Industrial and Organizational Psychology, **2** (2009), 11–13.
Copyright © 2009 Society for Industrial and Organizational Psychology. 1754-9426/09

COMMENTARIES

The International Test Commission Guidelines on Computer-Based and Internet-Delivered Testing

DAVE BARTRAM
SHL Group Ltd.

The International Test Commission (ITC) has a policy of promoting good practice in testing issues where international coordination of effort is most important. The ITC has devised guidelines to promote good practice in test adaptations (Hambleton, 2005) and good practice in test use (International Test Commission, 2001). Developments in the provision of Internet-delivered computer-based testing led the ITC to set up a project to develop guidelines specifically aimed at computer-based and Internet-delivered testing.

The intentions were to draw together common themes that run through existing guidelines, codes of practice, standards, research papers, and other sources and to create a coherent structure within which these guidelines can be used and understood. Contributions to the guidelines were made by the relevant committees of national psychological associations (including the Society for Industrial and Organizational Psychology and the American Psychological Association's Committee on Psychological Testing and Assessment) and psychological and educational testing specialists, including test developers, test publishers, and test users drawn from a large number of countries in the Americas, Europe, Australasia, and Africa (for further details of the process by which the guidelines were developed, see Coyne & Bartram, 2006). These guidelines are intended to complement other good practice guidelines such as the ITC (2001) *International Guidelines for Test Use* and the American Educational Research Association, American Psychological Association, and National Council on Measurement in Education (1999) *Standards for Educational and Psychological Testing* with a specific focus on Internet testing.

Development of the Guidelines

Although the guidelines[1] (ITC, 2006) specifically address the issue of unproctored Internet testing (UIT), it is important to realize that they do so within a much broader context of setting out good practice in the use of the Internet to deliver tests. The guidelines are structured around four general issues: (a) Technology: Ensuring that the technical aspects of CBT/Internet testing are considered, especially in relation to the hardware and software required to run the test at both the server and the client end;

Correspondence concerning this article should be addressed to Dave Bartram. E-mail: dave.bartram@shlgroup.com

Address: SHL Group Ltd., The Pavilion, 1 Atwell Place, Thames Ditton, Surrey, KT7 0XB, United Kingdom

Dave Bartram is past president of the International Test Commission (ITC) and is the International Association of Applied Psychology (IAAP) liaison member on the ITC Council. Together with Dr. Iain Coyne, University of Nottingham, UK, he was responsible for the ITC Computer-Based Testing and Internet-Delivered Testing Guidelines project.

Dave Bartram, SHL Group Ltd., UK.

1. www.intestcom.org/guidelines/guidelines/index.html

https://doi.org/10.1111/j.1754-9434.2008.01098.x Published online by Cambridge University Press

(b) Quality: Assuring the quality of testing and test materials and ensuring good practice throughout the testing process; (c) Control: Controlling the delivery of tests, test-taker authentication, and prior practice; and (d) Security: Security of the testing materials, privacy, data protection, and confidentiality. Within each of these four main sections, guidelines are written to provide advice to each of three main stakeholder groups: test developers, test publishers, and test users.

Four modes of test administration (open, controlled, supervised, and managed) form the basis for the ITC guidelines' discussion of good practice in test administration (see also Bartram, 2006a, 2006b). Of the four, controlled mode corresponds to what we are here identifying as UIT: that is, there is no face-to-face human supervision of the test session, but there is some control in that the test is only made available to known test takers. For the Internet, this is controlled through the requirement for the test taker to be provided with a unique and confidential logon username and password. It can also be extended to include control over the IP address they can use to access the test, the date and time they can access it, and the properties of the hardware or browser they need to use.

The guidelines make clear the need to establish the psychometric properties of the test under the conditions in which it will be used (Guideline 22.3), especially where these conditions involve the lack of proctors. Research has shown high levels of equivalence for proctored and unproctored administration of self-report measures (such as personality tests): See Bartram (2008) for a recent review. However, concerns over UIT focus primarily on ability testing and other tests of maximum performance, where there are "right" answers and where lack of proctoring is seen as increasing the risk of cheating. These include stealing items so that the answer key can be determined and provided to future test takers and cheating by proxy, where another person takes the test on behalf of the true applicant.

The guidelines that specifically address the issue of cheating in UIT are Guidelines 43, 44, and 45. Guideline 45.3 states that: "For moderate and high stakes assessment (e.g., job recruitment and selection), where individuals are permitted to take a test in Controlled mode (i.e., at their convenience in non secure locations), those obtaining qualifying scores should be required to take a supervised test to confirm their scores.

