



TEST USE IN SPAIN: THE PSYCHOLOGISTS' VIEWPOINT

José Muñiz¹, Ana Hernández² and José Ramón Fernández-Hermida¹

¹Universidad de Oviedo. ²Universidad de Valencia

Los test son los instrumentos de medida más utilizados por los psicólogos para obtener datos que les ayuden a tomar decisiones sobre las personas. Conocer las opiniones de los profesionales de la psicología acerca de la práctica de los test es fundamental para plantear acciones encaminadas a mejorar su uso en nuestro país. El objetivo del presente trabajo es conocer con detalle lo que piensan los psicólogos españoles sobre el uso de los test. Se aplicó un cuestionario de 30 ítems sobre distintos aspectos de la práctica de los test. Respondieron 1.248 colegiados, 73,7% mujeres, con una edad media de 46,31 años y una desviación típica de 11,08. Los resultados muestran unas actitudes favorables de los psicólogos al uso de los test, si bien se ponen de manifiesto algunos aspectos que es necesario mejorar. Se analizan los resultados y se comentan algunas ideas sobre el futuro de los test.

Palabras clave: Test, Uso de test, Opiniones de los psicólogos, Actitudes hacia los test.

Test Use in Spain: The psychologists' viewpoint. Tests are the measuring instruments most used by psychologists to obtain data that help them make decisions about people. It is essential to understand the opinions of psychology professionals regarding the practical implementation of tests, in order to propose and organize actions aimed at improving their use. The objective of this research is to understand in detail what Spanish psychologists think about the use of tests in our country. A 30-item questionnaire on the different aspects of testing practices was administered. The questionnaire was answered by 1,248 members of the Spanish Psychological Association, 73.7% women, with an average age of 46.31 years, and a standard deviation of 11.08. The results show favorable attitudes of Spanish psychologists to the use of tests in the exercise of their profession, although some aspects that need to be improved are presented. Results are discussed, and some ideas about the future of tests are analyzed.

Key words: Test, Use of tests, Psychologists opinions, Attitudes towards tests.

E EVALUATION PROCESS

Psychologists use tests to obtain behavioral samples that help them make decisions about people, whether in the clinical, work, educational, legal, or many other fields in which they carry out their work. Tests do not have a life of their own. They are not autonomous. They do not make the decisions themselves. Decisions are made by the professionals, naturally, supported by objective data, from evidence, and this is where tests enter the scene. As Paul Meehl (1954) alerted us early, in his disturbing little book, as he called it, psychologists must refrain from making diagnoses and decisions based solely on their clinical judgment; they must be backed up by objective data, and a key source for obtaining these data—while not the only one—is tests. A psychological evaluation process is broader and more complex than simply using the tests correctly (Fernández-Ballesteros, 2014; Fernández-Ballesteros et al., 2001; Groth-Marnat & Wright, 2016; Wright, 2011). It requires responding to many other crucial questions, including the following: *what* is being evaluated and *why*, *who* is doing the evaluating, *what* are the parties that are

legitimately involved in the assessment, *how* is the evaluation done, *how* will the information be provided to stakeholders (feedback), *what improvement plans and interventions* derive from the assessment, and finally, *how* do the different agents involved perceive the assessment. The first thing is to make it very clear *what* is being evaluated and *why*, as this will determine the type of instruments and strategies to follow. As our own Seneca said, favorable winds never blow for those who do not know which port they are heading to. Naturally, what is evaluated and why it is done will vary in each situation. There are no valid recipes for all cases; it will be the professionals who have to make decisions in each case. *Who* is responsible for the assessment is the second issue that should be clear; there will be situations where it is obvious, but in others there may be doubts. Whenever it is a question of evaluating psychological variables, the person responsible must be a psychology professional, although some auxiliary tasks can be delegated, for example the administration of the tests, but never the interpretation of the results or producing a diagnosis. Thirdly, in an evaluation process, *all the parties legitimately involved* in the evaluation must be identified, which will vary according to each situation. Failure to identify these parties correctly can ruin the evaluation, regardless of its technical design. Thus, for example, in an educational context it will be necessary to take into account,

Received: 15 November 2019 - Accepted: 2 December 2019

Correspondence: José Muñiz. Facultad de Psicología. Universidad de Oviedo. Plaza Feijoo, s/n. 33003 Oviedo. España.

E-mail: jmuniz@uniovi.es



of course, the students evaluated, but also parents, teachers, school management, educational administration, legal representatives of teachers, among others. Everyone plays their legitimate role. They have to be taken into account, since any of them can make an evaluation fail, it is an established fact that it is easier to destroy than to build. Fourth, there is the central point from the technical point of view, that is, *how* the evaluation is carried out, what measurement instruments, and what resources are to be used. It goes without saying that professionals must use tests that have adequate psychometric properties and they must use them correctly. We currently have numerous sources on the construction and analysis of test quality (American Educational Research Association, American Psychological Association, & National Council on Measurement in Education, 2014; Downing & Haladyna, 2006; Haladyna & Rodríguez, 2013; Irwing, Booth, & Hughes, 2018; Lane, Raymond, & Haladyna, 2016; Muñiz, 2018; Muñiz & Fonseca-Pedrero, 2019; Schmeiser & Welch, 2006). Many evaluations that successfully carry out the four phases described, unfortunately fail in the fifth, when it comes to providing adequate information (feedback) to all parties legitimately involved. When designing an assessment, it is necessary to be very clear from the beginning the type of report that is going to be made for each intervening party. Whilst it is a very common mistake to think that a report model is sufficient for everyone, one-size-fits-all does not work here. It is not uncommon for a well-designed and well-executed assessment to spoil when it comes to *communicating the results*. We have to be very careful with the reports, and make sure that the recipients understand them, because a good assessment poorly communicated automatically becomes a poor assessment. For example, the work of Goodman and Hambleton (2004), Ryan (2006), or Zenisky and Hambleton (2016) can be seen on the problem of reporting. Sixth, professionals should be aware that one does not assess for assessment's sake; the most logical thing is that the assessment leads to the development of an improvement or intervention plan. When improvement plans are carried out, they must be viable, evidence-based, and evaluable, in order to empirically prove their effectiveness. Finally, something that is too often forgotten is the need to objectively and systematically seek the opinion of the parties involved regarding the assessment. The participants' perception is key, as it will significantly modulate their success. Participants do not have to be experts in technical matters, they almost never will be, but that does not mean their opinion is less important; it is not necessary to be a shoemaker to give an opinion about shoes. In sum, in the preceding lines, the intention is to highlight that the development of an assessment process is much broader and more complex than the mere application of tests, although these are crucial in the process. That said, we will now focus precisely on the problem involved in the use of the tests as assessment and diagnostic tools.

TEST USE

Improper use of tests can cause serious harm to the people assessed. For this reason, different national and international organizations have been developing actions and programs to try to improve three essential aspects: the training of professionals, the quality of the tests, and the use made of them. The actions carried out by the Spanish Psychological Association (COP in Spanish) and by international organizations such as the International Test Commission (ITC), or the European Federation of Psychologists Associations (EFPA), can be grouped into two large strategies: restrictive and informative. *Restrictive* actions refer to the limiting or restricting of the use of the tests to people who have adequate training. This varies greatly from one country to another (Bartram, 1996; Bartram & Coyne, 1998; Evers et al., 2012, 2017; Muñiz, Prieto, Almeida, & Bartram, 1999), for example, Spanish publishers classify the tests into three categories (A, B, C) depending on their specialization, and those in categories B (collective cognitive and personality tests) and C (individual scales and projective tests) are sold to those who prove they are psychologists. This and other measures, although adequate, do not constitute a total guarantee, so they must be complemented with other *informative* ones. The idea is that both professionals and users have all the necessary information about the tests, in order to minimize their inappropriate use. These actions include the publication of ethical codes, such as that of the EFPA (2005), or the guidelines of the European Association of Psychological Assessment (Fernández-Ballesteros et al., 2001). A good compendium in Spanish can be seen in the monographic issue of the journal *Papeles del Psicólogo* (2009) dedicated to the subject, or in authors such as Knapp (2012), Koocher and Keith-Spiegel (2007), Lindsay, Koene, Ovreeide, and Lang (2008), Leach and Oakland (2007), or Muñiz (1997). As well as these ethical codes there is a whole set of guidelines to guide the construction, analysis, use, and quality assessment of the tests, including the technical standards developed by the American Psychological Association and two other organizations (AERA, APA, & NCME, 2014), as well as the guidelines developed by the ITC for the translation and adaptation of tests from one culture to another (Hambleton, Merenda, & Spielberger, 2005; ITC, 2017; Muñiz, Elosua & Hambleton, 2013), or for use in research, quality control, and security (Muñiz, Hernández, & Ponsoda, 2015). For other guidelines of interest, see, for example, the work of Muñiz and Bartram (2007), the websites of the ITC (www.intestcom.org), the EFPA (www.efpa.eu), or the COP (www.cop.es) in the section dedicated to the *Test Commission*.

