

DIFFERENCES IN OPINIONS REGARDING TESTING PRACTICES: THE ROLE OF AGE, GENDER, AND JOB SECTOR

Ana Hernández¹, M^a Dolores Hidalgo², José Muñiz³ and José Ramón Fernández-Hermida⁴

¹Universidad de Valencia. ²Universidad de Murcia. ³Universidad Nebrija. ⁴Universidad de Oviedo

Los test son importantes herramientas que ayudan a los profesionales de la Psicología a tomar decisiones sobre las personas. Para plantear acciones encaminadas a mejorar su uso, es importante determinar qué factores afectan a las opiniones que los psicólogos tienen sobre la práctica de los test. El presente estudio analiza el rol de tres factores: el género, la edad y el sector laboral (público vs. privado). Asimismo, se contrasta la opinión de los profesionales colegiados con las de los profesores universitarios que imparten asignaturas relacionadas con los test y su uso (Psicometría y Evaluación Psicológica o Psico-educativa). Tras analizar una muestra de 1.248 colegiados y 95 profesores de distintas universidades, los resultados muestran algunas diferencias de opinión en función de todos los factores considerados. Se analizan los resultados y sus implicaciones de cara al futuro.

Palabras clave: Uso de test, Opiniones sobre la práctica de los test, Género, Edad, Sector laboral.

Tests are important tools that help psychologists to make decisions about people. In order to propose actions aimed at improving the use of tests, it is important to determine what factors relate to psychologists' opinions about testing practices. In this study we assess the role of three factors: gender, age, and job sector (public vs. private). In addition, we compare practitioners' opinions with those reported by professors of test-related subjects (psychometrics, and psychological and psycho-educational assessment). After analyzing a sample of 1,248 members of the Spanish Psychological Association and 95 professors from different universities, results show some differences in opinions across all the factors analyzed. Results and future challenges are discussed.

Key words: Test use, Opinions about testing practices, Age, Gender, Job sector.

PSYCHOLOGICAL TESTS AND THEIR USE

Tests allow psychologists to obtain empirical evidence to help them in the process of psychological assessment and decision making. This has been recognized for years by psychology practitioners when asked about the use of tests in their daily practice (Muñiz & Fernández-Hermida, 2000, 2010; Muñiz et al., 2020). But for tests to be really useful, three basic conditions are needed: the psychometric properties of the scores obtained must be satisfactory, practitioners must have the appropriate training, and the tests must be used correctly (Elosua & Muñiz, 2013; Elosua, 2017; Muñiz et al., 2020). As Hernández et al. (2020a) point out, the first condition is guaranteed when, in the construction and analysis of tests, psychometric knowledge is applied rigorously (American Educational Research Association (AERA), American Psychological Association (APA), National Council on Measurement in Education (NCME), 2018; Haladyna & Rodríguez, 2013; Irwing et al., 2018; Muñiz & Fonseca-Pedrero, 2019). To ensure such rigor, different national and international organizations, such as the General Council of the Spanish Psychological Association (CGCOP), through its Test Commission, the International Test Commission (ITC), the European

Federation of Psychological Associations (EFPA), and test publishers, carry out different actions. These include the publication of guidelines for the creation and analysis of tests (AERA, APA, and NCME, 2018), and guidelines for the adaptation of tests from one culture to another (ITC, 2017; Muñiz et al., 2013; Hernández et al., 2020b). Also noteworthy is the development of models for assessing test quality, such as the EFPA model (Evers et al., 2013) or the CGCOP model (Prieto & Muñiz (2000), reviewed by Hernández et al. (2016)), as well as the application of these models to different tests to assess their quality. For example, in Spain, the CGCOP model, or test evaluation questionnaire (CET) and its subsequent revision (CET-R), has been applied to a total of 82 tests during various evaluation rounds (Muñiz et al., 2011; Ponsoda & Hontangas, 2013, Hernández et al., 2015; Elosua & Geisinger, 2016; Fonseca-Pedrero & Muñiz, 2017; Hidalgo & Hernández, 2019; Gómez-Sánchez, 2019; Viladrich et al., 2021), with new reviews underway. These reviews (with quantitative and qualitative reports) are published on the CGCOP website (<https://www.cop.es/index.php?page=evaluacion-tests-editados-en-espana>) and can be consulted by any interested professional. The aim is to help these practitioners to rigorously evaluate the tests they can use for a given assessment before making their choice.

However, although these actions are aimed at improving tests and curbing the use of tests of questionable quality, Lilienfeld et al. (2006) have already warned that there are

Received: 9 February 2020 - Accepted: 16 March 2020

Correspondence: Ana Hernández. IDOCAL. Universidad de Valencia, Av. Blasco Ibáñez, 21. 46010 Valencia. España.

E-mail: Ana.Hernandez@uv.es



tests that maintain their popularity despite being problematic. Some reasons are adduced. For example, the Barnum effect, referring to the tendency to accept general or vague interpretations offered by tests as valid descriptors for practically anyone, which erroneously increases confidence in the test used. Also, the alchemist's fantasy, related to the belief that the results of any test, when combined with other types of data, offer important clinical information, so the results provided by a bad test are taken as another source of valid information, even though this may reduce the ultimate quality of the evaluation. Finally, the clinical tradition and institutional inertia to use certain tests, despite the fact that there are other more updated and better ones. For this reason, it is necessary to continue to carry out actions that contribute to reducing these problems.

The selection of appropriate tests is closely related to the other two conditions mentioned to ensure the usefulness of psychological tests which, in turn, are closely linked: that the practitioners must have the appropriate training, and that the tests must be used correctly. Training in psychometrics and in the use of tests are crucial aspects for the proper use of tests. This training is mainly provided at university, in the subjects of Psychometrics and Psychological Assessment (Hidalgo & Hernández, 2019; Hernández et al., 2020a). When it comes to ensuring that test users have received such training, test publishers (in Spain CEPE, GiuntiEOS Psychometrics, Pearson, and TEA) play a crucial role, as they restrict the sale of psychological tests to those accredited with the required qualification. However, university training does not guarantee the proper use of any test, nor does it guarantee the correct use of tests for life. The aforementioned problems of using tests that are not suitable (Lilienfeld et al., 2006), as well as the lack of knowledge about new psychometric advances or about the impact of new technologies, or the lack of updated knowledge regarding the tests that are published, can lead to their misuse (Hernández et al., 2020a).

Therefore, in addition to the restrictive strategy that limits the use of tests, it is necessary to carry out informative and formative actions that promote the proper use of tests. Among the informative actions, as pointed out by Muñiz et al., (2020), in addition to the previously mentioned guidelines (APA, AERA, and NCME, 2018; ITC, 2017) and the publication of the reviews of different tests by using the CET-R (Hernández et al., 2016), also noteworthy is the publication of ethical codes, such as that of the EFPA (2005), and standards on the assessment process, such as the guidelines of the European Association of Psychological Assessment (Fernández-Ballesteros et al., 2001) or the ISO10667 standard (AENOR, www.aenor.es), which regulates the assessment of people in work and organizational contexts. Among the training actions, the courses on test construction and evaluation offered by the CGCOP through the Distance Continuing Education Program (FOCAD in Spanish) (Elosua, 2019; Muñiz & Fonseca-Pedrero, 2017) and the courses

offered by the main test publishers operating in Spain stand out.

