



FOURTH REVIEW OF TESTS PUBLISHED IN SPAIN: FORM AND CONTENT

Paula Elosua¹ and Kurt. F. Geisinger²

¹Universidad del País Vasco. ²Buros Center for Testing

La evaluación de tests y divulgación de las memorias resultantes se ha implantando en el panorama internacional con el claro objetivo de mejorar el uso de los tests, y con ello la práctica profesional. Se trata de una herramienta que ofrece al usuario de tests una revisión independiente sujeta a procedimientos y criterios definidos. En España se han llevado a cabo cuatro ediciones del proyecto de evaluación liderado por el Consejo General de la Psicología (COP). Básicamente existen dos modelos de evaluación de tests, el modelo Buros y el modelo de la European Federation Psychologists' Associations (EFPA), utilizado por el COP. Ambos, comparten objetivos y contenido, y excluyendo varias cuestiones de formato y gestión, el fondo en ambos modelos es el mismo. En este trabajo se ofrecen los resultados de la cuarta edición del proyecto de evaluación de tests del COP adecuándolos a las pautas generales que sigue un procedimiento de evaluación de tests, y mostrando los puntos críticos en cada fase del proceso.

Palabras clave: Tests, Uso de tests, Evaluación de tests, Psicometría.

Assessing the quality of tests and disseminating the results has been implemented in the international arena with the aim of improving testing and professional practice. Assessing tests provides users with an independent review, which follows clearly defined procedures and criteria. In Spain, the Spanish Association of Psychology (COP) leads the "test evaluation" project, which has recently completed its fourth edition. Basically, there are two primary assessment models; the Buros model and the European Federation of Psychologists' Associations (EFPA) model, which has been adopted by the COP. The two models share the same objectives and content; excluding the evaluation format and the management system, they are similar in essence. In this paper, we present the results of the fourth edition of the Spanish test evaluation program following the general guidelines for assessing the quality of tests, and we explain the critical points affecting each phase of the process.

Key words: Tests, Tests use, Assessing tests quality, Psychometrics.

The evaluation of tests and the dissemination of the resulting reports has been implemented internationally with the clear aim of improving the use of tests, and thus the professional practice. In our environment the action was implemented by the Test Commission of the General Council of the Spanish Psychological Association in 2010 (Muniz et al., 2011), and with the publication of this paper four editions have now been established (Ponsoda & Hontangas, 2013; Hernández, Tomás, Ferreres & Lloret, 2014) of a valuable professional tool, the main formal, procedural and content aspects of which we examine here.

There are basically two models of test evaluation: the model proposed by the European Federation of Psychologists' Associations (EFPA) (Evers et al., 2013), which has also been adopted by the General Council of the Spanish Psychological Association (COP), and the American model, epitomised by Buros (Carlson & Geisinger, 2012). Analysis of the content and processes

followed by these two models reveals that while they use different evaluative formats, the similarities between them are greater than the differences. They both share the same intention and starting point: that the publication of independent reviews subject to well-defined procedures and scientific criteria mean a boost in the process of continuous improvement in the construction/editing/use of tests.

The general assessment procedure can be summarised in 5 steps: (a) choosing the tests to be reviewed; (b) identifying the reviewers; (c) evaluating the test; (d) revising and editing; and (e) editorial comments and final report. Each stage counts towards the final result, and in each one there are risks that must be controlled to ensure the quality of the final report, thus ensuring that the assessment meets the intended objectives.

Before reviewing the test evaluation models, the structural differences between the bodies that implement them should be noted. Buros operates as an independent non-profit organisation, belonging to the University of Nebraska, in which 12 people work and it has 8 post-graduate collaborators. The Spanish review process rests on a general coordinator appointed for the purpose by the COP Test Commission. The coordinator is responsible

Correspondence: Paula Elosua. Facultad de Psicología. Av. Tolosa, 70. 20080 San Sebastian. España.

E-mail: paula.elosua@ehu.es





for the administrative and scientific implementation, although for the former task the coordinator has the support of the COP and is free to appoint a team of collaborators to assist in the administration.