Warranting special mention is the standard ISO 10667 (AENOR, www.aenor.es), which regulates everything related to the evaluation of people in work settings, and in whose preparation the COP actively participated, forming part of the international commission that developed it. The standard describes the competencies, obligations, and responsibilities



of the clients and the providers of the assessment service, before, during and after the assessment process. It also provides guidelines for all parties involved in the assessment process, including the person evaluated and those who receive the results of the assessment (Muñiz & Fernández-Hermida, 2010). The standard is having a great influence on psychological assessment in work contexts, because although it does not have the status of a law as such, it provides a rigorous protocol of action and gives prestige to the companies that adopt it. Companies and institutions can be certified in the corresponding agencies if they demonstrate to the auditors that they follow the evaluation processes established in the standard.

Another key activity within the strategy of disseminating rigorous information about the tests is the publication of the results of their assessment by experts to help professionals determine which are the most appropriate tests for their purposes. There are two major models for evaluating the quality of the tests, one of a more qualitative nature, used by the Buros Center for Testing in the United States, which is a literary review in which the strengths and weaknesses of the test being evaluated are discussed (Carlson & Geisinger, 2012; Elosua & Geisinger, 2016; Geisinger, 2012, 2016), and the other European, of a mixed qualitative-quantitative nature. Each of the psychometric characteristics of the test is scored from 1 to 5 and these numerical assessments are complemented by a qualitative evaluation of the main aspects of the test. The European model can be found on the EFPA website (www.efpa.eu), and the details of its revision can be found in Evers et al. (2013). The Spanish version of the Revised Test Review Questionnaire (CET-R) can be found on the COP website (www.cop.es) and the adaptation process followed can be seen in Hernández, Ponsoda, Muñiz, Prieto, and Elosua (2016). Using this CET-R model, seven annual test reviews have already been carried out in Spain, the eighth being currently underway (Elosua & Geisinger, 2016; Fonseca-Pedrero & Muñiz, 2017; Gómez-Sánchez, 2019; Hernández, Tomás, Ferreres, & Lloret, 2015; Hidalgo & Hernández, 2019; Muñiz et al., 2011; Ponsoda & Hontagas, 2013). To date, a total of seventy-two tests have been reviewed, and the reviews can be freely consulted on the COP's website. Ideally, most of the tests published in our country should be reviewed, as is the case in the Netherlands, however, in the European context we are one of the countries with the most advanced test evaluation projects. We must highlight the role of the Spanish test editors (CEPE, GiuntiEOS, Pearson, and TEA), who have at all times collaborated in the reviews, and offered their help and experience to improve the quality of the tests they edit, which will have repercussions for the benefit of all: the practitioners, the users, and the publishers themselves.

This review offers two great advantages, on the one hand, it provides first-hand information to professionals about the quality of the tests, and, on the other, it serves as a guide for constructors and editors of tests, guiding them on the

psychometric characteristics required for the tests. A third aspect that is not negligible is the review's importance in the teaching and training of future psychologists (Hidalgo & Hernández, 2019).

OBJECTIVE

Within the context of actions aimed at improving testing practice, the EFPA has promoted the application of a survey to garner the opinions of psychologists on the use of tests in each country. Understanding these opinions is key to promoting actions and projects on those aspects that are perceived as deficits by professionals. The first data, from six European countries, were obtained in 2000 (Muñiz et al., 2001), and a good overview of the results attained after that in Europe can be seen in Evers, Muñiz, Bartram, et al. (2012). The same survey was also used in other non-European countries (Evers, McCormick, Hawley, Muñiz, et al., 2017). In Spain, the first survey to determine the opinions of professionals about the tests was carried out in 1999 (Muñiz & Fernández-Hermida, 2000), and was repeated again ten years later, updating the questionnaire initially used (Muñiz & Fernández-Hermida, 2010). As can be seen in the works cited, in general the opinions of Spanish psychologists about the use of tests in our country are quite positive, although some aspects for improvement have been identified.

A decade after the last survey was applied, the main objective of this paper is to gather the opinions of Spanish psychologists about the use of tests in our country and their evolution. Knowing these opinions is a fundamental first step in order to be able to carry out actions and projects aimed at improving everything related to testing practices. It is, in short, about finding answers to questions such as what opinion do psychologists have about the tests? Do they think the tests are used properly? Do they think tests are expensive? What are the most important problems? Do they know about the annual test review carried out by the COP? What do they think about it?

SURVEY

The questionnaire used in 2019 is the same as the one that was used ten years ago with a number of modifications, some aimed at clarifying certain aspects of the tests that had not been included in the previous questionnaire (items 25, 26, 27, 28-9, 28-10, 28-11), and others to obtain the opinion of psychologists about the process of the annual test review carried out by the COP since 2011 (items 29-1, 29-2, 29-3, 29-4, 29-5). All closed response items have a Likert format with five categories, except for item 29-1 which is dichotomous. Precisely in this item, if the person responds that they do not know the annual test review carried out by the COP, the computer application skips the following four questions (29.2, 29.3, 29.4, 29.5), related to said review, and passes directly to item 30, which asks about the three most used tests by respondents. The survey used can be seen in Annex 1.



PARTICIPANTS

Of a total of 75,135 members of the COP, 1,248 responded to the survey. The main characteristics of the participants appear in Tables 1 and 2. The sample consists of 73.7% of women and 26.3% of men, and given that the COP is composed of 80.57% of women and 19.43% of men, there was a greater participation of men in the survey. Regarding age, 6.3% were between 23 and 29 years old, 24.1% between 30 and 39, 27.4% between 40 and 49, 29.1% between 50 and 59, 12, 4% between 60 and 69, and 0.7% were 70 years of age or older. The average age of the participants was 46.3 years, with a standard deviation of 11.1 years. This average is similar to the average age of the COP members, which is 44.19 years with a standard deviation of 12.2. These data seem to indicate that the sample used is a reasonably good representation of the population of members of the COP, although a greater number of participants would have been desirable. Regarding the specialties, Clinical-Health predominates (67.6%), followed by Educational (12.7%), and Work (5.5%). Other specialties, such as sports, legal, traffic, social services, etc., constitute the remaining 14.1%. It is interesting to note that in Clinical and Educational the percentage of men is 24.6% and 25.8%, respectively, while in the Work area this percentage increases to 48%. The majority work in the private sector (69.9%), regardless of the specialty: Clinical-Health (74.5%), Educational (66.7%), and Work (62.3%). Out of all the

respondents, 91.2% work as psychologists, and only 8.8% do other jobs.

PROCEDURE

The survey was applied online, through a link on the COP website. The computer system used verified that the person was registered with the COP, and it did not allow the possibility to answer the survey more than once. Different means such as Infocop, the Agenda, or the National Congress of Psychology were used to disseminate it among the members, its objectives were explained, and participation requested. An email was also sent to the members, informing them of the survey and encouraging them to answer it. Upon entering the survey, participants gave their informed consent and the confidential treatment of the data was guaranteed.

