Journal of Applied Testing Technology, *Vol 18(1), 13-20, 2017*

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

John A. Weiner* **and Gregory M. Hurtz**

PSI Services LLC, 611 N. Brand Blvd. 10th Floor, Glendale, CA 91203; john@psionline.com, ghurtz@psionline.com

Abstract

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. 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.

Keywords: Remote Proctoring, Remote Online Proctoring, Remote Invigilation, Equivalence of Proctoring Conditions, Test Security, Preventing Online Test Fraud, High-Stakes Testing, Computer-based Testing, Internet-based Testing

1. Introduction

In high-stakes professional testing environments, security is of paramount importance to protect test content, ensure authenticity of test-taker identity, and maintain validity and equivalent interpretation of scores and subsequent decisions. To this end, secure test delivery has traditionally necessitated an onsite proctor to attend to these concerns.

Recently, the practice of using remote online proctors has emerged on the high-stakes testing landscape. An already common practice in the distance education realm, surging interest in its use for professional licensing and certification testing prompted the issuance of a special report on the topic by the Research & Development Committee of the Institute for Credentialing Excellence (Plaus, Boren, Brazell, Wickett, & Weber, 2015). The report 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 such as found in certification and licensing testing. This highlights a critical need for research evaluating the comparability of remote proctoring systems to the traditional onsite test center proctoring model for high stakes examinations.

A focus on remote proctoring calls for differentiation among types of systems, as they vary on a continuum of approaches and security safeguards. Toward one end of the continuum is a system where test content is delivered to test-takers using their own computers, anytime/anywhere, and a video recording is captured via web camera without live supervision (e.g., record and review later). As an added level of security, live proctors in a remote location may interact with the examinee and observe the scheduled testing session in real time via video technology. Further toward the other end of the continuum is a system that uses real time remote proctors and, rather than test-takers using their own computers at their own location, the test is taken on specially equipped and configured computer kiosks that provide a standard test-taking experience, and are set up with equipment such as enhanced webcams and screen recorders for greater security. The kiosks may be located in any accessible and distraction-free location, such as libraries, office buildings, and community centers. The current study focuses on the latter scenario (live remote proctoring at kiosks) in a high-stakes testing environment.

* *Author for correspondence*

2. Literature Review

While there are essentially no published studies of online remote vs. onsite proctored high stakes exams, there is a noteworthy body of research addressing adjacent issues that can be organized into three categories. First is a number of recent articles and commentaries generally discussing research on cheating in online assessments and the potential benefits of safeguards such as remote proctoring and related practices that are used in both education (e.g., Berkey & Halfond, 2015; Cluskey, Ehlen, & Raiborn, 2011; Dunn, Meine, & McCarley, 2010; Tomasi, Figiel, & Widener, 2009) and employment testing contexts (Tippins et al., 2006). These articles raise relevant questions regarding test-taker cheating and other inappropriate behavior, how to combat such behavior, and whether remote proctoring provides an effective solution.

Second is research on usability and user reactions to remote proctoring systems (Bedford, Gregg, & Clinton, 2009, 2011; Karim, Kaminsky, & Behrend, 2014; Lilley, Meere, & Barker, 2016). This research examines the ability for users to launch the testing systems and take the test seamlessly without onsite assistance as well as more general reactions to the remotely-proctored testing environment. This small body of work has suggested that test-takers tended to perceive slightly more pressure and tension, and expressed some elevated concerns over privacy, under remotely proctored conditions relative to unproctored conditions (Karim et al., 2014). However, they did not tend to perceive that the testing conditions influenced their performance or the assessment experience (Karim et al., 2014; Lilley et al., 2016). Test-takers reported finding the systems easy to use, believed they were useful for reducing cheating (Bedford et al., 2009, 2011), and were supportive of the systems being more widely used (Bedford et al., 2009, 2011; Lilley et al., 2016). To the extent that these studies can be generalized, they suggest that test-takers do not react negatively to remote online proctoring. However, these samples were distance-learners and research participants, not candidates sitting for high-stakes certification or licensure tests.