Buros has extensive experience in evaluating tests, beginning in 1938 with the publication of the first *Mental Measurement Yearbook* (MMY, Buros, 1938). The MMYs are compendiums that present the collection of test evaluation reports carried out by this institute. To date, 19 volumes have been published, in which approximately 10,500 tests have been reviewed. Buros has also published two volumes related to test description, which are sold with the titles *Tests in Print* (TIP) and *Pruebas Publicados en Español* (Tests Published in Spanish) (PPE; Schlueter, Carlson, Geisinger, & Murphy, 2013). They are works that have an encyclopaedic format, providing descriptive lists of the tests that are available on the market. The increase in the Spanish-speaking population in the United States was the impetus for the creation and publication of PPE. The volume presents information about tests available in the United States that are either fully published in Spanish or include any part in this language. The current edition comprises 422 entries organised alphabetically by the test name (Carlson & Gonzalez, 2015).

CHOOSING THE TESTS TO REVIEW

Choosing the tests to be assessed generally begins with a review request from the editors. In Spain, the requests are processed through the Test Commission of the General Council of the Spanish Psychological Association (COP) (see Table 1; Tests reviewed in the 4th edition), which currently has within its membership representatives of three publishers. In Buros the requests are made directly from the Buros offices at the University of Nebraska. The Buros Center submits the applications received to a first screening in order to check that the tests meet certain conditions, including that they (a) provide psychometric data; (b) have an administration/correction/interpretation manual; and (c) present a technical manual. Without these requirements the tests are not evaluated. The direct consequence of this obligation is that non-commercialised tests fall outside the review process. In Spain the test commission has the power to propose the review of non-commercialised tests. In fact, a non-commercial test was evaluated in the second edition. Since the information and additional elements that accompany a publication in the format of a scientific paper are not comparable, in form or in content, with the material contained in a test sold by a publisher, the review of the test raised additional needs for the assessment process (Ponsoda & Hontangas, 2013).

After choosing the tests, the publishers submit three full copies to the body responsible for the review. The number of copies sent enables the coordinator or the administration office, where appropriate, to have a set of each test, which facilitates the checking of the evaluation information at any point in time. In this part of the process, Buros produces a descriptive sheet on the characteristics of the test, providing information about the title, objectives, scores, target population, administration time, acronym, publisher, publication date, authors, administration format and price.

IDENTIFICATION OF REVIEWERS

The evaluation process rests on the independent and unbiased judgment of the reviewers, on their capability

Acronym	Test	Publisher
ABAS-II	Sistema de Evaluación de la Conducta Adaptativa [Evaluation System for Adaptive Behaviour]	TTEA Ediciones
BADyG/M-r	Batería de Aptitudes Diferencial y Generales. Nivel M Renovado [Battery of Differential and General Aptitudes. Renovated Level M]	CEPE, S.L.
BETA	Batería para la Evaluación de los Trastornos Afásicos [Battery for the Assessment of Aphasic Disorders]	Instituto de Orientación Psicológica EOS
BSI-18	Inventario Breve de 18 Síntomas [Brief Symptoms Inventory of 18 Symptoms]	Pearson Educación
CECAD	Cuestionario Educativo-Clinico: Ansiedad y Depresión [Educational-Clinical Questionnaire: Anxiety and Depression]	TEA Ediciones
EHPAP	Evaluación de Habilidades y Potencial de Aprendizaje para Preescolares [Assessment of Skills and Learning Potential for Preschoolers]	Instituto de Orientación Psicológica EOS
PAIB-1	Prueba de Aspectos Instrumentales Básicos en Lenguaje y Matemáticas [Test of Basic Instrumental Aspects in Language and Mathematics]	CEPE, S.L.
PECC	Prueba para la Evaluación de la Cognición Cotidiana [Test for the Evaluation of Daily Cognition]	Instituto de Orientación Psicológica EOS
SCIP	Screening del Deterioro Cognitivo en Psiquiatría [Screening of cognitive Impairment in Psychiatry]	TEA Ediciones
WMS-IV	Escala de Memoria de Wechsler -IV [Wechsler Memory Scale -IV]	Pearson Educación
WPPSI-IV	Escala de Inteligencia de Wechsler para Preescolar y Primaria [Wechsler Preschool and Primary Scale of Intelligence]	Pearson Educación



and experience in analysing the documents and materials that make up the test, and in developing justified and fair arguments. The prudent identification of suitable professionals for this work is essential.

