

THE OPINION OF SPANISH PSYCHOLOGISTS ON THE USE OF TESTS

José Muñiz and José Ramón Fernández-Hermida

University of Oviedo

National and international psychological organizations interested in improving tests and testing practices follow two complementary strategies. On the one hand they try to restrict the use of tests to those professionals who have been properly trained in the field of tests and testing, and on the other, the dissemination of information on tests and testing is encouraged. In order to implement both strategies in a rigorous way it is essential to know the opinions of professional psychologists. To this end the European Federation of Psychologists' Associations (EFPA) has developed a questionnaire made up of 33 items. In this paper we present the answers of Spanish psychologists to the EFPA questionnaire. A total of 3126 psychologists responded to the questionnaire, 2235 women (71.5%), and 891 men (28.5%), all of them members of the Spanish Psychological Association (COP). The mean age was 41.92 years and the standard deviation 0.43. Mean number of years working as professionals was 12.5, with a standard deviation of 8.9. As regards the field of specialization, 69.6% work in Clinical Psychology, 13.6% in Educational Psychology, 6.4% in Work and Organizational Psychology, and 10.4% in other fields, such as sports, forensic, social services, or traffic, while 3.8% are unemployed. The results are articulated around eight main dimensions, which are discussed in detail, comparing the results obtained in the main fields of specialization, Clinical, Educational and Work and Organizational Psychology. Psychologists show a highly positive attitude toward tests and testing, though different aspects in which there is room for improvement are also pointed out. Finally the results are analyzed in detail, and some future perspectives considered.

Key words: Tests, Testing practices, Psychologists' opinions, EFPA.

Las organizaciones nacionales e internacionales interesadas en mejorar la práctica de los tests siguen dos líneas de actuación complementarias, por un lado se trata de restringir el uso de los tests a aquellos profesionales preparados para ello, y por otro se intenta difundir todo tipo de información técnica sobre los tests y su adecuada utilización. Para una correcta aplicación de estas dos estrategias es fundamental conocer de forma rigurosa las opiniones de los psicólogos profesionales sobre la práctica de los tests. Con tal fin la Comisión de Tests de la Federación Europea de Asociaciones de Psicólogos (EFPA) ha desarrollado un cuestionario compuesto por 33 ítems. En el presente trabajo se recogen las respuestas de los psicólogos españoles a la encuesta de la EFPA. Respondieron 3.126 psicólogos profesionales, 2.235 mujeres (71,5%) y 891 hombres (28,5%), todos ellos miembros del Colegio Oficial de Psicólogos. La edad media fue de 41,92 años y la desviación típica de las edades 10,43. La media de años en la profesión fue de 12,5, con una desviación típica de 8,9. El 69,6% pertenecían al ámbito de la psicología clínica, el 13,6% a educativa, el 6,4% a trabajo y el 10,4% a otras especialidades, tales como deporte, jurídica, tráfico, o servicios sociales, el 3,8% están desempleados. Los resultados se articulan en torno a ocho grandes dimensiones, que se analizan con detalle en función de las especialidades de Clínica, Trabajo y Educativa. Los psicólogos muestran una actitud general muy positiva hacia la utilización de los tests en el ejercicio de su profesión, si bien ponen de manifiesto algunos puntos débiles y limitaciones que deben ser mejorados cara al futuro. Se finaliza comentando los resultados en detalle y analizando las perspectivas de futuro.

Palabras clave: Tests, Uso de los tests, Opinión psicólogos, EFPA.

More than a decade ago, at the request of the Standing Committee on Tests and Testing of the European Federation of Psychologists' Associations (EFPA), a survey was carried out in six European countries, including Spain, to find out the

opinions of European psychologists on various aspects related to tests (Muñiz & Fernández-Hermida, 2000; Muñiz et al., 2001). The basic objective was to discover at first hand the opinions of psychology professionals in Europe about these aspects with a view to organizing actions and projects for improving test use. A summary of the projects and actions carried out in recent years by EFPA and by the International Test Commission (ITC) can be consulted in Muñiz and Bartram (2007). Ten years

Correspondence: José Muñiz. Facultad de Psicología. Universidad de Oviedo. Plaza Feijoo, s/n. 33003 Oviedo. España. E-mail: jmuniz@uniovi.es

having passed since the first survey on tests, the EFPA Tests Committee considered it time to sound out European psychologists' views again. One of the aims was to follow up the results previously obtained, but in view of the importance for psychology practice in general and test use in particular of two major technical advances – the Internet and the new information technologies – a further aim was to evaluate how these two factors affect the everyday practice of professionals in relation to tests. Thus, in addition to the key questions from the original survey, it was decided to include some items about the influence of these factors, such as: “Are computerized tests substituting paper-and-pencil tests in psychologists' everyday practice?”, or “Does Internet constitute a clear advance for psychological assessment?”

But before embarking on a description of the new survey and the results obtained, it is worth reviewing what has been done in recent years, at both the Spanish and European levels, with a view to improving test use. The ethical and deontological use of tests rests on two basic cornerstones, insofar as on the one hand tests must have appropriate psychometric properties, and on the other, they must be used correctly, from their application and scoring to the use that is made of the scores obtained. The organizations devoted to improving the use of tests, both national (Spanish Psychological Association, COP), and international (EFPA, ITC, or the American Psychological Association, APA), are responsible for a wide range of actions and projects framed within two broad strategies which we could call *restrictive* and *informative*.

The restrictive strategy refers to actions with the purpose of limiting the use of tests to those professionals properly qualified to use them. The systems employed vary from one country to another (Bartram, 1996; Bartram & Coyne, 1998; Muñiz, Prieto, Almeida, & Bartram, 1999), even if one of the most widely used across many countries, Spain included, involves classifying tests according to APA criteria in three categories (A, B, C), from least specialized to most specialized, with the use of those in categories B (group tests of a cognoscitive nature or assessing Personality) and C (individual scales and projective tests) being restricted to psychologists. Another common option is for professionals to obtain specific accreditation in which they attest their adequate knowledge of the instrument(s) in question. Although these

restrictions and others are to be recommended, they do not in themselves guarantee the appropriate use of tests (Moreland, Eyde, Robertson, Primoff, & Most, 1995; Simner, 1996); this strategy must be complemented by the dissemination of information to all the relevant parties, including professionals, users, institutions, and society in general.

