

Testing Practices in European Countries¹

José Muñiz¹, Dave Bartram², Arne Evers³, Dusica Boben⁴, Kruno Matesic⁵, Kathia Glabeke⁶, José R. Fernández-Hermida⁷, Jac N. Zaal³

EFPA Standing Committee on Tests and Testing¹, British Psychological Society², Dutch Association of Psychologists³, Slovenian Psychological Association⁴, Croatian Psychological Association⁵, Belgian Federation of Psychologists⁶, Spanish Psychological Association⁷

Address correspondence to:

José Muñiz
Facultad de Psicología
Universidad de Oviedo
Plaza Feijoo, s/n
33003 Oviedo (Spain)
E-mail: jmuniz@correo.uniovi.es

¹ Published in the *European Journal of Psychological Assessment*, 2001, 17(3), 201-211.

Testing Practices in European Countries

Abstract

Tests constitute one of the technologies most frequently used by psychologists in their professional practice. Like any other scientific technology, tests can be used in appropriate or inappropriate ways. Currently, various institutions, both national and international, are developing projects and activities aimed at improving testing practices. In Europe, the Task Force on Tests and Testing created by the European Federation of Psychologists' Associations (EFPA) designed a questionnaire to sample the opinions of professional psychologists about tests and testing in order to help us in making the right decisions for improving testing practices in Europe. In this paper, the results obtained using the EFPA Questionnaire are presented. A sample of 3,455 Spanish, 2,407 English, 2,079 Dutch, 321 Slovenian, 218 Croatian, and 210 Belgian psychologists responded to the questionnaire. In general, European psychologists show a positive attitude towards tests and testing, at the same time expressing the need for institutions to adopt a more active role in promoting good testing practices. Results show that the tests most frequently used are Intelligence tests, Personality questionnaires and Depression scales. Finally, the results are analyzed in detail, and some possible measures for improving testing practices in Europe are discussed.

Key words: Test use, Testing practices, Psychological assessment

Testing Practices in European Countries

Tests, like any other scientific technology, may be used correctly, in which case they constitute an invaluable resource for psychologists, or incorrectly, in which case they may seriously prejudice those involved in the assessment process. Used properly, tests offer a possibility for judging people in an egalitarian manner, on merit, rather than using criteria such as background, clan, family, appearance, letters of recommendation or the subjective judgments of supervisors and teachers. The quest for fair and objective methods of assessment represented the pioneering vision of those involved in the development of psychological testing. This spirit still prevails, despite the fact that some people, often without proper qualifications, misuse tests. The problem of misuse is not new and is far from uncommon in the history of our psychological science (Colom, 2000; Cronbach, 1975). With a view to the future, the inappropriate or incorrect use of tests and scales needs to be minimized as far as possible. We need to guard against the danger of adopting the intellectual posture of renouncing test use simply because in some situations, some users fail to use them as they should.

Test users and professionals frequently ask themselves certain questions, such as: Do the available tests have the necessary technical qualities? Do professionals use tests properly? What are the most commonly used tests? What can be done to improve tests and their use? This article looks at these questions and the opinions of European psychologists about them. However, before attempting to provide answers, let us first briefly consider what is being done in Europe in order to avoid the inappropriate or incorrect use of tests. If, as appears reasonable, we admit that test use in Europe can clearly be improved, the first question we must ask ourselves is: Why are tests used incorrectly? Anastasi (1987) asks just this, and suggests three possible reasons: a) deliberate intention to distort the truth, b) negligence, and c) insufficient knowledge of test use. She concludes that the third of these, insufficient knowledge of test use, is the most frequent cause of incorrect use. If this is the case, and there is fairly general agreement that it is, we need to consider what can be done to remedy the situation. Potential sources of error in the use of tests are diverse. Eyde and colleagues (1993), for example, identify some of the most common, such as: ignorance of the need to use various sources of convergent data, making important decisions based solely on test scores, failing to restrict test application to qualified personnel, photocopying copyrighted material, making interpretations

beyond the limits of the test, etc.

In the different European countries the organizations and institutions involved have adopted diverse measures aimed at improving the use of tests. It should not be forgotten that there is a wide variety of groups interested in tests being used correctly, such as test publishers, test authors, professionals, psychologists' associations, national psychological associations, other relevant professional bodies in the clinical, education or work area, those assessed by tests, consumers, and so on.

Attempts to improve test use revolve around two general complementary strategies: Restriction of access and training or providing information.

Restriction of access

The strategy based on restriction groups all those actions that attempt to restrict the use of tests to professionals with proven qualifications for doing so. The most widespread system is based on the old APA classification of tests in terms of three levels, A, B and C. This classification was developed to reflect difference in the degree of specialization and skill required to use the tests. Thus, group A includes tests of performance and knowledge, group B covers collective tests of aptitude and intelligence, whilst in group C we find individually-applied tests of intelligence and personality, as well as other complex tests, such as some projective tests. However, even supposing that such restriction of use were to be strictly observed, which is much to suppose, this scheme does not in itself guarantee the proper use of tests. As demonstrated empirically by Moreland et al. (1995) and Simner (1996), the approach of associating academic qualifications with these levels does not alone ensure correct use of tests. It is obvious that the mere fact of having a first degree in Psychology does not guarantee that a person is fully qualified to use correctly many of the tests in groups B and C. In view of this, in countries such as England there exist qualifications for specific categories of tests, and the British Psychological Society maintain a list of accredited users (Bartram, 1996). Restrictions imposed vary considerably from one country to another (Bartram & Coyne, 1998a, 1998b; Muñiz et al., 1999), but what appears evident is that restrictions alone are no guarantee of good practice – some type of training requirement is also necessary. Of course, a sensible professional would never use a test that s/he is not properly qualified to use, so that, in this case

as in others, the professional's ability to evaluate their own competence is a crucial factor.

