A Comparative Study of Online Remote Proctored versus Onsite Proctored High-Stakes Exams

John A. Weiner and Gregory M. Hurtz

☑ **Finalist for the Best Research Category**

☑ **Best Research sponsored by ACER UK**

Advances in technology have spurred innovations in secure assessment delivery. One such innovation, remote online proctoring, has become increasingly sophisticated and is gaining wider consideration for high-stakes testing, but many organizations still need convincing of its comparability to onsite proctoring. However, there is an absence of published research examining remote online proctoring and its effects on test scores and the examinee experience.

This paper describes a quasi-experimental field study carried out with three professional licensing examinations administered concurrently at different test sites that offered either onsite proctoring in testing centers or remote online proctoring in computer kiosks where the testing was proctored via Internet-connected video communication and surveillance.

Results using both classical test theory and item response theory methods revealed substantial reliability and a strong degree of measurement equivalence across proctoring conditions. Candidates revealed slightly less positive reactions to some of the remote proctored testing conditions, but reactions were positive overall and had virtually no relation to test performance. Overall, the results of this study support the equivalence of kiosk-based remote online proctored exams and exams proctored onsite in test centers.

Learn more at **psionline.com**

Study

In high-stakes testing, such as professional licensing and certification, security is of paramount importance, to protect test content, ensure authenticity of test-taker identity, and maintain validity and equivalent interpretation of scores. Consequently, secure test delivery has traditionally necessitated an onsite proctor to address these concerns.

Recently, there has been widespread interest in remote online proctoring. However, enthusiasm has been tempered by concerns over the extent to which remote proctoring can provide comparable results to test center delivery. This is particularly apposite to high stakes testing, where the consequences of an erroneous license or certification to practice could result in harm to the very public which the testing programs are designed to protect.[1]

Organizations need evidence-based research data to assess the efficacy of remote proctoring for their own needs, but a report published in 2015 by the Institute for Credentialing Excellence (Plaus, Boren, Brazell, Wickett, & Weber, 2015) characterized the state of research on remote online testing as severely lacking in volume and rigor as well as in relevance for high-stakes environments.

It was against this background that the authors set out to evaluate the comparability of remote versus onsite proctoring for high stakes delivery with regard to impact on scores, candidate experience and any correlation between candidate perception and test scores. By doing so, they could enable organizations considering remote proctoring, to take more informed decisions and therefore reduce barriers to adoption for this enabling technology.[2]

There was a valuable body of evidence addressing adjacent issues, including research on:

- Cheating in online assessments and the potential benefits of safeguards such as remote proctoring used in both education (e.g., Berkey & Halfond, 2015) and employment testing contexts (Tippins et al., 2006).

- The usability of and user reactions to remote proctoring systems (most recently Lilley, Meere, & Barker, 2016), however these used samples comprising distance learners and research participants, not candidates sitting for high-stakes tests.

- The impact of proctoring on the equivalence of assessments, but which compared proctored with unproctored delivery rather than different proctoring models.[3]

---

1   *Is there a genuine need for this research? How important is this need to the stakeholder community?*

2   *Does the research project encompass a recognized area of assessment that could be changed or reinforced by the outcome of the research findings? How will it achieve this?*

3   *Does the research materially contribute to the research body of knowledge? To what extent can the research be said to address areas not previously investigated? Particular weight will be given to the demonstration of original thinking.*

But the clear need was for evaluation research on the comparability of on-site and remote proctoring. Consequently, the aims were framed in terms of three key questions:

1. Do scores obtained with online remote proctoring exhibit sound psychometric properties, equivalent to scores obtained from the same tests administered at traditional test centers with onsite proctoring?

2. Do candidates' perceptions of general features of the testing conditions differ across online remote versus onsite proctoring conditions?

3. To what degree are candidate perceptions of the testing conditions related to variability in test scores? [4]

The study sample comprised candidates for three professional licensing exams in one US state over a 1-year period. These were multiple-choice licensure exams, developed according to professional standards (AERA et al., 2014) for content validity.

Cognizant of the broad continuum in remote proctoring approaches, a uniform model of delivery was maintained, using specially equipped testing kiosks offering a consistent candidate experience with enhanced webcams and screen recorders for greater security. These were located in colleges, testing labs, offices and community centers. Onsite proctoring was delivered in established test center locations.