Third is research exploring the impact of proctoring types on the equivalence of assessments. To date, all available research compares proctored to unproctored conditions and leaves unexplored comparisons among types of proctoring (e.g., in-person versus remote). Additionally, while research in educational and experimental contexts suggests that unproctored testing often leads to either score differences or item response patterns that are suggestive of a higher level of incidence of cheating than often found in proctored conditions (Brothen & Peterson, 2012; Karim et al., 2014; Prince, Fulton, & Garsombke, 2009; Wright, Meade, & Gutierrez, 2014), there is a growing body of research in an employment context which finds that scores obtained on completely unproctored internet tests do not differ appreciably from those taken under proctored test conditions (Arthur, Glaze, Villado, & Taylor, 2010; Kantrowitz & Dainis, 2014; Lievens & Burke, 2011; Nye, Do, Drasgow, & Fine, 2008). Though the latter finding may be reassuring for employee selection testing, it has little bearing on licensure and certification contexts where regulatory guidelines require assurance of measures taken to minimize inappropriate test-taker behaviors such as cheating (e.g., AERA, APA, NCME, 2014). In this setting, the consequences of an erroneous license or certification to practice could result in harm to the public which the testing programs are designed to protect.

3. Present Study

Given the lack of evaluation research on the comparability of results for remotely proctored exams versus traditional onsite proctored exams, the purpose of the current study is to begin to address this void by comparing kiosk-based remote online proctored examinations to the same examinations administered in test centers with onsite proctors. Three research questions guided the study: (1) Do scores obtained at kiosks 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 examinees' perceptions of general features of the testing conditions differ across online remote vs. onsite proctoring conditions? (3) To what degree are examinee perceptions of the testing conditions related to variability in test scores? The evaluation was replicated across three examinations in distinct professional licensure programs.

4. Methods

4.1 Participants

The study sample was comprised of examinees for three professional licensing exams in a mid-western state in the U.S. Test response data were extracted for all first-time examinees that were tested during a 1-year period at either a traditional test center with an onsite proctor, or at a testing kiosk with an online remote proctor. As seen in Table 1, the total sample included 14,623 cases across the three examinations. Gender and age of the test-takers sitting for each exam are also summarized in Table l, along with a tabulation of examinees tested via remotely proctored online/kiosks and onsite proctored test centers.

4.2 Materials and Apparatus

All examinations were multiple-choice licensure exams that had been developed according to professional standards (AERA et al., 2014) for content validity, reliability, and psychometric quality. They were administered via computer, at either a secure test center supervised by onsite proctors (Test Center), or at a remote kiosk computer supervised online by a live remote proctor using video communication and surveillance (Online/Kiosk).The kiosks were located in college libraries and testing labs, office buildings, and community centers; typical placement was in an enclosed room. Examinees also completed a post-examination survey to rate the testing conditions on a 4-point scale (ranging from 1=*poor* to 4=*excellent*). The ratings addressed conditions such as the testing system, noise, temperature, lighting, and testing staff.

Table 1. Characteristics of the research subjects

Exam	No. Examinees		
	Online Kiosk	Test Center	Total
Exam 1	4971	5728	10699
Exam 2	1036	515	1551
Exam 3	882	1491	2373
Total	**6889**	**7734**	**14623**
Gender			
Female	1889	4696	6585
Male	689	680	1369
Not Reported	4311	2358	6669
Age	Descriptive Statistics		
Mean	27.36	28.63	28.03
SD	8.60	8.92	8.80

4.3 Procedure

Examinees completed the tests under one of two conditions: Test Center, or Online/Kiosk. Test-takers selected their sessions in an online scheduling process by choosing a date and geographic location. They were blind to the testing conditions under investigation for this study, in that they were not told of the proctoring arrangement at their selected location when they scheduled their test sessions or at any time prior to arrival for their examination. Thus, the possibility of bias stemming from candidates choosing one condition over another based on personal preference was minimal. At both kiosk and test-center locations, candidates were required to verify their identities by presenting a valid form of identification, which was also captured digitally at the kiosk. At the end of their examination, and before receiving their results, examinees completed the survey of their testing environment.