Training and information

Training and information initiatives aimed at improving the use of tests are also quite varied, and are based on the plausible assumption that if a professional is sufficiently acquainted with test practice, the probability of his/her using them incorrectly will be markedly reduced. Ensuring test users are competent, through the provision of sufficient training and good information about the tests they use, appear to be the most effective strategy in the long term (Fremer, 1996; Martínez et al., 2000a, 2000b; Tyler, 1986). Some initiatives are focused on the tests themselves, providing information on their technical quality, as is the case of reviews by experts. In Holland, the National Psychological Association periodically reviews the tests on the market using a standard review model (Evers, 1996; Evers et al., 2000), and publishes the reviews in a book that is available to professionals. A similar system is applied in Britain (Bartram, 1996), though the reviews are directed specifically at test users who have met the British Psychological Society's standards of competence in test use. In Spain, the Test Commission of the Spanish Psychological Association (COP) has designed a test review model based on the Dutch and British models (Prieto & Muñiz, 2000), with the aim of reviewing, in the near future, the tests most commonly used in Spain. The European Federation of Professional Psychologists' Associations (EFPA) has established a Test Commission that is preparing, among other activities, a European model for the review of tests that will be available for adoption by member countries. In sum, the idea of this approach is to make available to the professional as much information as possible on the test, together with the reasoned judgments of experts.

In addition to this information on the test itself, various organizations provide documents in order to guide professionals in the use of tests. The most general of these are ethical and deontological codes, which always include some indications in relation to test use – see, for example, the European code (European Federation of Professional Psychologists' Associations, 1996), or the North American one (American Psychological Association, 1992). Another code of great interest in this regard is that which is being prepared by the European Association of Psychological Assessment (Fernández-Ballesteros, 1997), and which attempts to

Testing Practices in European Countries

cover not only the use of tests, but also other aspects involved in the assessment process in general. Of course, the most usual situation is that organizations draw up specific standards or guidelines aimed at providing orientation with regard to all aspects involved in the practice of tests, from their construction to their application. There are numerous sets of such guidelines (Muñiz, 1997), but those most commonly used throughout the world are the technical standards prepared by the American Psychological Association (American Educational Research Association et al., 1999). These standards are more oriented towards technical aspects, while the guidelines of the International Test Commission are geared more specifically to test use. Another important aspect is the translation/adaptation of tests from one culture to another – an aspect particularly relevant in Europe, where so many languages share the same continent. In this respect the guidelines published by the ITC on procedures for the proper adaptation of tests from one culture to another are especially pertinent (Hambleton, 1993, 1994; Hambleton & Bollwark, 1991). Here, we have cited just a few of the most well-known guidelines; a review of the different initiatives on guidelines for the use of tests can be found in Bartram (1998).

Objectives of the study

We have attempted to give an overview of the problem of test use, in order to provide a reference framework that will facilitate understanding of the different types of initiative undertaken to improve test practice. The conclusion that can be drawn is that restricting the use of tests to qualified professionals, in combination with the provision of information and training for users, appears to be the most effective formula for combating the incorrect use of tests. But, what do European psychologists think of all this? Are they in favor of using tests? Do they use them? Which ones? Given that we lack reliable answers to these and similar questions, the central aim of this work will be to try and respond to them in an empirical way; otherwise, it would be difficult to take any corrective measures that may be necessary. With the above aim, a questionnaire was administered to a sample of psychologists from the following countries: Belgium, Croatia, Holland, Slovenia, Spain and the United Kingdom.

Testing Practices in European Countries

Method

Questionnaire

The questionnaire used was designed by the EFPA Task Force on Tests and Testing, and designed to obtain the opinions of professional psychologists from the above-mentioned European countries on tests and their use. As can be seen in the Appendix, the questionnaire contains 25 Likert-type items and an open question on the most frequently used tests. The final seven items of the questionnaire refer to problems of test use, and are taken from the 86 elements identified by Eyde et al. (1988, 1993) in relation to the responsible use of tests. It also includes a section in which respondents can make any observations they consider pertinent. In the Likert-type items the scale runs from 1 (Total disagreement) to 5 (Total agreement).

Respondents

The questionnaire was sent by mail, together with a stamped addressed envelope, to all members of the national psychological associations of the countries included in the study. In the case of Belgium the envelope was not stamped addressed. The response rate can be considered satisfactory: Belgium 14.2%, Croatia 17.7%, Holland 20.8%, Slovenia 40%, Spain 12.4% and the United Kingdom 29.8%. Distribution of responses according to sex, age and professional speciality is shown in Tables 1 and 2.

As it can be observed in Tables 1 and 2, the predominance of women among professional psychologists is clear, being most pronounced in Croatia (82.6% versus 17.4%) and least in Belgium (54.3% versus 45.7%) and the United Kingdom (55% versus 45%). As far as mean age is concerned, it is highest in British psychologists, though this is probably due to the fact that the questionnaire was applied only to Chartered Psychologists, and not to all members of the British Psychological Society.

With regard to professional area, Clinical Psychology is clearly predominant in Belgium, Holland, Spain and the United Kingdom, with Educational Psychology being more common only in Slovenia.

Testing Practices in European Countries

Results

Opinion of professionals on the use of tests

Table 3 shows the means and standard deviations obtained for each of the questions in the different countries, the scale going from 1 to 5.