Before receiving their results, candidates completed a survey to rate the testing conditions on a 4-point scale (ranging from 1=poor to 4=excellent) addressing conditions such as the testing system, noise, temperature, lighting, and staff.

The results of the study indicated that candidate scores with online proctoring were psychometrically sound and comparable to the same exams administered with onsite proctoring. Candidates rated online kiosk-proctored exams favorably and ratings of testing conditions were virtually uncorrelated with exam performance, suggesting that candidates' experiences under the different testing conditions did not affect their performance. Overall, the results of the study supported the kiosk-based remote proctoring method for high-stakes test delivery, equivalent to onsite test center proctoring.

Building on this the authors believe an additional avenue for continued research is to explore the generalisability of the current findings across different examination programs and across different remote proctoring systems and protocols. [5]

---

4 *Were the aims of the research identified clearly at the outset? What were they? Was there a clear hypothesis to prove or disprove?*

5 *To what extent has the outcome of the research influenced improvements in the design, development, promotion or execution of assessments? Or has the research supported and even reinforced an existing approach?, i.e. where the research instead has perhaps laid a foundation for exploration for improvement elsewhere.*
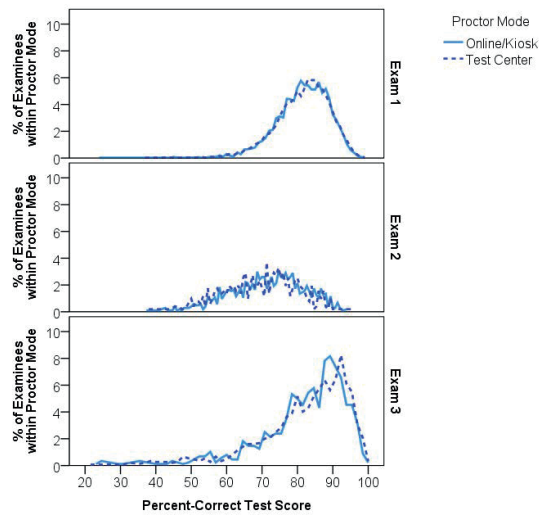
Figure 1. Test score distributions by proctoring mode condition.

The researcher team were keenly aware that these were live tests which could influence a candidate's career. Extreme care was taken throughout to ensure no individual candidate was adversely affected by participation, nor was there any impact on the test security or integrity of results. There were no external sponsors of this research. [6]

In terms of methodology, two key features of the approach were significant sample size and delivery as a single-blind study. The sample size was 14,623 cases across the three exams. Candidates were not told of the proctoring arrangements at their chosen location when booking, or at any time prior to arrival for their examination. Thus, the possibility of bias stemming from candidates acting on a preference for one arrangement over another was minimal; they would arrive on location and take their test using whichever of the two proctoring regimes was in place. [7]
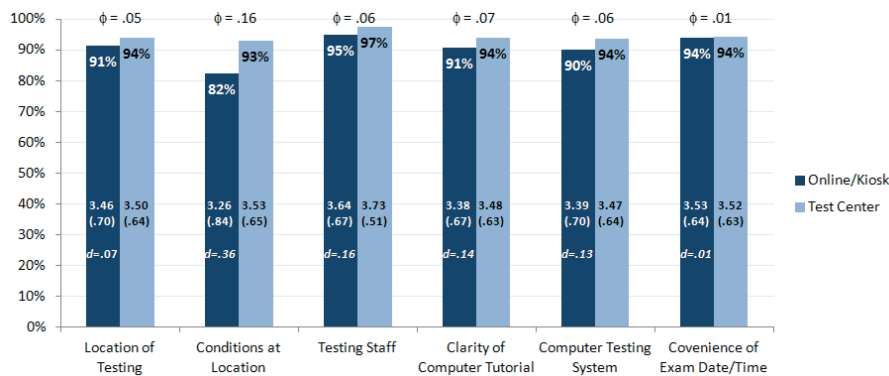
Figure 2. Percent of examinees rating testing conditions as good or excellent, by proctor type.

6   *What approach and good ethical practice was demonstrated by this research project? e.g. data security, confidentiality and privacy, conflicts of interest.*

7   *In what way was the methodology chosen appropriate for this research project and its stakeholders? Where there any problems with your choice of methodology which you had to address, or conversely particular benefits derived from the methodology chosen?*