

## Test commission: Twenty-five years working on test quality

Ana Hernández<sup>1</sup>, Paula Elosua<sup>2</sup>, José Ramón Fernández-Hermida<sup>3</sup> and José Muñiz<sup>4</sup>

<sup>1</sup>Universidad de Valencia, España. <sup>2</sup>Universidad del País Vasco, España. <sup>3</sup>Universidad de Oviedo, España. <sup>4</sup>Universidad Nebrija, España

*El objetivo del presente trabajo es mostrar las actividades y proyectos llevados a cabo por la Comisión de Test del Consejo General de Psicología de España para mejorar la calidad y el uso de los test. Se repasan los proyectos más relevantes de la comisión durante sus veinticinco años de existencia, describiendo los trabajos llevados a cabo para a) conocer las opiniones y actitudes de los psicólogos hacia los test, b) evaluar la calidad de los test editados en España, y c) desarrollar estándares, directrices y códigos éticos y deontológicos que mejoren la práctica de los test. También se resaltan otras actividades de la comisión de carácter formativo, así como sus relaciones con otras organizaciones internacionales con fines similares. Finalmente se llevan a cabo algunas reflexiones sobre los retos futuros de la Comisión de Test, haciendo hincapié en el papel clave que tendrán las nuevas tecnologías en la evaluación psicológica.*

**Palabras clave:** Test, Psicometría, Evaluación, Calidad de los test, Uso adecuado de los test.

*The objective of this paper is to present the activities and projects carried out by the Spanish Test Commission, established by the Spanish Psychological Association to improve the quality and use of tests. The most relevant projects of the commission during its twenty-five years of existence are reviewed, describing the work carried out: a) to understand the opinions and attitudes of psychologists towards the tests, b) to evaluate the quality of the tests published in Spain, and c) to develop standards, guidelines, and ethical and deontological codes to improve testing practices. The commission's other activities of a formative nature are also highlighted, as well as its relationships with other international organizations with similar purposes. Finally, several reflections on the future challenges of the Test Commission are shared, emphasizing the key role new technologies will have in psychological assessment.*

**Key words:** Test, Psychometrics, Evaluation, Test quality, Appropriate test use.



### RIGINS AND OBJECTIVES

The Test Commission was created in 1995 by the national association of Spanish psychology (Colegio Oficial de Psicólogos (COP), later Consejo General de Colegios Oficiales de Psicólogos) and acts as an advisory body to the governing board. Its main objective is to carry out a series of activities and projects aimed at improving the quality and use of tests in Spain. Tests constitute one of the most widely used tools by psychologists, both in their professional and research work. Therefore, ensuring their psychometric quality and appropriate use is essential to offer a quality service to users of psychology and to society in general (Hernández et al. 2015; Muñiz et al., 2020). The proper use of a measurement instrument requires, first of all, that it has proven psychometric quality, but also that it is used appropriately, for which the professionals and researchers who use it must have appropriate training. During its more than twenty-five years of existence, the Test Commission has developed different activities aimed at improving these three aspects: test quality, appropriate use, and training for those who use the tests. At the time the Test Commission was founded, psychology was already well established in Spain, both professionally and academically. By then, the psychology degree included psychometrics and psychological assessment as compulsory subjects in all Spanish universities, there was professional regulation of

psychology, and a wide variety of psychological tests were available on the market. It can be said, in short, that there were adequate conditions for the establishment of the test commission, as a complement to formal university training.

In order to gather all the points of view involved in the quality and use of tests, the Commission has always been formed, throughout these twenty-five years, by experts in Psychometrics and Psychological Assessment and by representatives of the Spanish test publishing companies. This convergence of professional and academic experts and publishers is essential to improve the quality and use of the tests, as they must work in a convergent manner. It also includes a representative of the COP, since many of the Commission's decisions have professional and institutional repercussions. From its foundation until 2020 the Commission was chaired by José Muñiz, Professor of Psychometrics, and from 2020 it has been chaired by Ana Hernández, Professor of Psychometrics at the University of Valencia. Table 1 includes all the people who have been members of the Commission from 1995 to the present.

In order to carry out its objectives, the Test Commission proposes and undertakes numerous initiatives, collaborating with international organizations that share its objectives, such as the Committee on Tests and Testing of the European Federation of Psychological Associations (EFPA) (currently called the Board of Assessment), or the International Test Commission (ITC). It should be noted that the European Test Commission was created at the proposal of the COP, and its first president was José Muñiz, who also chaired the Spanish commission.

### TEST COMMISSION PROJECTS

Activities and projects aimed at ensuring the quality and proper use of tests can be divided into two main strategies, one called *restrictive* and the other *informative*. The *restrictive* strategy includes all initiatives

Received: 1 November 2021 - Accepted: 10 December 2021

Correspondence: Ana Hernández. IDOCAL. Universidad de Valencia. Av. Blasco Ibáñez, 21. 4610 Valencia. España.

E-mail: Ana.Hernandez@uv.es

.....  
This work has been partially funded by the Ministry of Science and Innovation (PID2019-103859RB-100).

aimed at restricting the use of tests to professionals who have received specific training to do so, ranging from restrictions on the purchase and use of tests to legal provisions that prevent the use of tests by certain professionals. The *informative* strategy includes the dissemination of information to encourage the proper use of tests, such as ethical and professional codes, guidelines and recommendations, or the dissemination of information on the quality and characteristics of the tests available on the market (Muñiz & Bartram, 2007; Muñiz & Fernández-Hermida, 2010; Muñiz et al., 2015). These two strategies are complemented by a third one that is equally important, and closely linked to the previous ones, the *training* strategy, aimed at updating the knowledge and skills of professionals and their continuing education (Hernández et al., 2021). The following are some of the projects carried out by the Test Commission, which are included in some of the aforementioned strategies, especially in the formative and informative ones.