The first point to note is the opinion of professionals that the education received in obtaining a Psychology degree is not sufficient for the correct use of the majority of tests (Item 1). The immediate consequence of this is the need for postgraduate training. It is also clear (Item 2) that professionals require more technical information on tests. The call for the establishment of obligatory minimum technical requirements (Item 6) is evident, as is the demand for legislation to control potential abuse of tests (Item 7). There is fairly general agreement on the need for the respective national psychological associations to take a more active role in the regulation of test use (Item 15).

The clearly positive attitude of professionals towards the use of tests is reflected in Items 17 and 18: They consider them as an excellent source of information if employed in conjunction with other data and used correctly. While there are more similarities than differences among countries, it is worth mentioning the greater permissiveness among Dutch and British professionals with regard to the use of tests by professionals with no qualifications in psychology, as long as they are suitably qualified in general (Items 4, 5, 11). The answers given by the Spanish respondents to item 14 differ from those obtained in the others countries. Taken into the account the Spanish context, and the regulations to buy tests in Spain (A, B, C rules), this result seems to indicate that psychologists want to keep the right to buy tests, which does not mean they are in favour of using tests for which they are not properly trained.

Among the problems of test use most commonly found by respondents in their context, it should first be underlined that, in general, no serious problems were encountered (Item 10), even though problems do exist, as shown by mean scores of between 2.55 (Holland) and 3.37 (Spain). The commonest problems were identified as those of making photocopies (Item 19.1), failing to take into account local conditions (Item 19.6), and not keeping up with the field and failing to check one's own interpretations with others (Item 19.3). By countries, the highest means for test-use problems in general were found for Spain (3.10), followed by Croatia (2.93),

Testing Practices in European Countries

Belgium (2.54), United Kingdom (2.50), Slovenia (2.28) and Holland (2.04).

Tests most frequently used

Table 4 shows the tests most frequently used by professionals, according to their responses to Item 20 of the questionnaire. The percentages were calculated by dividing the number of times a test was reported by the total number of respondents.

As it can be observed, the commonest are the classic psychometric tests of intelligence and personality: WISC, WAIS, MMPI, 16PF, RAVEN, EPQ, BENDER, etc. In general, there is a clear predominance of psychometric tests over projective ones, with the exception of Belgium, in which three projective tests (TAT, RORSCHACH and CAT) appear among those more frequently used. These projective tests are used very frequently by the Walloon (French speaking community), and not so much by the Flemish. The percentages of use for both groups (Flemish/Wallon) are as follow: WISC (28/47%), WAIS (21/26%), WPPSI (13/23%), MMPI (12/4%), TAT (9/25%), CBCL (9/0%), RORSCHACH (7/25%), CAT (6/16%), RAVEN (3/6%), Rey (2/10%).

Only two tests are in the top ten of most used tests in all countries: The WAIS and the WISC. The MMPI is in the top ten four times, the Bender, Raven and RORSCHACH three times, and the 16PF, EPQ and BDI two times. Though all other tests are specific for the top ten in each of the countries, there is a clear predominance of psychometric tests over projectives ones. This is in sharp contrast to the results obtained by Muñiz et al. (1999) in Latin-American countries, where projective tests are in the majority. In general, there seem to be no marked differences in types of most frequently used tests between the countries, though it can be observed that the orientation on educational tests (tests for testing school performance and vocational guidance) is stronger in Slovenia than in the other countries.

On the basis of these results, it can be said that in the daily exercise of their profession, European psychologists make little use of instruments resulting from recent advances in psychometric techniques, such as those deriving from Item Response Theory models. The

Testing Practices in European Countries

reasons for this rather conservative attitude in the use of tests may be due either to the fact that professionals are satisfied with the functioning of the classic tests, or to ignorance of the new alternatives, which are in any case scarce. An interesting analysis of the reasons for the scarcity of new tests in the field of intelligence can be found in Sternberg and Kaufman (1996).

Factorial structure of the questionnaire

The construction of the questionnaire was guided by the need to include all aspects of test use that may be of relevance. Thus, in no way was it expected that the instrument would be unidimensional, as it did not seem reasonable to assume unidimensionality of the broad field of test practice. Exploratory factor analyses were carried out with the data obtained in each of the countries, though those of Belgium, Croatia and Slovenia are not included here due to the small number of subjects in their samples. Loadings ≤ 0.40 have been blanked, except in the cases in which the highest loading of the variable is ≤ 0.40 . Six factors with eigenvalues greater than 1 were found in the case of England, six in Spain, and seven in Holland. In order to make the factorial structures more meaningful and comparable, five factors were rotated in all three cases.

In the case of Spain (Table 5), the first factor grouped all the items related to problems of test use in the context of the professional psychologist. The second factor contains the items indicating psychologists' attitude towards tests. The third factor relates to the use of tests and the need for its regulation. The fourth factor is centered on training, information and knowledge about tests. The fifth factor groups the items relating to permissiveness in test use. The five factors explain 45.92% of the total variance. The factorial structure corresponds reasonably well to the dimensions considered in the construction of the items. Overall, an alpha coefficient of 0.63 was obtained. For a more detailed exposition of the Spanish results, see Muñiz and Fernández-Hermida (2000).

The British data also presented a first factor made up of all the items related to the use of tests, with the remainder of the items being grouped in four factors. The factorial structure is

Testing Practices in European Countries

slightly different from that obtained for Spain, though the same basic dimensions are reproduced. In addition to Factor 1, which contained all the items related to problems of use, the other four can be defined as: 2) Only psychologists should use tests, 3) Tests and test use have high utility, 4) Standards and competence are important, and 5) Information and training in tests and testing. The overall alpha coefficient was 0.59, slightly lower than that obtained for Spain.

