Internet-Based Proctored Assessment: Security and Fairness Issues

Thomas Langenfeld, ACT, Inc.

Abstract: The COVID-19 pandemic has accelerated the shift toward online learning solutions necessitating the need for developing online assessment solutions. Vendors offer online assessment delivery systems with varying security levels designed to minimize unauthorized behaviors. Combating cheating and securing assessment content, however, is not solely the responsibility of the delivery system. Assessment design practices also effectively minimize cheating and protect content. In developing online assessment solutions, organizations also must strive to ensure that all students have the opportunity to test.

Keywords: assessment design, equity, internet-based assessment

The delivery of internet-based cognitive ability assessments presents significant challenges, and testing companies are seeking effective solutions (ACT, Inc., 2020; College Board, 2020). The COVID-19 crisis has accelerated the shift to greater online learning and, subsequently, it has hastened the need for effective Internet testing. In this article, I review the research completed to date, and then apply it to offer guidance for implementing Internet-based testing solutions.

Studies evaluating the effectiveness of Internet-based testing have focused on college course assessments and certification/licensure exams (Berkey & Halfond, 2015; Bloemers, Oud, & van Dam, 2016). Although these contexts are different from large-scale educational assessments (i.e., SAT and ACT), because they assess cognitive abilities and have test taker stakes, their findings provide insights, understandings, and cautions.

Unproctored and Proctored Internet-Based Assessments

Internet-based testing was introduced in test centers in the mid-1990s. Shortly thereafter, schools and certification/licensure programs began exploring the use of different formats to achieve more convenient opportunities (i.e., anytime anywhere testing) (Bartram, 2009). Gradually, three formats emerged: unproctored Internet-based testing (UIT), live proctored testing, and artificial intelligence (AI) proctored testing.

In UIT, test takers receive a login and password, affirm their identity, and attest that they will not participate in unauthorized behaviors. UIT’s advantages include efficiency, convenience, and a high-tech appearance (Arthur, Glaze, Villado, & Taylor, 2010; Gibby, Ispas, McCloy, & Biga, 2009; Tippins, 2009). Despite these features, researchers have raised questions regarding score integrity and content security (Tippins et al., 2006). Studies evaluating UIT for high-stakes cognitive ability tests have generally found that score distributions tend to be higher than in proctored conditions (Arthur et al., 2010; Bloemers, Oud, & van Dam, 2016; Reynolds, Wasko, Sinar, Raymark, & Jones, 2009; Steger, Schroeders, & Gnambs, 2020). Bloemers et al. (2016) found that test takers occasionally violated UIT behavior agreements by receiving assistance from others, searching the Internet, and copying test content. Steger et al. (2020), in their meta-analysis of 49 studies with over 100,000 test takers, found an effect size of 0.20 standard deviation units favoring those testing in an UIT environment. Despite this finding, they found that effect sizes were reduced to near zero when (a) a test had strict time limits, (b) content was not Internet searchable, and (c) a lockdown browser was employed. Despite these safeguards, the overall effect sizes found by Steger et al. (2020) provide reasons for concern. Programs striving to develop an UIT programs need to carefully evaluate their program design. They should follow these evaluations with studies designed to access the extent that the delivery environment influences test scores and subsequent score interpretations.

Due to concerns around score integrity and content security, vendors developed live remote proctoring services. (For a comparison of eight different vendors, see Foster & Layman, 2013.) Remote proctoring vendors offer varying levels of service to (a) verify test taker identity, (b) observe test taker behavior to minimize cheating, and (c) secure test content. Vendors have begun to supplement live proctoring with AI proctoring. In both live and AI proctoring, services range from minimal security to high security. With minimal security, test takers are videoed, and the vendor downloads the video to the testing organization. Top-tier security includes either the proctor or AI bot authenticating test taker identity, providing standardized instructions, checking room conditions, and observing test taker behavior. With live proctoring, the proctor may record irregularities and may have the authority to stop testing. With AI proctoring, two protocols are offered: (a) if the algorithm identifies an irregularity, it flags the testing event and takes a specified action, or (b) after the algorithm identifies an irregularity, a human reviews the video to determine the appropriate follow up action (Foster & Layman,
Needs regular monitoring to ensure accuracy and fairness.