**PSYCHOLOGISTS’ OPINIONS AND ATTITUDES TOWARDS TESTS**

In order to improve the use of the tests, it is essential to know the opinions and attitudes of the professionals regarding them, as this will allow us to focus on those aspects that are perceived to be the most deficient. To this end, the Test Commission regularly conducts a survey among professionals. To date, three editions have been carried out, the first in 1999, the second in 2009, and the third in 2019. The

results (Muñiz & Fernández-Hermida, 2000, 2010; Muñiz et al., 2020; Hernández et al., 2021) allow us to establish an accurate diagnosis of the perception of professional psychologists about the use of tests, which is essential to promote actions for continuous improvement.

The questionnaire used in these studies provides information on several dimensions, and the participants are also asked to indicate the three tests they use most in their professional practice. In order to analyze the evolution of psychologists’ opinions and attitudes towards tests over time, the questionnaire used has always been the same, although with small variations adjusted to each moment in time. For example, in the last study (Muñiz et al., 2020, Hernández et al., 2021) the questionnaire was composed of 31 items grouped into 5 dimensions: Attitudes towards tests, Training and knowledge about tests, Need for control and regulation of tests and their use, Influence of new technologies on test use, and Problems in test use. In addition, participants were asked to indicate the three tests most commonly used in their professional practice, and the 2019 edition included, for the first time, items on the annual test review process carried out by the commission since 2011, as detailed in the following section.

The most relevant results of the last edition indicate (a) that psychologists recognize that the training received in the psychology degree may not be sufficient for the correct use of most tests, and they recognize the need for further training to use the tests properly; (b) that psychologists’ attitudes towards the use of tests are positive; (c) that, in general, psychologists are in favor of increasing the control and regulation of tests and their use; (d) that there is some skepticism about the incorporation of new technologies in testing practice; and (e) that in their professional field they recognize problems, such as the improper use of photocopies, or not being up to date with tests, although these problems occur with moderate frequency. The work of Muñiz et al. (2020) analyzes the differences that exist by specialty and Hernández et al. (2021) delves into those differences linked to age, sex, and work sector. Regarding the evolution of opinions and attitudes over time, the comparative study offers fairly stable results, although positive changes can be seen related to the increased use of tests in the practice of the profession, the improvement of the information available on the quality of the tests, and the decrease in problems or bad practices, such as the use of photocopies (Muñiz et al., 2020).

Finally, it should be noted that the results on the most widely used tests in Spain show that, over the years, projective tests have been completely relegated, with psychometric tests clearly predominating. In fact, in the last study, the 25 tests most used by Spanish psychologists are all psychometric, either tests adapted to the Spanish context or nationally developed tests (24%). These data are a clear indicator of the great activity carried out in our country in the construction, adaptation, and publication of tests (Muñiz et al., 2020).

**EVALUATION OF THE QUALITY OF TESTS PUBLISHED IN SPAIN**

When selecting a test for use, it is important that the professional evaluates or has access to external evaluations of the quality of the tests available. To this end, in 2010 the Test Commission initiated a review process of the tests used in Spain. These are both quantitative and qualitative evaluations carried out by experts, the aim of which is to help professionals choose the test that best suits their needs.

**TABLE 1  
PERSONS WHO HAVE BEEN MEMBERS OF THE COP TEST COMMISSION AT SOME POINT IN TIME**

NAME	AFFILIATION
Ana Hernández Baeza	University of Valencia
Ana Martínez Dorado	GiuntiEOS
Eduardo Montes Velasco	COP
Francisco José Abad	Autonomous University of Madrid
Frederique Vallar	Pearson
Gerardo Prieto	University of Salamanca
Jaime Pereña	TEA Ediciones
Javier Rubio	SHL Group
José Luis Galve	CEPE
José Muñiz	Nebrija University
José Ramón Fernández Hermida	COP
Miguel Martínez	EOS
Milagros Antón	TEA Ediciones
Nicolás Seisdedos	TEA Ediciones
Paula Elosua Oliden	University of the Basque Country
Rocío Fernández Ballesteros	Autonomous University of Madrid
Rosario Martínez Arias	GiuntiEOS
Vicente Ponsoda	Autonomous University of Madrid
Viviana Gutman Mariach	Pearson

The review process has been based on the test review model (CET, Cuestionario de Evaluación de Test) initially proposed by Prieto and Muñiz (2000). The CET was used in the first three editions of the project, with some minor modifications in the third edition. For the fourth edition, it was revised by Hernández et al. (2016), generating a new version: CET-R. This revision incorporated the most relevant psychometric and technological advances collected in the EFPA test evaluation model in its 2013 revision (Evers et al., 2013). Both models (CET and CET-R) are available on the website of the Test Commission: <https://www.cop.es/index.php?page=evaluar-calidad>. They both contain three major sections: (a) Technical description of the test, which includes aspects such as the purpose of the test, its classification, the mode of correction, the possibility of obtaining automated reports, and the price, among others. (b) Technical evaluation of the characteristics of the test, which includes general questions such as the quality of the substantive model, or the quality of the materials and reports generated, and questions related to its psychometric quality: psychometric analysis of items, evidence of validity, reliability, and adequacy of norms or scales. For all these issues, open-ended questions are included to allow reasoning of the scores assigned and to provide any other information that may be relevant. And finally, (c) overall quantitative and qualitative assessment of the test, the quantitative result of which is reflected in a technical sheet.