Before designing new actions to improve the use of psychological tests (whether restrictive, formative, or informative), it is important to understand the opinions of psychologists on the tests and their use. The CGCOP has been aware of this for years. Therefore, through its testing commission, it has promoted surveys among its members in order to know and assess the needs, challenges, and problems faced by psychology practitioners when using tests in their professional practice (Muñiz & Fernández-Hermida, 2000; 2010; Muñiz et al., 2020).

PSYCHOLOGISTS' OPINION ON TESTS

The latest study (Muñiz et al., 2020) offers interesting results, both in general and differentiating by specialty. Firstly, Spanish psychologists (especially those in the area of educational psychology, followed by those in the area of work psychology) show a positive attitude towards tests. However, they recognize the need for further training (and information) in order to use them appropriately. Likewise, although in general they are in favor of increasing the control and regulation of tests and their use, there are differences by specialty on some issues, such as the need to introduce a system to accredit the competence of test users or to introduce more controls, with the clinical specialty being the most reluctant to tackle these issues. On the other hand, practitioners are moderately skeptical about the incorporation of new technologies in the practice of testing, especially the clinicians. They also recognize that there are problems in the use of tests, although these occur with moderate frequency. The most important problem concerns the price of the tests, which is considered too high, especially among clinical-health and educational specialists. Finally, although the review of test quality carried out by the CGCOP Test Commission is little known (22.5% of the practitioners say they are aware of it), those who are aware of it rate it very positively.

OBJECTIVE

It is to be expected that psychologists' opinions about tests and their use will vary depending on other factors. In the present study, the role of gender, age, and work sector (public vs. private) is analyzed. Knowing these opinions will help us to better understand the reality of test use in different groups of practitioners and, if necessary, to propose measures aimed at specific groups.

For example, with regard to gender, there is some evidence that men and women perceive risk differently (which could carry over into the assessment of the consequences of using a test that is not the most appropriate) and also that women are more meticulous than men (Chan, 2011; Greenberg & Schneider, 1995; Gustafson, 1998). These differences could affect the way tests are used. Although these are stereotypes that may not generalize in all cases, it is interesting to assess



whether differences in opinion about test use are observed between men and women. In terms of sector, psychologists practicing in different sectors may have different resources, as well as different levels of autonomy to make decisions, etc., which could affect the tests used, and their use (e.g., making some inappropriate practices such as photocopying answer sheets). Finally, of particular interest is age, as it is often accompanied by greater professional experience. In addition, university education has also changed a lot with the incorporation of Spain into the European Higher Education Area, whose curricula began to be implemented between 2009 and 2010. Both aspects, training and experience, together with generational changes, may affect the opinions held about the tests. Finally, we will also contrast the opinions of practitioners with those of university professors who teach subjects related to tests and their use (psychometrics and psychological or psycho-educational assessment), since they play a crucial role in training (future) psychologists in the skills necessary for the proper use of tests.

PARTICIPANTS AND PROCEDURE

The members of the Spanish Psychological Society (COP) in our country were invited to participate in a survey through various media, providing information on the objectives and importance of the study. The survey was applied online, via a link presented on the CGCOP website. On entering the survey, participants gave their informed consent. Anonymity and confidential treatment of the data were guaranteed. The survey was answered by 1,248 members. A total of 73.7% were women and the average age was 46.3 years ($SD=11.1$). The majority (69.9%) worked in the private sector and 26.6% in the public sector, while 3.5% of participants were unemployed. Likewise, the main area of specialization was clinical-health (67.6%), followed by education (12.7%) and labor (5.5%). Other specialties, such as legal, sports, social services, etc., made up the remaining 14.1%. A more detailed description of the sample can be found in Muñiz et al. (2020).

Regarding the sample of university professors of psychometrics and psychological and psychoeducational assessment, these were identified through the websites of the psychology degree programs of the Spanish universities (a total of 316 professors), and they were invited by e-mail to participate in the study through an online survey, guaranteeing confidentiality. A total of 97 professors responded (30.4%), although the responses of two participants who indicated that they did not teach any of the subjects of interest were eliminated. Of the 95 professors who remained in the sample, 73.7% of the participants belonged to public universities, 51.6% were male and the majority (49.5%) were professors of psychometrics, followed by 35.8% of psychological assessment. Only two participants (2.1%) indicated that they were professors of psychoeducational assessment. The rest (12.6%) either taught

several of the aforementioned subjects or combined some of these subjects with others. The average length of teaching service was 17.5 years ($SD=11.1$).

THE SURVEY

To collect the psychologists' opinions on the tests and their use, we used the 31-item questionnaire described in detail and published in full in Muñiz et al. (2020), which was answered on a 5-point (1-5) graduated scale (see items in Table 1). Considering the results on the dimensionality of the questionnaire, 5 dimensions were distinguished: Attitudes towards tests, Training and knowledge about tests, Need for control and regulation of tests and their use, Influence of new technologies on the use of tests, and Problems in the use of tests. The reported reliability coefficients (Cronbach's alpha) ranged from 0.62 for the «Need for control and regulation of tests and their use» dimension to 0.91 for the «Problems in test use» dimension. The omega coefficient, which does not require the measures to be tau-equivalent (McDonald, 1999) presented values ranging from 0.80 for the «Need for control and regulation of tests and their use» dimension to 0.94 for the «Problems in test use» dimension.

Although all the dimensions were compared across the different interest groups, this study only offers the analyses differentiated by items that provided the most relevant results. The additional question on whether they were aware of the test evaluations carried out by the CGCOP test commission is also analyzed. Those who were aware of it were asked to rate these reviews (see items in Table 3).

In the sample of professors, collected previously (Hidalgo & Hernández, 2019), the questionnaire used was the one employed in 2010 (Muñiz & Fernández-Hermida, 2010), which did not include items 25, 26, or 27 of the "Attitudes towards tests dimension" of the 2020 questionnaire. In addition, some modifications were made. Item 6, «I received my current knowledge in relation to tests mainly during my psychology degree», was eliminated for obvious reasons. Likewise, all the problems of the last version of the questionnaire were included except the final one, referring to the high price of the tests. Finally, opinions on the test reviews carried out by the CGCOP test commission were not collected. Instead, professors were asked whether, in the classes they taught, they used the CET model and the evaluations performed (Hidalgo & Hernández, 2019).