The factorial structure of the Dutch data is similar to that of the data already mentioned, with the items related to problems of use again forming a first factor with strong cohesion, termed as “incorrect use”. The second factor is related to rules and regulations, the third may be labelled as appreciation of tests, the fourth relates to qualifications for test use, and the fifth meaning the knowledge about tests. The overall alpha coefficient was 0.64, practically the same as the one obtained for the Spanish data (0.63).

The very small variations observed in the factorial structures of the three countries supports the differences already observed on analyzing the patterns of the means of the items. The three factorial structures obtained are very similar. In all the three cases the first factor groups the items related to the use of tests. The second factor in the English structure basically corresponds to the second Dutch and the third Spanish. Factor number 3 in the English and Dutch structures, corresponds to factor number two in the Spanish structure, made up of items 16, 17 and 18. The Spanish factor 4, exactly corresponds to the Dutch factor 5 (items 1, 2, 3), and it is very close to the English factor 5, except for item 3. Finally, Dutch and English factor 4 are very close, with items 4, 5 and 11 in common; while this factor appears divided into two factors (3 and 5) in the case of the Spanish structure. In conclusion, it could be said that with some nuances the three factorial structures are essentially equivalent, and represent the main dimensions involved in psychologist’s attitudes relating to the use of tests.

Discussion and conclusions

The first observation from the data obtained that should be underlined is the positive attitude shown by psychologists towards the proper use of tests. These results coincide, in

Testing Practices in European Countries

general terms, with those obtained in a previous work by Poortinga et al. (1982). The data are also consistent with those obtained by Bartram and Coyne (1998a, 1998b) and by Muñiz et al. (1999). All of this appears to indicate that psychologists have no hesitation in using tests in the exercise of their profession, but that they do so judiciously, considering them as a helpful tool, and not as a kind of magic wand to solve problems or make decisions. This positive attitude, however, does not prevent psychologists from stating clearly that the day-to-day use of tests in Europe is far from perfect, and that there are several aspects that should be improved as soon as possible. They would like to see, for example, greater involvement of the professional organizations in the regulation of test use, as well as more information on technical aspects of tests, such as that provided by independent reviews, research, etc. There is also a clear demand for ongoing training, as that provided up to first-degree level is clearly insufficient for the correct use of all published tests.

Some of the professional and scientific organizations, at both national and international levels, have taken good note of this demand for a more active role, and are implementing some measures in this direction, as already mentioned in the introduction to this work. One of the measures currently being debated by the EFPA Standing Committee on Tests and Testing is that of introducing a “quality control” seal that would indicate the technical adequacy of tests.

As regards the other major demand – for specialized postgraduate training that would complement the inadequacies of that received up to first-degree level – attempts to satisfy it vary considerably from one country to another. There are differences in terms of who provides such training, the commonest sources being universities, professional associations, private companies, test publishers, and government institutions. In reality, there is no need for exclusivity as far as the provision of training is concerned; what is necessary is that it exists and that it is imparted by expert professionals. This not the place to discuss what types of program would be most appropriate, but it does appear that there would have to be a flexible offer in which at least three parameters were taken into account: the specialization required by the test, the type of professional for whom it is suitable and the specialist field in which it is used. In any case, account would have to be taken of the seven major factors (Eyde et al., 1988, 1993; Moreland et al., 1995) involved in the correct use of tests, namely, integral assessment, proper

Testing Practices in European Countries

use, psychometric knowledge, confidentiality of results, precision in scores, respect for rules, and feedback and interpretation of results for those assessed.

We have tried to offer an overview of the current situation with regard to the issue of test use and psychologists' opinions on it, and to what has been done to try and improve test use, what is being done and what remains to be done. However, interesting new challenges have arisen in the wake of recent developments in the study and measurement of psychological variables, driven by two important factors: a) the extensive development of the new psychometric models of Item Response Theory (IRT), which have permitted the proper treatment of problems for which no optimum solution existed within the classical theoretical framework, and b) the influence of the new technologies, especially the computer, on assessment. A third factor that can also be mentioned is the recent development of the Latent State-Trait Theory, that allows to incorporate situation and interaction effects in the model, and to control for these effects (Deinzer et al. 1995; Steyer et al. 1992, 1999, 2001).

Two of the most important developments that must be assimilated, because they are part of the here-and-now, are psychological assessment through the Internet and the increasingly generalized use of Computerized Adaptive Tests. These tests mark the end of the era in which the same test is applied to everyone, as adjusting and adapting the test for the person assessed brings notable improvements in all aspects: greater precision, economy of time and motivation of those assessed. We are not talking here about the future – this type of tests is already routinely used. For a good introduction to the issue see Wainer (1990), Sands et al. (1997), or Olea et al. (1999).

Finally, let us briefly mention the broad areas of psychological measurement in which the new technologies are having most impact, and which should be closely monitored. Excellent general expositions can be found in Prieto & Simon (1997). According to the recent authoritative work by Bennett (1999), there would be five main areas in which technological advances are producing irreversible changes. Firstly, the *design of tests* (Irvine & Kyllonen, 2001) which will become based more and more on cognitive models and principles, increasing their validity, impartiality, utility and credibility. Another area profoundly affected will be that of the *generation of items*, in which advances are being made almost daily with regard to the

Testing Practices in European Countries

automatic construction of items by computers, something that was unthinkable only a few years ago (Bejar, 1993; Embretson, 1998). The *presentation of items* constitutes one of the areas most affected by advances in information technology, which have made way for the multimedia context, making possible a marked improvement in item validity, given the similarity of test items to the real situations represented. Moreover, not only can items now be simulated and presented sequentially according to previous responses, but the possibility is also opened up of recording a range of data on responses, such as failed attempts, times, etc. The advantages over the classic paper and pencil format are innumerable. The *scoring of tests* is also being transformed thanks to the new technologies available, and in a variety of ways, which include rapid and automatic scoring of multiple-choice tests and the automatic preparation of auxiliary reports for professionals. Great strides are being made in this respect in the scoring of essays by computer, with several programs already available that attain higher levels of agreement with the scores of humans than those attained by other humans (Clauser, 2000; Braun et al., 1990; Bennett & Bejar, 1998; Clauser et al., 1995; Page & Petersen, 1995; Stevens et al., 1996). Finally, distance learning via the Internet and *remote assessment* itself present challenges and problems for psychological measurement that were unimaginable just a few years ago (Prieto & Simon, 1997). We could also mention other matters of great current importance with which experts grapple, such as item banks, information function of tests and items, bias, multidimensional models, and so on.

