Chapter 3.6
Improving Employee Selection with Online Testing

Jeffrey M. Pollack
Virginia Commonwealth University, USA

INTRODUCTION

Advances in technology now enable employers to utilize computers to administer online employee selection tests, which result in lower costs, increased efficiency, and fewer transcription errors (Richman, Kiesler, Weisband, & Drasgow, 1999; Tippins et al., 2006). Additionally, online employment testing software can effectively and efficiently assist in identifying individuals best suited to an occupation, reducing poor person-job fit, lowering turnover rates, decreasing training costs, and minimizing errors in hiring (Bingham, Ilg, & Davidson, 2002; Mooney, 2002). This article addresses issues related to online employment testing software including types of tests available, validity and reliability, proctoring, and social desirability. Additional terms are defined and implications and future directions for research are discussed.

BACKGROUND

For decades, the similarities and differences between written tests and computer-based tests have been assessed (Epstein & Klinkenberg, 2001). Early research investigated how a computer-based medical records keeping and interview system impacted patients (Slack & Van Cura, 1968; Slack, Hicks, Reed, & Van Cura, 1966). Additionally, research in this era investigated the use of computers as data gathering instruments (Evans & Miller, 1969; Vinsonhaler, Molineaux, & Rodgers, 1968). However, it is not until the 1990s that we see a research trend that begins to examine the equivalence of computer-based tests vs. conventional tests in an organizational setting (Donovan, Drasgow, & Probst, 2000; McHenry & Schmitt, 1994).

As technology has evolved, tests previously administered in a paper-and-pencil format have been changed to online versions. These tests include
clinical measurements, personality tests, attitude scales, cognitive ability tests and training inventories (Mead & Drasgow, 1993). Further examples of computer-administered assessments include medical admissions data, psychiatric evaluation exams, and consumer preference evaluations (Kiesler, Walsh, & Sproull, 1992; Richman et al., 1999; Synodinos & Brennan, 1988; Synodinos, Papacostas, & Okimoto, 1994).

The most simple and widely used type of computer-based test is computer assisted testing (CAT). These tests display a question on a computer screen and the respondent enters their response (Epstein & Klinkenberg, 2001). Computer assisted tests enable the online format of a test to very closely resemble the paper-and-pencil version and make the testing situation as similar as possible to a written one (Rozensky, Honor, Rasinski, Tovian, & Herz, 1986).

Another type of CAT program uses computer adaptive testing. Adaptive testing settings are different from assisted ones in that adaptive tests allow the computer to “go beyond a simple page turning function” (Epstein & Klinkenberg, 2001, p. 298). Adaptive tests allow a computer to receive a response, score it, and then choose the next appropriate question, either easier or harder based on a respondent’s answer (Green, Bock, Humphreys, Linn, & Reckase, 1984). In adaptive tests, which Epstein & Klinkenberg (2001) assert are similar to most non-computerized intelligence tests, there are multiple types that can “individualize” the testing experience and narrow the number of questions needed to assess the underlying trait (Burke, 1993; Weiss, 1985).

There are multiple options when using adaptive formats. A “two-stage” adaptive test is one in which a participant is given an initial pre-test called a routing exam. Based on his or her score on the routing exam, a test is then administered, which corresponds to their knowledge of the content (Epstein & Klinkenberg, 2001, p. 298). Other tests similar to the “two-stage” include the “pyramidal,” “flexilevel,” “stradaptive,” and “countdown” approaches (for reviews see Butcher, Keller, & Bacon, 1985; Epstein & Klinkenberg 2001; Weiss, 1985). More recent advances include test types called generating examples (GE) and are found described in reviews by Bennett (1999), Bennett et al., (1999) and Bennett, Steffen, Singley, Morley, and Jacquemin (1997).

Potential administrators of online tests must consider two main factors. First, the content area of the test must be identified. Whether it is desirable to test hard skills (e.g., proficiency in software or programming) and/or basic knowledge (e.g., ability to solve problems, communication skills) is the key consideration (Mooney, 2002). Second, online test administrators must consider who will compose the test. Many testing service companies will offer “authoring software” that enables users to compose their own questions as an alternative to the standard “menu” of tests available for various job classifications (Mooney, 2002). The authorship issue brings up questions of reliability and validity. This is especially true considering that the “in-house” development of tests may be the lower cost, but less reliable and valid option (Mooney, 2002). Companies often charge more per test administration for tests directly off the “menu” vs. charging a fee for test development but lowering the per test charge if it is modified by the end user (Mooney, 2002). It is important to note that most authoring software only allows for the creation of yes/no and true/false questions (Mooney, 2002). Additional considerations for potential administrators of online tests include the creation of user id’s, whether to use a timed test, whether respondents are proctored, whether respondents can backtrack to previous responses, and how to notify participants of results.