Actions carried out within the framework of the strategy we have called informative include all types of initiative aimed at disseminating information on test practice. It is understood that the more information available to professionals, users, families, and in general all those involved in test use, the lower the likelihood of tests being used inappropriately. With this in mind, various national and international organizations have developed ethical and deontological codes, as well as guidelines for the adequate use of tests. Important examples of the former would be the EFPA Meta-code of ethics (2005), the code developed for North America by the Joint Committee on Testing Practices (2002), or the guidelines of the European Association of Psychological Assessment (Fernández-Ballesteros et al., 2001). Good reviews can be found in authors such as Koocher and Keith-Spiegel (2007), Lindsay, Koene, Ovreeide and Lang (2008), or Leach and Oakland (2007), and not least in the latest special issue devoted to these questions in the Spanish journal *Papeles del Psicólogo* (2009). Apart from these codes, a set of guidelines is available indicating the steps to follow from the construction of a test through its application and interpretation to the application of the results (Bartram, 1998; Brennan, 2006; Downing & Haladyna, 2006; Muñiz, 1997). Worthy of special mention are the technical standards developed by the American Psychological Association and two other organizations (APA, AERA and NCME, 1999), as well as the guidelines drawn up by the International Test Commission (ITC) for the translation and adaptation of tests from one culture to another (Hambleton, Merenda, & Spielberger, 2005). Both sets of guidelines are currently undergoing a review process, so that it will not be long before new versions are available. To consult other guidelines on the use of tests in general, of computerized tests and Internet, or test use in the field of work and organizations see, for example, the work by Muñiz and Bartram (2007) or the websites of the ITC

(www.intestcom.org) and EFPA (www.efpa.eu). Relevant information can also be found on the website of the Spanish Psychological Association (COP), in the Tests Commission section (www.cop.es). In addition to the ethical codes and guidelines, two measures merit particular attention within the context of the informative strategy: on the one hand, a new ISO standard that is about to be published and will regulate all aspects related to person assessment in work contexts, and on the other, the test-assessment models developed in different countries, among them Spain. We shall now discuss both proposals.

ISO Standard 10667

The ISO acronym stands for the International Organization for Standardization (www.iso.org), which develops standards in all industrial and service sectors. The corresponding organization in Spain is AENOR (www.aenor.es). At the initiative of the German representation, a process was begun to draw up a new ISO standard for regulating everything related to the assessment of persons in the work context. Naturally, this new standard is of considerable interest to psychologists, given its central role for assessment in job situations, and the Spanish Psychological Association (COP) has actively participated in the international commission which is drawing up the standard, together with other national psychological associations, such as those of North America (APA) or Britain (BPS), to mention just two. After several meetings, a text based on a broad consensus has been drawn up, with only the finishing touches remaining. These ISO standards are very important, since once they are approved, companies can obtain certification a guarantee that they comply with them. Despite their lack of legal status in the strict sense, they do constitute an important regulatory standard for the market: it is not the same to be certified as it is not to be. Although the definitive text in this case is not yet published, the objective of the standard is to regulate the process of assessment of people in work and organizational contexts, covering the entire assessment process, from the setting up of the assessment contract through the assessment methodology itself to the use of the results. It will be applicable to the procedures and methods employed at the individual level (selection, advice,

training, etc.), the group level (work-team climate and cohesion) and the organizational level (work climate, company culture, satisfaction, etc.). The standard includes descriptions of the competencies, obligations and responsibilities of the clients and providers of the assessment service, before, during and after the assessment process. It also provides guidelines for all the parties involved in assessment, including the actual person assessed and those who receive its results. In sum, once it is published and the certification process begins, this new standard may represent a significant step in the direction of good practice in person assessment in work and organizational contexts.

The assessment of tests

Within the strategy of disseminating information about tests and testing practice, professional psychologists have highlighted, at every opportunity, the need for the availability of more technical information about tests (Muñiz & Fernández-Hermida, 2000; Muñiz et al., 2001). This has led to a call from the EFPA Tests Commission for the development of a European model of test assessment, inspired in previous models such as those of Britain (Bartram, 1996, 1998), Holland (Evers, 2001a, b) or Spain (Prieto & Muñiz, 2000). The European model can be consulted on the EFPA website (www.efpa.eu). The core idea of the model is to assess the psychometric properties of tests systematically and quantitatively and to offer this information – objective, up-to-date and obtained by experts – to potential test users. The work by Prieto and Muñiz (2000) describes the Spanish model, made up of three broad sections. The first involves a technical description of the test, and is made up of 31 items concerning the name of the test, its author, the construct measured, the area of application, and so on. The second section includes the technical appraisal of the instrument's characteristics, providing experts' ratings of characteristics such as its theoretical foundations, its adaptation/translation (if it was constructed in another country), its reliability, its validity, and so on. To this end, it includes 32 closed and 6 open items. In the majority of the closed items there are five categories ordered according to the quality of the characteristic assessed. In the open items respondents are required to provide a reasoned explanation of the responses to the closed items

and a rating of each characteristic. In the final section, respondents must provide an overall appraisal of the test and a summary of the first two sections, with the aim of summarizing all the information on a technical data sheet (Prieto & Muñiz, 2000). The model can be consulted in the cited work or on the COP website, in the Tests Commission section: www.cop.es.

What is really new in relation to this test assessment model is that at the latest meeting of the COP Tests Commission it was decided unanimously to begin applying the model to tests published in Spain. Starting with those most widely used, the idea is to assess a minimum of around twenty tests a year, with a view to having assessed the majority in the relatively near future, bringing Spain into line with other countries, such as Holland. For their assessment, tests will initially be sent to two experts in both psychometric aspects and the specific area addressed by the test. If the experts' opinions coincide, a final report based on them will be drawn up. Should there be divergences between the experts, the test will be sent to a third expert before the writing of the final report. The resulting assessments will be disseminated as widely as possible, being published in journals that reach all COP members, as well as on the Association's website.

It is in this context of the improvement of test practice that the EFPA survey for professional psychologists on different aspects of test use takes on real significance. Knowledge of their opinions can be used to develop measures aimed at improving the weak points identified by the professionals. We shall now present the results obtained in Spain. It is worth noting that a total of 17 countries participated in the 2009 edition, compared to just six in that of 1999, so that from the point of view of participation there has clearly been considerable progress.