DATA ANALYSIS

Descriptive statistics (means and standard deviations) were obtained for all items. To compare the scores, Student's t-tests were performed for gender, type of sample (practitioners vs. professors), and sector, differentiating between the private and public sectors, excluding unemployed individuals. In the case of age, 5 age groups were formed (see Table 2) and ANOVAs performed. When ANOVA results were statistically significant, comparisons were made between groups using



Tukey's test. When the assumption of homogeneity of variances required in the ANOVA was not met, the Brown-Forsythe robust test was used. When the results were statistically significant, the Games-Howell test was used for post-hoc comparisons. All analyses were carried out using the IBM-SPSS-26 program. In addition, effect size indicators were obtained: η^2_{partial} for ANOVAs and Cohen's d for t-tests (Cohen, 1992).

DIFFERENCES OF OPINION ON THE TESTS

The results of the comparisons for the questionnaire items according to the different groups formed are presented in Table 2, indicating the comparisons that were statistically

significant. The most relevant differences are discussed below. Those interested in the complete results, including the results by dimensions, can contact the first author.

Age

Firstly, with regard to the items relating to the dimension of training and knowledge about the tests (items 1, 2, 4, and 6), there were statistically significant differences only for item 6, «I received my current knowledge about tests mainly during my psychology degree». It is observed that the older the age, the lower the mean on the item ($p = 0.012$; $\eta^2_{\text{partial}} = 0.010$). Not surprisingly, practitioners continue to gain knowledge about the tests throughout their careers. However, there are

TABLE 1
QUESTIONNAIRE USED TO COLLECT FEEDBACK ON THE TESTS

Items

1. The training received in psychology is sufficient for the correct use of most of the tests.
2. The training received in courses and masters is sufficient for the correct use of most of the tests.
3. The European Federation of Psychologists' Associations (EFPA) should establish a system for accrediting the competence of test users.
4. Practitioners have sufficient information (independent reviews, research, documentation, etc.) on the quality of the tests published in our country.
5. In my professional field, computerized tests are progressively replacing paper-and-pencil tests.
6. My current knowledge of testing was acquired mainly during my psychology studies.
7. The application of the tests via the Internet has many advantages compared to the classic paper-and-pencil application.
8. The use of psychological tests should be restricted to qualified psychologists.
9. Although non-psychologists could administer and score the tests, the interpretation and reporting of the results should be restricted to psychologists.
10. The reports that are generated automatically by the computer are not valid.
11. Standards and guidelines defining the minimum technical qualities of a test should be mandatory [e.g., the standards of the European Federation of Psychologists' Associations (EFPA) or those of the Psychological Association].
12. Legislation is needed to control the most serious test abuses.
13. The administration of the tests via the Internet puts some test takers at a disadvantage.
14. Anyone who is capable of demonstrating competence in the use of the tests (whether a psychologist or not) should be allowed to use them.
15. If used properly, the Internet can greatly improve the quality of test application.
16. Controls on tests should be minimal, as they inhibit the development of new ideas and new assessment procedures.
17. The administration of tests via the Internet does not protect the privacy of users.
18. Publishers should be allowed to sell any test they deem appropriate.
19. The Spanish Psychological Association should play a more active role in regulating and improving the use of tests.
20. Internet test application opens up possibilities for fraud
21. In the performance of my profession, I regularly use tests
22. Tests are an excellent source of information when combined with other psychological data.
23. Used correctly, tests are of great help to the psychologist.
24. All things considered, I believe that in the last decade the use of tests has improved in my country.
25. Evidence-based professional practice enhances the use of tests as sources of information.
26. In professional decision making, I give significant weight to the data obtained by means of tests.
27. Before using a test I check its psychometric properties.
- 28-1. Making photocopies of copyrighted materials
- 28-2. Making assessments using inappropriate tests
- 28-3. Not being updated
- 28-4. Not contrasting interpretations with others
- 28-5. Not taking into account measurement errors in scores.
- 28-6. Not restricting the application of the tests to qualified personnel.
- 28-7. Not taking into account local conditions (country, region) that may affect validity.
- 28-8. Making interpretations that go beyond the limits of the test.
- 28-9. Using tests with inadequate scales
- 28-10. Using some items without the appropriate permits
- 28-11. Being very high priced



statistically significant differences only between the youngest group (aged 22-19 years) and the oldest group (aged 60 years or older) ($p = 0.013$). This difference suggests that current training may be more comprehensive in terms of testing than that received by earlier generations of psychologists.

Secondly, with respect to regulation and control in the use of tests, the results indicate that older psychologists tend to be more in favor of introducing control mechanisms, although the relationship is not always linear. For example, there are differences of opinion regarding the EFPA establishing an accreditation system for test users (item 3; $p = 0.017$; $\eta^2_{\text{partial}} = 0.026$). The differences reach statistical significance when comparing the group between 30 and 39 years of age, which was the least in favor, and the group aged 60 years and over, which was the most in favor ($p = 0.021$). Statistically significant differences were also observed in item 12 ($p = 0.001$; $\eta^2_{\text{partial}} = 0.012$). It is also the group between 30 and 39 years of age, together with the group between 40 and 49 years of age, who consider it to be less necessary to introduce legislation to control the most serious abuses, compared to the group between 50 and 59 years of age ($p = 0.006$ and $p = 0.008$, respectively). However, all three groups were in favor of such legislation, with means close to 4. It is also between the last two groups where statistically significant differences were observed on the role that the COP should play in regulating and improving the use of tests (item 19; $p = 0.001$; $\eta^2_{\text{partial}} = 0.014$). Again, it was the group between 50 and 59 years of age, compared to the group between 40 and 49 years of age, who gave the most importance to the COP playing a more active role ($p < 0.001$).

— INSERT TABLE 2 —

Thirdly, with regard to attitudes towards the use of tests (items 21-26, together with item 27, which presented a weak factorial saturation in the dimension), statistically significant differences were observed in item 25 ($p = 0.004$; $\eta^2_{\text{partial}} = 0.012$). It is noteworthy that it was the young people (the groups between 22 and 29, and between 30 and 39 years of age) who most rely on evidence-based professional practice to enhance the use of tests compared to practitioners aged 60 years or older ($p = 0.031$ and $p = 0.004$, respectively).

In fourth place, focusing on new technologies (computerization, automation, and Internet) (items 5, 7, 10, 13, 15, 17, and 20), as we might expect, young people were the most in favor of the incorporation of these technologies. In particular, statistically significant differences were observed in item 15 ($p = 0.007$; $\eta^2_{\text{partial}} = 0.011$). The two youngest groups (between 22 and 29, and between 30 and 39 years) were more confident that the Internet can greatly improve the quality of the application of the tests compared to the group aged 60 and over ($p = 0.023$ and $p = 0.008$, respectively). On the contrary, it was the older groups who were the most concerned about the protection of users' privacy when tests are administered over the Internet (item 17; $p < 0.001$; 2_{partial}

$= \eta 0.0026$), with statistically significant differences between the two youngest and two oldest groups and also between the group aged 40-49 years and the oldest group (with p -values between $p < 0.001$ and $p = 0.023$). Finally, statistically significant differences were also observed in item 5 ($p = 0.006$; $\eta^2_{\text{partial}} = 0.010$). It is curious that the youngest group, between 22 and 29 years of age, considers to a lesser extent that computerized tests are replacing paper-and-pencil tests compared to the opinions of the more senior practitioners, between 50 and 59 years of age and 60 or more ($p = 0.032$ and $p = 0.010$, respectively). Younger people probably expect this replacement to occur at a faster rate than it is actually occurring.