THE ACADEMIC AND FINANCIAL BOTTOM LINE

The bottom line is that, in most cases, online employment testing saves companies money and
Improving Employee Selection with Online Testing

Improving Employee Selection with Online Testing

reduces the recruitment cycle time (Bingham et al., 2002). As the field of online employment testing software evolves, it will be important to keep in mind the following concerns. Critics of online testing procedures bring up issues such as lack of computer familiarity, computer anxiety, lack of equivalency, as well as issues related to proctoring and social desirability as ways in which CAT is not an optimal testing method (Hofer & Green, 1985; Honaker, 1988).

Regarding the issue of computer anxiety, research indicates that there exist few correlates of age and gender with computer-based anxiety and that these are generally state, and not trait, characteristics (Chua, Chen & Wong, 1999). Furthermore, with brief mentoring sessions, computer-phobic individuals become easily and quickly more comfortable with the medium (Wilson, 1999). The bottom line with computer anxiety is not that individuals will score differently on the computer-based tests, but rather individuals who are computer-phobic will avoid taking the tests. This could limit or reduce the applicant pool for employers. Along these lines, a new and interesting area of research is evolving regarding high tech testing and organizational justice. This research focuses on how a testing situation is perceived by an applicant in terms of fairness, procedural ease, and appropriateness of testing method (Bauer et al., 2001; Truxillo, Steiner, & Gilliland, 2004). Many employers, for the time being, may still consider the option to offer an alternative to internet testing for certain jobs in order to not limit the potential applicant pool, even though the cost to administer written tests can be substantially higher than the online version (Mooney, 2002; Richman et al., 1999).

Another criticism, equivalence, falls into two main types: measurement equivalence and relationship equivalence. Measurement equivalence relates to how similar online testing compared to conventional paper-and-pencil testing assesses an underlying trait (Donovan et al., 2000). Alternatively, relational equivalence refers to how similarly the two test versions (online vs. conventional) relate to other variables the same way (Donovan et al., 2000). The following discussion focuses on measurement equivalence, and not relational equivalence, based on the prevalence of research on the former. Research indicates that there is no difference in respondent attitude or perception of tests in CAT settings vs. conventional tests (Richman-Hirsch, Olson-Buchanan, & Drasgow 2000). Measurement equivalence and reliability and validity have also been shown on various predictor scales such as the job descriptive index (Chang, 2005; Donovan et al., 2000; Layne, DeCristoforo, & McGinty, 1999) as well as through meta-analytic methods (Mead et al., 1993). The bottom line regarding equivalence is who designs the test. If it is in a company’s best financial interest, due to cost, to design a test on its own it will likely do so. However, the lower cost option may not always be the option which affords the highest level of equivalence, reliability, or validity.

Another consideration in administering online testing is the implication of using proctored tests vs. unproctored testing. When a test is taken in an unproctored setting, it is generally not possible to verify the test taker’s identity, to know if they received help or assistance, or to count how many times the test content was accessed before completion of the instrument (Tippins et al., 2006, p. 195). Generally, it is recommended that unproctored tests be used in low stakes testing situations only. Possible solutions such as encrypted software and password protection do not seem to alleviate the issues associated with unproctored testing. It is assumed, in unproctored high stakes testing situations, that the respondent may behave unethically and that the content of the test may be compromised. Issues related to cheating and security are not alleviated by simply utilizing an unproctored pencil-and-paper test. The bottom line in determining whether to use proctored or unproctored testing comes down to the financial resources of a company. Ideally, a company would
monitor all participants in testing situations. However, when it is not financially feasible to do so, outsourcing the process or leaving the applicants unproctored are options as well.

The final consideration is how socially desirable responses compare in computerized settings vs. conventional settings. Data are varied, yet ultimately conclusive. Some research finds that computer-based testing increases the propensity for a respondent to give socially desirable answers (Finegan & Allen, 1994; Lautenschlager & Flaherty, 1990; Potosky & Bobko, 1997). However, most research indicates that either there is no difference in social desirability or that computerized tests elicit less socially desirable distortion (Booth-Kewley, Larson, & Miyoshi, 2007; Feigelson & Dwight, 2000; Wilkerson, Nagao, & Martin, 2002). Computer-based testing elicits fewer socially desirable responses when respondents are alone and can backtrack to previous questions (Martin & Nagao, 1989; Richman et al., 1999). Thus, the key consideration is not whether online testing formats elicit fewer socially desirable answers but what testing format and design is used. The concern is that employers will assume that computer-based testing is more efficient at eliminating socially desirable answers when in fact there are other issues more important to consider such as proctoring or test-design.