In sum, the field of psychological measurement is today in a state of feverish innovatory activity the like of which has not been seen since its origins. This has given rise to new challenges for professionals, both in relation to technological advances and to the correct use of tests and the observation of codes of conduct. New tendencies will be obliged to coexist with more traditional forms of assessment, but European psychologists will do well to keep track of the rapidly changing scene in psychological assessment if they are to avoid a loss of competitiveness in the global context within which we are destined to operate.

Testing Practices in European Countries

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.
- Anastasi, A. (1987). What test users should know about the interpretation of test scores. Keynote address at Joint Committee on Testing Practices Second Test Publishers Conference, Rockville, Maryland. (Quoted from Fremer, 1996).
- American Psychological Association (1992). Ethical principles of psychologists and code of conduct. American Psychologist, *47*, 1597-1611.
- 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. (1998a). The ITC/EFPA survey of testing and test use in countries world-wide. Technical report for the ITC Council.
- Bartram, D. & Coyne, I. (1998b). Variations in national patterns of testing and test use: The ITC/EFPA international survey. European Journal of Psychological Assessment, *14*, 249-260.
- Bejar, I. I. (1993). A generative approach to psychological and educational measurement. In N. Frederiksen, R. J. Mislevy, & I. I. Bejar (Eds.), Test theory for a new generation of tests (pp. 323-357). Hillsdale, NJ: Erlbaum.
- Bennett, R. E. (1999). Using new technology to improve assessment. Educational Measurement: Issues and practice, *18*, 5-12.
- Bennett, R. E. & Bejar, I. I. (1998). Validity and automated scoring: It's not only the scoring. Educational Measurement: Issues and Practice, *17*, 9-17.
- Braun, H. I., Bennett, R. E., Frye, D., & Soloway, E. (1990). Scoring constructed responses using expert systems. Journal of Educational Measurement, *27*, 93-108.
- Clauser, B. E. (2000). Recurrent issues and recent advances in scoring performance assessments. Applied Psychological Measurement, *4*, 310-324.

Testing Practices in European Countries

- Clauser, B. E., Subhiyah, R. G., Nungenster, R. J., Ripkey, D. R., Clyman, D., & McKinley, D. (1995). Scoring a performance-based assessment by modelling the judgement process of experts. Journal of Educational Measurement, *32*, 397-415.
- Colom, R. (2000). Algunos mitos de la psicología: entre la ciencia y la ideología. Psicothema, *12*, 1-14.
- Cronbach, L. J. (1975). Five decades of public controversy over mental testing. American Psychologist, *30*, 1-14.
- Deinzer, R., Steyer, R., Eid, M., Notz, P., Schwenkmezger, P., Ostendorf, F., & Neubauer, A. (1995). Situational Effects in Psychological Assessment of Traits: The FPI, NEOFFI, and EPI questionnaires. European Journal of Personality, *9*, 1-23.
- Embretson, S. E. (1998). A cognitive design system approach to generating valid tests: Application to abstract reasoning. Psychological Methods, *3*, 380-396.
- European Federation of Professional Psychologists' Associations (1996). Meta-Code of ethics. European Psychologist, *1*, 151-154.
- Evers, A. (1996). Regulations concerning test qualifications and test use in The Netherlands. European Journal of Psychological Assessment, *12*, 2, 153-159.
- Evers, A., Van Vliet-Mulder, J. C. & Groot, C. J. (2000). Documentatie van tests en testresearch in Nederland. Amsterdam, NIP/Assen: Van Gorcum.
- Eyde, L. D., Moreland, K. L., Robertson, G. J., Primoff, E. S., & Most, R. B. (1988). Test user qualifications: A data-based approach to promoting good test use. Issues in Scientific Psychology, Washington, DC: American Psychological Association.
- Eyde, L. D., Robertson, G. J., Krug, S. E., Moreland, K. L., Robertson, A. G., Shewan, C. M., Harrison, P. L., Porch, B. E., Hammer, A. L., & Primoff, E. S. (1993). Responsible test use. Case studies for assessing human behavior. Washington, DC: American Psychological Association.
- Fernández-Ballesteros, R. (1997). Guidelines for the assessment process. European Psychologist, *2*, 352-355.
- Fremer, J. (1996). Promoting high standards for test use: Developments in the United States. European Journal of Psychological Assessment, *12*, 2, 160-168.
- Hambleton, R. K. (1993). Translating achievement tests for use in cross-national studies.

