The ITC Guidelines on Computer-Based and Internet-Delivered Testing: Where Do We Go From Here?

Filip Lievens
Department of Personnel Management and Work and Organizational Psychology
Ghent University, Belgium

The International Test Commission’s (this issue) Guidelines on Computer-Based and Internet-Delivered Testing constitute a comprehensive and excellent set of internationally agreed guidelines. This article looks forward and discusses what should be done to ensure that these guidelines will be used in educational, clinical, and organizational testing practice. Specifically, 4 recommendations are made. First, an abridged version that outlines the major issues for test-takers to attend to when completing a computer-based or Internet-delivered test should be constructed. Second, the practicality of the guidelines should be increased by regularly updating them, by consistently mentioning the statistical approaches for examining adherence to them, and by using them as a starting point for constructing formal standards. Third, one should not only use the traditional rational and technical way to persuade the relevant stakeholders to use the guidelines; one might also experiment with persuasion on the basis of short, “vivid,” and more comprehensible information (e.g., anecdotes or case studies). Fourth, some kind of premium might be given to test publishers, test developers, and test users who follow the guidelines (e.g., by placing certificates on their Web site or by listing the “10 Best Test Publishers”).

This is the sixth article in this special issue about the International Test Commission’s (ITC, this issue) Guidelines for Computer-Based and Internet-Delivered Testing. Whereas Bartram (this issue) cogently highlighted the need to consider the implications of the Internet on good practice at an international level, Coyne

Correspondence should be addressed to Filip Lievens, Department of Personnel Management and Work and Organizational Psychology, Ghent University, Henri Dunantlaan 2, 9000 Ghent, Belgium. E-mail: filip.lievens@ugent.be
and Bartram (this issue) gave a detailed overview of the design and development of the ITC guidelines, Foxcroft and Davies (this issue) applied the guidelines to a specific country (South Africa), and Sale (this issue) provided a practitioner’s viewpoint on the guidelines, I aim to look forward and provide some recommendations about how to ensure that these guidelines are actually used by test publishers, test developers, test users, and test-takers.

I chose to look forward instead of to look back because it is hard to criticize the excellent work that has been accomplished by the ITC. If one just considers the development process, it becomes clear that this has been a very thorough endeavor. The relevant international literature was reviewed, a dozen test publishers were surveyed, and a conference was organized. As a result, these guidelines are very much an international affair, as they are based on input from various countries (Australia, Canada, Croatia, Cyprus, Estonia, Slovenia, South Africa, The Netherlands, the United Kingdom, and the United States) and various organizations (British Psychological Society, British Standards’ Institute, American Psychological Association Internet Task Force, Society for Industrial and Organizational Psychology, National Board of Medical Examiners, European Federation of Psychologists’ Associations, Association of Test Publishers, etc.). Clearly, such an international outreach is needed for Internet-delivered testing, as this type of testing crosses, by definition, any physical borders. The guidelines are also quite comprehensive. Totaling 26 pages of text, they cover almost any imaginable issue related to computer-based and Internet-delivered testing identified to date, making a distinction between four main issues (technology used, psychometric quality, control over test administration, and security or privacy) and three key stakeholders (test developers, test publisher, and test users). Across these issues and stakeholders, a wide range of guidelines are presented, ranging from human factors to equivalency, adverse impact, or privacy.

Due to their comprehensiveness, I am sure that the guidelines will guide future research, especially on computer-based and Internet-delivered testing. In their recent overview of extant research with regard to Internet-based testing, Lievens and Harris (2003) noted

> The most striking examples are that, to the best of our knowledge, neither the criterion-related validity of Internet testing applications nor the possible adverse impact of Internet testing applications have been put to scrutiny. Moreover, most studies have treated Internet testing as a monolithic entity, ignoring the multiple dimensions of Internet testing discussed above. (p. 148)

The guidelines might serve as an excellent guide and basis for designing research studies to tackle these issues. For instance, to examine the impact of different modes of administration on the psychometric quality of different types of tests, the guidelines make the very relevant distinction between four modes of testing—namely, the open, controlled, supervised, and managed modes of testing.
So, where do we go from here? I believe the work is not finished. A potential shortcoming of any guidelines is that they might become a dead letter. So, the central issue should now be what we can do to improve the use and implementation of the guidelines for computer-based and Internet-delivered testing in practice. As mentioned by Coyne and Bartram (this issue), the consultation samples during the developmental process already gave some recommendations on how best to disseminate the guidelines. Experts in these samples suggested disseminating the guidelines via national organizations, mailing lists for the assessment community, conferences, workshops, symposia, newsletters, journal publications, the ITC Web site, and test sponsors and publishers. Undoubtedly, these are valuable suggestions. Yet, I would also like to delve into some other opportunities. Specifically, I provide four additional suggestions. The remainder of this article discusses these suggestions in more detail.