5. Results

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

Test score distributions for the three exams are displayed graphically in Figure 1 and psychometric properties are summarized in Table 2, separately for test center versus online/kiosk proctoring conditions. Overall, Figure 1 shows that the degree of overlap in score distributions was high and the patterns were not visually distinguishable. Table 2 reveals an acceptable level of reliability (KR-21) of test scores under the kiosk and test center conditions. Moreover, the standard deviations, reliabilities, and standard errors of measurement were nearly identical between proctoring conditions, and differences in means were essentially trivial and not systematically higher or lower for either condition. In the case of Exam 2, the difference in means (107.31 vs. 104.85) was slightly higher than was found for the other two exams, but the difference of 2.46 points was less than one-half (.45) the size of the pooled SEM across conditions (5.42), and the effect size *d* of .16 was below the common standard for what constitutes a "small" effect. The equivalent means, standard deviations, and reliabilities of scores across conditions aligns with the

parallel test model (Lord & Novick, 1968, pp. 48-49; Allen & Yen, 1979, pp. 59-60) and suggest that the tests were equivalent across conditions.

Further analysis of equivalence between the two proctoring conditions was carried out using the 2-parameter Item Response Theory (IRT) model estimated in BilogMG. Each exam was first calibrated in the group with the larger sample size, then all items were re-calibrated separately in the group with the smaller sample size while instructing BilogMG to set the mean and standard deviation of its latent ability distribution equal to the location and scale values taken from the larger group. By establishing a common latent scale across the groups, their item parameters are equated to a common scale and become directly comparable. The resulting test characteristic curves and conditional standard errors of measurement were then compared across conditions.

As seen in Figure 2, expected true scores and standard errors were virtually identical for Exams 1 and 2, and for Exam 3 were virtually identical in the mid to high ability regions. Of most importance is that for all three exams the proctoring conditions yielded virtually indistinguishable expected true scores and standard errors near the cut scores, where pass/fail decisions are made. This analysis further supports the equivalence of the Online/Kiosk and Test Center conditions. Note that all of the above analyses and plots in Figure 2 were also replicated using the 1-parameter IRT model with nearly identical results. The 3-parameter model was also attempted but the solution would not converge satisfactorily in all cases.

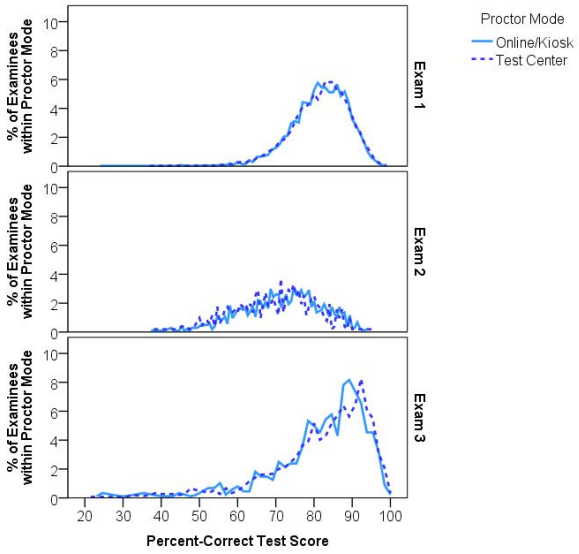
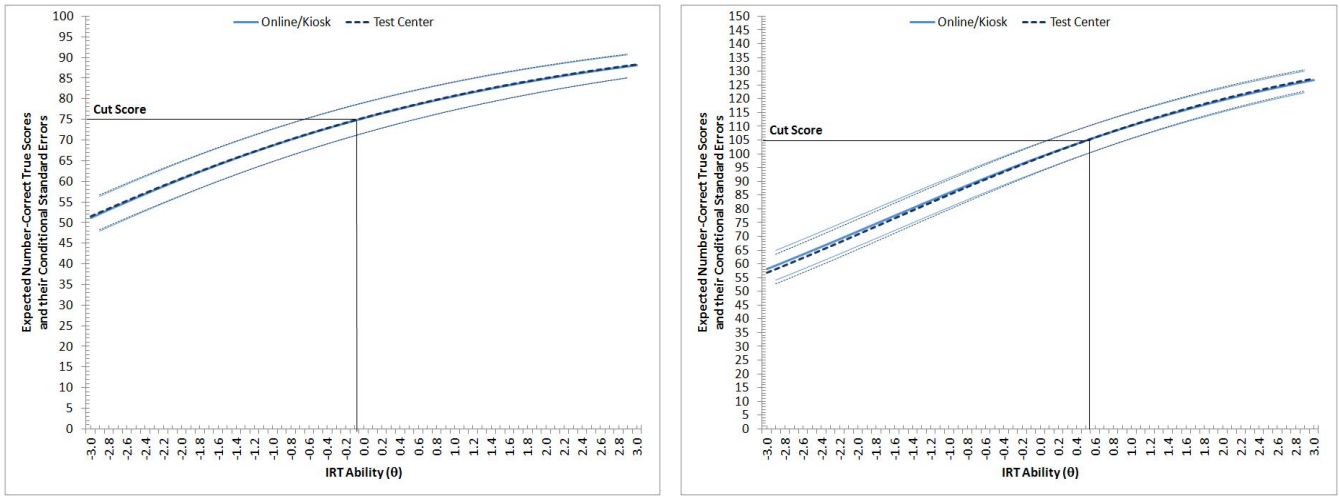