In the phase of identifying reviewers, for each test the most common strategy is to select a pair of specialists with complementary profiles; one with a strong background in measurement, and a second with a high level of experience in the area measured in the test. Having the two professional profiles allows for sources of opinion built on different perspectives and experiences. Together, their contributions help us to obtain more comprehensive assessments that include formal, substantive and applied aspects.

According to the criteria used by the COP and Buros, the reviewers must demonstrate experience in the use/construction of tests. Buros has a database of 900 authors from which to select those it considers most appropriate, and upon whom it imposes certain restrictions in order to avoid bias; for example, two reviewers of the same test must not belong to the same university or publishing company; and attempts are made to combine expert reviewers with young authors. In Spain the work of identification and selection is the responsibility of the coordinator who, using the literature studies, tries to

match the necessary profiles to guarantee good evaluations (Table 2; Reviewers of tests in the 4th edition). As compensation for the time and effort dedicated to the review, Buros sends a free copy of the MMY in which they have participated; the COP offers 50 to each reviewer. In both cases the evaluators keep a copy of the reviewed test.

The initial contact with the reviewers is carried out by email; in the email, prospective reviewers are informed about the test evaluation project, they are invited to review a particular test, they are advised to reject the review in the case of conflicting situations, and a recommended three month period is set for the review to take place. In the 4th edition, reviewers were explicitly asked to sign a declaration of conflict of interest to ensure the impartiality and independence of the review.

Once the reviewers have agreed to participate in the evaluation, they are sent the material. The reviewers carry out their work with the same material that a user would receive upon buying the test. The test evaluation questionnaire is sent together with the material provided by the publisher, (CET; Prieto & Muñiz, 2000), which in the third edition had a number of modifications and a document with information clarifying some of the issues contained in the questionnaire; (Hernández, Tomás, Ferreres & Lloret, 2014). Buros sends a guide to the review process and the expectations it involves.

TEST EVALUATION

The evaluation of the tests is the most delicate phase of the process. The reviewers have to produce an evaluative report on the test based on the detailed analysis of the information provided in the manual, and the materials accompanying it.

The evaluation consists of completing a questionnaire that guides the reviewer on the most important aspects to consider whilst carrying out their work. The formats of the evaluation questionnaires used by Buros and the COP are different. Buros' work is carried out using a form with test items for which it is expected that the reviewer will express a coherent and informative judgement. The form includes five points to be developed: (a) general description of the test; (b) construction; (c) technical report with an analysis of standardisation, reliability and validity; (d) general comments conveying the following: (1) the strengths and weaknesses of the test; (2) references to current research that supports the theoretical model on which the test is built, and (3) evidence to support its use. Finally (e) there is a brief summary of six or seven sentences in which the reviewer summarises their general conclusions and recommendations regarding the correct use of the test.

TABLE 2
REVIEWERS WHO CARRIED OUT THE EVALUATION OF THE TESTS

Reviewers	Affiliation
Juan Antonio Amador	Universidad de Barcelona
Sonia Alfonso Gil	Universidad de Vigo
Constantino Arce	Universidad de Santiago de Compostela
Juan Ramón Barrada	Universidad de Zaragoza
Roberto Colom	Universidad Autónoma
Ana Delgado	Universidad de Salamanca
Eduardo Doval	Universidad Autónoma de Barcelona
Sergio Escorial	Universidad Complutense
Sara Fernández Guinea	Universidad Complutense
María Paloma González Castro	Universidad de Oviedo
Francisco Gutierrez Martínez	Universidad Nacional de Educación a Distancia
José Antonio López Pina	Universidad de Murcia
Urbano Lorenzo Seva	Universitat Rovira i Virgili
Luis María Lozano	Universidad de Granada
Ignacio Montorio	Universidad Autónoma de Madrid
Juan Antonio Moriana	Universidad de Córdoba
Cristino Pérez	Universidad de Granada
Luz Pérez Sánchez	Universidad Complutense
Pedro Prieto Marañón	Universidad de La Laguna
Antonio José Rojas	Universidad de Almería
Bonifacio Sandín	Universidad Nacional de Educación a Distancia
Paula Samper	Universidad de Valencia



