Industrial and Organizational Psychology, **2** (2009), 49–51.
Copyright © 2009 Society for Industrial and Organizational Psychology. 1754-9426/09

Research on Unproctored Internet Testing

BEN-ROY DO
Roosevelt University

This article will focus on two areas of research: measurement equivalence of proctored and unproctored administration of cognitive measures and the effect of cheating through the use of item preknowledge and two-step testing.

Score Equivalence and Measurement Equivalence

Prior research on the score equivalence of Internet testing across proctored and unproctored settings gives us somewhat inconsistent results. One group of researchers found comparable results across proctored and unproctored conditions. Oswald, Carr, and Schmidt (2001) found no differences in internal structure or mean test scores for cognitive ability, although they found differences for noncognitive measures. Other researchers found higher mean scores in unproctored than in proctored conditions (e.g., Beaty, Fallon, Shepherd, & Barrett, 2002; Do, Shepherd, & Drasgow, 2005).

Yet, other researchers found that although there was significant score difference, the proctored condition, not the unproctored condition, has better scores. For instance, Shepherd, Do, and Drasgow (2003) found that proctored participants scored higher on the working with information scale than unproctored participants, with a small effect size. In another study using two-step testing where qualified unproctored participants were invited to participate in a proctored confirmation test, Nye, Do, Drasgow, and Fine (2008) found that participants in the proctored condition scored higher than those in the unproctored condition, with small to moderate effect sizes.

How do we reconcile the inconsistency among those three groups of results? In unproctored Internet tests (UITs), although job applicants may cheat with help from a smart friend or experienced worker to receive higher test scores, they may also be subject to distractions in a nonstandardized environment and thus obtain lower test scores. These factors could be the reason we are seeing inconsistent results from those three groups of studies.

To ensure test results from UIT are comparable to proctored counterparts, it is important to look beyond score equivalence and focus on the psychometric properties of the test and items across test administration modes. Absent random assignment of people to conditions, observed mean score differences may reflect actual sample differences of the groups, as opposed to measurement nonequivalence or the presence of some effect across groups. In addition, it is important to ensure that testing conditions exhibit no bias, such that a test or item does not unfairly favor one group of examinees over another. Measurement equivalence is one way to address these issues.

Correspondence concerning this article should be addressed to Ben-Roy Do. E-mail: benroydo@gmail.com
Address: Department of Psychology, Roosevelt University, 1400 North Roosevelt Boulevard, Schaumburg, IL 60173
Ben-Roy Do, Department of Psychology, Roosevelt University.

https://doi.org/10.1111/j.1754-9434.2008.01107.x Published online by Cambridge University Press

To establish measurement equivalence across proctored and unproctored test administration modes, persons with identical scores on the latent construct would have the same expected raw score or true score at either the item level, the subscale total score level, or both (Drasgow & Kanfer, 1985). By using item response theory, items with differential item functioning (DIF) can be removed from future testing, and the absence of differential test functioning (DTF) indicates that the test is not biased against one population (e.g., unproctored applicants) and the test is indeed measuring equivalently across test administration modes. With many organizations using UIT as a screening tool to remove unqualified candidates from further consideration, it is important to ensure that no test bias is observed in UIT. If little evidence of DIF and DTF is found, there should be less concern that test scores in unproctored settings have been influenced by the administration mode.

In a study conducted by Do et al. (2005), data from 12,175 applicants for entry-level management positions with a retail organization were analyzed. In terms of score equivalence, they found that people in the unproctored group scored higher on the cognitive ability test ($t = -22.12$) with moderate effect size ($d = -.46$). Such mean differences may occur because of the lack of random assignment rather than the effect of the UIT. To establish measurement equivalence, they examined DTF using the DFITD4 (Raju, 1999) program, as well as the effect size measure for DTF (Stark, Chernyshenko, & Drasgow, 2004). They found the effect size was quite small ($d_{DTF} = -.07$). In other words, moderate effect size found through test scores may result because of true group differences, whereas the trivial effect size for DTF indicates that measurement properties of the scale were generally equivalent across conditions.

Effects of Cheating

Prior research suggested that cheating during college ranges from 3% of students (Karlins, Michaels, & Podlogar, 1988) to 87% (McCabe, 1992). Although this wide range may reflect confusion about what cheating is and may be overestimated because of peer pressure or researcher expectations (Chapman, Davis, Toy, & Wright, 2004), it is nonetheless clear that cheating does occur in proctored settings. Some work has already been conducted to understand the characteristics of people who cheat on tests (Whitley, 1998) and why they cheat (Baird, 1980; Smith, Ryan, & Diggins, 1972; Stevens & Stevens, 1987). For instance, students who are high in achievement motivation are more likely to cheat (Eisenberger & Shank, 1985; Johnson, 1981), and lenient attitudes toward cheating are positively related to tolerance of deviance and negatively related to self-restraint (Jensen, Arnett, Feldman, & Cauffman, 2002).

What is less known, however, is the effect of cheating and test performance. An experimental study conducted by Do, Brummel, Chuah, and Drasgow (2006) showed that when research participants were given cheat sheets with correct and incorrect item preknowledge, their scores were significantly different from regular test takers, with moderate effect sizes. However, will similar results be observed when two-step testing is used with unproctored screening and proctored confirmation tests?