DATA ANALYSIS

Descriptive statistics of the items and the factor dimensions obtained were calculated. Reliability was assessed using the Alpha coefficient (Cronbach, 1951). Analysis of variance was used to study the differences between the means of the specialties. The dimensional structure was estimated by means of a Principal Components analysis, using *Parallel Analysis* on a polychoric correlation matrix, with oblique rotation (Calderón, Navarro, Lorenzo-Seva, & Ferrando, 2019; Ferrando & Lorenzo-Seva, 2017; Ferrando & Anguiano, 2010; Izquierdo, Olea, & Abad, 2014; Lloret, Ferreres, Hernández, & Tomás, 2014). The analyses were carried out with the programs SPSS (IBM, 2016) and FACTOR (Ferrando & Lorenzo-Seva, 2017; Lorenzo-Seva & Ferrando, 2013).

DIMENSIONS EVALUATED BY THE QUESTIONNAIRE

Five main dimensions were obtained (Table 3) that explain 51.63% of the total variance.

The *first* dimension refers to the attitudes of psychologists towards the tests, as they are considered a useful tool (items 21-26, plus item 27, which has a weak factor loading). The *second* dimension groups the items related to the problems of use of the tests and covers the 11 problems presented in the questionnaire (items 28.1 to 28.11). The *third* dimension includes the items related to the regulation of the tests and their use. As expected, the items referring to the need for legislation and regulation of the use of the tests presented positive loadings, while those referring to permissiveness in their use presented negative loadings. The *fourth* dimension refers to training and knowledge about the tests (items 1, 2, 4, and 6). Finally, the *fifth* dimension includes the use of internet, computerization, and automation (items 5, 7, 10, 13, 15, 17, and 20). The items that present positive loadings are those that call into question the validity and potential of these technological advances, while those that present negative loadings reflect a recognition of the advantages of computerized tests and the use of the internet. The aforementioned fourth component, referring to training and

**TABLE 1
DESCRIPTION OF THE SAMPLE AND THE
POPULATION SURVEYED**

	Sample	Spanish Psychological Association
Participants	1,248	75,135
Women	73.7%	80.57%
Men	26.3%	19.43%
Average age (SD)	46.31 (11.08)	44.19 (12.20)
Years in the association (DT)	17.45 (10.96)	12.80 (10.38)

SD: standard deviation

**TABLE 2
DESCRIPTION OF THE SAMPLE ACCORDING TO DISTRIBUTION
BY PROFESSIONAL FIELD AND SECTOR**

		%
Professional field	Clinical and healthcare	67.6
	Education	12.7
	Work	5.5
	Others	14.1
	Sector	Public
	Private	69.9
	Unemployed	3.5



knowledge about the tests, is the one with the lowest reliability, and can be attributed to a large extent to the small number of items that make up this dimension. In sum, taking into account the results obtained, it seems reasonable to think that the items of the questionnaire are organized around five broad dimensions, which could be described as: attitudes towards the tests, problems of use, regulation of the use of the tests, training-knowledge, and the use of internet, computerization, and automation. It should be noted that in the previous editions of the survey, eight dimensions were obtained instead of five as now, because the method of extracting dimensions used then (Kaiser's rule) tends to generate more fragmented factor structures. Although it is not included here due to parsimony, the data has also been analyzed using this method and in general the results confirm those obtained in previous editions.

OPINIONS REGARDING THE USE OF THE TESTS

Table 4 shows the means and standard deviations of the participants' responses to the questionnaire items. The data are provided for the total sample and divided by the professional specialties of Clinical-Health, Educational, and Work. The differences between the means of the specialties were studied using analysis of variance. When the assumption of variance homogeneity was not fulfilled (items 5, 11, 15, 17, 19, 21, 22, 23, 24, 26, and 28-2), the robust Brown-Forsythe test was used for comparisons. The asterisk after the text of the item indicates that there were significant differences between specialties ($p < 0.05$), and the pairs of comparisons that were statistically significant are in brackets (the C refers to Clinical-Health, the E to Education and the T to Work).

The data in the table speak for themselves, but some notable aspects are worth mentioning. The first (items 1 and 6) is that the members consider, with very good criteria, that the knowledge obtained in the psychology degree is not sufficient for the correct use of most of the tests, and therefore, it is necessary to continue training. Professionals also warn that computerized tests (item 5) are gradually replacing the paper-and-pencil classics, and this is occurring with greater intensity in the educational field, and especially in the workplace. There is a clear consensus on the need for it to be psychologists who use and interpret the test results (items 8 and 9). Psychologists are very aware of the possibilities of fraud that open up with the online application of tests (items 17 and 20), and demand that the COP play a more active role in improving the use of tests (item 19). Psychologists routinely use tests in the exercise of their profession (item 21), especially in the educational field, followed by clinical and work, and they have no doubts about their usefulness if they are used properly (items 22 and 23). They also have the perception that in the last decade the use of tests has improved in Spain (item 24). As for the problems in test use in our country (items 28.1 to 28.11), the perception of respondents is quite positive, whilst there are, of course, complaints about the high price of the tests, especially in clinical and educational fields.

Table 5 shows the means and standard deviations of the participants' responses to the five dimensions of the questionnaire. Items that presented negative loadings in dimensions three and five were reversed before

**TABLE 3
FACTOR ANALYSIS OF THE ITEMS OF THE QUESTIONNAIRE**

Dimensions					
Ítems	I	II	III	IV	V
21	0.86				
22	0.85				
23	0.85				
24	0.61				
25	0.72				
26	0.81				
27	0.27				
28-1		0.65			
28-2		0.84			
28-3		0.84			
28-4		0.78			
28-5		0.82			
28-6		0.76			
28-7		0.78			
28-8		0.83			
28-9		0.84			
28-10		0.80			
28-11		0.35			
3			0.55		
8			0.61		
9			0.52		
11			0.66		
12			0.78		
14			-0.28		
16			-0.52		
18			-0.54		
19			0.72		
1				0.89	
2				0.73	
4				0.45	
6				0.74	
5					-0.43
7					-0.76
10					0.48
13					0.66
15					-0.72
17					0.64
20					0.69
% variance	17.53	14.64	8.18	5.87	5.41
% accumulated	17.53	32.17	40.35	46.22	51.63
Coefficient Alpha	0.84	0.91	0.69	0.62	0.70

Note. Weights below 0.45 were eliminated, except when the variable did not reach that weight.



TABLE 4
MEAN AND STANDARD DEVIATION OF EACH OF THE ITEMS OF THE SURVEY BY SPECIALTY
(CLINICAL-HEALTH, EDUCATIONAL, WORK) AND GLOBAL