The review process begins with the appointment of a coordinator by the Test Commission, and with the selection of the tests to be evaluated. Each test is evaluated by two independent reviewers, usually a psychometrician and a professional expert in the construct measured by the test. The independent reviews are integrated by the coordinator in a report that is sent to the publishers for their comments. Finally, with all the information at his or her disposal, the review coordinator prepares a final report that is posted on the Test Commission's website and is freely available to all professionals. Table 2 shows the coordinators of the different reviews.

To date (November 2021), nine editions of this review process have been carried out, with a total of 89 tests reviewed, in their different versions. The tenth edition is currently underway. All the detailed revisions can be consulted on the COP website: <http://www.cop.es/index.php?page=evaluacion-test-editados-en-espana>. Based on them, different works have been carried out that summarize both the results and the review process, also making interesting complementary and innovative contributions (by successive edition: Muñiz et al., 2011; Ponsoda & Hontangas, 2013; Hernández et al., 2015; Elosua & Geisinger, 2016; Fonseca & Muñiz, 2017; Hidalgo & Hernández, 2019; Gómez-Sánchez, 2019; Viladrich et al., 2021).

We believe that the evaluation of test quality is a very relevant project, with very positive repercussions on the improvement of the use of tests and, therefore, on the professional practice of psychology. Proof of this is that, in the last survey of psychologists' opinion on tests (Muñiz et al., 2020), those who are aware of the reviews generally consider them important and necessary. They also indicate that they consult the reports and that these help them to make decisions about the tests to be used. However, only 22.5% of participants said they were aware of these reviews (Muñiz, et al., 2020); therefore, a clear dissemination effort still needs to be carried out.

### ETHICAL AND PROFESSIONAL STANDARDS, GUIDELINES AND CODES

As part of its information dissemination work, the commission is responsible for disseminating, and in many cases translating and summarizing, the most relevant international guidelines on testing. The following guidelines are available to professionals on its website:

- ✓ An adaptation of the codes of ethics and good practice of the American Psychological Association (APA, 1992).
- ✓ Minimum standards for the proper use of tests (adapted from Moreland et al., 1995).
- ✓ The ITC guidelines refer to: (a) ethical and appropriate use of tests; (b) translation and adaptation of tests from one culture to another (see Muñiz et al., 2013 and Hernández et al., 2020); (c) quality control of test scores, their analysis and reporting; (d) security of tests, examinations, and other assessments; and (e) use of tests and other assessment instruments in research. A summary of the last three guidelines can be found in Muñiz et al. (2015).

With regard to the psychological assessment process, which includes tests as a fundamental tool, it is worth mentioning the participation of the Test Commission in the development of the European Guide to the Assessment Process (Fernández-Ballesteros et al., 2001), translated into Spanish on the commission's website, and of the ISO 10667 Standard (ISO, 2011), in the drafting of which the then president of the commission, José Muñiz, collaborated. The standard, which was revised in 2020, came into operation in Spain in 2013 and is especially relevant to the field of Work and Organizational Psychology, as it regulates everything related to the process of assessing people in the work and organizational environment. According to information provided by AENOR, the agency responsible for the standard in Spain (November 4, 2021), there are currently only 10 certified companies. However, the survey carried out by the Working Group on Psychology and Good Practices in Recruitment and Selection of People of the COP of Madrid (COP-Madrid, 2015), suggests that the standard may be being adopted at a less formal level by a larger number of organizations. According to this survey, and as recommended by the standard: (a) most companies keep a record of the technical documentation of the

**TABLE 2**  
**COORDINATORS OF TEST EVALUATIONS**  
**CARRIED OUT BY THE COP TEST COMMISSION**

EDITION	NAME	AFFILIATION
1	José Muñiz	Nebrija University
2	Vicente Ponsoda	Autonomous University of Madrid
3	Ana Hernández Baeza	University of Valencia
4	Paula Elosua Oliden	University of the Basque Country
5	Eduardo Fonseca Pedrero	University of La Rioja
6	M. Dolores Hidalgo Montesinos	University of Murcia
7	Laura E. Gómez Sánchez	University of Oviedo
8	Carme Viladrich	Autonomous University of Barcelona
9	Luis Manuel Lozano	University of Granada

assessment methods they use in their recruitment and selection processes (56%), (b) the people conducting the assessments have the required technical/specific training (74%), (c) the people being assessed are informed about the different phases of the process and the implications of each of these phases (90%), and (d) the security and confidentiality of the information collected during the assessment process is guaranteed (91%). These figures indicate that we are on the right track, although, in some respects, there is still room for improvement.

### OTHER ACTIONS AND PROJECTS

The organization of round tables and symposia focused on testing in national and international congresses, or the organization of specific training activities on the construction, use, and evaluation of tests are complementary actions organized by the test commission. Specifically, the courses organized by the Spanish Psychological Association through the Distance Continuing Education Program (FOCAD) stand out (Elosua, 2019; Muñiz & Fonseca, 2017), or those offered by the main test publishers operating in Spain (TEA, Pearson, GiuntiEOS, or CEPE). It is the latter that ensure that the only restrictive action exercised in our country for the use of tests is applied: test publishers ensure that only qualified persons have access to tests depending on their classification according to the APA categories (A, B, and C) (APA, 2014). Those tests classified as B (collective cognitive and personality tests) or C (individual tests and projective tests) can only be acquired by psychologists.