> Procedures should be used to check whether the test-taker's original responses are consistent with the responses from the confirmation test.
> Test-takers should be informed in advance of these procedures and asked to confirm that they will complete the tests according to instructions given (e.g., not seek assistance, not collude with others etc.).
> This agreement may be represented in the form of an explicit honesty policy which the test-taker is required to accept. (ITC, 2006, p. 164)

This makes clear the need for a mixed approach to managing the risks associated with unproctored administration. Key to this is the need for some process that checks the data obtained in the proctored condition with that subsequently obtained under proctored or supervised conditions. It is also important that test takers know that their unproctored data will be checked in this way and that they make an explicit undertaking to complete the unproctored test as instructed (this is the so-called "honesty contract").

There are various ways in which scores obtained in UIT can be confirmed. Clearly, the most direct is by a follow-up proctored parallel-form test. This can be either used to confirm the score obtained on the first unproctored test or used in place of the first test (depending on how the various tests have been designed and the role of the test scores in the decision-making process).

The ITC recognizes that there are a number of ethical issues associated with UIT that need to be considered by test users. It is important that test users do not interpret the failure to confirm a score obtained in UIT as proving cheating or dishonesty on the part of the applicant. As in a proctored

https://doi.org/10.1111/j.1754-9434.2008.01098.x Published online by Cambridge University Press

environment, there needs to be positive evidence of cheating rather than just circumstantial evidence. Remote monitoring technologies can provide positive evidence and may be more widely employed in the future, but for the present, data forensics can only indicate aberrant scores or response patterns. Failure of confirmation simply raises a query that needs to be addressed by hiring managers. Any score confirmation procedure will be associated with a false-positive rate, whereby some applicants who have completed the UIT without cheating will not have their scores confirmed. Other contributors to this set of commentaries have commented on how this issue can be dealt with and on how good practice is being developed around this issue.

A second issue arises in the way cut scores are used. It is important to ensure that if someone were to cheat on an UIT and then be selected for further assessment, this would not exclude someone who had not cheated. This entails avoiding the use of a top-down selection approach in screening, whereby the "best" 50 people are selected if there are 50 places in the next stage of the assessment. The correct procedure is to set a cut score on screening that should yield 50 people for the second stage if there is no cheating. To the extent that cheating occurs, the number passing the cut will be higher than expected, but the additional numbers will be excluded by confirmation testing.

There are many more issues to consider in managing potential cheating, not least of which concern maintenance of test security (which are addressed in Guidelines 46 and 47). However, ultimately, the decision of whether to use UIT is one that brings together issues of practice, science, and policy. The practice issues relate to why client organizations want to use UIT in the first place. The economic argument for early use of high-validity instruments (such as ability tests) in a recruitment procedure is a compelling one, as it results in considerable cost savings as far smaller numbers of applicants have to be put through expensive short-list assessment procedures. It is this economic argument that has driven the rapid increase in the demand for and use of UIT and which has created the need for science and policy to find ways of delivering this in a fair, effective, and efficient way. Science is developing methods of ensuring test security through use of remote authentication, randomized test construction, data forensic analysis, and methods of score confirmation; and policy is developing good practice in relation to the procedures test providers and hiring managers should follow to ensure UIT is used in an ethical and fair way.

It is paradoxical that the concerns raised over risks of cheating in UIT, despite the fact that cheating is also an issue in proctored assessment, have resulted in the development of technologies, policies, and procedures that potentially make online UIT more secure than traditional paper-and-pencil proctored tests.

References

American Educational Research Association, American Psychological Association & National Council on Measurement in Education. (1999). *Standards for educational and psychological testing*. Washington, DC: American Psychological Association.

Bartram, D. (2006a). Testing on the Internet: The need for international guidelines. *International Journal of Testing, 6*, 121–132.

Bartram, D. (2006b). Testing on the Internet: Issues, challenges and opportunities in the field of occupational assessment. In D. Bartram & R. K. Hambleton (Eds.), *Computer-based testing and the Internet: Issues and advances* (pp. 13–37). Chichester, UK: John Wiley & Sons.

Bartram, D. (2008). The advantages and disadvantages of on-line testing. In S. Cartwright & C. L. Cooper (Eds.), *The Oxford handbook of personnel psychology* (pp. 234–260). Oxford, UK: Oxford University Press.

Coyne, I., & Bartram, D. (2006). Design and development of the ITC guidelines on computer-based and Internet delivered testing. *International Journal of Testing, 6*, 133–142.

Hambleton, R. (2005). Issues, designs and technical guidelines for adapting tests into multiple languages and cultures. In R. K. Hambleton, P. Merenda, & C. Spielberger (Eds.), *Adapting educational and psychological tests for cross-cultural assessment* (pp. 3–38). Mahwah, NJ: Erlbaum.

International Test Commission (2001). International guidelines for test use. *International Journal of Testing, 1*, 93–114.

International Test Commission (2006). International guidelines on computer-based and Internet delivered testing. *International Journal of Testing, 6*, 143–172.

https://doi.org/10.1111/j.1754-9434.2008.01098.x Published online by Cambridge University Press