Testing Practices in European Countries

- European Journal of Psychological Assessment, 9, 54-65.
- Hambleton, R. K. (1994). Guidelines for adapting educational and psychological tests: A progress report. European Journal of Psychological Assessment, 10, 229-240.
- Hambleton, R. K. & Bollwark, J. (1991). Adapting tests for use in different cultures: Technical issues and methods. Bulletin of the International Test Commission, 18, 3-32.
- Irvine, S. & Kyllonen, P. (Eds.) (2001). Item generation for test development. New Jersey: Lawrence Erlbaum.
- Martínez, J., García, E., & Muñiz, J. (2000a). Efecto del entrenamiento sobre las propiedades psicométricas de los tests. Psicothema, 12, Suppl. 2, 358-362.
- Martínez, J., Muñiz, J., & García, E. (2000b). Mejora de las puntuaciones de los tests mediante el entrenamiento. Psicothema, 12, Suppl. 2, 63-367.
- Moreland, K. L., Eyde, L. D., Robertson, G. J., Primoff, E. S., & Most, R. B. (1995). Assessment of test user qualifications. American Psychologist, 5, 14-23.
- Muñiz, J. (1997). Aspectos éticos y deontológicos de la evaluación psicológica. In A. Cordero (Ed.), Evaluación psicológica en el año 2000 (pp. 307-345). Madrid: TEA Ediciones.
- Muñiz, J. & Fernández-Hermida, J.R. (2000). La utilización de los tests en España. 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, 151-157.
- Olea, J., Ponsoda, V., & Prieto, G. (1999). Tests informatizados: fundamentos y aplicaciones. Madrid: Pirámide.
- Page, E. B., & Petersen, N. S. (1995). The computer moves into essay grading: Updating the ancient test. Phi Delta Kappa, 76, 561-565.
- Poortinga, Y., Coetsier, P., Meuris, G., Miller, K.M., Samsonowitz, V. & Seisdedos, N. (1982). A survey of attitude towards tests among psychologists in six European countries. International Review of Applied Psychology, 31, 7-34.
- Prieto, G. & Muñiz, J. (2000). Un modelo para evaluar la calidad de los tests utilizados en España. Papeles del Psicólogo, 77, 65-71.
- Prieto, J. M. & Simon, C. (1997). Network and its implications for assessment. In N. Anderson and P. Herriot (Eds.), International handbook of selection and assessment

Testing Practices in European Countries

- (pp. 97-124). Chichester: Wiley.
- Sands, W. A., Waters, B. K., & McBride, J. R. (Eds.) (1997). Computerized adaptive testing: From inquiry to operation. Washington: APA.
- 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.
- Sternberg, R. J. & Kaufman, J. C. (1996). Innovation and Intelligence Testing: The curious case of the dog that didn't bark. European Journal of Psychological Assessment, *12*, 175-182.
- Stevens, R., Lopo, A., & Wang, P. (1996). Artificial neural networks can distinguish novice and expert strategies during complex problem solving. Journal of the American Medical Informatics Association, *3*, 131-138.
- Steyer, R., Ferring, D., & Schmitt, M. J. (1992). States and traits in psychological assessment. European Journal of Psychological Assessment, *8*, 79-98.
- Steyer, R. & Partchev, I. (2001). Latent state-trait modeling with logistic item response models. In R. Cudeck, S. Du Toit, & D. Sörbom (eds.), Structural equation modeling: Present and future (pp. 481-520). Chicago: Scientific Software International.
- Steyer, R., Schmitt, M., & Eid, M. (1999). Latent state-trait theory and research in personality and individual differences. European Journal of Personality, *13*, 389-408.
- Tyler, B. (1986). Responsibility in practice: Some implications of the BPS survey on test use. Bulletin of the British Psychological Society, *39*, 410-413.
- Wainer, H. (1990). Computerized adaptive testing: A primer. Hillsdale, NJ: LEA.

Testing Practices in European Countries

Appendix: EFPA Questionnaire

On the initiative of the *EFPA Task Force on Tests and Testing*, a survey is being carried out in European countries in order to find out the opinion of psychologists on various aspects related to tests. This information will be of assistance in making decisions to improve test use in our country. We would greatly appreciate your devoting a few minutes to giving us your sincere opinion about a series of matters related to tests. Your responses should relate to your understanding of the situation in your area of professional specialisation. Thank you very much for your co-operation.

Your responses are to be made on a scale of 1-5: if you *totally disagree* with the statement, circle 1; if you *totally agree* with the statement, circle 5. Use the numbers 2, 3 and 4 for intermediate opinions. The survey is anonymous.

General data:

Age: Sex: Male/Female
Professional speciality: Clinical Educational Work Other (indicate)

Questionnaire

1. The training received in psychology degree courses is sufficient for the correct use of most tests
2. Professionals are provided with sufficient information (independent reviews, research, etc.) on the quality of tests published in our country
3. My current knowledge with regard to tests is basically that which I learned on my psychology degree course
4. The use of psychological tests should be restricted to qualified psychologists
5. While non-psychologists may administer and score tests, interpretation and feedback should be restricted to psychologists
6. There should be an enforceable standard defining the minimum technical qualities of a test
7. Legislation is needed to control the more serious abuses of testing
8. People who have been tested have a right to have their results explained to them
9. Abuse and misuse in testing is best countered through training and increasing public understanding of tests
10. Abuse and misuse of tests and testing procedures is a serious problem in our country
11. Anyone who can demonstrate their competence, as a test user should be allowed to use tests
12. Controls on tests and testing should be minimal, as controls discourage the development of new ideas and new procedures
13. Publishers should be allowed to sell whatever tests they think fit
14. Psychologists should only be allowed to use those tests for which they have demonstrated their competence
15. Our National Psychological Association should take a more active role in the regulation and improvement of test use
16. I use tests regularly in the exercise of my profession
17. Tests constitute an excellent source of information if they are combined with other psychological data
18. Used correctly, tests are of great help to the psychologist
19. Indicate the frequency with which the following test-use problems occur in your professional environment (1: very rarely; 5: very frequently)
 1. Making photocopies of *copyrighted* materials
 2. Making evaluations using inappropriate tests
 3. Not keeping up with the field and failing to check one's own interpretations with others
 4. Not considering errors of measurement of a test score
 5. Not restricting test administration to qualified personnel
 6. Not taking into account conditions that cast doubt on reported validity for a local situation
 7. Making interpretations which go beyond the limits of the test
20. Name the three tests you most frequently use in the exercise of your profession:

Testing Practices in European Countries

Observations. Please make any additional comments you think appropriate (you may include extra sheets, if necessary)

Please put the questionnaire in the envelope provided and post it (no stamp needed). *Thank you very much for your co-operation.*

Acknowledgements:

The authors would like to thank all members of the EFPA Task Force on Tests and Testing for their invaluable help in the design of the questionnaire used.