### INTERNATIONALIZATION OF THE TEST COMMISSION

As it could not be otherwise in a globalized world, the projects in which the Commission works go beyond the limits of our borders, and are framed in an international vision of collaboration and cooperation for the improvement of the use of tests. The Test Commission works regularly with international organizations such as the European Federation of Psychologists' Associations (EFPA) and the International Test Commission (ITC), which offer a framework of reference and an international scope of application in the regulation of the use of tests (Bartram, 2011; Muñiz & Bartram, 2007; Muñiz et al., 2001).

In this regard, it is important to highlight the work of the ITC in the development and dissemination of guidelines related to different aspects of test construction, adaptation, and use. The guidelines attempt to respond to the demands and challenges regarding the correct use of tests. In addition to the guidelines already mentioned, the ITC has made recommendations on: (a) large-scale assessments in linguistically and/or culturally diverse populations (ITC, 2018), (b) use of test updates and outdated tests (ITC, 2015), (c) computerized and internet-based assessment (ITC, 2005) and, to be published soon, (d) technology-supported assessments, developed jointly with the Association of Test Publishers (ATP).

The Test Commission, as a member of the EFPA Committee on Tests and Testing (known as the Board of Assessment since 2011), has participated in the international studies on the attitudes and opinions of psychologists towards tests. In the first edition of the project, 6 European countries participated (Muñiz et al., 2001), and in the second edition the number of participants increased to 17 (Evers et al., 2012), to which 12 were later added, including countries from Africa, America, and Oceania, such as Brazil, Lebanon, Indonesia,

Nigeria, or New Zealand (Evers et al., 2017). Currently, information continues to be collected in order to undertake a third edition of the project. The results of these studies allow us to compare the opinions of psychologists from different countries. The latest data conclude that in all countries the opinion towards the tests is generally positive, with a slight skepticism towards the use of the Internet, except for China, which scored significantly high in this aspect (Evers et al., 2017). Once the study is completed, it will be interesting to analyze the evolution of opinions and attitudes over the last 30 years.

Another important EFPA project that has gained traction in the Commission is the development and updating of a model for assessing test quality. The first model was published on the EFPA website in 2002 (Bartram, 2002), and it has subsequently been revised twice: 2008 (Lindley et al., 2008) and 2013 (Evers et al., 2013). The last update served as inspiration for the Spanish CET-R model. At this moment the EFPA has formed a new working group to study a new update, incorporating novel aspects, such as gamification, which are being incorporated in psychological and educational assessment.

Finally, we highlight the creation of the competency standards required to be able to use tests in the different fields of psychology. These standards are mandatory for EFPA member countries interested in obtaining European certifications for test users. This project was launched together with the EAWOP (European Association of Work and Organizational Psychology) for the field of Work and Organizational Psychology and was later generalized to the areas of Education and Health. These European certifications, which accredit the competencies of test users, are already being issued in countries such as the United Kingdom and Sweden.

All documentation on the EFPA's various testing projects can be found at <http://assessment.efpa.eu/documents/>. ITC guidelines and recommendations are available at <https://www.intestcom.org/page/28>.

### LOOKING TO THE FUTURE

The Test Commission has been going strong for twenty-five years now, and on its silver jubilee, in addition to looking back at what has already been carried out, we would like to make some reflections on the future. We do so with extreme caution, knowing that the future, as Seneca (2013) warned us, lies in uncertainty, and that its nature is liquid, evanescent and plagued by black swans, i.e., highly improbable and unpredictable events that change our lives; the COVID-19 pandemic was the last one (Bauman, 2002; Taleb, 2008). In the coming years the commission will continue to strengthen and enhance the lines of work that already have a solid track record. However, to enhance their effectiveness and usefulness we must try to improve their dissemination among psychology professionals. As we have pointed out, in the last survey of psychologists' opinions on tests (Muñiz, et al., 2020), only 22.5% of respondents said they were aware of a central activity of the commission, such as the evaluation of the quality of the tests that are made public on the COP website. Nor does the publication of standards and recommendations on the use of tests seem, in itself, sufficient to change assessment practices (Rios & Sireci, 2014). One way to improve the impact of guidelines related to the use of tests would be their explicit inclusion in psychology curricula, mainly within the subjects of psychometrics and psychological assessment. Some universities already do this, incorporating the CET-R model in Psychometric practices (Viladrich et

al., 2021). In any case, further work should be carried out to increase the visibility and impact of the actions of the Test Commission among professionals, which will result in a shortening of the distance between academia and research, on the one hand, and professional practice, on the other (Elosua, 2012).

But surely the main future challenges for the committee have to do with the growing incorporation of new technologies in evaluation. Technological advances are having a decisive influence on the different phases of assessment and on the measuring instruments themselves. To cite just a few examples, the development of items, with multimedia content, or based on games, the test scoring and the preparation of reports, which can be done in an automated way, or the way tests are applied, through the internet, by means of adaptive applications, etc. (Nieto et al., 2018; Parshall et al., 2010; Sanz et al., 2020; Seelow et al., 2019; Sorrel et al., 2021; Wan & Henly, 2012). All this generates new needs and situations to which we must respond. The recent pandemic originating from COVID-19 and its impact on psychological assessment, for example, show the continuing need for adaptation to new environments and the role that technology can play in that process. Confinement and social distancing forced the need for remote assessments and remote test administration (Elosua, 2021), which would have been unthinkable decades ago. But not only that, technology, especially through cell phones and other portable devices, is also contributing to the rise of ambulatory assessment, which collects information about people's behaviors, emotions, thoughts, etc., in a personalized, dynamic, contextual, and ecological way (Muñiz & Fonseca-Pedrero, 2019). And these advances enable the use of new indicators, such as physiological responses, physical indicators, eye movements, or information provided by social networks. These types of indicators generate such a large amount of information that traditional data analysis techniques are not capable of handling them, so new analysis methodologies are emerging, grouped under the denomination of *big data* (Kosinski et al., 2013). This type of data also requires new psychometric models, such as network models (Borsboom & Cramer, 2013; Fonseca-Pedrero, 2018), or dynamic systems models (Nelson et al., 2017).