PARTICIPANTS IN THE SURVEY

The sample is made up of 3126 professional psychologists who responded to a survey sent to 51,545 members of the Spanish Psychological Association. The most relevant descriptive data are shown in Tables 1 and 2. On comparing some of the data of the sample with those corresponding to the population (Table 1), we find very similar values, so that there do not appear to be large biases in relation to the sample used, which

represents 6% of the population. It should be pointed out that women predominate in the psychology profession, accounting for 78% of the total, even though in the sample the percentages vary slightly, with 71.5% women and 28.5% men; this would reflect a greater disposition to respond in men. By specialization, in Clinical Psychology there are 29% men, in Educational, 22%, and in Work and Organizational, 45%, indicating that the last of these clearly attracts more men than the other two. In terms of specialization in the sample as a whole, the three classic fields continue to predominate (Table 2), Clinical Psychology being a clear leader with 69.6%, followed by Educational (13.6%) and Work and Organizational Psychology (6.4%); the remaining fields (sports, forensic, traffic, social services, and others) account for 10.4%. A total of 32.6% work in the public sector, and 63.6% in the private sector, with 3.8% unemployed. Currently, psychology professionals in Spain are a relatively young group, with 14% aged between 20 and 29, 28.9% between 30 and 39, 30.8% aged 40 to 49, 21.7% aged between 50 and 59, 3.9% between 60 and 68, and 0.7% aged 70 or over.

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

	Sample	Spanish Psychological Association
Participants	3.126	51.545
Women	71.5%	78.1%
Men	28.5%	21.9%
Mean Age (SD)	41.92 (10.43)	40.58(10.13)
Years of Professional Practice (SD)	12.50(8.90)	10.33(8.60)

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

Professional Field	%
Clinical	69,6
Educational	13,6
Work and Organizational	6,4
Other	10,4
Sector	
Public	32,6
Private	63,6
Unemployed	3,8

QUESTIONNAIRE

In order to obtain the opinions of psychologists about tests and testing practice, a 33-item questionnaire was used (see Appendix), originally developed in English by the Tests Commission of the European Federation of Psychologists' Associations (EFPA). The first 32 items are Likert-type with five categories, scored from 1 to 5, whilst the final item was open, for the professionals to indicate

which tests they used most in their everyday practice. They were constructed on the basis of the original scale used in 1999, some items being removed and others – related to the use of computerized tests and Internet – added. Twenty of the items from the 1999 survey were maintained, which made it possible to compare the results from that time with those obtained now. It was translated into Spanish and back-translated to English to enable it to

TABLA 3
ANÁLISIS DE COMPONENTES PRINCIPALES

Componentes								
Items	I	II	III	IV	V	VI	VII	VIII
Item 25-5	0.773							
Item 25-2	0.770							
Item 25-8	0.764							
Item 25-7	0.750							
Item 25-4	0.747							
Item 25-3	0.740							
Item 25-6	0.704							
Item 25-1	0.463							
Item 23		0.868						
Item 22		0.862						
Item 21		0.728						
Item 24		0.539						
Item 3			0.700					
Item 12			0.689					
Item 8			0.624					
Item 11			0.616					
Item 19			0.599					
Item 17				0.705				
Item 20				0.680				
Item 13				0.625				
Item 10				0.564				
Item 1					0.833			
Item 6					0.736			
Item 2					0.618			
Item 7						0.680		
Item 5						0.610		
Item 15						0.593		
Item 9						0.402		
Item 14							0.739	
Item 18							0.615	
Item 16							0.481	
Item 4								0.503
% of	13.27	8.06	7.88	7.10	5.54	4.80	4.49	3.68
Variance								
%								
Accumulated	13.27	21.33	29.21	36.32	41.85	46.65	51.15	54.82

Note. The eight components with eigenvalues over 1 were orthogonally rotated. Weights under 0.45 were removed to make it easier to read the table, except where the variable did not attain that weight.

be checked that the two versions, the original and the one generated from the Spanish version, were essentially equivalent, as recommended in the ITC guidelines (Hambleton et al., 2005). With the Spanish version, various qualitative and quantitative pilot studies were carried out to ensure that the survey items could be perfectly understood and interpreted by the population they addressed (Wilson, 2005).

DATA COLLECTION AND ANALYSIS

For the collection of data, a questionnaire was sent to all psychologist members of the Spanish Psychological Association (COP), preceded by a letter of introduction from Francisco Santolaya, President of the COP, in which he explained the reasons for the survey. A stamped addressed envelope was also included, in which respondents could place the completed questionnaire before posting it in any letter box.

Descriptive statistical analyses were carried out on the items and on the general data requested from participants. The dimensional structure of the scale items was determined by means of a principal components analysis with orthogonal rotation (varimax) of the components with eigenvalues greater than one, following the Kaiser criterion. Although from a technical point of view the Maximum Likelihood method may be more advisable (Ferrando & Anguiano, 2010), this strategy is maintained in order to permit better comparison with previous results (Muñiz et al. 2001). Reliability of the test was estimated by means of Cronbach’s (1951) alpha coefficient, and comparisons of the means of the items were made through variance analysis. All analyses were carried out using SPSS-15.

DIMENSIONS ASSESSED BY THE QUESTIONNAIRE

The alpha coefficient of the scale was 0.665, indicating that the internal consistency of the scale is not very high. This is to be expected, since at no time was it attempted to obtain a scale with high internal consistency, but rather to assess different aspects involved in test practice.

As can be seen in Table 3, the items of the scale revolve around eight dimensions, obtained through principal components analysis, which explain 54.82% of the total variance. The first component groups all the items related to problems with the use of tests. The second component is made up of items referring to psychologists’ attitudes

towards tests. The third component emerges as very clear, referring to the need to regulate test use, either legally or through the national or European professional organizations. These first three components fully coincide with those obtained for the instrument applied ten years previously (Muñiz & Fernández-Hermida, 2000). The fourth component is made up of items relating to the use of Internet and computerized reports. This dimension is new, given that these items had not been included in the 1999 version of the questionnaire. The fifth component refers to training and knowledge in relation to tests. The sixth is again constituted by items related to Internet and computerized tests, with some weight also from an item related to test use by non-psychologists. The seventh, with three items, focuses on permissiveness in the use of tests, and the eighth covers the item related to the technical information about tests possessed by professionals. The structure is very clear, and reflects well the basic dimensions to take into account in assessments of test use. It is highly similar to the structure found in the 1999 application, two new dimensions about Internet and computerized tests having been added.