Items	Clinical-healthcare N=844		Education N=159		Work N=69		Global N=1.248	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD
1.- The training received in Psychology Bachelor's Degree courses is sufficient for the correct use of most tests	2.46	1.16	2.62	1.14	2.64	1.19	2.47	1.16
2.- The training received in Psychology Master's Degree courses is sufficient for the correct use of most tests	3.05	1.07	3.13	0.96	2.94	1.11	3.05	1.05
3.- The European Federation of Psychologists' Associations (EFPA) should establish a system to accredit the skills of test users* (C-E)	3.34	1.29	3.60	1.24	3.58	1.36	3.39	1.30
4.- Professionals are provided with sufficient information (independent reviews, research, etc.) on the quality of tests published in our country	3.01	1.10	3.21	.96	2.91	1.10	3.04	1.09
5.- In my professional field, computerized tests are gradually replacing paper-and-pencil tests* (C-E, C-T)	3.27	1.30	3.62	1.10	3.90	1.07	3.36	1.29
6.- My current knowledge with regard to tests is basically that which I learned on my psychology degree courses	2.30	1.20	2.42	1.21	2.57	1.25	2.33	1.21
7.- The administration of tests on the Internet has many advantages compared to the classic paper-and-pencil administration* (C-T, E-T)	3.14	1.18	3.20	1.08	3.59	1.10	3.16	1.18
8.- The use of psychological tests should be restricted to qualified psychologists	4.23	1.02	4.28	0.98	4.41	1.12	4.24	1.03
9.- While non-psychologists may administer and score tests, interpretation and feedback should be restricted to psychologists	4.39	1.13	4.36	1.13	4.43	1.12	4.38	1.14
10.- Reports generated automatically by computer have no validity	2.81	1.15	2.68	0.97	2.81	1.14	2.78	1.12
11.- The standards (e.g., European Federation of Psychologists' Association [EFPA], American Psychological Association [APA]) defining the minimum technical qualities of a test should be enforceable	4.15	0.89	4.18	0.77	4.32	0.81	4.18	0.87
12.- Legislation is needed to control the more serious abuses of testing* (C-T, E-T)	3.99	1.01	3.91	1.04	4.36	0.79	4.02	1.01
13.- The application of the tests on the Internet puts some people evaluated at a disadvantage	3.45	1.09	3.28	1.04	3.23	1.02	3.41	1.10
14.- Anyone who can demonstrate their competence as a test user (whether a psychologist or not) should be allowed to use tests	2.09	1.30	2.05	1.18	2.46	1.40	2.08	1.28
15.- If used properly, the Internet can greatly improve the quality of test application* (C-T, E-T)	3.33	1.09	3.30	0.95	3.83	0.82	3.36	1.08
16.- Controls on tests and testing should be minimal, as controls discourage the development of new ideas and new procedures	1.84	0.97	1.87	0.93	1.74	0.97	1.83	0.96
17.- Online test application does not allow users to protect their privacy	2.45	1.19	2.33	1.09	2.16	1.12	2.41	1.18
18.- Publishers should be allowed to sell whatever tests they think fit	1.75	1.01	1.80	1.03	1.74	1.08	1.76	1.02
19.- The Spanish Psychological Association should take a more active role in the regulation and improvement of test use	4.09	0.99	3.99	0.91	4.25	0.85	4.09	0.99
20.- The online application of tests opens up possibilities of fraud	3.57	1.14	3.36	1.18	3.52	0.99	3.52	1.14
21.- I use tests regularly in the exercise of my profession* (C-E, E-T)	3.97	1.19	4.32	0.96	3.72	1.35	4.02	1.18
22.- Tests constitute an excellent source of information if they are combined and complemented with other psychological data (C-E, C-T)	4.41	0.90	4.61	0.72	4.68	0.58	4.47	0.85
23.- Used correctly, tests are of great help to the psychologist* (C-E, C-T)	4.40	0.90	4.63	0.71	4.67	0.56	4.46	0.85
24.- Taking into account all aspects, I believe that in the last decade the use of tests has improved in my country* (C-E)	3.67	0.95	3.86	0.87	3.75	0.88	3.73	0.93
25.- The professional practice based on evidence enhances the use of tests as sources of information*	3.89	0.96	4.07	0.82	4.13	0.78	3.96	0.92



26.- In professional decision making I give an important weight to the data obtained through tests* (C-E)	3.36	1.04	3.86	0.78	3.57	0.81	3.46	1.00
27.- Before using a test I check its psychometric properties	3.83	1.16	3.87	1.09	4.12	1.04	3.89	1.15
28-1. Making photocopies of copyrighted materials	3.11	1.43	3.03	1.33	2.78	1.49	3.05	1.42
28-2 Making evaluations using inappropriate tests	2.39	1.29	2.48	1.29	2.68	1.50	2.44	1.31
28-3 Not keeping up with the field	3.06	1.26	3.08	1.20	3.16	1.36	3.08	1.26
28-4 Failing to check one's own interpretations with others	3.06	1.27	2.92	1.27	3.20	1.28	3.04	1.28
28-5 Not considering errors of measurement of a test score	2.98	1.26	3.00	1.16	3.04	1.27	3.01	1.25
28-6 Not restricting test administration to qualified personnel	2.88	1.49	2.95	1.40	3.12	1.59	2.88	1.48
28-7 Not taking into account conditions that cast doubt on reported validity for a local situation	2.86	1.36	2.96	1.23	3.00	1.44	2.89	1.33
28-8. Making interpretations that go beyond the limits of the test	2.88	1.39	2.82	1.27	2.96	1.38	2.87	1.36
28-9. Using tests with inappropriate scales	2.51	1.30	2.43	1.27	2.59	1.33	2.50	1.30
28-10. Using some items without the corresponding permissions	2.26	1.30	2.14	1.17	2.30	1.22	2.25	1.28
28-11. Having a very high price* (C-T, E-T)	4.25	1.06	4.25	1.04	3.65	1.21	4.17	1.11
SD: standard deviation Note. The asterisk indicates that the differences in means between the specialties are statistically significant (p <0.05). The letter pairs in parentheses indicate between which specific specialties there are significant differences: Clinical (C), Educational (E), and Work (T).								

calculating the total score. In this way, a higher score indicates a more positive opinion towards the control and regulation of the tests and their use, and a more negative opinion towards the use of technology (use of internet, computerized tests, automated reports). The global results are offered as well as those for the three specialties, whose averages are compared using an ANOVA, with the exception of the first dimension, for which the Brown-Forsythe test was used because the assumption of variance homogeneity was not met.

As can be seen in Table 5, the attitudes of Spanish

psychologists towards the tests are quite favorable, with a global average of 4 points on a scale ranging from 1 to 5. Specialists in education and work are somewhat more favorable than clinicians, which has its logic, given the nature of the professional practice of the specialties. As already mentioned in our comments on the results broken down by items, Spanish psychologists do not have serious problems using the tests, with an average score of 2.93. They show a clear position in favor of greater control and regulation of the use of the tests, with an average score of 4.07 in that dimension. They clearly demand a need for continuous

TABLE 5
MEAN AND STANDARD DEVIATION OF THE DIMENSIONS BY SPECIALTIES
(CLINICAL-HEALTH, EDUCATIONAL, WORK)

Dimensions	Clinical-healthcare N=844		Education N=159		Work N=69		Global N=1.248	
	Media	DT	Media	DT	Media	DT	Media	DT
1. Attitudes towards the tests* (C-E)	3.94	0.74	4.17	0.57	4.09	0.56	4.00	0.70
2. Problems in the use of the tests	2.93	0.94	2.91	0.89	2.95	1.07	2.93	0.94
3. Control and regulation of test and its use	4.05	0.57	4.08	0.61	4.14	0.56	4.07	0.58
4. Training and knowledge about the tests	2.70	0.78	2.85	0.72	2.76	0.79	2.72	0.77
5. Internet, computerization and automation* (C-E, C-T)	2.93	0.70	2.79	0.64	2.63	0.59	2.89	0.69
Note. The asterisk indicates that the differences in means between the specialties are statistically significant (p <0.05). The letter pairs in brackets indicate specific specialties between which there are significant differences: Clinical (C), Educational (E), and Work (T).								



training, as shown by an average of 2.72 in dimension 4. And they are not very enthusiastic about the use of new technologies (internet, computerization, automation). In particular, clinical specialists are somewhat more skeptical about the usefulness of technological advances and perceive more problems in their use (2.93), than specialists in Education (2.79) and Work (2.63).

KNOWLEDGE OF THE TEST REVIEW

Regarding the question of whether they know about the annual test review carried out by the COP, it is surprising to see that only 22.5% of the members say they know about it. Therefore, there are 77.5% who do not know this important review, so it is clear the need to persist in its dissemination. Among those who are familiar with it, they think it is important and necessary, as shown in the data in Table 6.

EVOLUTION OF OPINIONS

Table 7 shows the average scores for each item in the three reviews carried out in the years 2000, 2010, and 2020. Since the surveys used in the three reviews do not have exactly the same items, the results are only provided for the items common to all three. As it can be seen, the most outstanding general feature is stability, but with some nuances worth mentioning. The correlations between the results of the three evaluations are very high: 0.986 between the 1st and 2nd, 0.955 between the 1st and 3rd, and 0.982 between the 2nd and 3rd.