All these advances present great opportunities for psychological assessment and have important advantages. As Simmering et al. (2019) point out, first, they facilitate the collection of dynamic, real-time data and allow the inclusion of contextual information. All this can contribute to increase the ecological validity of assessments and avoid the need to respond retrospectively or to generalize from a one-off assessment using a traditional test. Secondly, it facilitates the recording of process information (item response times, pupillary measurements, social interactions) that can be very useful for obtaining evidence of the validity of measures of certain constructs, such as attentional control or, at the group level, trust or leadership. Third, it allows not only the creation of adaptive tests, but can also give individuals personalized feedback based on their responses or overall test scores, which in itself is an intervention that can improve learning and certain behaviors. Also, with outpatient assessment, the almost immediate identification of states or situations at risk for the individual will allow immediate actions to be taken that increase the chances of intervention success. Finally, the possibility of presenting audiovisual or virtual reality items allows the creation of more realistic

situations that increase the ecological validity of the assessments and decrease certain response biases such as social desirability (Woods et al., 2020). In addition, these formats tend to be more motivating for people, who become more involved in the tests, which could contribute to reduce measurement error.

Some of the technological advances mentioned above are already being implemented in our country. For example, there are more adaptive tests marketed (Barrada, 2012), and mobile applications are being created for outpatient assessment, such as, for example, the app for telematic monitoring of children and young people at risk of psychological problems during the recent confinement by COVID-19: <https://www.fbbva.es/noticias/una-app-para-el-seguimiento-de-jovenes-con-problemas-de-salud-mental-durante-el-confinamiento/>). However, according to the survey conducted by Muñiz et al. (2020), most psychologists are still somewhat reluctant to incorporate these types of technological advances. Such reluctance may be justified by the issues that are still not well resolved and the challenges posed by all the aforementioned advances (Iliescu & Greiff, 2019; Simmering et al., 2019; Tonidandel et al., 2002).

One of the main challenges is to ensure that such promising technological advances do not jeopardize the essential: construct validity (Iliescu & Greiff, 2019; Simmering et al., 2019; Woods et al., 2020). No matter how motivating and realistic the tests are, no matter how much objective information they incorporate, it is necessary to ensure the psychometric quality of the scores generated. Some reflections and examples of the psychometric rigor that should accompany technologically innovative tests can be found in the works of this monograph (Abad et al., 2022; Andrés, et al., 2022; Elosua, 2022; Fonseca-Pedrero et al., 2022; Santamaría & Sánchez-Sánchez, 2022; Suárez-Álvarez et al., 2022). For its part, psychometrics will have to advance in the design of new ways of assessing psychometric quality. In addition, the incorporation of technology opens up a whole series of ethical and equity issues, depending on the technological resources and the degree of familiarity with the technology of the people being assessed, without ruling out possible negative reactions to the assessment. Another focus of new problems are security issues, such as the control of items exposed online, the identity of participants who are assessed online, privacy (for example, when data from social networks are used), the maintenance and storage of large amounts of data, etc.

Despite these problems and risks, tests that employ some of the methodological advances presented (computerized tests, automatic scoring, adaptive tests, automated reports, ambulatory assessment) are becoming more and more common in our country. And we have no doubt that many more will be incorporated. However, before adopting a particular technology, psychometricians, test authors, and test publishers must assess the potential costs and benefits of that technology for assessing a given construct in the population of interest. And, above all, the tests that implement these advances must be accompanied by guarantees of psychometric quality, with rigorous studies to back them up.

The Commission should be very attentive to how the different advances are implemented in our country and adopted by publishers and professionals. As the various technologies are incorporated into the professional practice of psychology, it will be necessary to update the test evaluation model (CET-R). Moreover, the translation and

dissemination of the guidelines governing computerized and internet-based assessment (ITC, 2005) and the forthcoming ITC guidelines governing technology-based assessments, and their follow-up, will allow for increased rigor in the construction and implementation of these new-generation tests. However, it should be noted that many of these new forms of assessment use complex algorithms and models that are often beyond the skills and knowledge of psychologists, requiring collaborative work with other professionals, such as programmers and software and hardware engineers. Technology alone is no guarantee of anything (Liem et al., 2018), it is crucial that the incorporation of technology into the field of assessment is always done with psychological and psychometric guarantees, not allowing oneself to be dazzled by the fireworks of technology for technology's sake. Neither data nor technology know psychology, and that knowledge is provided by psychologists; the clear definition of the psychological construct being evaluated, the reliability, and the validity are non-negotiable, to cite just three essential issues. The COP Test Commission, like a hundred-eyed Argus, will do its utmost to ensure that these psychological and psychometric guarantees are met.