Nye et al. (2008) used a perceptual speed test to determine the extent of cheating on an UIT with a parallel version of the test administered in a proctored confirmation test. Had cheating occurred, one would expect to find higher mean in the unproctored condition and some indication of different factor structures. However, after accounting for regression to the mean, they found that participants in the proctored condition scored higher than their unproctored screening test results, probably because of practice effects. It appears that the use of UIT seems appropriate for this type of test.

In other words, cheating may not be very prevalent in UIT. As UIT is mainly used in low-stakes situations, job applicants may not be motivated to spend extra time and effort to cheat. In academic testing, a single test score may have a large influence on course grade or admission decisions. However, in employee

https://doi.org/10.1111/j.1754-9434.2008.01107.x Published online by Cambridge University Press

selection, a multiple hurdle approach is often used such that other assessment tools, such as proctored confirmation tests, job interviews, and background information, are usually considered before making a hiring decision. By using UIT as a screening tool with follow-up selection hurdles, job applicants may be less willing to cheat.

Conclusion

Organizations could benefit from UIT through increased efficiency, availability of new assessment tools, and reduced cost of testing (Chapman & Webster, 2003). However, to ensure the integrity of the tests, examining measurement equivalence and monitoring cheating and aberrant responses should be an ongoing process. Further research on using UIT on different cognitive domains could provide us better insights into how to use UIT as a useful screening tool in low-stakes testing situations.

References

Baird, J. S., Jr. (1980). Current trends in college cheating. *Psychology in the Schools, 17*, 515–522.

Beaty, J. C., Fallon, J. D., Shepherd, W. J., & Barrett, C. (2002, April). *Proctored versus unproctored web-based administration of a cognitive ability test*. Paper presented at the 17th Annual Conference of the Society for Industrial and Organizational Psychology, Toronto, Ontario, Canada.

Chapman, D. S., & Webster, J. (2003). The use of technologies in the recruiting, screening, and selection processes for job candidates. *International Journal of Selection and Assessment, 11*, 113–120.

Chapman, K. J., Davis, R., Toy, D., & Wright, L. (2004). Academic integrity in the business school environment: I'll get by with a little help from my friends. *Journal of Marketing Education, 26*, 236–249.

Do, B.-R., Brummel, B. J., Chuah, S. C., & Drasgow, F. (2006, May). *Item preknowledge on test performance and item confidence*. Paper presented at the 21st Annual Conference of the Society for Industrial and Organizational Psychology, Dallas, TX.

Do, B.-R., Shepherd, W., & Drasgow, F. (2005, April). *Measurement equivalence across proctored versus unproctored testing with job incumbents*. Paper presented at the 20th Annual Conference of the Society for Industrial and Organizational Psychology, Los Angeles.

Drasgow, F., & Kanfer, R. (1985). Equivalence of psychological measurement in heterogeneous populations. *Journal of Applied Psychology, 70*, 662–680.

Eisenberger, R., & Shank, D. M. (1985). Personal work ethic and effort training affect cheating. *Journal of Personality and Social Psychology, 49*, 520–528.

Jensen, L. A., Arnett, J. J., Feldman, S. S., & Cauffman, E. (2002). It's wrong, but everybody does it: Academic dishonesty among high school and college students. *Contemporary Educational Psychology, 27*, 209–228.

Johnson, P. B. (1981). Achievement motivation and success: Does the end justify the means? *Journal of Personality and Social Psychology, 40*, 374–375.

Karlins, M., Michaels, C., & Podlogar, S. (1988). An empirical investigation of actual cheating in a large sample of undergraduates. *Research in Higher Education, 29*, 359–364.

McCabe, D. L. (1992). The influence of situational ethics on cheating among college students. *Sociological Inquiry, 62*, 365–374.

Nye, C., Do, B.-R., Drasgow, F., & Fine, S. (2008). Two-step testing in employee selection: Is score inflation a problem? *International Journal of Selection and Assessment, 16*, 112–120.

Oswald, F. L., Carr, J. Z., & Schmidt, A. M. (2001, April). *The medium and the message: Dual effects of supervision and web-based testing on measurement equivalence for ability testing and personality measures*. Paper presented at the 16th Annual Conference of the Society for Industrial and Organizational Psychology, San Diego, CA.

Raju, N. S. (1999). *DFITD4: A Fortran program for calculating dichotomous DIF/DTF* [Computer program]. Chicago: Illinois Institute of Technology.

Shepherd, W., Do, B.-R., & Drasgow, F. (2003). *Equivalence of proctored versus unproctored online pre-employment assessments*. Paper presented at the 18th Annual Conference of the Society for Industrial and Organizational Psychology, Orlando, FL.

Smith, C. P., Ryan, E. R., & Diggins, D. R. (1972). Moral decision making: Cheating on examinations. *Journal of Personality, 40*, 640–660.

Stark, S., Chernyshenko, S. C., & Drasgow, F. (2004). Examining the effects of differential item (functioning and differential) test functioning on selection decisions: When are statistically significant effects practically important? *Journal of Applied Psychology, 89*, 497–508.

Stevens, G. E. & Stevens, F. W. (1987). Ethical inclinations of tomorrow's managers revisited: How and why students cheat. *Journal of Education for Business, 61*, 24–29.

Whitley, B. E. (1998). Factors associated with cheating among college students: A review. *Research in Higher Education, 39*, 235–274.

https://doi.org/10.1111/j.1754-9434.2008.01107.x Published online by Cambridge University Press