The test evaluation questionnaire (CET; Prieto & Muñiz, 2000) used by the COP combines closed answers and open answer essay questions. The CET is divided into three sections: (a) general description of the test; (b) evaluation of the characteristics of the test and (c) overall evaluation of the test. The first section is a descriptive technical form on the test that provides information on aspects such as the title, objectives, scores or the application time. It corresponds, largely, with the descriptors that Buros prepares in the phase of test selection. The second section contains 37 items with graduated responses (inadequate, adequate but with deficiencies, adequate, good, excellent) with the aim of quantifying the quality of the feasibility studies, validation and weighting. It also includes essay items for these three aspects in which the reviewer produces an informed judgment. The last part of the CET is dedicated to a single open question that asks the reviewer to comment on the test, including its strengths/weaknesses and a summary of the most relevant information for the correct use of test.

Whether the number of sections in the questionnaire is three or five, both formats address the same points in the assessment, which can be interpreted in terms of concurrent validation. As the two are similar in content, it is worth asking whether there are advantages and disadvantages in using one format or the other. The answer to this question should be given by the effects of using both types of item on any standardised test. The nature of the item affects, among other things, the score, the representation of the construct and the ease / difficulty of constructing items (Haladyna, 2004). Assuming they are well constructed, closed answer items: (a) are versatile; (b) guarantee the evaluation of the significant points of the area under evaluation; (c) enable fast and efficient quantification; (d) are easier to answer; and (e) permit item analysis. However, they also have disadvantages. In closed response formats and complex assessments, occasionally the depth and specificity of the evaluation may not be perfectly reflected in the response options. For example, with regards to the CET, the valuation of the quality of a weighting sample is analysed in the same way for a questionnaire on intellectual skills intended for the community population as for a diagnostic questionnaire used with clinical groups. In both cases the number of participants is analysed in the research, but the values are difficult to compare as they come from different populations. Perhaps, in these situations, a note could be added in reference to the type of sample/population.

The CET adds, to the advantages of the evaluation via essay questions, a quantitative assessment of the characteristics of the test, which can be read as a quick

photograph of the points evaluated. It is an aspect that is complementary to the essay questions and adds value to the evaluation, but does not replace it. Providing quantitative values for each of the sections does not legitimise the calculation of a final mean for assessing the quality of a test. The validation of a test must always take into account the proposed use of the scores (AERA, APA & NCMEA, 2014; Elosua, 2003), and the selection of a test by a potential user must consider the context of application; if the inference of the scores is related to the prediction of future behaviour, however excellent the quality of the materials or the weighting sample, and no matter how high the scores are in those sections, the aspect related to the predictive evidence of validation is always one of the first criteria to observe; and an arithmetic mean calculated across all of the aspects evaluated does not give this information.

With respect to the items of the CET in the four editions carried out to date by the COP, a number of items have been detected whose current wording raises doubts for the assessor and which should be reviewed, as has been noted (Hernández, Tomás, Ferreres & Lloret, 2014; Muñiz, Fernandez-Hermida, Fonseca-Pedrero, Campillo-Álvarez & Peña-Suarez, 2011; Ponsoda & Hontangas, 2013). To ensure a better understanding of the content of the items, in the third edition (Hernandez et.al, 2013) a supporting document was drafted with some modifications also included at that time. We opted to improve the CET series in this way, notably by reviewing some of the items and by adding explanatory statements as was done in the original version of the EFPA questionnaire (<http://www.efpa.eu/professional-development>). To do this, it would be advisable, as Ponsoda and Hontangas indicate (2013), to work on a computerised version of the CET to facilitate its completion and to allow the addition of hyperlinks with explanatory statements.