OPINIONS ABOUT THE USE OF TESTS

Table 4 shows the means and standard deviations of participants’ responses to the questionnaire items. It shows the data for the total sample and broken down by the professional specializations of Clinical, Work and Organizational, and Educational Psychology. An asterisk after the text of the item denotes that the differences between the means of the three specializations were statistically significant at the 95% confidence level.

The detailed results can be seen in Table 4; here we shall discuss some of the most noteworthy data for each one of the dimensions of the questionnaire (Table 3). In the first dimension, related to problems with test use, it can be seen that although the situation is not serious, the mean rating being 3.12, there are clearly some aspects in which there is room for improvement. People are still making photocopies of tests (3.51), and in the opinion of psychologists they are not always up to date with their knowledge (3.25) or in the habit of discussing their interpretations with other professionals (3.33). It is also found that, as suspected, the problems with test use are more accentuated in the field of Work and

TABLE 4
MEAN AND STANDARD DEVIATION OF EACH SURVEY ITEM BY SPECIALIZATION
(CLINICAL, EDUCATIONAL AND WORK/ORGANIZATIONAL) AND OVERALL

Items	Clinical		Educational		Work and Organizational		Overall	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD
1.- The training provided in the Psychology degree is sufficient for the correct use of most tests	2.41	1.18	2.44	1.14	2.61	1.23	2.43	1.18
2.- The training provided on courses and Masters programmes is sufficient for the correct use of most tests*	3.12	1.08	3.07	1.01	2.90	1.00	3.09	1.07
3.- The European Federation of Psychologists' Associations (EFPA) should set up a system for accrediting test-user competency*	3.34	1.37	3.40	1.34	3.89	1.23	3.39	1.36
4.- Professionals have sufficient information (independent reviews, research, documentation, etc.) on the quality of tests published in Spain	2.74	1.14	2.72	1.09	2.72	1.09	2.73	1.12
5.- In my professional field computerized tests are gradually replacing paper-and-pencil tests*	2.89	1.35	2.96	1.24	3.53	1.26	2.94	1.35
6.- My current knowledge in relation to tests is basically what I learned on my Psychology degree course	2.57	1.36	2.48	1.28	2.74	1.37	2.59	1.36
7.- The application of tests via the Internet has many advantages compared to application using the classical paper-and-pencil format*	2.75	1.21	2.64	1.14	3.11	1.21	2.78	1.20
8.- The use of psychological tests should be restricted to qualified psychologists*	4.12	1.19	4.15	1.11	4.39	.98	4.12	1.17
9.- Although non-psychologists might apply and score tests, interpreting the results and providing information about them should be left to psychologists	4.39	1.17	4.41	1.06	4.52	1.05	4.39	1.16
10.- Reports generated automatically by computer have no validity	2.96	1.14	2.94	1.14	2.87	1.12	2.94	1.14
11.- Standards and guidelines that define 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 American Psychological Association (APA)].*	4.07	.98	4.12	.91	4.27	.87	4.10	.96
12.- Legislation is necessary to control the most serious abuses related to tests*	3.99	1.04	4.03	.99	4.26	.96	4.01	1.04
13.- The application of tests via Internet puts some examinees at a disadvantage	3.53	1.10	3.51	1.13	3.46	1.11	3.54	1.10
14.- Anyone capable of demonstrating their competence in the use of tests (be they a psychologist or not) should be authorized to use them	2.11	1.34	2.07	1.29	2.22	1.35	2.10	1.32
15.- If used appropriately, the Internet can greatly improve the application of tests*	3.04	1.12	3.00	1.08	3.34	1.12	3.08	1.11
16.- Controls on tests should be minimal, since they inhibit the development of new ideas and new assessment procedures	1.94	1.06	1.93	1.01	1.89	.99	1.93	1.04
17.- The application of tests via Internet does not permit the protection of user privacy*	2.95	1.25	3.09	1.18	2.73	1.25	2.95	1.24
18.- Publishers should be allowed to sell any test as they see fit	1.80	1.11	1.67	.95	1.71	1.01	1.77	1.09
19.- The Spanish Psychological Association (COP) should play a more active role in regulating and improving the way tests are used*	4.09	1.07	4.22	.89	4.23	1.03	4.13	1.03
20.- The application of tests via Internet makes them vulnerable to fraud	3.80	1.09	3.78	1.11	3.63	1.15	3.78	1.10
21.- In the course of my profession I use tests regularly*	3.77	1.29	3.98	1.28	3.78	1.16	3.76	1.30
22.- Tests constitute an excellent source of information if combined with other psychological data*	4.44	.89	4.59	.79	4.49	.81	4.46	.87
23.- Used correctly, tests are of great help to the psychologist*	4.38	.89	4.55	.76	4.53	.74	4.41	.88
24.- Taking into account all aspects, I think that test use in my country has improved over the last decade*	3.58	.97	3.69	.87	3.42	.97	3.58	.96
25.- Please estimate the frequency in your professional context of the following problems related to test use (1: highly infrequent; 5: very common)								
(1) Making photocopies of material protected by copyright	3.53	1.36	3.50	1.35	3.48	1.38	3.51	1.38
(2) Making assessments using inappropriate tests*	2.62	1.31	2.47	1.23	3.13	1.25	2.64	1.31
(3) Not being up to date*	3.23	1.25	3.08	1.24	3.51	1.16	3.25	1.23
(4) Not discussing one's interpretations with others*	3.32	1.26	3.14	1.26	3.58	1.20	3.33	1.25
(5) Not taking into account the measurement errors of scores *	3.10	1.22	2.97	1.20	3.30	1.19	3.10	1.22
(6) Not restricting the application of tests to qualified personnel*	2.93	1.49	2.76	1.44	3.39	1.45	2.92	1.49
(7) Not taking into account local conditions (country, region) that may affect validity*	3.19	1.31	3.15	1.30	3.47	1.25	3.21	1.31
(8) Making interpretations that go beyond the scope of the test*	2.96	1.37	2.86	1.32	3.24	1.36	2.97	1.36

Note. The asterisk denotes statistically significant differences between the means of the item according to specialization, $p < 0.05$.