The first aspect to highlight is that Spanish psychologists appreciate that the information they have about the quality of the tests (item 4) has improved, going from an average value of 2.73 in 2010 to an average of 3.04 in 2020. Undoubtedly, the work of the test reviews carried out by the COP over the last decade is in line with this perception. Another aspect to highlight is the increase in the use of tests in the exercise of the profession, going from a value of 3.76 in 2010 to 4.02 in 2020. This constitutes an indirect indicator of the positive attitude towards tests, as well as a

clear commitment to an exercise of the profession based on empirical evidence obtained from the tests. It should be noted, thirdly, that all items related to problems in the use of tests (item 28) have lowered their average values, which means that professionals perceive a general improvement in test use. The most significant change is in item 28.1, related to photocopies of the tests, which went from 3.51 in 2010 to 3.05 in 2020. It is likely that the increase in online tests and computerized corrections has favored this perception. There is also the perception that the tests are better suited to the local conditions (country, region) than a decade ago, going from an average of 3.21 to 2.89 when assessing the problem of lack of fit. The efforts of national organizations such as the COP, international organizations, such as ITC, and the editors themselves, to translate and adapt the tests to the conditions of the populations evaluated seem to be taking effect.

TESTS MOST USED IN SPAIN

Table 8 shows the results referring to the most commonly used tests in the professional daily practice of psychologists. It should be noted, in the first place, that the 25 most used tests are all psychometric, with projective tests relegated to much lower positions. As can be observed, first of all there is the WISC intelligence test for children, followed by the Millon Clinical Multiaxial Inventory (MCMI) and the Spanish adaptation of the Personality Assessment Inventory (PAI). It is

TABLE 6
OPINIONS OF PSYCHOLOGISTS ON THE TEST QUALITY REVIEW CARRIED OUT ANNUALLY BY THE SPANISH PSYCHOLOGICAL ASSOCIATION

Items	Mean	SD
I think it is important to improve the quality of the tests	4.46	.72
I check the test reviews on the Spanish Psychological Association website	3.81	1.15
This helps me in choosing the tests I use	3.65	1.10
This is unnecessary; psychologists already have enough information to know which test to use	1.62	1.01

SD: standard deviation

TABLE 7
MEANS OF THE ITEMS OBTAINED IN THE SURVEYS OF THE YEARS 2000, 2010, AND 2020

Items	Evaluation-1 2000	Evaluation-2 2010	Evaluation-3 2020
1	2.41	2.43	2.47
4	2.38	2.73	3.04
6	2.57	2.59	2.33
8	4.23	4.12	4.24
9	4.34	4.39	4.38
11	4.33	4.10	4.18
12	4.29	4.01	4.02
14	2.42	2.10	2.08
16	1.85	1.93	1.83
18	1.57	1.77	1.76
19	4.15	4.13	4.09
21	3.56	3.76	4.02
22	4.41	4.46	4.47
23	4.37	4.41	4.46
28-1	3.60	3.51	3.05
28-2	2.63	2.64	2.44
28-5	3.07	3.10	3.01
28-6	2.91	2.92	2.88
28-7	3.28	3.21	2.89
28-8	2.99	2.97	2.87



also noteworthy that all the tests have been adapted to the Spanish context, and 6 of the 25 have been developed by Spanish authors, a representation of 24%, which is not bad. Comparing this evaluation with that of ten years ago, ten new tests have been added to the list, indicated in Table 8 with an asterisk (PAI, SENA, D2, ENFEN, LSB-50, CARAS, CompeTEA, RIAS, BAI, GDS), some of which were published after the 2010 survey (PAI, SENA, LSB-50). These changes in the list of the most used tests constitute a good indicator of the great activity that is taking place in the field of the construction and publication of tests in Spain.

When the use is analyzed by specialty (Table 9), it is observed, as it cannot be otherwise, that the tests vary significantly. Thus, for example, in the Clinical and Educational areas, the evaluation of cognitive aspects (WISC) appears first, while in Work it is Personality (16PF) that is first.

SOME GENERAL CONCLUSIONS

This is the third survey of Spanish psychologists about their opinions on different issues related to test practice in our country, the other two dating from 2000 and 2010. A total of

**TABLE 9
THE TEN MOST USED TESTS BY SPANISH PSYCHOLOGISTS BY SPECIALTY**

	Clinical-Healthcare N=844		Education N=159		Work N=69	
1	WISC	19.9%	WISC	64.8%	16PF	27.5%
2	MCMI	14.1%	PROLEC	26.4%	CompeTEA	21.7%
3	BDI	14.0%	SENA	12.6%	CTC	8.7%
4	PAI	12.4%	D2	11.3%	PAI	7.2%
5	WAIS	12.4%	WAIS	10.1%	BAT-7	5.8%
6	16PF	10.9%	BADYG	9.4%	BIP	5.8%
7	MMPI	10.6%	ENFEN	7.5%	MMPI	5.8%
8	STAI	10.1%	TAMAI	6.2%	TPT	5.8%
9	SENA	8.8%	CARAS	5.7%	MCMI	4.3%
10	CUIDA	7.2%	16PF	5.0%	PAPI	4.3%

Notes: The acronyms of the tests in this table that are not described in Table 8 are identified below: BADYG (Battery for evaluation of differential and general aptitudes); BAT-7 (TEA Skills Battery); BIP (Bochum Personality and Skills Inventory); CTC: Clinical TEA Questionnaire; PAPI (The Personality and Preference Inventory); TAMAI (Multifactor Self-Assessment Test of Child Adaptation); TPT (TEA Personality Test)

**TABLE 8
THE 25 TESTS MOST USED BY SPANISH PSYCHOLOGISTS**

Name of the test		N	%
WISC	Weschler Intelligence Scale for Children	289	23.2%
MCMI	Millon Clinical Multiaxial Inventory	157	12.6%
PAI*	Personality Assessment Interview	156	12.5%
16PF	16 Personality Factors	142	11.4%
WAIS	Weschler Adult Intelligence Scale	138	11.1%
BDI	Beck Depression Inventory	122	9.8%
MMPI	Minnesota Multiphasic Personality Inventory	120	9.6%
CUIDA	Evaluación de Adoptantes, Cuidadores, Tutores y Mediadores [Evaluation of Adopters, Caregivers, Tutors, and Mediators]	110	8.8%
SENA*	Sistema de Evaluación de Niños y Adolescentes [Evaluation System for Children and Adolescents]	107	8.6%
STAI	State Trait Anxiety Inventory	99	7.9%
PROLEC	Batería de Evaluación de los Procesos Lectores [Battery for the Assessment of Reading Processes]	71	5.7%
MMSE	Mini Mental State Examination	54	4.3%
SCL-90	Symptom Checklist 90	51	4.1%
D2*	Test de Atención [Attention Test]	44	3.5%
BASC	Behavior Assessment System for Children	42	3.4%
ENFEN*	Evaluación Neuropsicológica de las Funciones Ejecutivas en Niños [Neuropsychological Evaluation of Executive Functions in Children]	35	2.8%
LSB-50*	Listado de Síntomas Breve [Brief Symptoms List]	30	2.4%
RAVEN	Matrices Progresivas de Raven [Raven Progressive Matrices]	28	2.2%
CARAS*	Test de Percepción de Diferencias [Difference Perception Test]	27	2.2%
CompeTEA*	Evaluación de Competencias [Skills assessment]	26	2.1%
NEO-PI	NEO Personality Inventory	24	1.9%
RIAS*	Reynolds Intellectual Assessment Scales	23	1.8%
BAI*	Beck Anxiety Inventory	20	1.6%
GDS*	Geriatric Depression Scale	20	1.6%
ISRA	Inventory of Situations and Anxiety Response	20	1.6%

Notes: The tests that have appeared new since the 2010 survey are marked with asterisks. The different versions of the same test are grouped under the original name. For example, the abbreviated forms of a test are grouped, as is the case of RIAS and its abbreviated version RIST. The responses on the MEC are also grouped together, which is the version of the MMSE adapted and validated in Spain, or the different versions of the WISC and the WAIS, such as WISC-R and WISC-V.



1,248 members responded to the thirty questions in the survey. From the results already discussed in detail in the preceding lines, a number of general conclusions can be drawn.

The factor analysis of the survey questions indicates that these are articulated around five main dimensions: *Attitudes towards the tests*, *Problems in the use of the tests*, *Control and regulation of the tests and their use*, *Training and knowledge about the tests*, and finally, a fifth dimension related to the *influence of new technologies in the practice of tests*, such as computerization, internet, and the automation of assessments.