REVISING AND EDITING

Once the two evaluations of each of the tests have been received, the editorial process proceeds to revise the reports received. Buros and the COP work differently at this point, which is logical given that the final reports that the two bodies produce are different. Buros publishes the two reviews that have been carried out for each test; the COP produces a single report that synthesises the work of the two reviewers who analysed the test.

Buros pays attention to verifying the accuracy of the data and matters of style, with the aim of achieving a final report that is in accordance with its editorial direction. Once the texts have been revised and corrected, Buros publishes in the MMY the reports that have been



submitted by the reviewers. For the factual review Buros has two or three PhD students who inspect and correct aspects such as the references or the format.

The COP elaborates a unique text-report for each test, which condenses the quantitative information and summarises the evaluations provided by the two reviewers. Editing the evaluation basically means reviewing and enriching it, making sure there are no mistakes regarding the objective data, and ensuring that the language used is homogeneous, correct, appropriate and not hurtful. The editor analyses and corrects, where appropriate, the tone of the review, ensuring the correct use of psychometric terms and the appropriateness of the language to the profile of the potential user; the use of technical terms or, in certain cases, overly methodological jargon is avoided. Without rejecting the viewpoint of the reviewer, negative comments are eliminated in the editing process, and the revisions are shaped to offer a summary that is clear and accurate on the characteristics of the test that may be important for the potential user. The editing does not alter the nature of the professional judgments made by the evaluators, but it does correct observations that are poorly descriptive. The aim of the review process is not to highlight the limitations or shortcomings of a test, but to put them in the appropriate context.

In producing the test report, we assume that the ultimate goal of the review is not to give the test an absolute rating that the user will adopt as a measure of quality and justification for its use, but rather we uphold a critical and active reading, based on a professional opinion and in which the context of the test use is important. As Thorndike (1999) noted, the reviews must be considered with mature judgment and knowledge of the situation in which the test is to be applied; both judgments fall to the user.

Once the reports have been edited, Buros sends them to the reviewers for them to approve the modifications. In the model followed by the COP this step is omitted because the reviews are not published in their original format.

EDITORIAL COMMENTS AND FINAL REPORT

Once the editing process has finished, the reports of the tests that have been evaluated are sent to the publishers for review. The publishers usually send the reports to the authors of the tests. Both the authors and the editors write their comments and submit them to the coordinator. This stage is important to ensure that the reports do not contain errors or inaccuracies; if any are detected, they are corrected before publication.

Buros has established that if major changes are required at this stage, these are forwarded directly to the reviewer

for his or her approval or rejection. The COP does not include this option. Moreover, publishers and authors in the Buros process are only permitted to object to errors of fact; revising the opinions of the reviewers is not permitted.

The objective of this stage is to ensure the precision and accuracy of the information included in the report. Our model, which is still young, does not have defined guidelines on the role of publishers at this stage, and there is variability in the responses of the editors; from those who simply correct any inaccuracies, to those who send additional material such as scientific articles or dissertations in which the test in question has been used. It would be important to establish guidelines for the author/editor to concentrate on the factual information; and focus his or her comments on the material evaluated by the reviewer.

After correcting the mistakes and analysing the suggestions, the final report is sent for each of the tests, which will form part of the MMY or be published on the webpage of the COP (<https://www.cop.es/index.php?page=evaluacion-tests-editados-en-espana>)

DISCUSSION AND CONCLUSIONS

It is an increasingly widespread idea that good professional practice includes the use of tests that have been evaluated by external bodies (Carlson & Geisinger, 2012; Evers, Sijtsma, Lucassen, & Meijer, 2010; Ponsoda & Hontangas, 2013) and that therefore the evaluation of tests is a support tool in the professional work of the psychologist.

If the purpose of the evaluation is to improve the professional practice, the question is, what information does a potential user need in order to consider the use of a test in an applied or research situation? This is the fundamental issue in the evaluation process, and derived from it is the importance of: (a) a clear and well-structured handbook; (b) the quality of the materials; (c) the technical description of the test with accurate information concerning the analysis of its construction, reliability, validation and weighting, (d) the need to justify the use of the scores in a particular context, and (e) references to recent investigations with the theoretical model on which the test was built.