Figure 1. Test score distributions by proctoring mode condition.

Table 2. Psychometric properties of exams by condition

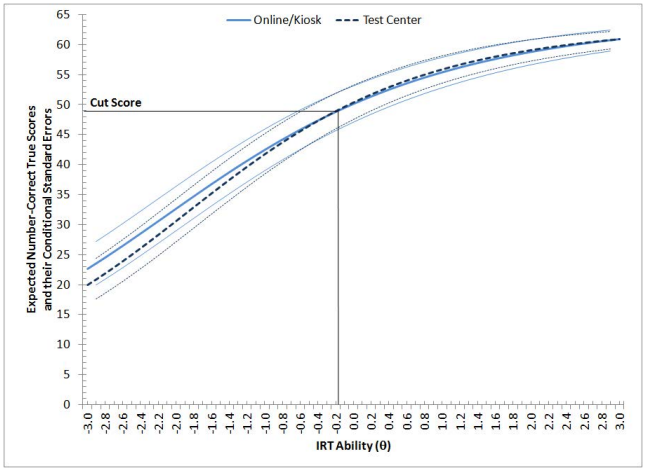
	k	N	M	SD	Mean p	Mean r_{pb}	KR-21	SEM
Exam 1								
Online/Kiosk	100	4971	81.47	7.57	0.81	0.22	0.74	3.83
Test Center	100	5728	81.58	7.50	0.81	0.21	0.74	3.82
d-value			-0.02	0.02				
Exam 2								
Online/Kiosk	150	1036	107.31	15.64	0.70	0.23	0.88	5.39
Test Center	150	515	104.85	15.48	0.69	0.23	0.87	5.49
d-value			0.16	0.06				
Exam 3								
Online/Kiosk	65	882	53.33	7.88	0.82	0.34	0.86	2.96
Test Center	65	1491	53.35	8.35	0.82	0.35	0.88	2.94
d-value			-0.00	0.04				

Note: *k* = number of items; N = number of test-takers; *p* = item difficulty; r_{pb} = point-biserial item-total correlation adjusted for overlap; *d* = Cohen's *d* effect size.



a. Exam 1

b. Exam 2



c. Exam 3

Figure 2. Expected true scores and conditional standard errors from the 2-parameter IRT model, by proctoring condition.

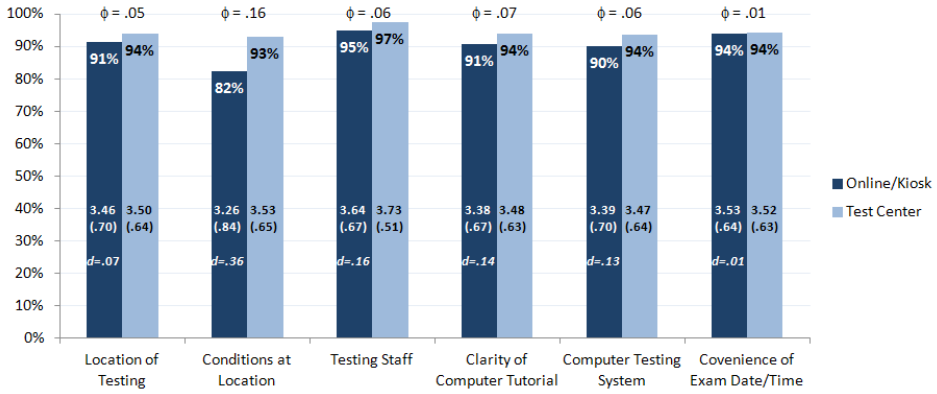


Figure 3. Percent of examinees rating testing conditions as good or excellent, by proctor type.