Finally, with regard to problems in test use (items 28.1 to 28.11), statistically significant differences were observed in items 28.1 ($p = 0.005$; $\eta^2_{\text{partial}} = 0.012$) and 28.11 ($p < 0.001$; $2_{\text{partial}} = \eta 0.020$). In particular, there was a tendency to use fewer photocopies of copyrighted materials (item 28.1) as age increases, although the only statistically significant differences were between the 30-39 and 50-59 age groups ($p = 0.043$). Along the same lines, it was the youngest people who complained the most about the high price of tests (item 28.11), with statistically significant differences between the 3 youngest groups and the group aged 60 years and over ($p = 0.005$, $p < 0.001$, and $p = 0.008$, respectively).

Gender

Both men and women considered university training in post-graduate courses and master's degrees, and especially in the psychology degree, not to be sufficient for the correct use of most of the tests. However, men rated this training more positively than did women, both in terms of training during the degree (item 1; $p = 0.001$; $d = 0.22$) and training in courses and Master's degrees (item 2; $p = 0.028$; $d = 0.14$). Regarding practices in the use of tests and the problems they encounter in their professional environment, while men indicated that they check the psychometric properties of the tests they are going to use to a greater degree than women (item 27; $p = 0.002$; $d = 0.20$), men reported that, in their environment, interpretations are made that go beyond the limits of the test (item 28.8; $p = 0.014$; $d = 0.16$) and inappropriate scales are used (item 28.9; $p < 0.001$; $d = 0.26$) more frequently than women. Finally, compared to men, women considered the price of the tests to be very high (item 28.11; $p = 0.001$; $d = 0.22$).

Sector

The results show that public sector practitioners were more favorable to the establishment of systems of regulation and control (with statistically significant differences in 4 of the 9 items of the dimension: items 3, 11, 12, and 14). In particular, public sector Psychological Association members were more in favor of the implementation of a system of accreditation of the competence of test users by the EFPA (item 3; $p = 0.038$; $d =$



0.13), compared to those in the private sector. The public sector members also advocated to a greater extent that the standards and guidelines marking the minimum qualities of a test (such as those of the EFPA or the CGCOP) should be mandatory (item 11; $p = 0.038$; $d = 0.14$). In addition, they gave more importance to the need to introduce legislation to control the most serious abuses of tests (item 12; $p = 0.007$; $d = 0.16$) and, finally, they were more reluctant to allow anyone capable of demonstrating competence in the use of tests (whether a psychologist or not) to be authorized to use them (item 14; $p =$

0.046; $d = 0.13$). Interestingly, it was in the private sector where computerized tests were considered to be replacing paper-and-pencil tests to a greater extent (item 5; $p = 0.005$; $d = 0.15$), and it was also private sector practitioners who reported using the tests more frequently (item 21; $p = 0.026$; $d = 0.15$). Finally, with regard to problems in the use of the tests, although the means were always below 3, in the private sector it was considered that interpretations that go beyond the limits of the tests are made more frequently than in the public sector (item 28.8; $p = 0.009$; $d = 0.17$) and participants in this sector