The points covered by the evaluation of tests, whether the format is more or less open, have become the gold standard for constructing/editing a test. As it constitutes an external evaluation standard, the CET must meet requirements concerning its validity and provide a complete representation of the domain being evaluated. However as well as the source of criterial evidence, it is



important to consider consequential aspects of test use. The evaluation of tests has direct consequences for all phases involved in the construction/editing of tests, insofar as it promotes improvements in the construction process and encourages publishers to adopt increasingly rigorous criteria for the publication of tests and manuals. This is not the only consequence, however. The professional impact of the evaluation of tests has consequences for the undergraduate and graduate training of the psychologist. Given the importance of the tool, and the need for an active reading of the evaluation reports, the reviewing of tests is being incorporated into the academic curricula of psychology in subjects related to psychometrics and assessment. Also, perhaps especially in the United States, fairness of tests and assessments for all subcomponents of the population is an important consequence and reviewers are encouraged to comment on the appropriateness of measures for different ethnic, cultural, ability, and other like groups. The acceptance of the test review by academia, by the authors, the publishers and the users is positive. The publishers are allies in this process, and we are certain, both because of their attitude and their willingness that they agree with the guidelines set by the COP.

The test review plan is confined to an ambitious program of projects to improve the use of tests (see Elosua and Muñiz, 2013), projects that have been designed and

brought to fruition at a moment that is clearly defined by the development of psychometrics, which is beginning to be referred to outside Spain as the Spanish school of psychometrics. The four editions published to date allow us to conclude that the balance of the project has been positive. But the task is not finished; the review of tests as a tool and gold standard requires continuous work to improve procedures, formal and substantive, in order to fulfil its goal: to provide relevant information about the tests available to the professional psychologist. It is therefore important to remember the observations made in the previous issues affecting specific content and proposals for improvement, which are now being introduced in a new version of the CET.

Based on the COP's still fledgling experience of reviewing tests, we believe it would be appropriate to establish a clear editorial direction to collect and systematise objectives and expectations for each of the five stages into which we have divided the evaluation. It is necessary to formalise a theoretical and procedural framework for each of the actors involved (the coordinator, the evaluator, the publishing company/the author), which defines each of their roles in every phase of the process. Since we are working with a questionnaire, the objective should be to standardise the conditions of application and to ensure that the objectives of the assessment and the procedure are understood by

TABLE 3
SUMMARY OF THE MEAN QUALIFICATIONS OF THE EVALUATED TESTS

TABLE 3 SUMMARY OF THE MEAN QUALIFICATIONS OF THE EVALUATED TESTS											
Tests											
Characteristics	ABAS-II	BADyG/M-r	BETA	BSI-18	CECAD	EHPAP	PAIB-1	PECC	SCIP	WMS-IV	WPPSI-IV
Quality of the materials and documentation	4.5	4	3	3	4.5	4	4	3.5	4	5	4.5
Theoretical fundamentation	4	3	3.5	2	4	3	3	4	4	4.5	5
Spanish adaptation	4	NOT APPLICABLE	NOT APPLICABLE	3	NOT APPLICABLE	3	NOT APPLICABLE	NOT APPLICABLE	NOT APPLICABLE	4	5
Item analysis	--	4	--	--	5	--	--	3	3	--	3.5
Content validity	3.5	3	3	3	4	2	3	3	3	4	5
Construct validity	5	3	4	2	5	3	2	3.5	4	4	4
Analysis of bias	--	--	--	--	--	--	--	--	--	--	2
Predictive validity	--	3.5	2	--	--	3	4	2	3.5	4	3.5
Reliability: internal consistency	5	4.5	3.5	5	5	3.5	5	3	4	4	3.5
Reliability: stability	2.5	--	--	3.5	--	--	--	2.5	3.5	3	3.5
Reliability: equivalence	--	--	--	--	--	--	--	--	3	--	--
Weighting	4.5	4.5	2.5	3.5	4.5	2.5	3.5	2.5	4	4.5	4.5

Note. -- Information not provided.
Evaluation scale: 1 (Inadequate) – 2 (Adequate but with deficiencies) – 3 (Adequate) – 4 (Good) – 5 (Excellent)



all. The need for a clear editorial direction is particularly relevant in the COP model because, although the final report maintains the intent and tone provided by the reviewers, the format is altered and this can sometimes generate confusion for the professional who carried out the review.