Research Question 2: Do examinees' perceptions of general features of the testing conditions differ across online remote vs. onsite proctoring conditions?

Examinee ratings of the testing conditions aggregated across the three exams (N = 11,699 to 11,623) are summarized in Figure 3. The bars in the figure represent the percentage of test-takers who chose "good" or "excellent" (versus "average" or "poor") on the 4-point response scale, and phi coefficients are placed atop each bar to index the strength of association between proctoring type and these dichotomized ratings. Means and standard deviations on the original scale are also provided on the bars as well as Cohen's *d* effect size values to evaluate standardized mean differences. The overall pattern in the figure reveals a high percentage of examinees rating testing conditions as good or excellent across the different survey items and high average ratings on the 4-point scale, although there was a tendency for test centers to fall slightly higher on both metrics. Most phi coefficients and *d* values were trivial, while the largest difference, with a Φ of .16 and *d* of .36, was in ratings of conditions at the location (noise, temperature, distractions). For this item, 93% rated the test centers as good or excellent while 82% rated the online/kiosk conditions as good or excellent. Overall, the vast majority of examinees expressed a high degree of satisfaction with both test center and online/kiosk proctoring, with means ranging from 3.26 to 3.73 on the 4-point scale across items and conditions.

Research Question 3: To what degree are examinee perceptions of the testing conditions related to variability in test scores?

A multiple regression analysis was conducted to examine the associations between percent-correct scores on the exams, proctoring type, and examinees' ratings of the testing conditions, controlling for exam difficulty differences. As seen in Table 3, examinee perceptions of the testing conditions were essentially unrelated to their test scores, after accounting for average performance differences between exams (R^2 = .16) and proctor type (R^2 = .00). Adding examinee perceptions of the testing conditions to the model revealed trivial associations with examination performance (R^2=.01). In addition, interactions between proctor type and examinee perceptions were null (R^2=.00). These results indicate that the proctoring and test delivery conditions experienced by candidates, as reflected by their ratings, were unrelated to their performance on the exams. That is, there was no evidence of impact on test scores due to remote proctoring.

Table 3. Regression of Examinee Performance (Percent-correct Scores) on Ratings of Testing Conditions, Controlling for Exam Difficulty Differences and Proctor Type

Predictor	ΔR^2	Beta	r
Exam 2 (Dummy)	.16	-.40	-.40
Exam 3 (Dummy)		.02	.09
Proctor Type (0 = Test Center; 1 = Online/Kiosk)	.00	.01	-.05
Location of testing	.01	-.03	.05
Conditions at the location		-.02	.06
Test center staff		.07	.08
Clarity of computer tutorial		-.05	.06
Computer testing system		.06	.14
Convenience of the exam date and time		.05	.11
Interaction Block (all Proctor Type x Rating products)	.00		

Note: *N* = 11,446; Due to extremely high power all *r*'s were statistically significant at $p<.001$, as were all Beta values except those for Location of Testing (p = .013), Conditions at the Location (p = .054), and Proctor Type (p = .520)

6. Discussion

The purpose of the current study was to compare online remote proctoring of high-stakes licensure examinations to traditional test center proctoring of the same examinations to assess the equivalence of, or potential differences between, these alternative test administration conditions. The results of the study indicated that examinee scores obtained under online kiosk-based proctoring conditions were psychometrically sound and comparable to the same exams administered under traditional test center proctoring conditions. Further, examinees rated online kiosk-proctored exams favorably and ratings of testing conditions were virtually uncorrelated with exam performance, suggesting that examinees' 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.

This study helps address the call for research on remote proctoring that is directly relevant to high-stakes examinations, particularly in the credentialing and licensure realm (Plaus et al., 2015), as opposed to past research which has been outside this realm and has not directly compared remote proctoring to traditional in-person proctoring. Of particular relevance to high-stakes testing is the fact that the kiosk-based remote proctoring setup mimics in many ways the traditional test center environment, if the kiosk is placed in a sufficiently controlled setting. This allows for greater flexibility in test administration while not sacrificing psychometric quality, as indicated in the results of this study. However, it should be recognized that the current supportive evidence relates to a specific remote proctoring approach that utilizes standard technology and real-time proctors, which may be quite different from other remote proctoring scenarios.