Testing Practices in European Countries

Countries	N	Sex		Age	
		Female	Male	O	F
Belgium	210	54.3%	45.7%	40.68	11.46
Croatia	218	82.6%	17.4%	39.11	8.69
Holland	2079	63.5%	36.5%	41.23	11.04
Slovenia	321	73.0%	27.0%	41.80	9.62
Spain	3455	72.7%	27.3%	35.64	7.99
UK	2407	55.0%	45.0%	46.74	9.73

Table 1. Sample description

Country		Belgium	Croatia	Holland	Slovenia	Spain	UK
Professional Field	Clinical	56.19%	28.90%	60.40%	30.03%	58.90%	45.04%
	Educational	23.81%	23.40%	7.20%	33.00%	19.70%	24.84%
	Work	10.95%	18.30%	20.10%	28.71%	14.70%	17.37%
	Others	9.05%	29.40%	12.30%	8.25%	6.70%	12.75%

Testing Practices in European Countries

Table 2. Professional field

Item	Belgium		Spain		UK		Croatia		Slovenia		Holland	
	0	F	0	F	0	F	0	F	0	F	0	F
1. The training received in psychology degree courses is sufficient for the correct use of most tests	2.97	1.12	2.41	1.19	2.18	1.18	3.11	1.25	2.67	1.17	2.70	1.09
2. Professionals are provided with sufficient information (independent reviews, research, etc.) on the quality of tests published in our country	2.37	0.92	2.38	0.99	2.65	1.05	2.51	1.02	2.88	1.01	2.75	0.95
3. My current knowledge with regard to tests is basically that which I learned on my psychology degree course	2.74	1.32	2.57	1.34	1.92	1.25	2.37	1.25	2.15	1.22	2.55	1.27
4. The use of psychological tests should be restricted to qualified psychologists	4.29	0.92	4.23	1.06	3.88	1.29	4.71	0.72	4.73	0.73	4.17	0.98
5. While non-psychologists may administer and score tests, interpretation and feedback should be restricted to psychologists	3.84	1.28	4.34	1.18	3.79	1.37	4.78	0.64	4.71	0.76	3.98	1.14
6. There should be an enforceable standard defining the minimum technical qualities of a test	4.18	0.82	4.33	0.84	4.09	1.04	4.23	0.99	4.09	0.97	3.96	0.87
7. Legislation is needed to control the more serious abuses of testing	4.06	0.91	4.29	0.89	3.72	1.18	4.08	1.04	4.26	0.92	3.61	1.07
8. People who have been tested have a right to have their results explained to them	4.64	0.60	4.55	0.78	4.65	0.78	4.44	0.85	4.64	0.68	4.79	0.48
9. Abuse and misuse in testing is best countered through training and increasing public understanding of tests	3.00	1.05	4.01	1.16	4.08	0.99	3.88	1.11	3.63	1.24	3.21	1.07
10. Abuse and misuse of tests and testing procedures is a serious problem in our country	2.72	1.05	3.37	0.99	2.92	0.89	3.06	1.03	2.61	0.91	2.55	0.80
11. Anyone who can demonstrate their competence as a test user should be allowed to use tests	2.62	1.15	2.42	1.32	2.51	1.29	1.69	1.02	1.73	1.07	2.81	1.16
12. Controls on tests and testing should be minimal, as controls discourage the development of new ideas and new procedures	2.43	0.98	1.85	0.95	1.98	1.06	1.91	0.95	1.89	0.95	2.09	0.83
13. Publishers should be allowed to sell whatever tests they think fit	2.54	1.14	1.57	0.93	2.07	1.23	2.28	1.16	1.88	1.15	2.75	1.15
14. Psychologists should only be allowed to use those tests for which they have demonstrated their competence	3.23	1.13	1.39	0.83	3.64	1.21	2.98	1.17	3.75	1.10	3.28	1.05
15. Our National Psychological Society should take a more active role in the regulation and improvement of test use	4.00	0.72	4.15	0.93	4.03	0.99	4.48	0.78	4.13	0.83	3.71	0.85
16. I use tests regularly in the exercise of my profession	3.28	1.29	3.56	1.22	3.98	1.31	3.90	1.38	3.68	1.37	3.08	1.39
17. Tests constitute an excellent source of information if they are combined with other psychological data	4.47	0.63	4.41	0.83	4.42	0.86	4.85	0.40	4.59	0.69	4.41	0.69
18. Used correctly, tests are of great help to the psychologist	4.33	0.71	4.37	0.83	4.42	0.85	4.77	0.49	4.58	0.71	4.28	0.73
19-1. Making photocopies of <i>copyrighted</i> materials	2.96	1.31	3.60	1.32	2.44	1.36	3.28	1.30	2.04	1.20	2.15	1.11
19-2. Making evaluations using inappropriate tests	2.07	1.17	2.63	1.28	2.18	1.14	2.78	1.17	1.94	0.99	1.78	0.93
19-3. Not keeping up with the field and failing to check one's own interpretations with others	2.98	1.01	3.23	1.22	2.84	1.09	2.88	1.10	2.89	1.11	2.47	0.94
19-4. Not considering errors of measurement of a test score	2.69	1.22	3.07	1.22	2.86	1.20	2.97	1.16	2.50	1.11	2.32	1.12
19-5. Not restricting test administration to qualified personnel	2.20	1.23	2.91	1.46	2.14	1.26	2.70	1.23	2.09	1.24	1.59	0.95
19-6. Not taking into account conditions that cast doubt on reported validity for a local situation	2.49	1.17	3.28	1.31	2.48	1.19	2.96	1.15	2.29	1.09	2.01	0.98
19-7. Making interpretations which go beyond the limits of the test	2.43	1.16	2.99	1.37	2.57	1.26	2.95	1.12	2.25	1.15	1.98	0.98