First, I suggest making an abridged version of the guidelines available for test-takers. As they are now, the guidelines distinguish between four main categories—namely, test publishers, test developers, and test users. Clearly, these stakeholders are the most relevant ones, as it is a condition sine qua non that they follow some codes of practice to guarantee the quality, performance, control, and security of computer-based and Internet-delivered testing. I believe, however, that test-takers should also be considered. After all, in both clinical, educational, and employment settings, people usually take tests without a thorough knowledge and understanding of what constitutes good computer-based or Internet testing. In the guidelines, test-taker perceptions and issues are definitely taken into account (e.g., the section on providing information on practice items, the section on human factors, or the section on the security of test-takers’ data). My concrete suggestion, however, is that a very brief synopsis of the guidelines is also made. For instance, the guidelines might be formulated in the form of “10 things one should pay attention to when taking a computer-based or Internet-delivered test.” I believe that “educating” test-takers in this way and giving them voice might further protect them against so-called cowboy test operators. It might also encourage the other stakeholders (test publishers, test developers, and test users) to actually use the guidelines. Paying attention to test-takers also fits nicely with relatively recent developments in selection research. Over the last 15 years, a considerable body of research has focused on applicants’ and test-takers’ perceptions (Hausknecht, Day, & Thomas, 2004; Ryan & Ployhart, 2000). Finally, to paraphrase a famous Patti Smith song, “Applicants have the power.” In fact, when applicants feel that they are not treated fairly, they might challenge test users, test publishers, and test developers in the courts. In the end, this might also encourage test users, test publishers, and test developers to pay attention to the guidelines and use them. Let me be clear here. I am not arguing in favor of a proliferation of legal cases with regard to computer-based and Internet-delivered testing. Yet, it should be noted that in the United States over the years, some selection procedures have become popular be-
cause they could withstand legal attacks. For example, assessment centers and structured interviews have made rapid inroads in practice due to the fact that they are seldom challenged in the courts (Terpstra, Mohammed, & Kethley, 1999; Williamson, Campion, Malos, Roehling, & Campion, 1997).

As a second suggestion, it would help to increase the “practicality” of the guidelines. This could be done in various ways. One way might consist of indicating in some instances more concretely how test publishers, test developers, and test users might put some of the guidelines to the test. Granted, in most cases, the guidelines indicate the statistical approaches that should be conducted. For example, the guidelines clearly refer to differential item functioning in the context of adverse impact. In other instances, however, test users are left guessing how to actually implement the guidelines. For instance, Guideline 20.5 suggests “to verify that the computer-based/Internet test does not require knowledge, skills, or abilities that are irrelevant to the construct being assessed.” I am not sure that many test publishers will know how to examine this. Similarly, Guideline 22.1 does not posit that Cxxxx Fxxxx Axxxx or Ixxxx Rxxxx Txxxx models can be used to examine equivalence (invariance) of tests across different conditions (see also Guideline 40.4). Another suggestion to bolster the practicality of the guidelines might consist of mentioning more specific formal standards. Although I acknowledge that the provision of rules-of-thumb might be dangerous and is not the objective of the guidelines per se (they are meant to be an overarching framework that can be implemented locally), I expect the relevant parties in the testing process to ask questions about cut-off points, desirable values, and so forth. Furthermore, the practicality of the guidelines might be enhanced by amending current codes of ethics (e.g., British Psychological Association) and adding rather specific guidelines to them directly focusing on online testing (Barak & Buchanan, 2004). Finally, it should be clear that we should ourselves take some advice from the guidelines to heart. Specifically, the guidelines posit that “the system should be designed to accommodate likely advances in technology” (Guideline 1.4). The same is true for the guidelines themselves. We cannot expect users to follow guidelines that seem outdated to them. This will not be an easy task given the rapid speed at which technology is changing nowadays. Similarly, a couple of years from now, a formal evaluation should take place to verify whether the guidelines are actually used in practice and what factors impede their implementation. On the basis of comments and suggestions of the relevant stakeholders, revisions should be made to the guidelines.

A third issue is which strategies we can deploy to persuade test publishers, test developers, and test users to use the guidelines. One way of persuading the relevant parties involved might consist of documenting the need for guidelines by relying on “hard” evidence such as validity studies and adverse impact statistics. Generally, one might refer to this strategy as the traditional central route to persuasion. Johns (1993) referred to this strategy as promoting new practices as rational technical innovations. It might be fruitful, however, to draw some lessons from the broader models of social
persuasion, belief change, and attitude change (e.g., Carson, Becker, & Henderson, 1998; Highhouse, 1996; Morley, 1987; Petty & Cacioppo, 1981). For example, this large literature suggests the importance of emphasizing the credibility of the information source when presenting the information. The thorough process of developing the guidelines and the involvement of multiple international organizations demonstrates that we have covered this credibility issue with the guidelines. Hence, it should be communicated accordingly to the stakeholders. Apart from credibility, the social persuasion literature suggests provision of information that is straightforward, short, “vivid,” and comprehensible to the parties involved. Examples here are anecdotes or (legal) case studies that might be presented in the popular press and media. I believe that these approaches should be considered much more frequently, as they might have more impact than technical validity reports.

As a last recommendation, I suggest that some kind of premium is given to test publishers, test developers, and test users who follow the guidelines. Along these lines, Naglieri et al. (2004) mentioned that a mental health consumer advocacy and education program started a process for checking credentials of online counselors. If online therapists pass the check conducted by this advocacy’s group, they are allowed to put an icon on their Web site. Users can verify this icon by going to the mental health consumer advocacy Web site. Naglieri et al. then suggested that a consortium of test publishers could do the same. I agree, and I believe the ITC guidelines could be fruitfully used to this end. Relatedly, a list of “10 Best Test Publishers” could be made. Although such a project is not without problems (which countries would it cover, what are the criteria, who makes the decision, what redress do companies not on the list have, what happens if a publisher takes out a legal action for not being on the list?), I believe it could be a success. It suffices to look at the tremendous impact of the “Best Companies to Work For” or “Great Place to Work For” lists. Each company that is included in these lists is proud to mention this in its job ads and on its Web site.

In conclusion, the ITC (this issue) Guidelines on Computer-Based and Internet-Delivered Testing constitute a comprehensive and excellent set of internationally agreed guidelines. Now we should ensure that they are used. Thus, I would take the liberty to add a third purpose to the guidelines. It is not only necessary to produce a set of international guidelines that highlight good practice in computer-based and Internet-delivered testing and to raise awareness among all stakeholders in the testing process of this. It is also crucial to undertake efforts to disseminate the guidelines as best as possible. To this end, this article provided various recommendations.

REFERENCES