TABLE 2
COMPARISONS OF THE ITEM MEANS FOR THE DIFFERENT GROUPS ANALYZED

No. of items and groups with differences	AGE										GENDER				SECTOR				PRACTITIONERS vs. PROFESSORS			
	22-29 years (N=79) A		30-39 years (N=300) B		40-49 years (N=342) C		50-59 years (N=362) D		60 or more (N=163) E		Male (N=328)		Female (N=920)		Public (N=332)		Private (N=872)		Global Practitioners (N=1,248)		Professors (N=95)	
	Mesn	SD	Mesn	SD	Mesn	SD	Mesn	SD	Mesn	SD	Mesn	SD	Mesn	SD	Mesn	SD	Mesn	SD	Mesn	SD	Mesn	SD
1. **Gender, **Prof.	2.35	1.17	2.49	1.16	2.43	1.11	2.49	1.18	2.54	1.17	2.66	1.21	2.41	1.13	2.57	1.14	2.44	1.16	2.47	1.16	2.97	0.99
2. *Gender	3.08	1.07	3.09	1.09	2.99	1.05	3.04	1.00	3.07	1.03	3.16	1.04	3.01	1.05	3.08	0.97	3.03	1.07	3.05	1.05	3.20	0.95
3. *Age (B-E), *Sector, **Prof.	3.28	1.22	3.20	1.32	3.44	1.25	3.46	1.32	3.58	1.27	3.34	1.38	3.42	1.26	3.51	1.22	3.34	1.33	3.39	1.30	3.75	1.07
4. **Prof.	2.90	0.96	3.12	1.12	3.00	1.10	3.06	1.08	3.00	1.07	3.09	1.09	3.03	1.08	3.00	1.05	3.06	1.10	3.04	1.09	2.77	0.88
5. **Age (A-D) (A-E), **Sector, *Prof.	3.00	1.24	3.30	1.31	3.30	1.29	3.46	1.27	3.58	1.28	3.36	1.22	3.36	1.31	3.20	1.30	3.44	1.29	3.36	1.29	3.08	1.20
6. *Age (A-E)	2.68	1.24	2.42	1.27	2.32	1.18	2.27	1.23	2.15	1.11	2.43	1.16	2.30	1.23	2.37	1.20	2.29	1.20	2.33	1.21	—	—
7.	3.30	1.15	3.21	1.17	3.14	1.17	3.19	1.16	2.99	1.25	3.11	1.15	3.18	1.19	3.23	1.16	3.12	1.18	3.16	1.18	3.37	1.09
8.	4.25	0.85	4.18	1.01	4.20	1.07	4.25	1.07	4.38	0.98	4.20	1.10	4.25	1.01	4.23	1.06	4.23	1.02	4.24	1.03	4.21	0.97
9.	4.43	0.96	4.38	1.09	4.30	1.22	4.40	1.15	4.45	1.12	4.38	1.14	4.38	1.14	4.34	1.16	4.38	1.14	4.38	1.14	4.28	1.07
10.	2.76	1.08	2.71	1.11	2.77	1.14	2.87	1.11	2.80	1.17	2.85	1.15	2.76	1.11	2.81	1.10	2.77	1.13	2.78	1.12	2.62	1.14
11. **Age (C-D), *Sector	4.04	0.84	4.18	0.79	4.10	0.92	4.32	0.82	4.10	0.99	4.17	0.90	4.18	0.86	4.27	0.79	4.15	0.89	4.18	0.87	4.34	0.78
12. **Age (B-D) (C-D), **Sector	3.96	0.94	3.90	1.04	3.92	1.00	4.17	0.99	4.13	1.02	4.10	0.99	3.99	1.02	4.12	0.89	3.96	1.06	4.02	1.01	4.01	0.97
13.	3.51	1.07	3.45	1.08	3.38	1.14	3.38	1.07	3.40	1.13	3.32	1.16	3.44	1.07	3.34	1.06	3.42	1.11	3.41	1.10	3.60	0.98
14. *Sector	2.16	1.19	2.14	1.25	2.11	1.28	1.93	1.24	2.15	1.42	2.11	1.29	2.07	1.28	1.95	1.19	2.11	1.31	2.08	1.28	2.34	1.30
15. **Age (A-E) (B-E), **Prof.	3.52	0.90	3.47	1.10	3.35	1.10	3.33	1.03	3.11	1.13	3.39	1.05	3.34	1.09	3.34	1.07	3.35	1.08	3.36	1.08	3.60	0.98
16. **Prof.	1.95	0.93	1.78	0.87	1.85	0.91	1.77	1.00	1.93	1.11	1.76	0.93	1.85	0.97	1.87	0.98	1.80	0.95	1.83	0.96	1.53	0.73
17. **Age (A-D) (A-E) (B-D) (B-E) (C-E), *Prof.	2.10	1.08	2.19	1.11	2.37	1.13	2.57	1.21	2.70	1.30	2.42	1.25	2.41	1.16	2.37	1.14	2.42	1.20	2.41	1.18	2.14	1.02
18.	1.66	0.86	1.82	1.07	1.80	1.06	1.70	0.95	1.74	1.04	1.84	1.06	1.73	1.00	1.75	1.01	1.76	1.03	1.76	1.02	1.94	0.91
19. *Age (C-D), *Prof.	4.14	0.80	4.11	0.99	3.92	1.05	4.22	0.92	4.13	1.03	4.09	1.05	4.09	0.96	4.17	0.91	4.05	1.02	4.09	0.99	3.91	0.83
20.	3.59	1.13	3.52	1.12	3.42	1.13	3.54	1.18	3.68	1.15	3.49	1.16	3.53	1.14	3.48	1.08	3.53	1.18	3.52	1.14	3.52	0.97
21. **Age (B-E), *Sector	3.75	1.20	4.12	1.15	4.05	1.14	4.07	1.16	3.76	1.32	3.93	1.24	4.05	1.16	3.90	1.23	4.08	1.16	4.02	1.18	4.04	1.14
22. **Prof.	4.53	0.77	4.49	0.83	4.47	0.86	4.46	0.86	4.44	0.89	4.47	0.89	4.48	0.83	4.41	0.86	4.49	0.85	4.47	0.85	4.74	0.55
23. **Prof.	4.53	0.75	4.46	0.86	4.46	0.85	4.49	0.82	4.35	0.92	4.44	0.91	4.47	0.83	4.39	0.85	4.47	0.86	4.46	0.85	4.68	0.57
24.	3.59	0.84	3.73	0.93	3.75	0.92	3.76	0.95	3.71	0.95	3.66	1.01	3.76	0.90	3.73	0.86	3.73	0.96	3.73	0.93	3.88	0.87
25. **Age (A-E) (B-E)	4.10	0.83	4.05	0.84	3.97	0.92	3.92	0.94	3.74	1.03	4.00	0.98	3.94	0.90	3.98	0.89	3.94	0.94	3.96	0.92	—	—
26-26.	3.37	0.94	3.52	1.05	3.43	0.97	3.49	0.95	3.37	1.08	3.53	1.06	3.43	0.97	3.42	0.97	3.47	1.02	3.46	1.00	—	—
27. **Gender	3.71	1.16	3.88	1.11	3.80	1.18	3.97	1.16	4.01	1.08	4.06	1.10	3.83	1.16	3.98	1.09	3.84	1.18	3.89	1.15	—	—
28-1. **Age (B-D), **Prof.	3.39	1.44	3.24	1.40	3.04	1.46	2.93	1.41	2.88	1.36	3.05	1.40	3.05	1.43	3.05	1.43	3.03	1.42	3.05	1.42	3.48	1.31
28-2. **Prof.	2.49	1.31	2.38	1.26	2.54	1.36	2.34	1.26	2.53	1.41	2.55	1.34	2.40	1.30	2.40	1.28	2.43	1.32	2.44	1.31	3.14	1.21
28-3. **Prof.	3.20	1.16	3.10	1.30	3.16	1.28	3.02	1.21	3.00	1.30	3.10	1.26	3.08	1.26	3.05	1.25	3.07	1.27	3.08	1.26	3.62	1.07
28-4. **Prof.	3.04	1.23	3.04	1.29	3.08	1.31	2.99	1.23	3.09	1.30	3.09	1.29	3.02	1.27	3.00	1.28	3.02	1.28	3.04	1.28	3.77	0.95
28-5. **Prof.	3.04	1.26	3.09	1.25	3.06	1.26	2.90	1.24	2.98	1.26	3.10	1.30	2.97	1.24	2.92	1.26	3.02	1.25	3.01	1.25	3.93	1.17
28-6. **Prof.	2.99	1.46	2.92	1.48	2.87	1.49	2.80	1.49	2.96	1.44	2.99	1.44	2.84	1.49	2.83	1.53	2.88	1.47	2.88	1.48	3.41	1.26
28-7. **Prof.	2.97	1.40	2.85	1.33	2.94	1.35	2.88	1.32	2.89	1.32	2.95	1.36	2.87	1.32	2.82	1.31	2.89	1.34	2.89	1.33	3.40	1.19
28-8. *Gender, **Sector, **Prof.	3.05	1.34	2.90	1.34	2.90	1.37	2.75	1.37	2.94	1.33	3.03	1.39	2.81	1.34	2.69	1.33	2.92	1.36	2.87	1.36	3.69	1.12
28-9. **Gender, **Prof.	2.71	1.29	2.41	1.24	2.57	1.33	2.42	1.30	2.62	1.28	2.75	1.34	2.41	1.27	2.45	1.28	2.51	1.30	2.5	1.30	3.63	1.16
28-10. **Prof.	2.32	1.35	2.18	1.20	2.35	1.36	2.19	1.24	2.31	1.31	2.36	1.32	2.21	1.26	2.19	1.29	2.26	1.28	2.25	1.28	3.29	1.26
28-11. **Age (A-E) (B-D) (B-E) (C-E), **gender, **Sector	4.37	1.16	4.33	1.05	4.20	1.14	4.10	1.10	3.85	1.11	3.99	1.16	4.23	1.09	4.01	1.14	4.20	1.11	4.17	1.11	—	—

NOTE: Age, gender, and sector refer only to the sample of practitioners. Prof. (Profession): Practitioners vs. professors See item statements in Table 3.
p< 0.05; ** p< 0.01; SD= Standard Deviation



also complained to a greater extent about the price of the tests (item 28.11; $p = 0.008$; $d = 0.17$).