In conclusion, we believe that, for future editions, a multiannual editorial team should be established, responsible for conveying the COP's editorial direction in the review process, and to establish the objectives and expectations for each stage of the process. We agree with Carlson and Geisinger (2012) that the review of tests must stimulate progress towards professional standards of test construction in which good work prevails and poor work is rejected, at the same time as it encourages the authors and publishers to develop manuals that include detailed information on construction, standardisation, psychometric properties, and appropriate and inappropriate uses of the test. The continuous incorporation of psychometric advances must not be left out however, in order to bridge the distance between theory and practice (Elosua, 2012). Clearly, psychology and professional practice are no longer the same as when they began. The psychometric models have progressed, the theories have evolved, and the problems and needs of the psychological practice change and adapt to the needs of each moment. Since the publication of the first tests, concepts such as reliability and validity have been enriched, and the scientific, professional and ethical requirements demanded of tests have adapted to new needs (De Boeck & Elosua, in press). In this evolution, the process of continuous assessment of tests plays an important role as a stimulus and incentive for the rapprochement between theory and practice.

ACKNOWLEDGEMENTS

We would like to thank the reviewers of the fourth edition of tests and the publishers and authors of the tests evaluated for their willingness, dedication and professionalism.

REFERENCES

- American Educational Research Association, American Psychological Association, & National Council on Measurement in Education (2014). *Standards for educational and psychological tests*. Washington DC: American Educational Research Association.
- Buros, O. (1938). *The 1938 Mental Measurements Yearbook*. Highland Park, NJ: Gryphon.
- Carlson, J. F., & Geisinger, K. F. (2012). Test Reviewing at the Buros Center for Testing. *International Journal of Testing*, 12, 122-135.
- Carlson, J. F., & Gonzalez, S. E. (2015). Using Pruebas Publicadas en Español to enhance test selection. In K.F. Geisinger (Ed.), *Psychological testing of Hispanics (2nd ed.): Clinical and intellectual issues* (pp. 11-27). Washington, DC: American Psychological Association.
- De Boeck, P., & Elosua, P. (in press). Reliability and Validity: History, Notions, Methods, Discussion. In D. Iliescu (ed.). *International Handbook of Testing and Assessment*.
- Elosua, P. (2003). Sobre la validez de los tests [On the validity of tests]. *Psicothema*, 15, 315-321.
- Elosua, P. (2012). Tests publicados en España: Usos, costumbres y asignaturas pendientes [Tests Published in Spain: Uses, Customs and Pending Matters]. *Papeles del Psicólogo*, 33, 12-21.
- Elosua, P., & Muñiz, J. (2013). Proyectos españoles para una mejora en el uso de los tests [Spanish projects for improving the use of tests]. *Revista Latinoamericana de Ciencia Psicológica*, 5, 139-143.
- Evers, A., Sijtsma, K., Lucassen, W. & Meijer, R. R. (2010). The Dutch review process for evaluating the quality of psychological tests: History, procedure and results. *International Journal of Testing*, 10, 295-317.
- Evers, A., Muñiz, 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
- Haladyna, T. M. (2004). *Developing and validating multiple-choice test items*. Mahwah, NJ: Lawrence Erlbaum Associates.
- Hernández, 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.
- Muñiz, J., Fernández-Hermida, J.R., Fonseca-Pedrero, E., Campillo-Álvarez, A. & Peña-Suárez, E. (2011). Evaluación de tests editados en España [Review Of Tests Published In Spain]. *Papeles del Psicólogo*, 32, 113-128.
- 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 to Evaluate The Quality of the Tests Used in Spain]. *Papeles del Psicólogo*, 77, 65-71.
- Schlueter, J., Carlson, J. F., Geisinger, K. F., Murphy, L. L. (Eds.) (2013). *Pruebas publicadas en Español [Tests Published in Spanish]*. Lincoln, NE: Buros Center of Testing