Organizational Psychology (3.39) than in Clinical (3.07) and Educational Psychology (2.99) – at least in the perceptions of professionals. These differences, now confirmed in the Spanish case, are what motivated the EFPA Tests Commission to launch a pilot project to explore the possibility of accreditation for test users in the field of Work and Organizations. An active part in this project is played by the COP, which named Dr Ana Hernández from the University of Valencia as a representative to the EFPA committee. The idea is the establishment of a European accreditation in the field of tests (Eurotest) similar to the Europsy (Bartram & Roe, 2005; Lunt, 2005; Peiró, 2003). Another initiative aimed at improving test use in this field is ISO standard 10667, mentioned previously, which sets out to regulate person-assessment processes in work contexts.

The second factor refers to the attitudes of psychologists toward tests, the data from the 1999 survey being confirmed (Muñiz & Fernández Hermida, 2000) insofar as there is a favourable attitude toward tests when they are combined with other psychological data (4.46). It is interesting and encouraging to note that professionals consider the use of tests to have improved in Spain over the last decade (3.58), since although there are still much to do, it seems things are going in the right direction. By specialization, it is those working in the Educational field who make most use of tests in their everyday practice (3.98), and once more it is in that of Work and Organizational Psychology where the perceived improvement in test use over the last decade is lowest (3.42), even if it is above the mean of the scale.

The third dimension refers to the need for regulation of test use. Here we can find a highly favourable opinion of professionals toward both the introduction of legal measures and the intervention of professional associations for improving the use of tests. There is support for the idea of the EFPA establishing a system for the accreditation of test-user competence, with most enthusiasm for such measures coming from the Work and Organizational field (3.89). In all five items making up this factor, indeed, it is professionals from this field who most emphatically call for measures to regulate test use, endorsing the intervention of professional associations at the national and international levels.

The fourth factor focuses on the use of Internet and computerized reports. It would seem clear that psychologists are quite sceptical about the use of computerized reports, as well as about the arrival of Internet in the field of assessment. This should be interpreted not as a defensive attitude toward the new technologies, but rather as one of precaution about issues arising in relation to Internet, such as the threat to privacy or fraud, or in relation to the disadvantages associated with users who are unfamiliar with the web. As regards reports generated automatically by computer, their validity is not rejected out of hand (2.94), but nor are they given an unqualified blessing; what is clear is that such reports can be of great help to psychologists but in no way can they substitute them, being nothing more than tools which the professional should use where appropriate. Of the four items making up this factor, only in one are there statistically significant differences between specializations – that which refers to the maintenance of privacy on using Internet –, with those from the field of Work and Organizational Psychology considering that an adequate level of privacy can be maintained. Those belonging to this sector are obviously more accustomed than Clinical and Educational Psychologists to working in remote-assessment contexts, since the systems employed today permit high levels of privacy when the web is used for these purposes.

The fifth factor is made up of three items referring to psychologists' training and knowledge in relation to tests. Here, a clear need for specific training is expressed by the professionals, given that neither the Psychology degree (2.43), nor subsequent Masters courses (3.09) meet such training needs. On this point there is a consensus between those from the three specialization groups, who all obviously recognize that if technical knowledge in general remains valid for some five years, the area of tests is no exception, so that ongoing programmes and continual updating are essential. New tests are developed, new techniques and new models emerge, and while the knowledge acquired on degree courses and some Masters programmes constitutes an essential base, this must be periodically complemented and brought up to date through specialized training. This represents a substantial challenge for professional associations and universities, as well as other institutions related to the profession.

The sixth dimension involves, like the fourth, aspects related to Internet and computerized tests. It does not appear, for now, that computerized tests are replacing paper-and-pencil tests, though it is in the Work and Organizational field in which the greatest progress has been made. It is observed that the use of Internet is still scarce among professionals, Work and Organizational Psychologists being those most likely to use it. Respondents' opinion on the use of tests by non-psychologists is unequivocal, in the sense that although they accept application and scoring by those without qualifications in the discipline, the interpretation of scores must be the exclusive preserve of psychologists (4.39). It

is one thing to apply a test and score it, and quite another to be capable of making the appropriate inferences about human behaviour on the basis of the scores, for which extensive psychological knowledge is required.

The seventh factor refers to permissiveness in the use of tests. The professionals make it abundantly clear that the publishing and use of tests must be controlled, and there is consensus among those from the three fields of specialization.

Finally, a single item makes up the eighth factor, related to the technical information about tests possessed by professionals. It is the unanimous view of the three major specialization groups that professionals should have more information of this type. This is in confirmation of what emerged from the 1999 survey, and has led the COP, through its Tests Commission, to launch a project for assessing all tests published in Spain, the results being made available to professionals. It was estimated that the first appraisals would be published by 2011.

TABLE 5
THE 25 TESTS MOST COMMONLY USED BY SPANISH PSYCHOLOGISTS

Name of the test	N	%
WISC* (Wechsler Intelligence Scale for Children)	649	22.70
16PF (16 Personality Factors)	609	22.37
MCMI (Millon Clinical Multiaxial Inventory)	489	17.96
MMPI (Minnesota Multiphasic Personality Inventory)	480	17.63
BDI (Beck Depression Inventory)	372	13.66
WAIS* (Wechsler Adult Intelligence Scale)	370	12.93
STAI (State-Trait Anxiety Inventory)	316	11.60
RORSCHACH (Rorschach)	154	5.66
SCL-90 (Symptom Checklist 90)	143	5.25
RAVEN (Raven Progressive Matrices)	137	5.03
TAMAI (<i>Test Autoevaluativo Multifactorial de Adaptación Infantil</i> ; Multi-Factor Self-Assessment Test for Child Adjustment)	120	4.41
MMSE (Mini Mental State Examination)	113	4.15
MSCA (McCarthy Scales of Children's Abilities)	95	3.49
BADYG (<i>Batería de Aptitudes Diferenciales y Generales</i> ; Battery of Differential and General Aptitudes)	93	3.42
TALE (<i>Test de Análisis de Lecto-Escritura</i> ; Reading-Writing Analysis Test)	92	3.38
HTP (House-Tree-Person Test)	88	3.23
EPQ (Eysenck Personality Questionnaire)	84	3.08
BENDER (Bender Visual Motor Gestalt Test)	80	2.94
ISRA (<i>Inventario de Situaciones y Respuesta de Ansiedad</i> ; Inventory of Situations and Anxiety Response)	72	2.64
PROLEC (<i>Batería de Evaluación de los Procesos Lectores</i> ; Battery for the Assessment of Reading Processes)	68	2.50
MACI (Millon Adolescent Clinical Inventory)	59	2.17
BASC (Behavior Assessment System for Children)	57	2.09
CUIDA (<i>Eval. de Adoptantes, Cuidadores, Tutores y Mediadores</i> ; Assessment of Adoptive Parents, Foster Carers, Guardians and Mediators)	51	1.87
ITPA (Illinois Test of Psycholinguistic Abilities)	51	1.87
CAQ (Clinical Analysis Questionnaire)	48	1.80

*Included under the acronyms WISC and WAIS are the different versions of both tests which are available, such as the WISC-R or the WISC-IV.