**FUTURE TRENDS**

Practitioners and academics can benefit from research along the following lines: ethical behavior, relational equivalence, organizational surveys, and the three market-based issues described below.

The topic of ethical behavior in testing situations can benefit from research comparing the similarities and differences between computer-based proctoring vs. human proctoring. For instance, can companies or test vendors conserve time and/or money by utilizing computer-based monitoring systems? In what ways would an individual behave differently if a computer were monitoring their activities and verifying their identity vs. if a human were doing the same job? Answering this question and quantifying how much money, if any, could be saved may prove to be a worthy practical and theoretical line of inquiry.

Much attention has been directed towards measurement equivalence. However, future research should look to verify the relational equivalence of online testing software vs. conventional tests. As Donovan et al. (2000) note, the issue of relational equivalence is an important concept and is one-half of the equation when considering equivalence between online and paper-and-pencil test.

One potential overlooked benefit of online testing software is the ability for companies to gather data internally for employee satisfaction, customer satisfaction, and evaluation of personnel and procedures. Limited research has examined organizational surveys in terms of who responds to internal online surveys, how confidential the surveys are considered, and how ratings differ between online and paper versions (Rosenfeld, Booth-Kewley, & Edwards, 1993; Thompson, Surface, Martin, & Sanders, 2003). Additional research may prove beneficial in answering questions of who responds, and why, so as not to “inadvertently silence important segments of the workforce” (Thompson et al., 2003, p. 226).

Finally, three market-related issues need attention. First, little if any regulation exists in the test vendor marketplace. Vendors often need no proof of the validity or reliability of a test or measure, and sadly, most employers do not know to ask for such evidence. Second, there is the issue of who owns the data. Test vendors use data they collect to provide proof of their testing preparation methods and ease of use of their instruments. However, there are issues of confidentiality and consent related to the use of the data. Third, as demand for online employment testing software grows, so may the prices associated with these services. Increased prices may prohibit smaller firms from
utilizing the more efficient and effective methods. Instead, in this situation small firms may return to less effective and less efficient pencil-and-paper tests, or in a worst case scenario they may utilize in-house instruments, which could be considered inaccurate, at best.

**CONCLUSION**

Advances in technology will enable the further proliferation of online employment testing software. Academics and practitioners must provide consistent assessments regarding the validity, reliability, practical utility, and credibility of online testing formats as these processes and procedures evolve. Though the many benefits of online testing such as lower costs, increased efficiency, and fewer transcription errors may be tempting for employers, these benefits must not be traded for reduced reliability and validity of the testing instrument.

**REFERENCES**


**KEY TERMS**

**Computer Adaptive Tests (Tailored Tests):** Adaptive tests allow a computer to receive a response, score it, and then select a next question (either harder, easier, or similar) based on the respondent’s knowledge or expertise.

**Computer Assisted Testing (CAT):** A setting in which a respondent views a selection instrument, or analysis measure, on a computer screen is considered a computer assisted test.

**Conventional Tests:** Paper-and-pencil tests, or tests not computer-based, are considered conventional tests.

**High Stakes Testing:** A testing situation where the consequences may impact people other than the respondent. Examples of high stakes testing situations include, but are not limited to, testing for selection, hiring or, promotion.

**Low Stakes Testing:** A testing situation where the consequences likely do not impact people other than the respondent. Examples of low stakes testing situations include, but are not limited to, testing for training purposes, continuing education, or self-diagnosis.
**Proctored Testing:** A testing situation in which an individual is monitored by a human test administrator.

**Recruitment Process Outsourcing:** When an employer contracts with a vendor to handle some, or all, of the selection and hiring process including the initial stages of data gathering, screening for applicants, interviewing, doing background checks, hiring, and negotiating salary and/or benefits.

**Un-Proctored Internet Testing (UIT):** A setting in which a respondent is not monitored by a human test administrator when completing an assessment.

*This work was previously published in Encyclopedia of Human Resources Information Systems: Challenges in e-HRM, edited by T. Torres-Coronas & M. Arias-Oliva, pp. 504-509, copyright 2009 by Information Science Reference (an imprint of IGI Global).*