Practitioners vs. professors

The results indicate that professors have a more favorable view of the training received during the degree (item 1; $p < 0.001$; $d = 0.43$), although in neither case did the score reach 3, the midpoint of the scale. Interestingly, however, it was the professors, compared to the practitioners, who most questioned the availability of sufficient information (independent reviews, research, documentation, etc.) on the quality of the tests (item 4; $p = 0.005$; $d = 0.25$). Secondly, considering the items referring to regulation and control over tests and their use, professors were more in favor of the establishment of a system of accreditation of test users (item 3; $p = 0.003$; $d = 0.28$), and considered to a lesser extent that controls over the use of tests should be minimal (item 16; $p < 0.001$; $d = 0.32$). However, the practitioners were slightly more in favor of the CGCOP playing a more active role in regulating and improving the use of tests (item 19; $p = 0.039$; $d = 0.18$). Thirdly, with regard to attitudes towards tests, professors showed even more favorable opinions than practitioners on two of the items; in particular, they indicated that tests are an excellent source of information combined with other data (item 22; $p < 0.001$; $d = 0.32$), and that, used correctly, they are of great help to the psychologist (item 23; $p = 0.001$; $d = 0.26$). Fourthly, with respect to the items referring to new technologies, the professors were more favorable to the use of the Internet in the application of tests. They considered that if used appropriately it can improve the quality of the application (item 15; $p = 0.032$; $d = 0.22$), and they questioned to a lesser extent the issue of privacy being exposed when using the Internet (item 17; $p = 0.013$; $d = 0.23$). In the academic setting, it seems that computerized tests are replacing paper-and-pencil tests at a slower pace than in the professional setting (item 5; $p = 0.041$; $d = 0.22$). Finally, as far as problems are concerned, it is curious that all the problems regarding test use (items 28.1 to 28.10 in Table 1) were reported more frequently by professors than by practitioners (Table 2) ($p = 0.004$ for item 28.1, and $p < 0.001$, for items 28.2 to 28.10, with effect sizes ranging from $d = 0.30$, for 28-1 to $d = 0.88$ for 28.9).

KNOWLEDGE AND ASSESSMENT OF THE TEST REVIEWS PROMOTED BY THE CGCOP

As regards knowledge of the test reviews carried out by the CGCOP, as the age of the groups increases, the higher the percentage of practitioners who said they know the reviews ($p < 0.001$; $2_{\text{partial}} = \eta^2 0.028$), with proportions varying between 0.05 for the group between 22 and 29 years of age, and 0.31 for the group aged 60 years or more (Table 3). Among those who know them, there were no statistically significant differences in the ratings according to age. In terms of gender, the proportion of men who were aware of these reviews

(0.34) differs from the proportion of women (0.18) ($p < 0.001$; $d = 0.39$), although among those who were aware of the reviews, there were no statistically significant differences in their evaluations. With regard to the sector (public vs. private), there were no statistically significant differences in the percentage of practitioners who were aware of the reviews in each case. However, it is the practitioners in the public sector who considered these reviews to be most important for improving test quality ($p = 0.004$; $d = 0.32$). Finally, the differences between the proportions of professors and practitioners who were aware of the reviews were very marked: 0.71 for professors and 0.23 for practitioners ($p < 0.001$; $d = 1.13$).

SOME CONCLUSIONS AND REFLECTIONS

The results obtained reveal some differences in the opinions of psychology practitioners on the use of the tests according to age, gender, and work sector. Before analyzing these differences, it should be noted that a) the opinions about tests were positive (Muñiz et al., 2020) and, when statistically significant differences were observed, they were of small magnitude, since the effect sizes were small according to traditional criteria (Cohen, 1992); and b) the trends we point out are general, since within the same category of analysis there is a certain degree of variability in the opinions and practices of test use.

The results of the analysis according to age indicate, firstly, that practitioners, over the years, are continuing to gain knowledge about the tests. The training and information strategies promoted by the CGCOP should contribute, at least in part, to this continuous development throughout the professional career. We believe that some of the topics that require special attention because of their impact on testing and its use are the psychometric models of item response theory (IRT), computerized adaptive testing (CAT), continuous norming, clinical diagnostic models, network models, and outpatient assessment.

In addition, it is noted that younger practitioners place more importance on evidence-based practice to enhance the use of tests. If this trend continues, practitioners will be even more careful in the selection of tests, based on empirical evidence of their quality and usefulness. This will hopefully minimize problems such as the Barnum effect, the alchemist's fantasy, or tradition and institutional inertia to use certain tests (Lilienfeld et al., 2006). Efforts should continue to be made to promote practice based on scientific evidence, and specific actions should be directed to the most senior practitioners. The aspects to consider when selecting a test are well known regardless of the assessment objective (Canivez, 2019; Carretero-Dios & Pérez, 2007; Davis & Baillie, 2019), from the conceptual delimitation of the construct to be assessed to the quality and appropriateness of the interpretative rules. All these issues are considered in the test evaluation model (Hernández et al., 2016) and are presented in the reviews published on the



CGCOP website. Therefore, it is necessary to continue disseminating these reviews, since only 22.5% of the Psychology Association members say they are aware of them. Encouraging the selection and use of tests based on scientific evidence will also contribute to the continuous improvement of the image and confidence in the profession, which is not a trivial matter.

With the new generations, the replacement with computerized tests and most likely those administered online is predicted, as younger people have a more positive opinion towards new technologies. Thus, automation in scoring and generating reports, innovation in new item/test formats, with multimedia and game-based content (gamification), online assessment, and ambulatory assessment will become more and more frequent (Muñiz & Fonseca-Pedrero, 2019; Parshall et al., 2010; Sanz et al., 2020; Seelow, 2019; Wan & Henly, 2012). All these trends are part of what is known as Psychology 2.0 (Armayones et al., 2015), which is impacting all aspects of psychological assessment, and is being facilitated by the use of cell phones and other portable devices (Armayones et al., 2015; Chernyshenko & Stark, 2016; Muñiz & Fonseca-Pedrero, 2019). Even social networks have been used on occasion to perform evaluations based on «likes» (Kosinski et al., 2013). New psychometric models, such as network models (Borsboom & Cramer, 2013; Fonseca-Pedrero, 2018) or dynamic systems models (Nelson et al., 2017), are needed to analyze this type of data.

We believe that these technological advances open up tremendous opportunities in the world of assessment, but they also have their risks. For example, in a market dominated by

face-to-face testing, the COVID-19 pandemic has accelerated the process of tele-assessment, it has led to the emergence of remote test administration, and it has highlighted the need for tests that can be administered with guarantees in this context. In this regard, Elosua (2021) presents the main risks of the remote administration of tests (especially those that have not been created specifically for online administration) and offers a series of recommendations to follow. Muñiz et al. (2020) also warn of the need to be cautious and avoid some very sophisticated computerized proposals for online assessment of psychological and educational constructs that, despite being very attractive and driven by strong marketing campaigns, have no empirical evidence to support them. The new generations, more favorable to incorporating new technologies, should be especially attentive to these possible deceptions.