TESTS MOST WIDELY USED IN SPAIN

Survey participants were asked to indicate the three tests they used most in their everyday practice. Table 5 shows the results obtained. As it can be seen, in first place is the WISC intelligence scale for children, followed by the personality test 16PF. All the commonly used instruments are classic psychometric tests well established in psychology, the Rorschach projective test appearing in eighth place. Among the 25 most widely used tests are six developed by Spanish authors (24%), indicating the substantial progression of work in this area in our country. Table 6 shows the ten most often used tests by specialization; as would be expected, the differences are notable, reflecting the different tasks corresponding to each field. It is worth highlighting the widespread use of the Beck Depression Inventory (BDI), which appears in fifth place, not least because this is a test not marketed in Spain, so that, clearly, photocopies of the instrument are being used, as well as criteria taken from studies and publications referring to it. It would be highly advisable for this test so widely used by professionals to be subject to a more systematic and rigorous validation process in our country, which we assume has yet to be applied for reasons related to commercial and intellectual property aspects.

Table 7 shows the means of the items common to the 2000 and 2010 versions. As it can be seen, they are quite similar, with few differences across the decade: the correlation between the two is 0.986. We might perhaps point out a slight development in the desired direction in item 4, insofar as respondents' opinion on how much information psychologists have about test quality has improved over this period, from a mean of 2.38 to 2.73. It is scant consolation – the mean is still low – but at least there is movement in the right direction. Even so, at this rate it would take some 50 years to arrive at a reasonable situation, so clearly more has to be done, and more quickly.

FUTURE PERSPECTIVES

It is clear that tests, which emerged in psychology over a century ago, are here to stay, and that a great deal of water has flowed under the bridge since those first sensory-motor tests developed by Galton at the end of the nineteenth century, or since Binet and Simon (1905) proposed the first individual scale of intelligence. Nobody since then has been capable of predicting which way tests would go, and we do not pretend to do so here; what follows are some reflections on the current situation of tests, and on some aspects which are likely to influence their future. It may sound something of a cliché, but the great force that is reshaping psychological assessment today are the new information technologies, especially those related to computers, multimedia and Internet. In the view of authors such as Bennett (1999, 2006), Breithaupt, Mills and Medican (2006) or Drasgow, Luecht and Bennett (2006), the new technologies are influencing all aspects of psychological assessment, such as test design, item construction and presentation, test scoring and remote assessment. All of this is bringing about changes in the format and content of assessment, leading to the quite reasonable doubt over whether paper-and-pencil tests as we know them today will be able to resist the onslaught of the technological progress we are witnessing. In this sense, what has been said recently in discussions about the future of books and newspapers could well be applied to the case of tests. New forms of assessment are emerging, such as authentic assessment in the educational field (portfolios, written compositions, projects), though psychometric tests

will continue to be fundamental tools, given their objectivity and their economy of resources and time (Phelps, 2005, 2008). In the opinion of a specialist such

TABLE 6
THE 10 TESTS MOST COMMONLY USED BY SPANISH PSYCHOLOGISTS ACCORDING TO SPECIALIZATION

	Clinical	Educational	Work and Organizational
1	MCMI	WISC	16PF
2	16PF	BADYG	PAPI
3	MMPI	TALE	DAT
4	BDI	MSCA	TPT
5	WISC	16PF	IPV
6	WAIS	RAVEN	MMPI
7	STAI	PROLEC	IGF
8	RORSCHACH	BENDER	BFQ
9	SCL-90	ITPA	MCMI
10	MMSE	TAMAI	NEO PI

Note. The acronyms for the tests not appearing in Table 5 are as follows: PAPI (The Personality and Preference Inventory), DAT (Differential Aptitude Test), TPT (*Test de Personalidad de TEA*; TEA Personality Test), IPV (*Inventario de Personalidad para Vendedores*; Personality Inventory for Sales Personnel), IGF (*Inteligencia General de TEA*; TEA General Intelligence), BFQ (Big Five Questionnaire), NEO PI (NEO Personality Inventory).

TABLE 7
MEANS OF THE ITEMS OBTAINED IN 2000 AND IN 2010

Items	Results 2000 (Mean)	Results 2010 (Mean)
1	2,41	2,43
4	2,38	2,73
6	2,57	2,59
8	4,23	4,12
9	4,34	4,39
11	4,33	4,10
12	4,29	4,01
14	2,42	2,10
16	1,85	1,93
18	1,57	1,77
19	4,15	4,13
21	3,56	3,76
22	4,41	4,46
23	4,37	4,41
25-1	3,60	3,51
25-2	2,63	2,64
25-5	3,07	3,10
25-6	2,91	2,92
25-7	3,28	3,21
25-8	2,99	2,97

as Professor Ronald K. Hambleton (Hambleton, 2004, 2006) from the University of Massachusetts, six broad areas will attract the attention of researchers and professionals in the coming years. The *first* is the use of tests on an international scale, given increasing globalization and ease of communication, which raises the spectre of a whole series of problems related to the adaptation of instruments from certain countries to others (Byrne et al., 2009; Hambleton et al., 2005). The *second* refers to the use of new psychometric models and technologies for generating and analyzing tests. We might mention here the whole new psychometrics derived from Item Response Theory (IRT) models, which succeed in solving some problems that were intractable within the classical framework, though as often occurs, while some problems are solved there arise others which were unforeseen. The *third* area relates to the appearance of new item formats deriving from advances in IT and multimedia. From modest matrices in black and white we have moved to today's interactive screens, with animation and sound, capable of reacting to the responses of examinees (Irvine & Kyllonen, 2002; Shermis & Burstein, 2003; Sireci & Zenisky, 2006; Zenisky & Sireci, 2002). Of course, though, it is not a case of innovating for the sake of innovation: before substituting the old formats by the new ones it must be empirically demonstrated that they are an improvement, and psychometric properties such as reliability and validity are certainly non-negotiable. The *fourth* area that will attract attention concerns everything related to computerized tests and the way they are related to Internet. Special mention in this connection should be reserved for Computerized Adaptive Tests (CAT), which permit adjustment of the test to the characteristics of the person assessed, without a loss of objectivity or comparability between examinees, thereby opening up highly promising perspectives in assessment (Olea, Ponsoda, & Prieto, 1999). Remote assessment is another technique that is fast becoming widely employed, and this raises serious security issues (relating to both the data itself and the person tested) – after all, it has to be guaranteed that the person being assessed is actually the same person who he or she claims to be, especially in contexts of personnel selection or of tests with important repercussions for the future life of the person under