In both protocols, the AI algorithm needs regular monitoring to ensure accuracy and fairness.

Testing organizations need to determine security levels based on test takers’ motivation to cheat and the consequences stemming from cheating or breached security. After a thorough analysis, the organization should select the security level commensurate with the intended use and consequences stemming from test scores.

Assessment Design to Minimize Unauthorized Behaviors

Beyond determining the appropriate level of vendor security, assessment design is equally important for maintaining a fair and secure assessment (Bartram, 2006; Foster, 2009; Weiner & Hurtz, 2017). Studies of remote proctored conditions have found that test takers generally score at the same level as test takers in live proctored conditions (Berkey & Halfond, 2015; Karim, Kaminsky, & Behrend, 2014; Lilley, Meere, & Barker, 2016; Weiner & Hurtz, 2017). This finding is attributed in part to the remote proctored environment, but it is also attributable to design methods that minimize cheating and protect content (Bartram, 2009). Organizations should select from design practices that minimize unauthorized behaviors (Bartram, 2009; Foster, 2009; ITC, 2013; ProctorTrack by Verificient, 2020), and implement a combination of the following practices:

- Offer the test at a single time for all test takers
- Offer the test within a strict time window to reduce the amount of shared information
- Have a timed test, and the time allotment should be relatively strict
- Allow test takers to access the test only once
- Randomly sequence items on the test form
- Provide multiple forms
- Computer adaptive test (CAT) design to minimize the likelihood that two test takers receive the same items
- Do not allow test takers the opportunity to change their responses or return to earlier items once they have proceeded to the next item
- Administer tests using a locked-down browser

The Duolingo test of English proficiency is designed to be taken anytime anywhere (Brenzel & Settles, 2017; LaFlair & Settles, 2019). The test becomes high stakes when it is used by colleges and universities as part of the admissions criteria for foreign students. Duolingo’s multi-faceted security approach utilizes (a) lockdown browser, (b) technology-based test taker authentication, (c) eye tracking technology, (d) CAT design with a deep item pool, and (e) AI proctoring with a 48-hour review period before the release of scores. With these design features, Duolingo is an exemplar for secure Internet-based testing.

Internet-Based Proctored Testing

Test Taker Authentication

Test taker authentication is perhaps the greatest safeguard against cheating. Bloemers et al. (2016) found that cheaters frequently have someone else take the assessment or receive assistance during testing. Authentication refers to the process of ensuring that the person beginning the test—and remaining at the workstation—is the person who is supposed to be there (Foster & Layman, 2013). Vendors offer sophisticated methods to authenticate test taker identity including keystroke analytics, facial recognition, voice recognition, fingerprint reader, and iris reader. At relatively low costs, authenticating test taker identity employs technologies that provide enhanced security that would be expensive, cumbersome, and time consuming in traditional test centers (Berkey & Halfond, 2015; Foster & Layman, 2013). Privacy Interests

Live remote proctoring has been criticized for creating additional test anxiety (Karim et al., 2014; Lilley et al., 2016; Stowell, & Bennett, 2010), violating personal privacy (Karim et al., 2014; Lilley et al., 2016; Weiner & Hurtz, 2017), and leading to test taker withdrawal (Karim et al., 2014). As with all Internet-based transactions, individuals are concerned about the safeguarding of personal data. Remote proctoring privacy issues start with the loss of privacy some individuals feel while being observed and videoed (Foster & Layman, 2013; Karim et al., 2014; Lieberman, 2018; Lilley et al., 2016). Still, privacy issues extend beyond the observation of the test taker. Individuals are anxious about privacy as it relates to the (a) use of personally identifiable information (Association of Test Publishers [ATP], 2020; International Test Commission [ITC], 2013; McPartland, Dyer, Schoenig, & Nicosia, 2020), (b) secure data transfer to a host server (Foster & Layman, 2013; ITC, 2013; McPartland et al., 2020), (c) application of data forensics to evaluate unauthorized behaviors (ITC, 2013; McPartland et al., 2020), and (d) storage of test takers’ data including video (McPartland et al., 2020). Video recordings of test takers are legally defined as personal data, and thereby are covered by data protection laws and regulations (ATP, 2020; Kagan, 2019).

