The online examination dilemma: to proctor or not to proctor?

Mohammad Dadashzadeh
Oakland University

ABSTRACT

The Covid-19 shutdown of in-person university classrooms in Spring 2020 brought the question of how to conduct online examinations into urgent focus. Although, online education has been around since 1984 and online exams have been routinely built into online course syllabi, the mid-semester transition to online teaching confronted many instructors – depending traditionally on proctored examinations to assess student performance – with the quandary of how to rely on an unproctored examination. This paper presents an overview of published pedagogical research on proctored versus unproctored testing and reviews existing technological solutions for proctored online examinations.

Keywords: Learning Assessment, Cheating, Examination Malpractice, Online Education, Unproctored Testing, Remote Proctoring, Pedagogical Design
INTRODUCTION

Examinations are the classic assessment tool used in education. They perform many interrelated functions. While an important function of testing is to ensure that the learning objectives of the course are being met, it can also serve as a motivational agent making some students extremely competitive. Nevertheless, all students benefit from the formative function of examinations in providing feedback about their performance highlighting their strengths and weaknesses (Heywood, 2016). Of course, constructing a well-designed test can be challenging and using its results for actionable pedagogical correction, even more so.

Research has shown that taking practice tests on studied material before taking a final test on the same material leads to gains in both learning and retention (Roediger & Karpicke, 2006). Therefore, this positive testing effect can be taken advantage of by mid-course formative assessments designed to improve student learning (Adesope, Trevisan, & Sundararajan, 2017). A well-constructed multiple choice test with clear questions and plausible answers (i.e., no all- or none-of-the-above choices) administered with extra time to encourage deeper thinking on more complex questions can be very useful to assess student learning while offering additional lessons learned opportunities in a quick follow up review by the teacher (Berwick, 2019). The more individualized the feedback by the instructor, the greater is the opportunity for reflection and growth for the student. Summative assessments, such as final examinations, where there is no provision for such formative assessment have been found to be less effective (Sheehan, 1985).

Quizzes, interim tests, and final examinations comprise only one category of techniques available to the instructor for assessment of student learning. Indeed, learning assessment should be regarded as a multidimensional attempt to sample, observe, and judge according to criteria, the individual learner in action (Alverno College, 1994). Since no single examination technique can be deemed satisfactory in terms of both reliability and validity (Ager & Weltman, 1967), multiple techniques encompassing both direct and indirect methods of assessment should be used (Luce & Kirnan, 2016). Indirect evidence of student achievement requires that faculty infer actual student skills, abilities, knowledge, and values, rather than observe direct evidence of learning or achievement. Often this involves students self-reporting their learning or growth. Parsons (2008) provides a detailed analysis of literature supporting a variety of styles of assessments and assessment activities including simulations, behavioral observations, and performance appraisal.

The weighting of learning assessment components is an essential part of course design which along with how the final grade is determined must be clearly stated in the syllabus. An instructor can consider nearly everything as a grading component towards the final mark. This can include student class participation, perceived effort, progress over the period of the course, comportment in group projects and team activities, in addition to graded efforts on assignments, term papers, projects, presentations/performances, quizzes, interim tests, and final examination. In assigning the final course grade, however, it is universally accepted that to preserve academic integrity and fairness there must be adequate assurance that a student’s individual work product is indeed his or hers.

Insofar as examinations are concerned, the assurance of individual work has long been established by proctoring the individual. However, distance education and online education present new challenges for proctored examinations. In the next section we present an overview of published pedagogical research on proctored versus unproctored testing. This is followed by a review of existing technological solutions that support proctored online examinations. We conclude
with a summary of pedagogical course design considerations vis-à-vis the decision to proctor or not to proctor the exams.