assessment. Great progress is being made in this area in both the basic and applied contexts (Bartram & Hambleton, 2006; Leeson, 2006; Mills et al., 2002; Parshall et al., 2002). In *fifth* place is an area that may appear peripheral, but which is taking on considerable importance: that of the systems used for giving the results to users and stakeholders. It is essential that they both understand unequivocally the results of assessments, and it is not obvious which is the best way of achieving this, especially if results have to be sent to a professional for their interpretation and explanation, as frequently occurs in situations of personnel selection or educational assessment (Goodman & Hambleton, 2004). This aspect obviously has less influence in clinical contexts. *Finally*, it is highly probable that in the future there will be great demand for training from different professionals – in addition to psychologists – working in assessment, such as teachers, doctors or nurses. It is not a question of such professionals using and interpreting strictly psychological tests, but rather of their demanding information so as to be able to understand and participate in the processes of assessment and certification that take place in their work contexts. These are some of the aspects around which assessment activity will quite likely revolve in the not-too-distant future, though we should stress that this is by no means an exhaustive list, but is rather intended to help readers orient themselves in the changing world of psychological assessment.

ACKNOWLEDGEMENTS

The authors would like to express their sincere thanks to the members of the COP Tests Commission for their help and cooperation with this article: Rocío Fernández Ballesteros, Miguel Martínez, Eduardo Montes, Jaime Pereña and Javier Rubio. Thanks are also due to the COP (Spanish Psychological Association) in general, without whose help this work would have been impossible. Furthermore, we are grateful to Ángela Campillo, Eduardo Fonseca and Elsa Peña for all their help with the data processing. Finally, we should not forget to offer our sincere gratitude to all those professionals who responded to the survey; without them we could have done nothing. Funding for the article was provided by the COP and the Spanish Ministry of Science and Innovation (*Ministerio de Ciencia e Innovación*) (Ref. nº PSI2008-03934).

APPENDIX

QUESTIONNAIRE USED FOR OBTAINING THE OPINIONS OF PSYCHOLOGISTS ON TEST PRACTICE

+++++

GENERAL DATA

Age: Sex: Man Woman
 Year in which you obtained your Psychology degree:
 Years as a member of the Spanish Psychological Association:
 Professional specialization: Clinical-Health Educational Work and Organizational Other (indicate)
 Do you currently work as a Psychologist? Yes No
 Which sector do you work in? Public Private Unemployed
 How many years have you been in your current job?

INSTRUCTIONS

The items below are designed for you to respond on a scale of 1 to 5. If you totally disagree with the statement, mark 1. If you totally agree with it, mark 5. Use the numbers 2, 3 and 4 to indicate your degree of agreement in between. Your completed questionnaire will remain totally anonymous.

QUESTIONNAIRE

- 1.- The training provided in the Psychology degree is sufficient for the correct use of most tests
- 2.- The training provided on courses and Masters programmes is sufficient for the correct use of most tests
- 3.- The European Federation of Psychologists' Associations (EFPA) should set up a system for accrediting test-user competency
- 4.- Professionals have sufficient information (independent reviews, research, documentation, etc.) on the quality of tests published in Spain
- 5.- In my professional field computerized tests are gradually replacing paper-and-pencil tests
- 6.- My current knowledge in relation to tests is basically what I learned on my Psychology degree course
- 7.- The application of tests via the Internet has many advantages compared to application using the classical paper-and-pencil format
- 8.- The use of psychological tests should be restricted to qualified psychologists
- 9.- Although non-psychologists might apply and score tests, interpreting the results and providing information about them should be left to psychologists
- 10.- Reports generated automatically by computer have no validity
- 11.- Standards and guidelines that define 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 American Psychological Association (APA)].
- 12.- Legislation is necessary to control the most serious abuses related to tests
- 13.- The application of tests via Internet puts some examinees at a disadvantage
- 14.- Anyone capable of demonstrating their competence in the use of tests (be they a psychologist or not) should be authorized to use them
- 15.- If used appropriately, the Internet can greatly improve the application of tests
- 16.- Controls on tests should be minimal, since they inhibit the development of new ideas and new assessment procedures
- 17.- The application of tests via Internet does not permit the protection of user privacy
- 18.- Publishers should be allowed to sell any test as they see fit
- 19.- The Spanish Psychological Association (COP) should play a more active role in regulating and improving the way tests are used
- 20.- The application of tests via Internet makes them vulnerable to fraud
- 21.- In the course of my profession I use tests regularly
- 22.- Tests constitute an excellent source of information if combined with other psychological data
- 23.- Used correctly, tests are of great help to the psychologist
- 24.- Taking into account all aspects, I think that test use in my country has improved over the last decade
- 25.- Please estimate the frequency in your professional context of the following problems related to test use (1: highly infrequent; 5: very common)
 - (1) Making photocopies of material protected by copyright
 - (2) Making assessments using inappropriate tests
 - (3) Not being up to date
 - (4) Not discussing one's interpretations with others
 - (5) Not taking into account the measurement errors of scores
 - (6) Not restricting the application of tests to qualified personnel
 - (7) Not taking into account local conditions (country, region) that may affect validity
 - (8) Making interpretations that go beyond the scope of the test
26. Please indicate the three tests you most commonly use in the exercise of your profession:
 1.
 2.
 3.