#### ACKNOWLEDGMENTS

The authors wish to express their gratitude to all the people who have collaborated with the Test Commission in the development of its projects and activities, with special mention to those who were part of it at some point. Many thanks to all of them, without their help it would not have been possible to develop the activities described in the article.

#### CONFLICT OF INTEREST

The authors declare that there is no conflict of interest.

#### REFERENCES

- Abad, F. J., Schames, R., Sorrel, M., Nájera, P., García-Garzón, E., Garrido, L. E., & Jiménez, M. (2022). Construyendo tests adaptativos de elección forzosa "on the fly" para la medición de la personalidad [Building adaptive forced choice tests "on the fly" for personality measurement]. *Papeles del Psicólogo*, 43(1), 29-35. <https://doi.org/10.23923/pap.psicol.2982>
- American Psychological Association (1992). Ethical principles of psychologists and code of conduct. *American Psychologist*, 47(12), 1597-1611. <https://doi.org/10.1037/0003-066X.47.12.1597>
- American Educational Research Association, American Psychological Association, and National Council on Measurement in Education. (2014). *Standards for educational and psychological testing*. American Psychological Association.
- Andrés, J. C., Aguado, D., & de Miguel, J. (2022). ¿Qué hay detrás de LinkedIn? Midiendo a través de rúbricas las LinkedIn Big Four Dimensions [What's behind LinkedIn? Measuring the LinkedIn Big Four Dimensions through rubrics]. *Papeles del Psicólogo*, 43(1), 12-20. <https://doi.org/10.23923/pap.psicol.2979>
- Barrada, J. R. (2012). Tests adaptativos informatizados: una perspectiva general [Computerized adaptive tests: an overview]. *Anales de Psicología*, 28(1), 289-302.
- Bartram, D. (2002). *Review model for the description and evaluation of psychological tests*. Brussels: European Federation of Psychologists' Associations (EFPA)
- Bartram, D. (2011). Contributions of the EFPA Standing Committee on Test and Testing (SCTT) to standards and good practice. *European Psychologist*, 16(2), 149-159. <https://doi.org/10.1027/1016-9040/a000093>
- Bauman, Z. (2002). *Modernidad líquida [Liquid modernity]*. México: FCE.
- Borsboom D., & Cramer, A. O. J. (2013). Network analysis: An integrative approach to the structure of psychopathology. *Annual Review of Clinical Psychology*, 9, 91-121. <https://doi.org/10.1146/annurev-clinpsy-050212-185608>
- Colegio Oficiales de Psicólogos de Madrid [Psychological Association of Madrid] (COP-Madrid) (2015). *II Encuesta del Observatorio de Reclutamiento y Selección de Personas [II Survey of the Observatory of the Recruitment and Selection of Personnel]*. [https://issuu.com/colegiooficialpsicologosmadrid/docs/presentaci\\_n\\_de\\_resultados\\_2014\\_de](https://issuu.com/colegiooficialpsicologosmadrid/docs/presentaci_n_de_resultados_2014_de)
- Elosua, P. (2012). Tests publicados en España: Usos, Costumbres y Asignaturas Pendientes [Tests published in Spain: Uses, customs, and unresolved matters]. *Papeles del Psicólogo*, 33(1), 12-21.
- Elosua, P. (2019). *Los test psicológicos en la práctica profesional [Psychological tests in professional practice]*. FOCAD. Consejo General de Psicología de España [Spanish Psychological Association].
- Elosua, P. (2021). Aplicación remota de test: riesgos y recomendaciones [The remote application of tests: risks and recommendations]. *Papeles del Psicólogo*, 42, 33-37. <https://doi.org/10.23923/pap.psicol.2952>
- Elosua, P. (2022). Impacto de la TIC en el entorno evaluativo. Innovaciones al servicio de la mejora continua [Impact of ICT on the assessment environment. Innovations for the purpose of continuous improvement]. *Papeles del Psicólogo*, 43(1), 3-11. <https://doi.org/10.23923/pap.psicol.2985>
- 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(2), 82-88.
- Evers, A., Muñoz, J., Bartram, D., Boben, D., Egeland, J., Fernández-Hermida, J.R., & col. (2012). Testing practices in the 21st Century: Developments and European psychologists' opinions. *European Psychologist*, 17(4), 300-319. <https://doi.org/10.1027/1016-9040/a000102>
- Evers, A., Muñoz, J., Hagemester, C., Høstmælingen, A., Lindley, P., Sjöberg, A., & Bartram, B. (2013). Assessing the quality of test: Revision of the EFPA review model. *Psicothema*, 25(3), 283-291. <https://doi.org/10.7334/psicothema2013.97>
- Evers, A., McCormick, C. M., Hawley, L. R., Muñoz, J., Balboni, G., Bartram, D., ... & Zhang, J. (2017). Testing practices and attitudes toward tests and testing: An international survey. *International Journal of Testing*, 17(2), 158-190. <https://doi.org/10.1080/15305058.2016.1216434>
- Fernández-Ballesteros, R., DeBruyn, E.E.J., Godoy, A., Hornke, L., Ter Laak, J. Vizcarro, C., Westhoff, K., Westmeyer H. & Zacagnini, J.L. (2001) Guidelines for the Assessment Process (GAP): A proposal for Discussion. *European Journal of Psychological Assessment*, 17(3), 187-200. <https://doi.org/10.1027/1015-5759.17.3.187>
- Fonseca-Pedrero, E. (2018). Análisis de redes en psicología [Network analysis in psychology]. *Papeles del Psicólogo*, 39(1),