Differences of opinion according to gender and sector (public or private) are not substantial. The private sector is generally more reluctant to introduce control and regulation mechanisms in the use of tests (although it is in favor of doing so), while the public sector is slower to incorporate new technologies. With respect to gender, what is most striking is that women are the ones who consider the price of the tests to be too high. This suggests the possible existence of a gender-based wage gap. Although it would be necessary to control for many variables before affirming that there is a wage gap, some studies suggest this. Crothers et al. (2010) reported that American female psychologists are paid less than their male colleagues, controlling for years of experience and training, despite being a profession in which women predominate.

TABLE 3
KNOWLEDGE AND ASSESSMENT OF THE TEST REVIEWS CARRIED OUT BY THE COP

	AGE					GENDER				SECTOR				PRACTITIONERS vs. PROFESSORS								
	22-29 years (N=79) A		30-39 years (N=300) B		40-49 years (N=342) C		50-59 years (N=362) D		60 years (N=163) E		Men (N=328)		Women (N=920)		Public (N=332)		Private (N=872)		Practitioners (N=1,248)		Professors (N=95)	
No. of items and groups with differences	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD
29. Knows the annual test review carried out by the COP Test Commission (**Gender, **Age (A-C) (A-D) (A-E) [B-D] (B-E), **Prof.)	0.05	0.22	0.16	0.36	0.23	0.42	0.28	0.45	0.31	0.47	0.34	0.48	0.18	0.39	0.26	0.44	0.22	0.41	0.23	0.42	0.71	0.46
If they know the review:																						
29-1. I think it is important for improving the quality of tests (**Sector).	4.50	1.00	4.32	0.66	4.41	0.80	4.61	0.60	4.37	0.82	4.48	0.66	4.45	0.76	4.62	0.51	4.39	0.79	4.46	0.72	—	—
29-2. I consult the test reviews on the COP website.	3.75	1.89	3.72	1.28	3.73	1.19	3.89	1.14	3.84	0.90	3.94	1.05	3.72	1.20	3.94	1.03	3.75	1.18	3.81	1.15	—	—
29-3. It helps me in the choice of tests I use.	3.75	1.26	3.53	1.20	3.60	1.17	3.72	1.05	3.71	1.03	3.71	1.12	3.62	1.09	3.72	1.03	3.62	1.13	3.65	1.10	—	—
29-4. It is unnecessary; psychologists already have enough information to know which test to use.	1.50	1.50	1.79	1.16	1.63	0.98	1.58	1.03	1.53	0.95	1.60	1.15	1.63	0.92	1.56	1.03	1.65	1.02	1.62	1.01	—	—

p < 0.05, ** p < 0.01 for the variables indicated. Prof.: Practitioners vs. Professors Professors were not asked the evaluative questions about test evaluation. Question 29 is answered by the dichotomous response «Yes» [1]/ «No» [0], so the mean represents the proportion of people who are aware of the assessment. The remaining questions (29-1 to 29-4), are answered on a graduated scale of 5 alternatives (1-5). Age, gender, and sector refer only to the sample of practitioners. SD= Standard Deviation



Future studies should investigate this important aspect.

It should be noted that in this study we have considered the relationship between three important sociodemographic and occupational variables—age, gender, and sector (public or private)—and opinions on tests and their use. For future studies, it would be interesting to collect and analyze the role of other variables such as the autonomous community, the university graduated from, or the educational level (bachelor's degree, master's degree, or doctorate), among others.

Finally, it should be noted that the most marked differences, with larger effect sizes, are between practitioners and professors, especially when assessing the frequency with which different problems related to test use occur, with higher scores in general for professors. Although the samples differ substantially in size, the results suggest various options, which future studies should explore. It may be that there are more problems in the academic setting, where the consequences of inappropriate test use do not have such important direct consequences on the test takers, since in the academic world tests are mainly used for research purposes (although there are obviously minimum quality criteria that must be maintained). Another possibility is that, because they are experts in assessment and psychometrics, they are more demanding in the quality standards that should be met and therefore value the problems as more important. Or finally, it may indicate that there is a gap between the academic world and the professional world, as has been highlighted on some occasions by test publishers (Hidalgo & Hernández, 2019).

Before concluding, it is necessary to point out that the present study has a number of limitations. Firstly, although Muñoz et al. (2020) found that the sample represented the population of COP members reasonably well in some of the main variables, such as age, the data collection procedure employed does not guarantee that the sample is representative of the total number of psychologists in our country. This criticism extends to the sample of professors used. A second limitation, related to the previous one, is the small number of participants. The total of 1,248 participants in the sample of practitioners constituted only 1.6% of the practitioners registered with the Psychological Association at the time of the survey, although among professors, participation was somewhat higher (30%). Finally, the sample size is related to the last limitation. Some of the results presented could be due not only to the variables analyzed but also to the interaction with other variables (for example, the differences observed according to age may depend on the type of sector). Although it would have been interesting to analyze the interactions, this analysis was discarded, since some of the categories resulting from crossing the different variables were not sufficiently represented in our sample (with extreme cases in which there were no observations at all). For future opinion surveys on tests, it would be advisable to reinforce participation through different strategies in order to obtain larger samples to ensure that all the categories of interest are sufficiently represented.

In short, despite the limitations mentioned above, the results of the study allow important conclusions to be drawn. Although the measures promoted by the different associations and institutions to improve tests and their use seem to be having an impact on practitioners, there is still a long way to go, especially considering the changes that new technologies are bringing about in the way tests are created, administered, scored, and even interpreted. For this reason, the CGCOP test commission must continue to work on improving the skills and knowledge of test users, with training and informative actions, but also on continuing to improve the quality of the tests so that practitioners have quality tools at their disposal, without forgetting the importance of the evaluation process. At the same time, the dissemination of the different actions carried out and the tools available to psychologists should also continue to be promoted. In this sense, universities, through their professors, are fundamental agents for reaching the younger generations.

CONFLICT OF INTEREST

There is no conflict of interest.

REFERENCES

- American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (2018). *Estándares para pruebas educativas y psicológicas [Standards for educational and psychological testing]* (M. Lieve, Trans.). Washington, DC: American Educational Research Association. (Original work published 2014)
- Armayones, M., Boixadós, M., Gómez, B., Guillamón, 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.
- 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. doi: 10.1146/annurev-clinpsy-050212-185608
- Canivez, G. L. (2019). Evidence-based assessment for school psychology: Research, training, and clinical practice. *Contemporary School Psychology*, 23, 194-200. doi: 10.1007/s40688-019-00238-z
- Carretero-Dios, H., & Pérez, C. (2007). Standards for the development and review of instrumental studies: Considerations about test selection in psychological research. *International Journal of Clinical and Health Psychology*, 7, 863-882.
- Chan (2011) Feminising and masculinising primary teaching: a critical examination of the interpretive frameworks of male primary school principals in Hong Kong. *Gender and Education*, 23, 745-759. doi: 10.1080/09540253.2011.611041