Observations: Please mention any other aspect you consider relevant (you can enclose more pages if necessary)



REFERENCES

- American Educational Research Association, American Psychological Association, and National Council on Measurement in Education. (1999). *Standards for educational and psychological testing*. Washington, DC: American Psychological Association.
- Bartram, D. (1996). Test qualifications and test use in the UK: The competence approach. *European Journal of Psychological Assessment, 12*, 62-71.
- Bartram, D. (1998). The need for international guidelines on standards for test use: A review of European and international initiatives. *European Psychologist, 2*, 155-163.
- 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.
- Bartram, D. & Hambleton, R. K. (Eds.) (2006). *Computer-based testing and the Internet*. Chichester: John Wiley and Sons.
- Bartram, D. & Roe, R. A. (2005). Definition and assessment of competences in the context of the European diploma in psychology. *European Psychologist, 10*, 93-102.
- 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).
- Binet, A. & Simon, T. H. (1905). Méthodes nouvelles pour le diagnostic du niveau intellectuel des anormaux [New methods for diagnosing the intellectual level of the abnormal]. *L'année Psychologique, 11*, 191-244.
- 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.
- Brennan, R. L. (Ed.) (2006). *Educational measurement*. Westport, CT: ACE/Praeger.
- Byrne, B. M., Leong, F. T., Hambleton, R. K., Oakland, T., van de Vijver, F. J., & Cheung, F. M. (2009). A critical analysis of cross-cultural research and testing practices: Implications for improved education and training in psychology. *Training and Education in Professional Psychology, 3*(2), 94-105.
- 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., Luecht, R. M., & Bennett, R. E. (2006). Technology and testing. In R. L. Brennan (Ed.), *Educational measurement*. Westport, CT: ACE/Praeger. (pp. 471-515).
- European Federation of Professional Psychologists' Associations (2005). *Meta-Code of ethics*. Brussels: Author (www.efpa.eu).
- Evers, A. (2001a). Improving test quality in the Netherlands: Results of 18 years of tests ratings. *International Journal of Testing, 1*, 137-153.
- Evers, A. (2001b). The revised Dutch rating system for test quality. *International Journal of Testing, 1*, 155-182.
- 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*(1), 18-33.
- 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.
- Hambleton, R. K. (2004). Theory, methods, and practices in testing for the 21st century. *Psicothema, 16*, 696-701.
- Hambleton, R. K. (2006). *Testing practices in the 21st century*. Keynote Address, University of Oviedo, Spain, March 8th.
- Hambleton, R. K., Merenda, P. F., & Spielberger, C. D. (Eds.) (2005). *Adapting educational and psychological tests for cross-cultural assessment*. London: LEA.



- Irvine, S. & Kyllonen, P. (Eds.) (2002). *Item generation for test development*. Mahwah, NJ: Lawrence Erlbaum.
- Joint Committee on Testing Practices. (2002). *Ethical principles of psychologists and code of conduct*. Washington DC: Joint Committee on Testing Practices.
- Koocher, G. & Kith-Spiegel, P. (2007). *Ethics in psychology*. New York: Oxford University Press.
- 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.
- Leeson, H. V. (2006). The mode effect: A literature review of human and technological issues in computerized testing. *International Journal of Testing*, 6, 1-24.
- Lindsay, G., Koene, C., Ovreeide, H., & Lang, F. (2008). *Ethics for European psychologists*. Göttingen and Cambridge, MA: Hogrefe.
- Lunt, I. (2005). The implications of the "Bologna process" for the development of a European qualification in psychology. *European Psychologist*, 10, 86-92.
- Mills, C.N., Potenza, M.T., Fremer, J.J., & Ward, W.C. (Eds.) (2002). *Computer-based testing: Building the foundation for future assessments*. Hillsdale, NJ: LEA.
- Moreland, K. L., Eyde, L. D., Robertson, G. J., Primoff, E. S., & Most, R. B. (1995). Assessment of test user qualifications. *American Psychologist*, 5, 1, 14-23.
- Muñiz, J. (1997). Aspectos éticos y deontológicos de la evaluación psicológica [Ethical and deontological aspects of psychological assessment]. In A. Cordero (Ed.), *Evaluación psicológica en el año 2000* [Psychological assessment in the year 2000] (pp. 307-345). Madrid: TEA Ediciones.
- 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. & Fernández-Hermida, J.R. (2000). La utilización de los tests en España [Test use in Spain]. *Papeles del Psicólogo*, 76, 41-49.
- 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.
- Olea, J., Ponsoda, V., & Prieto, G. (1999). *Tests informatizados: fundamentos y aplicaciones* [Computerized tests: Foundations and applications]. Madrid: Pirámide.
- Papeles del Psicólogo (2009). *Número monográfico sobre Ética Profesional y Deontología* [Special Issue on Professional Ethics and Deontology]. Vol. 30, 182-254.
- Parshall, C.G., Spray, J.A., Kalohn, J.C., & Davey, T. (2002). *Practical considerations in computer-based testing*. New York: Springer.
- Peiró, J.M. (2003) La enseñanza de la Psicología en Europa. Un proyecto de Titulación Europea [The teaching of Psychology in Europe. A European Diploma project]. *Papeles del Psicólogo*, 86, 25-33
- Phelps, R. (Ed.) (2005). *Defending standardized testing*. London: LEA.
- Phelps, R. (Ed.) (2008). *Correcting fallacies about educational and psychological testing*. Washington: APA.
- Prieto, G. & Muñiz, J. (2000). Un modelo para evaluar la calidad de los tests utilizados en España [A model for assessing the quality of the tests used in Spain]. *Papeles del Psicólogo*, 77, 65-71.
- Shermis, M. D. & Burstein, J. C. (Eds.) (2003). *Automated essay scoring*. London: LEA.
- Simner, M. L. (1996). Recommendations by the Canadian Psychological Association for improving the North American safeguards that help protect the public against test misuse. *European Journal of Psychological Assessment*, 12, 72-82.
- Sireci, S.G. & Zenisky, A.L. (2006). Innovative item formats in computer-based testing: In pursuit of construct representation. In S. M. Downing, & T. M. Haladyna (Eds.), *Handbook of test development*. Hillsdale, NJ.: LEA.
- Wilson, M. (2005). *Constructing measures: An item response modeling approach*. Mahwah, NJ: Lawrence Erlbaum Associates.
- Zenisky, A.L. & Sireci, S.G. (2002). Technological innovations in large-scale assessment. *Applied Measurement in Education*, 15, 337-362.