An additional avenue for continued research, then, is to explore the generalizability of the current findings across different examination programs and across different remote proctoring systems and protocols. For example, are there differences between the online remote proctoring kiosk approach in the current study and other remote proctoring systems and approaches, with respect to score equivalence and candidate perceptions? Alternative methods that use the examinees' own computer hardware and video cameras, or record-and-review later systems may yield different results. Moreover, the "remote proctoring" category should be further differentiated in the design of future studies, as all remote proctoring systems and strategies should not be assumed to be equal in their effectiveness in addressing guidelines such as those identified in the Institute for Credentialing Excellence report (Plaus et al., 2015). This is an important avenue for future research to explore.

A strength of this study was the simultaneous administration of the same examinations across the different proctoring modalities with efforts to prevent selection bias by keeping test-takers blind to the proctoring conditions at the time they signed up for their examination. While there is some possibility that across the year of test administrations participants had gained knowledge that certain test-centers were remotely-proctored, it is unclear whether this occurred and how such knowledge might have impacted results. In future research, a post-administration questionnaire may include a question asking whether examinees knew about the proctoring conditions prior to examination day.

Additional development of a post-administration questionnaire might look to the voluminous body of research on applicant reactions to employee selection procedures (e.g., Bauer et al., 2001; Bauer et al, 2006; Gilliland, 1993, 1994; Karim et al, 2014), in order to more fully explore differences in test-takers' reactions and perceptions across proctoring modalities. Models of test attitudes and motivation (e.g., Arvey, Strickland, Drauden, and Martin, 1990) may provide a useful framework for future comparisons of examinee reactions to proctoring conditions.

With respect to the kiosk-based online remote proctoring system investigated in the current study, our results support its psychometric quality by demonstrating that examinations administered under this remote proctoring system were reliable and equivalent to the same examinations administered under a more traditional proctoring environment at a certified test center. Any potential differential effects of security for online/kiosk vs. onsite proctoring were not revealed in the analyses. These findings help address the critical need for such research evidence and provide important information to advance the use of technology in administering high-stakes examinations.

7. Conflict of Interest Disclosure

There were no external sponsors of this research.

8. Acknowledgements

The authors would like to thank Nelia Vazquez of PSI Services LLC for her assistance with data extraction and processing for the analyses carried out in this study.

9. References

AERA, APA, NCME (2014). Standards for educational and psychological testing. Washington DC: American Educational Research Association.

Allen, M. J., & Yen, W. M. (1979). *Introduction to measurement theory*. Monterey: Brooks-Cole.

Arthur, W. A, Jr., 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,* 1–16.

Arvey, R. D., Strickland, W., Drauden, G., & Martin, C. (1990). Motivational components of test taking. *Personnel Psychology, 43,* 695–716.

Bauer, T. N., Truxillo, D. M., Sanchez, R. J., Craig, J. M., Ferrara, P., & Campion, M. A. (2001). Applicant reactions to selection: Development of the selection procedural justice scale (SPJS). *Personnel Psychology, 54*, 387–419.

Bauer, T. M., Truxillo, D. M., Tucker, J. S., Weathers, V., Bertolino, M., Erdogan, B., & Campion, M. A. (2006). Selection in the information age: The impact of privacy concerns and computer experience on applicant reactions. *Journal of Management, 32,* 601–621.

Bedford, W., Gregg, J., & Clinton, S. (2009). Implementing technology to prevent online cheating: A case study at a small southern regional university (SSRU). *MERLOT Journal of Online Learning and Teaching, 5*, 230–238.

Bedford, D. W., Gregg, J. R., & Clinton, M. S. (2011). Preventing online cheating with technology: A pilot study of remote proctor and an update of its use. *Journal of Higher Education Theory and Practice, 11*, 41–58. Retrieved August 11, 2016 from http://www.na-businesspress.com/jhetpopen.html

Berkey, D., & Halfond, J. (2015, Jul 20). Cheating, student authentication and proctoring in online programs. *New England Journal of Higher Education*.

Brothen, T., & Peterson, G. (2012). Online exam cheating: A natural experiment. *International Journal of Instructional Technology and Distance Learning, 9(2)*, 15–20.