- Chernyshenko, O. S. & Stark, S. (2016). Mobile psychological assessment. In F. Drasgow (Ed.), *Technology and testing* (pp. 206-216). Routledge. doi: 10.4324/9781315871493
- Cohen, J. (1992). A power primer. *Psychological Bulletin*, 112, 155-159. doi: 10.1037/14805-018
- Crothers, L. M., Schmitt, A. J., Hughes, T. L., Lipinski, J., Theodore, L. A., Radliff, K., & Ward, S. (2010). Gender differences in salary in a female dominated profession. *Gender in Management: An International Journal*, 25, 605-626. doi: 10.1108/17542411011081392
- Davis, H. (H.) IV, & Baillie, P. (2019). Cognitive and personality assessment in applied sport psychology: Evidence-based test selection. In M. H. Anshel, T. A. Petrie, & J. A. Steinfeldt (Eds.), *APA handbook of sport and exercise psychology, Vol. 1. Sport psychology* (p. 99-110). American Psychological Association. doi: 10.1037/0000124-000
- Elosua, P. (2017). Avances, proyectos y retos internacionales ligados al uso de tests en psicología [International advances, projects, and challenges related to the use of tests in psychology]. *Estudios de Psicología*, 34, 201-220. doi: 10.1590/1982-02752017000200002
- Elosua, P. (2019). *Los test psicológicos en la práctica profesional [Psychological tests in professional practice]*. FOCAD. Consejo General de Psicología de España.
- Elosua, P. (2021). Aplicación remota de test: riesgos y recomendaciones [The remote application of tests]. *Papeles del Psicólogo*, 42, 33-37. doi: 10.23923/pap.psicol2021.2952
- 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, 82-88.
- Elosua, P. & Muñiz, J. (2013). Proyectos españoles para una mejora en el uso de los tests [Spanish projects for an improvement in the use of tests]. *Psiencia. Latin American Journal of Psychological Science*, 6, 5-9. doi: 10.5872/psiencia/5.2.75
- EFPA (2005). *Meta-Code of ethics*. (www.efpa.eu).
- Evers, A., Muñiz, J., Hagemeister, C., Høstmælingen, A., Lindley, P., Sjöberg, A., & Bartram, D. (2013). Assessing the quality of tests: Revision of the EFPA review model. *Psicothema*, 25, 283-291. doi: 10.7334/psicothema2013.97
- Fonseca-Pedrero, E. (2018). Análisis de redes en psicología [Network análisis in psychology]. *Papeles del Psicólogo*, 39, 1-12. doi: 10.23923/pap.psicol2018.2852
- 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, 161-16. doi: 10.23923/pap.psicol2017.2844
- Gómez-Sánchez, L. E. (2019). Séptima evaluación de test editados en España [Seventh review of tests published in Spain]. *Papeles del Psicólogo*, 40, 205-210. doi: Séptima evaluación de test editados en España
- Greenberg, M. R., & Schneider, D. F. (1995). Gender differences in risk perception: Effects differ in stressed vs. non stressed environments. *Risk Analysis*, 15, 503-511. doi: 10.1111/j.1539-6924.1995.tb00343.x
- Gustafsson, P. E. (1998). Gender Differences in risk perception: Theoretical and methodological perspectives. *Risk analysis*, 18, 805-811. doi: 10.1111/j.1539-6924.1998.tb01123.x
- Haladyna, T. M. & Rodríguez, M. C. (2013). *Developing and validating test items*. Routledge. doi: 10.4324/9780203850381
- Hernández, A., Elosua, P., Wechsler, S.M. & Muñiz, J. (2020a) El uso de los Test en España y Latinoamérica: Perspectivas actuales y retos futuros [Test use in Spain and Latin America: Current perspectives and future challenges]. *Anuario Internacional de Revisiones en Psicología*.
- Hernández, A., Hidalgo, M. D., Hambleton, R. K., & Gomez-Benito, J. (2020b). International test commission guidelines for test adaptation: A criterion checklist. *Psicothema*, 32, 390-398. doi: 10.7334/psicothema2019.306
- 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.
- 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, 1-8
- 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. doi: 10.23923/pap.psicol2019.2886
- Irwing, P., Booth, T. & Hughes, D. J. (Eds.) (2018). *The Wiley handbook of psychometric testing: A multidisciplinary reference on survey, scale and test development*. John Wiley & Sons Ltd. doi: 10.1002/9781118489772
- ITC Guidelines for Translating and Adapting Tests (Second Edition) (2017). *International Journal of Testing*, 17, 1-34. (www.InTestCom.org)
- 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. doi: 10.1073/pnas.1218772110
- Lilienfeld, S., Wood, J., & Garb, H. (2006). Why questionable psychological tests remain popular. *Scientific Review of Alternative Medicine*, 10, 6-15.
- McDonald R. P. (1999). *Test theory: A unified treatment*. Mahwah, Lawrence Erlbaum. doi: 10.4324/9781410601087



- 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, 108-121.
- Muñiz, J., & Fonseca, E. (2017). *Construcción de instrumentos de medida en psicología [The construction of measurement instruments in psychology]*. Madrid: FOCAD. Consejo General de Psicología de España
- Muñiz, J. & Fonseca, E. (2019). *Diez pasos para la construcción de un test [Ten steps in the construction of a test]*. *Psicothema*, 31, 7-16. doi: 10.7334/psicothema2018.291
- Muñiz, J., Elosua, P. & Hambleton, R.K. (2013). Directrices para la traducción y adaptación de los tests: segunda edición [Guidelines for the translation and adaptation of tests: second edition]. *Psicothema*, 25, 151-157. doi: 10.7334/psicothema2013.24
- Muñiz, J., Hernández, A., & Fernández-Hermida, J. R. (2020). La utilización de los test en España: el punto de vista de los psicólogos [The use of tests in Spain: the point of view of psychologists]. *Papeles del Psicólogo*, 41, 1-15. doi: 10.23923/pap.psicol2020.2921
- Muñiz, J., Fernández-Hermida, J. R., Fonseca, E., Campillo, A. & Peña, E. (2011). Evaluación de test editados en España [Review of tests published in Spain]. *Papeles del Psicólogo*, 32, 113-128.
- 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. doi: 10.1001/jamapsychiatry.2017.0001
- Parshall, C. G., Harnes, 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. doi: 10.1007/978-0-387-85461-8
- Prieto, G., & Muñiz, J. (2000). Un modelo para evaluar la calidad de los tests utilizados en España [A model to evaluate the quality of the tests used in Spain]. *Papeles del Psicólogo*, 77, 65-72.
- Sanz, S., Luzardo, M., García, C., & Abad, F. J. (2020). Detecting cheating methods on unproctored Internet tests. *Psicothema*, 32, 549-558. doi: 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, 1-16.
- 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 evaluation of tests published in Spain: a participatory experience]. *Papeles del Psicólogo*, 42, 1-9. doi: 10.23923/pap.psicol2020.2937
- Wan, L., & Henly, G. A. (2012). Measurement properties of two innovative item formats in a computer-based test. *Applied Measurement in Education*, 25, 58-78. doi: 10.1080/08957347.2012.635507