The General Data Protection Regulation (GDPR) requires that institutions and organizations articulate a “legitimate interest” to justify the use of video observation and recording in test taking (ATP, 2020; Kagan, 2019). For high-stakes assessments, testing organizations can articulate the imperative of ensuring the integrity of scores and the consequences of decisions stemming from these scores.

Opportunity to Test

Opportunity to test parallels the concept of opportunity to learn (Boykin & Noguera, 2011). It refers to all students having the opportunity to demonstrate through testing their knowledge, skills, and abilities. As the Internet becomes an important pathway enabling both learning and assessment, quality Internet access is critical for all students. Three components are required to achieve equal opportunity for online learning and assessment: (a) reliable Internet access, (b) suitable electronic devices, and (c) accessible content.

Inequities in Technology Access

The popular media has written extensively about the “digital divide.” The digital divide produces a “homework divide” where the gap between those who have adequate access to technology and those who do not results in student inequities (Moore, & Vitale, 2018; Moore, Vitale, & Stavinoga, 2018). ACT surveyed a sample of students who took the ACT test (N = 7,233) to learn about their access to technology. Moore et al. (2018) found that nearly 15% of the students taking the ACT rated their home Internet service as either unpredictable or terrible. This finding demonstrates that,
regarding reliable Internet access, a significant portion of students are disadvantaged (Moore et al., 2018).

ACT also surveyed students to identify the number of electronic devices that they may access at home. One percent of students reported that they had access to zero home devices; 14% of students reported that they had access to one home device (Moore et al., 2018). Of the students reporting that they had access to one device, 56% reported that it was a smartphone. Due to inadequacies in Internet connectivity and suitable devices, as organizations move to Internet-based testing solutions, they also need to develop strategies to assist underserved students in obtaining reliable connectivity and suitable testing devices. Assessment organizations are not equipped to address this challenge single-handedly; however, they should look to partner with other stakeholders (i.e., educational institutions, grant-making agencies) to address students’ opportunity to learn and test.

The crisis created by the COVID-19 pandemic has followed a second crisis regarding societal inequities and the need for greater social justice. The calls for greater social justice will not be answered educationally if the digital divide continues to exist resulting in large numbers of students struggling to access online learning and assessment materials. In combination, the pandemic coupled with the need for greater social justice provide policy makers, educational institutions, and assessment organizations the opportunity to re-envision the educational and learning environment.

Accessible Test Content
Test administrations have endeavored to develop accessible test content for traditional in-person administrations. In this new environment, they need to strive to achieve accessibility in remote proctored environments. Many policies and accommodations used to provide accessible content in traditional administrations apply directly to Internet-based testing. Undoubtedly, Internet-based testing will present special circumstances that must be addressed and accommodated. Administrators must carefully analyze and review various accessibility needs and determine effective strategies for implementing them into their solutions.

Concluding Thoughts
The COVID-19 crisis will accelerate changes in learning and assessment. Educational learning and assessment will emerge from the pandemic having explored and used different technologies and platforms. Several of the technologies and platforms that emerged during the pandemic will take on a more prominent role in the post-pandemic landscape. Because test takers and test users are attracted to the convenience of anytime anywhere testing, proctored Internet-based proctored assessment will likely become a greater part of the educational experience.