Spanish psychologists' positive attitude towards the tests, when used properly, has been confirmed, with an average of 4 points on a scale of 1 to 5 in this dimension. The most favorable professionals are those in the Educational area (4.17), followed by Work (4.09), and Clinical-Health (3.94).

Regarding the problems of using the tests, Spanish psychologists recognize that there are aspects in need of improvement, but they consider the existing problems to be of moderate severity, with an average of 2.93, slightly below the theoretical average of the scale which stands at 3.

Respondents are clearly in favor of increasing the control and regulation of test use by the institutions involved (average 4.07), such as the European Federation of Psychologists Associations (EFPA), the International Test Commission (ITC), and the COP at the national level.

The need for continuous training in everything related to the tests (average of 2.72) is highlighted, since it is recognized that what has been learned in previous studies, such as bachelor's, or master's levels, is not enough to use the current tests successfully.

A certain containment, perhaps prudence, of the professionals has been noted regarding the incorporation of new technologies in the practice of tests, such as online assessment through the internet or the automation of reports (average of 2.89). This skepticism is greater in the Clinical dimension, followed by Educational and Work.

Psychologists believe that the tests are priced too high (average 4.17), especially the Clinical and Educational dimensions, and to a lesser extent in that of Work. Editors should take this aspect into consideration.

The test quality review carried out by the COP Test Commission is still little known among professionals, with only 22.5% knowing about it. It is clear that efforts must be made in its dissemination. The good news is, however, that those who know about it consider it very important and necessary.

Regarding the evolution of the opinions of Spanish psychologists on test practice in the last two decades, it should be noted that the professionals have noticed an improvement in the information they have about the quality of the tests, an increase in use of tests in the exercise of the profession, and an improvement in the perception of the correct use of the tests. These are indicators in the right direction, although there is ample room for improvement, a point on which we will have to continue insisting, within all the institutions, both national and international.

As for the tests most used by professionals, it stands out that the 25 most cited ones are all psychometric, the projective ones appearing much lower in the list. The professionals of the Clinical-Health, Educational, and Work specialties use very different tests, as expected, those of a cognitive nature predominating in the Clinical and Educational areas (WISC), and those of Personality in Work (16PF).

LOOKING TO THE FUTURE

The opinions of the professionals on the practice of tests are a still picture of their current perceptions and, seen in perspective, they reflect a positive trend in relation to the opinions from the years 2000 and 2010. We will see what happens in 2030 and onwards. Everything changes very quickly, and psychological evaluation is no exception. As Heraclitus advised us twenty-five centuries ago, everything flows, nothing remains. We will reflect briefly on the future of the review and the tests following the recent works on the subject by Muñiz (2018) and Muñiz and Fonseca-Pedrero (2019). The most important driving force in the change of evaluation processes are the new information and communication technologies, and especially computer, multimedia, and Internet advances. Experts such as Bennet (1999, 2006), Breithaupt, Mills, and Melican (2006), Drasgow (2016), Drasgow, Luecht, and Bennet (2006), Sireci and Zenisky (2016), or Sireci and Faulkner-Bond (2014), among many others consider that these technologies are influencing all aspects of psychological evaluation, such as test design, the construction and presentation of items, test scores, and distance assessment. Although new forms of evaluation are emerging, psychometric tests will continue to be fundamental tools, given their objectivity and economy of means and time (Phelps, 2005, 2008). In this context of technological change, what is known as Psychology 2.0 is emerging (Armayones et al., 2015), which aims to extend psychology through the facilities offered by the Internet and social networks. The evaluation cannot be alien to these new trends, with new psychometric approaches appearing that are connected to the analysis of big data that are currently available (Markovetz, Blaszkiewicz, Montag, Switala, & Schlaepfer, 2014). For example, the potential advantages of using mobile phones and other portable devices as terminals for assessment open up new possibilities for future assessment (Armayones et al., 2015; Chernyshenko & Stark, 2016; Insel, 2017; Miller, 2012). In addition, works such as the pioneering one by Kosinski, Stillwell, and Graepel (2013) successfully analyze the possibility of using Facebook "likes" as predictors of different human characteristics, including personality traits, which makes us wonder if our trails on social networks will one day replace questionnaires and tests as we know them currently (Muñiz & Fonseca-Pedrero, 2019). We feel that the answer will be no, but you never know. Another issue that is gaining strength is that of ambulatory assessment which, while not new in psychology, has resurfaced strongly today, driven by



information and communication technologies (Chernyshenko & Stark, 2016; Myin-Germeys et al., 2018; Trull & Ebner-Priemer, 2013; van Os, Delespaul, Wigman, Myin-Germeys, & Wichers, 2013). The ambulatory assessment covers different methods and procedures that aim to study people's experiences (emotions, feelings, thoughts, psychological symptoms) in their natural environment and in daily life. This methodology enables the evaluation of psychological constructs from a more dynamic, personalized, contextual, micro-longitudinal, and ecological perspective. This is achieved by conducting assessments several times a day to sufficiently capture the temporal variability of the phenomena. The questions are activated by a beep and can be presented randomly or at predetermined time intervals. It is a complementary approach to the traditional paper-and-pencil procedures of psychometric assessment (Fonseca-Pedrero & Muñiz, 2017; Muñiz & Fonseca-Pedrero, 2019). To analyze this type of data, new psychometric models are needed, among which the network models stand out (Borsboom & Cramer, 2013; Fonseca-Pedrero, 2017, 2018), as well as models from dynamic systems theory or chaos theory (Nelson, McGorry, Wichers, Wigman, & Hartmann, 2017). These new approaches and technologies are surely here to stay in the field of the evaluation, and they are welcome, but they will have to pass the ultimate test and convince the implacable judge: validity. All of the fantasies about the technological advances demonstrate that they contribute improvements to the measure of the construct under evaluation, otherwise they are no more than a mere firework display, because what is essential does not change: it will always be necessary to provide empirical evidence of reliability and validity, and thus guarantee that measuring instruments evaluate objectively, equitably, and rigorously, which is the north star that must not be lost sight of (Muñiz & Fonseca-Pedrero, 2019). In fact, we already pay attention too often to sophisticated proposals to evaluate online psychological and educational constructs that have no empirical evidence behind them. Some of them dare to evaluate such complex constructs as dyslexia or ADHD and generate very apparent graphics with various colors and traffic lights. They even venture to diagnose, but there is nothing behind them. No evidence of validity is provided. They are mere appearance, albeit very lucrative, certainly. For the sake of the users and the profession we must all contribute to unmask these scams, enshrouded in marketing and technology, but without any validity. Every measuring instrument, no matter what technology it uses, must provide a detailed manual stating the evidence of validity that supports it, and if it fails to do so, it is worth nothing. It is that simple.

ACKNOWLEDGMENTS

The authors wish to express their gratitude to the many researchers and professionals who participated in the development of the questionnaire used. Deserving special

mention is the president of the Spanish Psychological Association, Francisco Santolaya, who supported the project clearly and decisively, the members of the Spanish Psychological Association Test Commission (Milagros Antón, José Luis Galve, Rosario Martínez Arias, Vicente Ponsoda, and Frédérique Vallar) and those who coordinated the annual test reviews: Paula Elosua Oviden, Laura E. Gómez Sánchez, María Dolores Hidalgo, Eduardo Fonseca Pedrero, and Carme Viladrich. The computer implementation of the survey was carried out by José Ramón Correas González. Thank you very much everyone. Without your help this work would not have been possible.

CONFLICT OF INTERESTS

There is no conflict of interests.

