EMA) in clinical assessment and clinical research: Introduction to the special section. *Psychological Assessment*, 21, 457–462.
- Trull, T. J., & Ebner-Priemer, U. W. (2013). Ambulatory assessment. *Annual Review of Clinical Psychology*, 9, 151–176.
- van Os, J., Delespaul, P., Wigman, J., Myin-Germeys, I., & Wichers, M. (2013). Beyond DSM and ICD: introducing “precision diagnosis” for psychiatry using momentary assessment technology. *World Psychiatry*, 12, 113–117.
- Wells, C. S., & Faulkner-Bond, M. (2016). *Educational Measurement. From Foundations to Future*. New York, NY: The Guilford Press.
- Zenisky, A. L., & Luecht, R. M. (2016). The future of computer-based testing: Some new paradigms. In C. S. Wells and M. Faulkner-Bond (Eds.), *Educational measurement. From foundations to future* (pp. 221–238). New York: Guilford Press.