General findings currently support the use of live and AI remote proctoring in that they minimize cheating, secure test content, and provide comparable score distributions (Karim et al., 2014; LaFlair & Settles, 2019; Weiner & Hurtz, 2017). Top-tier security levels are costly, and testing organizations need to seek security levels commensurate with the use and consequences stemming from test scores. Test design practices further aid in protecting score integrity and test content (Bartram, 2006; Foster, 2009; ITC, 2013; ProctorTrack by Veriforce, 2020). Throughout the process of developing and implementing a solution, testing organizations need to carefully evaluate privacy issues so as to develop policies that protect and ensure test taker personally identifiable information and data.

Equity or opportunity to test is a concern that organizations need to address. Calls for greater social justice will not be answered if educational institutions and assessment organizations do not step up to find solutions to inequities stemming from the current digital divide. The lack of reliable Internet connectivity and suitable electronic devices affect a significant proportion of students (approximately 15% of the students who normally take the ACT test) resulting in these students being placed at a major disadvantage with any Internet-based testing solution (Moore et al., 2018). As testing organizations devise solutions, they must also look to partner with other stakeholders to devise strategies that afford all students the opportunity for convenient Internet-based learning and testing.

References
ACT, Inc. (2020). ACT technical guide for online testing. Retrieved from https://www.act.org/content/dam/act/unsecured/documents/TechnicalGuidefortheACTTakenOnline.pdf.
Arthur, W., Glaze, R. M., Villado, A. J., & Taylor, J. E. (2010). The magnitude and extent of cheating and response distortion effects on unproctored internet-based tests of cognitive ability and personality. International Journal of Selection and Assessment, 18(1), 1–16.
Association of Test Publishers (ATP). (2020). Testing guidelines: Guidelines for computer-based testing. Retrieved from https://www.testpublishers.org/assets/documents/CBTGGuidelines.pdf.
Bartram, D. (2006). Testing on the internet: Issues, challenges and opportunities in the field of occupational assessment. In D. Bartram & R. Hambleton (Eds.), Computer-based testing and the internet: Issues and advances (pp. 19–37). New York, NY: Wiley & Sons.
Bartram, D. (2009). Commentaries: The international test commission guidelines on computer-based and internet-delivered testing. Industrial and Organizational Psychology, 2, 11–13.
Brenzel, J. & Settles, B. (2017). The Duolingo English test—Design, validity, and value. Duolingo Whitepaper. Retrieved from https://englishtest.duolingo.com/research.
Berkey, D., & Halfond, J. (2015, July 20). Cheating, student authentication and proctoring in online programs. New England Journal of Higher Education. Retrieved from https://nejhe.org/journal/cheating-student-authentication-and-proctoring-in-online-programs/.
Bloomers, W., Oud, A., & van Dam, K. (2016). Cheating on unproctored Internet intelligence tests: Strategies and effects. Personal Assessment and Decision, 2, 21–29.
Boykin, A. W., & Noguera, P. (2011). Creating the opportunity to learn: Moving from research to practice to close the achievement gap. Arlington, VA: ASCD.
College Board. (2020). 2019–2020 AP exams online: What you need to know. Retrieved from https://blog.collegeboard.org/2019-20-ap-exams-online-what-you-need-know.
Foster, D. (2009). Secure, online, high-stakes testing: Science fiction or business reality? Industrial and Organizational Psychology, 2(1), 31–34.
Foster, D. & Layman, H. (2013). Online proctoring systems compared. Retrieved from https://www.caveon.com/wp-content/uploads/2014/03/Online-Proctoring-Systems-Compared-Mar-12-2014.pdf.
Gibby, R. E., Ispas, D., McCloy, R. A., & Biga, A. (2009). Moving beyond the challenge to make unproctored internet testing a reality. Industrial and Organizational Psychology, 2, 64–68.
International Test Commission (ITC). (2010). ITC guidelines on computer-based and internet delivered testing. Retrieved from https://www.intestcom.org/files/guideline_computer_based_testing.pdf.