REFERENCES

- American Educational Research Association, American Psychological Association y National Council on Measurement in Education. (2014). *Standards for educational and psychological testing*. Washington, DC: American Psychological Association.
- Armayones, M., Boixadós, M., Gómez, B., Guillaumon, N., Hernández, E., Nieto, R., Pousada, M. & Sara, B. (2015). Psicología 2.0: oportunidades y retos para el profesional de la psicología en el ámbito de la e-salud [Psychology 2.0: opportunities and challenges for the psychology professional in the field of ehealth]. *Papeles del Psicólogo*, 36, 153-160.
- Bartram, D. (1996). Test qualifications and test use in the UK: The competence approach. *European Journal of Psychological Assessment*, 12, 62-71.
- Bartram, D. & Coyne, I. (1998). Variations in national patterns of testing and test use: The ITC/EFPPA international survey. *European Journal of Psychological Assessment*, 14, 249-260.
- Bennett, R. E. (1999). Using new technology to improve assessment. *Educational Measurement: Issues and Practice*, 18(3), 5-12.
- Bennett, R. E. (2006). Inexorable and inevitable: The continuing story of technology and assessment. In D. Bartram and R. K. Hambleton (Eds.), *Computer-based testing and the Internet*. Chichester: John Wiley and Sons. (pp. 201-217).
- 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.
- Breithaupt, K. J., Mills, C. N., & Melican, G. J. (2006). Facing the opportunities of the future. In D. Bartram and R. K. Hambleton (Eds.), *Computer-based testing and the Internet* (pp. 219-251). Chichester: John Wiley and Sons.
- Calderón, C., Navarro, D., Lorenzo-Seva, U. & Ferrando, P. J. (2019). Multidimensional or essentially unidimensional? A multi-faceted factor-analytic approach for assessing the



- dimensionality of tests and items. *Psicothema*, 31(4), 450-457.
- Carlson, J. F. & Geisinger, K. F. (2012). Test reviewing at the Buros Center for Testing. *International Journal of Testing*, 12, 122-135.
- Chernyshenko, O. S. & Stark, S. (2016). Mobile psychological assessment. In F. Drasgow (Ed.) (2016). *Technology and testing*. New York: Routledge.
- Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika*, 16, 297-334.
- Downing, S. M. & Haladyna, T. M. (Eds.) (2006). *Handbook of test development*. Hillsdale, NJ: LEA.
- Drasgow, F. (Ed.) (2016). *Technology and testing*. New York: Routledge.
- Drasgow, F., Luecht, R. M., & Bennett, R. E. (2006). Technology and testing. In R. L. Brennan (Ed.), *Educational measurement*. Westport, CT: ACE/Praeger. (págs. 471-515).
- Elosua, P. & Geisinger, K. F. (2016). Cuarta evaluación de test editados en España: forma y fondo [Fourth Review of Tests Published in Spain: Form and Content]. *Papeles del Psicólogo*, 37(2), 82-88.
- European Federation of Professional Psychologists' Associations (2005). *Meta-Code of ethics*. Brussels: Author (www.efpa.eu).
- 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., Bartram, D., Boben, D., Egeland, J., Fernández-Hermida, J. R., et al. (2012). Testing practices in the 21st Century: Developments and European Psychologists opinions. *European Psychologist*, 17(4), 300-319.
- Evers, A., Muñiz, J., Hagemester, C., Hostmaelingen, A., Lindley, P., Sjöberg, A., & Bartram, D. (2013). Assessing the quality of tests: Revision of the EFPA review model. *Psicothema*, 25(3), 293-201.
- Fernández-Ballesteros, R. (2014). *Evaluación psicológica [Psychological assessment]*. Madrid: Pirámide.
- Fernández-Ballesteros, R., De Bruyn, E., Godoy, A., Hornke, L., Ter Laak, J., & Vizcarro, C. et al. (2001). Guidelines for the assessment process (GAP): A proposal for discussion. *European Journal of Psychological Assessment*, 17, 187-200.
- Ferrando, P.J. & Anguiano, C. (2010). El análisis factorial como técnica de investigación en Psicología [Factor analysis as a research technique in psychology]. *Papeles del Psicólogo*, 31, 18-33.
- Ferrando, P. J. & Lorenzo-Seva, U. (2017). Program factor at 10: Origins, development and future directions. *Psicothema*, 29, 236-240. doi: 10.7334/psicothema2016.304.
- 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*, 10, 183-224.
- Fonseca-Pedrero, E. (2018). Análisis de redes en psicología [Network analysis in psychology?]. *Papeles del Psicólogo*, 39, 1-12.
- Fonseca-Pedrero, E. & Muñiz, J. (2017). Quinta evaluación de test 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), 161-16.
- Geisinger, K. F. (2012). Worldwide test reviewing at the beginning of the twenty-first century. *International Journal of Testing*, 12, 103-107.
- Geisinger, K. F. (2016). Test evaluation. In S. Lane, M. R. Raymond and T. M. Haladyna (eds.), *Handbook of test development*. New York: Routledge.
- Goodman, D.P. & Hambleton, R.K. (2004). Student test score reports and interpretive guides: Review of current practices and suggestions for future research. *Applied Measurement in Education*, 17, 145-220.
- 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.
- Groth-Marnat, G. & Wright, A. J. (2016). *Handbook of psychological assessment*. Hoboken, NJ: Wiley.
- Haladyna, T. M. & Rodríguez, M. C. (2013). *Developing and validating test items*. Londres: Routledge.
- Hambleton, R. K., Merenda, P. F. & Spielberger, C. D. (Eds.) (2005). *Adapting educational and psychological tests for cross-cultural assessment*. London: LEA.
- Hernández, A., Tomás, I., Ferreres, A. & Lloret, S. (2015). Tercera evaluación de test editados en España [Third evaluation of tests published in Spain]. *Papeles del Psicólogo*, 36(1), pp. 1-8
- Hernández, A., Ponsoda, V., Muñiz, J., Prieto, G. & Elosua, P. (2016). Revisión del modelo para evaluar la calidad de los test utilizados en España [Assessing the quality of tests in Spain: revision of the Spanish test review model]. *Papeles del Psicólogo*, 37, 192-197.
- Hidalgo, M. D. & Hernández, A. (2019). Sexta evaluación de test editados en España: resultado e impacto del modelo en docentes y editoriales [Sixth review of tests published in Spain: Results and impact of the model on lecturers and publishers]. *Papeles del Psicólogo*, 40, 21-30.
- IBM Corp. (2016). *IBM SPSS Statistics for Windows, Version 24.0*. Armonk, NY: IBM Corp.
- Insel, T. R. (2017). Digital phenotyping: Technology for a new science of behavior. *JAMA*, 318, 1215-1216.
- Irwing, P., Booth, T. & Hughes, D. J. (Eds.) (2018). *The Wiley handbook of psychometric testing: A multidisciplinary reference on survey, scale and test development*. UK: John Wiley & Sons Ltd.
- ITC Guidelines for Translating and Adapting Tests (Second Edition) (2017). *International Journal of Testing*, 17, 1-34, (www.InTestCom.org)