- 1-12. <https://doi.org/10.23923/pap.psicol2018.2852>
- Fonseca, E., & Muñiz, 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(3), 161-168. 7. <https://doi.org/10.23923/pap.psicol2017.2844>
- Fonseca-Pedrero, E., Ródenas, Gabriel, Pérez-Albéniz, A., Al-Halabí, S., Pérez, M., & Muñiz, J. (2022). La hora de la evaluación ambulatoria [The Time of Ambulatory Assessment]. *Papeles del Psicólogo*, 43(1), 21-28 <https://doi.org/10.23923/pap.psicol.2983>
- Gómez Sánchez, L. E. (2019). Séptima evaluación de test editados en España [Seventh review of tests published in Spain]. *Papeles del Psicólogo*, 40(3), 205-210. <https://doi.org/10.23923/pap.psicol2019.2909>
- Hernández, A., Hidalgo, M. D., Hambleton, R. K. & Gómez-Benito, J. (2020). International Test Commission guidelines for test adaptation: A criterion checklist. *Psicothema*, 32(3), 390-398. <https://doi.org/10.7334/psicothema2019.306>
- Hernández, A., Hidalgo, M. D., Muñiz, J., & Fernández-Hermida, J. R. (2021). Diferencias de opinión sobre el uso de los test: su relación con la edad el género y el sector laboral [Differences in opinions regarding testing practices: The role of age, gender and job sector]. *Papeles del Psicólogo*, 42(2), 94-104. <https://doi.org/10.23923/pap.psicol.2960>
- 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), 1-8.
- 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(3), 192-197.
- 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(1), 21-30. <https://doi.org/10.23923/pap.psicol2019.2886>
- Iliescu, D., & Greiff, S. (2019). The impact of technology on psychological testing in practice and policy. *European Journal of Psychological Assessment*, 35(2), 151-155. <https://doi.org/10.1027/1015-5759/a000532>
- International Test Commission (2005). *International Guidelines on Computer-Based and Internet Delivered Testing*. [www.intestcom.org].
- International Test Commission (2015). *International Guidelines for Practitioner Use of Test Revisions, Obsolete Tests, and Test Disposal*. [www.intestcom.org].
- International Test Commission. (2018). *ITC Guidelines for the Large-Scale Assessment of Linguistically and Culturally Diverse Populations*. [www.InTestCom.org]
- ISO (2011). *Procedures and methods to assess people in work and organizational settings (part 1 and 2)*. Geneva: Author.
- Kosinski, M., Stillwell, D., & Graepel, T. (2013). Private traits and attributes are predictable from digital records of human behaviour. *Proceedings of the National Academy of Sciences (PNAS)*, 110, 5802-5805. <https://doi.org/10.1073/pnas.1218772110>
- Liem, C. C. S., Langer, M., Demetriou, A., Wicaksana, A.S., Hiemstra, A. M. F., Born, M. P., & König, C. J. (2018). Psychology meets machine learning: Interdisciplinary perspectives on algorithmic job candidate screening. In H. J. Escalante, S. Escalera, I. Guyon, X. Baró, Y. Güçlütürk, U. Güçlü, & M. A. J. v. Gerven (Eds.), *Explainable and interpretable models in computer vision and machine learning* (pp. 197-251). Springer. [https://doi.org/10.1007/978-3-319-98131-4\\_9](https://doi.org/10.1007/978-3-319-98131-4_9)
- Lindley, P.A., Bartram, D., & Kennedy, N. (2008). *EFPA Review Model for the description and evaluation of psychological tests: Test review form and notes for reviewers: Version 3.42*. Brussels: EFPA Standing Committee on Tests and Testing (September, 2008).
- Moreland, K. L., Eyde, L. D., Robertson, G. J., Primoff, E. S., & Most, R. B. (1995). Assessment of test user qualifications: A research-based measurement procedure. *American Psychologist*, 50(1), 14-23. <https://doi.org/10.1037/0003-066X.50.1.14>.
- Muñiz, J., & Bartram, D. (2007). Improving international test and testing. *European Psychologist*, 12(3), 206-219. <https://doi.org/10.1027/1016-9040.12.3.206>
- Muñiz, J., & Fernández-Hermida, J.R. (2000). La utilización de los test en España [Test use in Spain: the psychologists' viewpoint]. *Papeles del Psicólogo*, 76, 41-49.
- Muñiz, J., & Fernández-Hermida, J.R. (2010). La opinión de los psicólogos españoles sobre el uso de los test [The opinion of Spanish psychologists on the use of tests]. *Papeles del Psicólogo*, 31(1), 108-121.
- Muñiz, J., & Fonseca, E. (2017). *Construcción de instrumentos de medida en psicología [Construction of measurement instruments in psychology]*. Madrid: FOCAD. Consejo General de Psicología de España [Spanish Psychological Association]
- Muñiz, J., Elosua, P., & Hambleton, R.K. (2013). Directrices para la traducción y adaptación de los test: segunda edición [International Test Commission Guidelines for test translation and adaptation: Second edition]. *Psicothema*, 25(2), 151-157. <https://doi.org/10.7334/psicothema2013.24>
- Muñiz, J., Hernández, 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(3), 161-173.
- Muñiz, J., Hernández, A., & Fernández-Hermida, J. R. (2020). Utilización de los test en España: el punto de vista de los psicólogos [Test use in Spain: the psychologists' viewpoint]. *Papeles del Psicólogo*, 41(1), 1-15. <https://dx.doi.org/10.23923/pap.psicol2020.2921>.
- Muñiz, J., Fernández-Hermida, J.R., Fonseca-Pedrero, E., Campillo-Álvarez, A., & Peña-Suárez, E. (2011). Evaluación de test editados en España [Review of tests published in Spain]. *Papeles del Psicólogo*, 32(2), 113-128.
- Muñiz, J., Bartram, D., Evers, A., Boben, D., Matesic, K., Glabeke, K., Fernández-Hermida, J.R., & Zaal, J. (2001). Testing practices in European countries. *European Journal of Psychological Assessment*, 17(3), 201-211. <https://doi.org/10.1027/1015-5759.17.3.201>
- Nelson, B., McGorry, P. D., Wichers, M., Wigman, J. T., & Hartmann, J. A. (2017). Moving from static to dynamic models of the onset of mental disorder. *JAMA Psychiatry*, 74(5), 528-534. <https://doi.org/10.1001/jamapsychiatry.2017.0001>