Cluskey, G. R., Jr, Ehlen, C. R., & Raiborn, M. H. (2011). Thwarting online exam cheating without proctor supervision. *Journal of Academic and Business Ethics, 4*. Retrieved from http://www.aabri.com/jabe.html.

Dunn, T. P., Meine, M. F., & McCarley, J. (2010). The remote proctor: An innovative technological solution for online course integrity. *The International Journal of Technology, Knowledge and Society, 1,* 1–7.

Gilliland, S. W. (1993). The perceived fairness of selection systems: An organizational justice perspective. *Academy of Management Review, 18*, 694–734.

Gilliland, S. W. (1994). Effects of procedural and distributive justice on reactions to a selection system. *Journal of Applied Psychology, 79,* 691–701.

Kantrowitz, T. M., & Dainis, A. M. (2014). How secure are unproctored pre-employment tests? Analysis of inconsistent test scores. *Journal of Business and Psychology, 29,* 605-616.

Karim, M. N., Kaminsky, S. E., & Behrend, T. S. (2014). Cheating, reactions, and performance in remotely proctored testing: An exploratory experimental study. *Journal of Business and Psychology, 29, 555–572.* DOI: 10.1007/s10869-014-9343-z.

Lievens, F., & Burke, E. (2010). Dealing with the threats inherent in unproctored internet testing of cognitive ability: Results form a large-scale operational test program. *Journal of Occupational and Organizational Psychology, 84,* 817-824.

Lilley, M., Meere, J., & Barker, T. (2016). Remote live invigilation: A pilot study. *Journal of Interactive Media in Education, 1,* 1–5. DOI: http://dx.doi.org/10.5334/jime.408.

Lord, F. M., & Novick, M. R. (1968). *Statistical theories of mental test scores*. Reading, MA: Addison-Wesley.

Luxton, D. D., Pruitt, L. D., & Osenbach, J. E. (2014). Best practices for remote psychological assessment via telehealth technologies. *Professional Psychology: Research and Practice, 45*, 27–35.

Nielssen, O., Dear, B. F., Staples, L. G., Dear, R., Ryan, K., Purtell, C., & Titov, N. (2015). Procedures for risk management and a review of crisis referrals from the MindSpot Clinic, a national service for the remote assessment and treatment of anxiety and depression. *BMC Psychiatry, 15:304*, 1–6. DOI: 10.1186/s12888-015-0676-6.

Nye, C. D., Do, B., 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.

Okrainec, A., Vassiliou, M., Jimenez, M. C., Henao, O., Kaneva, P., & Ritter, E. M. (2016). Remote FLS testing in the real world: Ready for "prime time." *Surgical Endoscopy, 30,* 2697–2702.

Plaus, K., Boren, L., Brazell, T. P., Wickett, J., & Weber, A. (2015). *Remote proctoring test delivery: A report on options and considerations.* Washington, DC: Institute for Credentialing Excellence.

Prince, D. J., Fulton, R. A., & Garsombke, T. W. (2009). Comparisons of proctored versus non-proctored testing strategies in graduate distance education curriculum. *Journal of College Teaching & Learning, 6*(7), 51–62.

Rodchua, S., Yiadom-Boakye, G., & Woolsey, R. (2011). Student verification system for online assessments: Bolstering quality and integrity of distance learning. *Journal of Industrial Technology, 27*(3), 2–8.

Settle, J. R., Robinson, S. A., Kane, R., Maloni, H. W., &Wallin, M. T. (2015). Remote cognitive assessments for patients with multiple sclerosis: A feasibility study. *Multiple Sclerosis Journal, 21,* 1072–1079.

Tippins, N. T., Beaty, J., Drasgow, F., Gibson, W., Pearlman, K., Segall, D. O., & Shepherd, W. (2006). Unproctored internet testing in employment settings. *Personnel Psychology, 59,* 189–225.

Tomasi, L. F., Figiel, V. L., & Widener, M. (2009). I've got my virtual eye on you: Remote proctors and academic integrity. *Contemporary Issues in Educational Research, 2*, 31–35.

Wright, N. A., Meade, A. W., & Gutierrez, S. L. (2014). Using invariance to examine cheating in unproctored ability tests. *International Journal of Selection and Assessment, 22,* 12–22.