Testing Practices in European Countries

Table 3. Item means and standard deviations

Belgium	Croatia	Holland	Slovenia	Spain	UK
WISC/ WISC-R/ WISC-III (32%)	EPQ (29%)	WISC/ WISC-R (17%)	WISC (37%)	16PF/16PF-5 (27%)	WISC (33%)
WAIS, WAIS-R (23%)	BENDER (26%)	MMPI/MMPI-II (16%)	BENDER (28%)	WISC/ WISC-R/WPPSI (23%)	WAIS (26%)
WPPSI/WPPSI-R (18%)	SPM-RAVEN (21%)	SCL-90: Symptom Check List (15%)	WAIS (28%)	WAIS (18%)	Individual Wechsler Scales (22%)
TAT (16%)	MMPI (20%)	NVM: Dutch abridged MMPI (14%)	Emotional Profile Index (22%)	MMPI/MMPI-II (15%)	British Ability Scales (14%)
RORSCHACH (15%)	WISC (19%)	WAIS (13%)	Aptitude Tests for Elementary School Children (18%)	Beck Depression Inventory (9%)	MBTI (9%)
CAT (10%)	Revised BETA (16%)	NPV: Dutch Personality Questionnaire (13%)	RORSCHACH (16%)	STAI (8%)	16PF (8%)
MMPI/MMPI-II (9%)	Index of Emotion (15%)	UCL/UCL-R: Utrechtse Coping List (6%)	Test of School Maturity (12%)	RORSCHACH (6%)	Beck Depression Inventory (8%)
Complex Figure REY (5%)	CORNELL Index (12%)	GIT/V-GIT: Groninger Intelligence	Vocational Interests Test (11%)	RAVEN (5%)	WPPSI (8%)

Testing Practices in European Countries

		Test (5%)			
RAVEN (5%)	WAIS (10%)	NEO-PI-R/NEO-FFI (Big Five) (4%)	Self Directed Search (7%)	BENDER (4%)	Neale Analysis of Reading Test (6%)
CBCL (Children Behavior Checklist) (5%)	Problem T (7%)	DAT (4%)	EPQ/EPQ-R (7%)	ISRA (4%)	Occupational Personality Questionnaire (4%)

Testing Practices in European Countries

Table 4. Tests more frequently used by professionals

Items	Factors				
	1	2	3	4	5
19-2	.79				
19-7	.78				
19-4	.78				
19-3	.77				
19-6	.73				
19-5	.70				
19-1	.50				
18		.86			
17		.85			
16		.71			
14		-.54			
7			.71		
6			.60		
15			.57		
10			.51		
4			.48		
8			.29		
1				.85	
3				.79	
2				.46	
11					.62
12					.55
13					.51
9					.51
5					.32
Eigenvalues	3.97	2.73	1.80	1.64	1.34
Percent of Variance Explained	15.87	10.93	7.21	6.55	5.36

Extraction Method: Principal Component Analysis, 5 factors extracted. Rotation Method:

Varimax with Kaiser Normalization. Loadings $\leq .40$ blanked, except if highest loading $\leq .40$

Testing Practices in European Countries

Table 5. Factor Analysis of the Questionnaire (Spain)

Items	Factors				
	1	2	3	4	5
19-6	.83				
19-7	.82				
19-2	.79				
19-4	.76				
19-5	.73				
19-3	.69				
19-1	.49				
15		.64			
8		.61			
6		.60			
14		.58			
9		.53			
7		.49			
18			.85		
17			.85		
16			.76		
3			-.44		.41
4				.74	
11				-.73	
5				.63	
13				-.44	
12				-.43	
1					.72
2					.53
10					-.33
Eigenvalues	3.95	2.52	2.39	2.34	1.54
Percent of Variance Explained	15.82	10.09	9.57	9.35	6.18

Extraction Method: Principal Component Analysis, 5 factors extracted. Rotation Method: Varimax with Kaiser Normalization. Loadings $\leq .40$ blanked, except if highest loading $\leq .40$.

Testing Practices in European Countries

Table 6. Factor Analysis of the Questionnaire (England)

Items	Factors				
	1	2	3	4	5
19-7	.81				
19-6	.80				
19-2	.75				
19-3	.74				
19-4	.73				
19-5	.69				
19-1	.42				
7		.67			
6		.64			
15		.62			
12		-.57			
14		.50			
10		.49			
13		-.39			
8		.26			
18			.83		
17			.81		
16			.64		
11				.64	
4				-.59	
5				-.41	
9				.39	
1					.77
3					.55
2					.47
Eigenvalues	4.19	2.53	2.21	1.33	1.30
Percentage of Variance Explained	16.12	9.72	8.50	5.12	4.99

Extraction Method: Principal Component Analysis, five factors extracted. Rotation Method: Varimax with Kaiser Normalization. Loadings $\leq .40$ blanked, except if highest loading $\leq .40$.

Testing Practices in European Countries

Table 7. Factor Analysis of the Questionnaire (Holland)