- Nieto, M. D., Abad, F. J., & Olea, J. (2018). Assessing the Big Five with bifactor computerized adaptive testing. *Psychological Assessment, 30*(12), 1678–1690. <https://doi.org/10.1037/pas0000631>
- Parshall, C. G., Harmes, J. C., Davey, T., & Pashley, P. J. (2010). Innovative Items for Computerized Testing. In W. J. van der Linden, & C. A. W. Glas (Eds.), *Elements of Adaptive Testing, Statistics for Social and Behavioral Sciences* (pp. 215–230). Springer. <https://doi.org/10.1007/978-0-387-85461-8>
- Ponsoda, V., & Hontangas, P. (2013). Segunda evaluación de test editados en España [Second evaluation of tests published in Spain]. *Papeles del Psicólogo, 34*(2), 82-90.
- Prieto, G., & Muñoz, J. (2000). Un modelo para evaluar la calidad de los test utilizados en España [A model for assessing the quality of the tests used in Spain]. *Papeles del Psicólogo, 77*, 65-71.
- Rios, J. A., & Sireci, S. G. (2014). Guidelines versus practices in cross-lingual assessment: A disconcerting disconnect. *International Journal of Testing, 14*(4), 289–312. <https://doi.org/10.1080/15305058.2014.924006>
- Santamaría, P., & Sánchez-Sánchez, F. (2022). Cuestiones abiertas en el uso de las nuevas tecnologías en la evaluación psicológica [Open questions in the use of new technologies in psychological assessment]. *Papeles del Psicólogo, 43*(1), 48-54. <https://doi.org/10.23923/pap.psicol.2984>
- Sanz, S., Luzardo, M., García, C., & Abad, F. J. (2020). Detecting cheating methods on unproctored Internet tests. *Psicothema, 32*(4), 549-558. <https://doi.org/10.7334/psicothema2020.86>
- Seelow, D. (2019). The art of assessment: Using game based assessments to disrupt, innovate, reform and transform testing. *Journal of Applied Testing Technology, 20*(S1), 1-16.
- Séneca, L. A. (2013). *Sobre la brevedad de la vida, el ocio y la felicidad [On the Shortness of Life, Leisure, and Happiness]*. Acanalado.
- Simmering, V. R., Ou, L., & Bolsinova, M. (2019) What Technology Can and Cannot Do to Support Assessment of Non-cognitive Skills. *Frontiers of Psychology, 10*: 2168. <https://doi.org/10.3389/fpsyg.2019.02168>
- Sorrel, M. A., Abad, F. J., & Nájera, P. (2021). Improving Accuracy and Usage by Correctly Selecting: The Effects of Model Selection in Cognitive Diagnosis Computerized Adaptive Testing. *Applied Psychological Measurement, 45*(2), 112-129. <https://doi.org/10.1177/0146621620977682>
- Suárez-Álvarez, J., Fernández-Alonso, R., García-Crespo, F. J., & Muñoz, J. (2022). El uso de las nuevas tecnologías en las evaluaciones educativas: La lectura en un mundo digital [The use of new technologies in educational assessments: Reading in a digital world]. *Papeles del Psicólogo, 43*(1), 36-47. <https://doi.org/10.23923/pap.psicol.2986>
- Taleb, N. (2008). *El cisne negro [The black swan]*. Paidós.
- Tonidandel, S., Quiñones, M. A., & Adams, A. A. (2002). Computer-adaptive testing: The impact of test characteristics on perceived performance and test takers' reactions. *Journal of Applied Psychology, 87*(2), 320–332. <https://doi.org/10.1037/0021-9010.87.2.320>
- Viladrich, C., Doval, E., Penelo, E., Aliaga, J., Espelt, A., García-Rueda, R., & Angulo-Brunet, A. (2021). Octava evaluación de test editados en España: Una experiencia participativa [Eighth review of tests published in Spain: A participative experience]. *Papeles del Psicólogo, 42*(1), 1-9. <https://doi.org/10.23923/pap.psicol.2937>
- Wan, L., & Henly, G. A. (2012). Measurement properties of two innovative item formats in a computer-based test. *Applied Measurement in Education, 25*(1), 58-78. <https://doi.org/10.1080/08957347.2012.635507>
- Woods, S. A., Ahmed, S., Nikolaou, I., Costa, A. C., & Anderson, N. R. (2020). Personnel selection in the digital age: A review of validity and applicant reactions, and future research challenges. *European Journal of Work and Organizational Psychology, 29*(1), 64-77. <https://doi.org/10.1080/1359432X.2019.1681401>