- Izquierdo, I., Olea, J. & Abad, F. J. (2014). Exploratory factor analysis in validation studies: Uses and recommendations. *Psicothema, 26*, 395-400.
- Joint Committee on Testing Practices. (2002). *Ethical principles of psychologists and code of conduct*. Washington DC: Joint Committee on Testing Practices.
- Knapp, S.J. (2012). *APA handbook of ethics in psychology*. Washington, DC: American Psychological Association.
- Koocher, G. & Kith-Spiegel, P. (2007). *Ethics in psychology*. New York: Oxford University Press.
- 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.
- Lane, S., Raymond, M.R. & Haladyna, T. M. (2016). *Handbook of test development (2nd edition)*. New York, NY: Routledge.
- Leach, M. & Oakland, T. (2007). Ethics standards impacting test development and use: A review of 31 ethics codes impacting practices in 35 countries. *International Journal of Testing, 7*, 71-88.
- Lindsay, G., Koene, C., Ovreeide, H., & Lang, F. (2008). *Ethics for European psychologists*. Gottingen and Cambridge, MA: Hogrefe.
- Lloret-Segura, S., Ferreres-Traver, A., Hernández-Baeza, A., & Tomás-Marco, I. (2014). El análisis factorial exploratorio de los ítems: una guía práctica, revisada y actualizada [Exploratory factor analysis of items: A practical, revised, and updated guide]. *Anales de Psicología, 30*(3), 1151-1169.
- Lorenzo-Seva, U. & Ferrando, P. J. (2013). FACTOR 9.2 a comprehensive program for fitting exploratory and semi-confirmatory factor analysis and IRT models. *Applied Psychological Measurement, 37*, 497-498. doi: 10.1177/0146621613487794
- Markovetz, A., Blaszkiewicz, K., Montag, C., Switala, C. & Schlaepfer, T. E. (2014). Psycho-Informatics: Big Data shaping modern psychometrics. *Medical Hypotheses, 82*, 405-411.
- Meehl, P. E. (1954). *Clinical versus statistical prediction: A theoretical analysis and a review of the evidence*. Minneapolis: University of Minnesota Press.
- Miller, G. (2012). The smartphone psychology manifesto. *Perspectives on Psychological Science, 7*, 221-237.
- Muñiz, J. (1997). Aspectos éticos y deontológicos de la evaluación psicológica [Ethical and deontological aspects of psychological evaluation]. In A. Cordero (Ed.), *Evaluación psicológica en el año 2000 [Psychological assessment in the year 2000]* (pp. 307-345). Madrid: TEA Ediciones.
- Muñiz, J. (2018). *Introducción a la psicometría [An introduction to psychometrics]*. Madrid: Pirámide.
- Muñiz, J. & Bartram, D. (2007). Improving international tests and testing. *European Psychologist, 12*(3), 206-219.
- 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.
- 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*(2), 151-157.
- Muñiz, J. & Fernández-Hermida, J.R. (2000). La utilización de los test en España [Test use in Spain]. *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., Fernández-Hermida, J. R., Fonseca, E., Campillo, A. & Peña, E. (2011). Evaluación de test editados en España [Evaluation of tests published in Spain]. *Papeles del Psicólogo, 32*(2), 113-128.
- Muñiz, J. & Fonseca-Pedrero, E. (2019). Diez pasos para la construcción de un test [Ten steps to building a test] *Psicothema, 31*, 7-16. doi:10.7334/psicothema2018.291.
- Muñiz, J., Hernández, A., & Ponsoda, V. (2015). Nuevas directrices sobre el uso de los test: 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., Prieto, G., Almeida, L., & Bartram, D. (1999). Test use in Spain, Portugal and Latin American countries. *European Journal of Psychological Assessment, 15*(2), 151-157.
- Myin-Germeys, I., Kasanova, Z., Vaessen, T., Vachon, H., Kirtley, O., Viechtbauer, W., & Reininghaus, U. (2018). Experience sampling methodology in mental health research: New insights and technical developments. *World Psychiatry, 17*, 123-132.
- 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*, 528-534.
- Papeles del Psicólogo (2009). *Número monográfico sobre Ética Profesional y Deontología [Monographic number on Professional Ethics and Deontology]*. Vol. 30, 182-254.
- Phelps, R. (Ed.) (2005). *Defending standardized testing*. London: LEA.
- Phelps, R. (Ed.) (2008). *Correcting fallacies about educational and psychological testing*. Washington: APA.
- Ponsoda, V. & Hontagas, P. (2013). Segunda evaluación de test editados en España [Second evaluation of tests published in Spain]. *Papeles del Psicólogo, 34*(2), 82-90.
- Ryan, J. M. (2006). Practices, issues, and trends in student test score reporting. In S. M. Downing & T. M. Haladyna (Eds.) (2006). *Handbook of test development*. Hillsdale, NJ.: LEA. (pp. 677-710).
- Schmeiser, C. B. & Welch, C. (2006). Test development. In R. L. Brennan (Ed.), *Educational measurement (4th edition)*



- (pp. 307-353). Westport, CT: American Council on Education/Praeger.
- Sireci, S. & Faulkner-Bond, M. (2014). Validity evidence based on test content. *Psicothema*, *26*, 100-107.
- Sireci, S. & Zenisky, A. L. (2016). Computerized innovative item formats: Achievement and credentialing. In S. Lane, M. R. Raymond, & T. M. Haladyna (Eds.), *Handbook of test development*. New York: Routledge.
- 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., Mying-Germays, I. & Wichers, M. (2013). Beyond DSM and ICD: Introducing precision diagnosis for psychiatry using momentary assessment technology. *World Psychiatry*, *12*, 113-117.
- Wright, A. J. (2011). *Conducting psychological assessment: A guide for practitioners*. Hoboken, NJ: Wiley.
- Zenisky, A. L. & Hambleton, R. K. (2016). A model and good practices for score reporting. In S. Lane, M. R. Raymond & T. M. Haladyna (eds.), *Handbook of test development*. New York: Routledge.

Annex 1

Survey used to collect the opinions of psychologists on the use of the tests

General data

Age: Gender: Male Female

Year in which you obtained your Bachelor's Degree/Honor's Degree in Psychology:

Years you have been a member of the Spanish Psychological Association:

Professional Specialty: Clinical-Educational Health Work Other (state which)

Currently working as a psychologist Yes No

Work in the sector: Public Private Unemployed

Number of years in current job

Instructions

The questions listed below are formulated to be answered on a scale of 1 to 5. If you *totally disagree* with the phrase, indicate 1, if you *totally agree*, indicate 5. Use numbers 2, 3, and 4 for intermediate situations. The survey is completely anonymous.

Questionnaire

1. The training received in Psychology Bachelor's Degree courses is sufficient for the correct use of most tests
2. The training received in Psychology Master's Degree courses is sufficient for the correct use of most tests
3. The European Federation of Psychologists' Associations (EFPA) should establish a system to accredit the skills of test users
4. Professionals are provided with sufficient information (independent reviews, research, etc.) on the quality of tests published in our country
5. In my professional field, computerized tests are gradually replacing paper-and-pencil tests
6. My current knowledge with regard to tests is basically that which I learned on my psychology degree courses
7. The administration of tests on the Internet has many advantages compared to the classic paper-and-pencil administration
8. The use of psychological tests should be restricted to qualified psychologists
9. While non-psychologists may administer and score tests, interpretation and feedback should be restricted to psychologists
10. Reports generated automatically by computer have no validity
11. The standards (e.g., European Federation of Psychologists' Association [EFPA], American Psychological Association [APA]) defining the minimum technical qualities of a test should be enforceable
12. Legislation is needed to control the more serious abuses of testing
13. The application of the tests on the Internet puts some people evaluated at a disadvantage



14. Anyone who can demonstrate their competence as a test user (whether a psychologist or not) should be allowed to use tests
15. If used properly, the Internet can greatly improve the quality of test application
16. Controls on tests and testing should be minimal, as controls discourage the development of new ideas and new procedures
17. Online test application does not allow users to protect their privacy
18. Publishers should be allowed to sell whatever tests they think fit
19. The Spanish Psychological Association should take a more active role in the regulation and improvement of test use
20. The online application of tests opens up possibilities of fraud
21. I use tests regularly in the exercise of my profession
22. Tests constitute an excellent source of information if they are combined and complemented with other psychological data
23. Used correctly, tests are of great help to the psychologist
24. Taking into account all aspects, I believe that in the last decade the use of tests has improved in my country
25. The professional practice based on evidence enhances the use of tests as sources of information
26. In professional decision making I give an important weight to the data obtained through tests
27. Before using a test I check its psychometric properties
28. Estimate the frequency of the following test use problems in your professional environment (1: very rare; 5: very common)
 1. Making photocopies of copyrighted materials
 2. Making evaluations using inappropriate tests
 3. Not keeping up with the field
 4. Failing to check one's own interpretations with others
 5. Not considering errors of measurement of a test score
 6. Not restricting test administration to qualified personnel
 7. Not taking into account conditions that cast doubt on reported validity for a local situation
 8. Making interpretations that go beyond the limits of the test
 9. Using tests with inappropriate scales
 10. Using some items without the corresponding permissions
 11. Having a very high price
29. In relation to the annual test review carried out by the Spanish Psychological Association Test Commission:
 1. I know about it Yes No (if the answer is No, **do not** go to question 30, in fact in the computerized survey already does this automatically)
 2. I think it is important to improve the quality of tests
 3. I check the test reviews on the Spanish Psychological Association website
 4. It helps me in choosing the tests I use
 5. It is unnecessary, psychologists already have enough information to know which test to use
30. Name the three tests you use most frequently in the exercise of your profession:
 1.
 2.
 3.

Observations: Comment on any other aspect you deem appropriate