**PROCTORED VERSUS UNPROCTORED TESTING**

Abundant evidence indicates that cheating is a significant problem at colleges and universities (Barthel, 2016; McCabe, Treviño, & Butterfield, 2012). In surveys conducted by McCabe et al. (2001) at institutions with no academic honor code, 71% of students self-admitted to serious cheating either on written work (including plagiarism, fabricating or falsifying a bibliography, turning in work done by someone else, or copying a few sentences of material without footnoting them in a paper), or on tests (including copying on an exam with or without another student’s knowledge, using crib notes, or helping someone else to cheat on a test). However, the existence of academic honor code was associated with lower rate of cheating (54%) suggesting that students recognize that they belong to a special community demanding compliance with higher standards in exchange for the many privileges associated with honor codes including unproctored exams (McCabe et al., 1999).

An important function of examinations is to aid the credentialing process, that is, the certification of competency both of knowledge and performance in passing a course towards successful completion of a university degree. Indeed, if a university degree is a valuable credential signaling employers to differentiate employees, cheating diminishes that signal and erodes the value of university degrees while reducing cheating would strengthen the credibility of university education as a signal in the labor market (Carrell, Malmstrom, & West, 2008).

Examination malpractice or cheating may occur at any stage in the examination process including examination development, student test preparation, actual test administration, grading, and issuance of the results. Kellaghan and Greaney (2019) enumerate the following security vulnerabilities during administration of exams:

- Impersonation – when someone else takes the place of the student
- External Assistance – when unauthorized assistance is obtained either by accessing information sources (e.g., using search engines) or by connecting the student to outside “helpers” (e.g., using two-way smart glasses)
- Smuggling of Foreign Materials – when unauthorized material such as ultraviolet pens are brought to the examination to access information written previously in invisible ink on what appears to be scrap paper
- Copying – when another student’s answer/work is reproduced with or without permission
- Collusion – when unauthorized passing of information between students occurs during the examination (e.g., using smart watches)
- Theft – when examination questions/scoring keys are stolen or “hacked” during preparation, printing/publishing, or distribution
- Fake Examination Centers – when fictitious testing centers are established where students can complete the examination without supervision and with the assistance of information sources and/or “helpers”

It is reasonable to assert that the higher the stakes/rewards associated with success in an examination, the greater the incidence of exploiting the security vulnerabilities in administering the exam. It is also fair to expect that most students will not resort to unfair means to succeed in an examination. Nevertheless, university administration and college faculty must remain vigilant in
their efforts to reduce cheating and to ensure that evaluation of students is fair, valid, and educationally beneficial.

Control of malpractice during various stages of the examination process is a never-ending battle requiring continual monitoring of security procedures. The advances in smart devices pose special security threats during the administration of exams that should be addressed. Extreme measures have indeed been taken for high-stakes national examinations to thwart cheating. In 2018, Algerian authorities used metal detectors, surveillance cameras, and phone jammers to prevent collusion using smart devices smuggled into examination halls and went even further by shutting down Internet access during exam time (Bradbury, 2018).

The transition from traditional pencil-and-paper testing (PPT) to computer-based testing (CBT) increased test security by eliminating the greatest risk of test exposure – multiple printed copies of the test. However, CBT supports additional security benefits including randomizing question order and locking down the computer preventing the test-taker from accessing other applications. In general, computer-based assessments require fewer proctors and less proctor training to administer the tests. Combined with other cost savings in terms of labor and supplies, streamlined administration, scoring accuracy and efficiency, immediate reporting and feedback, CBT quickly became the preferred method for delivering high-stakes examinations (e.g., Graduate Record Examinations (GRE), Medical College Admission Test (MCAT), etc.) scheduled on-demand in dispersed testing centers.

While CBT utilizes desktop application software to deliver the assessment, learning management systems (LMS) such as Blackboard, Canvas, D2L, and Moodle are web-based and can utilize any web browser for the user interface to deliver tests. The transition from traditional CBT to Internet-Based Testing (IBT) has reflected that requirement. Typically, a customized web browser is used that locks down the testing environment within the LMS preventing the test-taker from moving away from the assessment by opening another browser tab, accessing other applications including messaging, screen-sharing, virtual machines, or remote desktops, while at the same time disabling printing and screen capture functions as well as copying and pasting anything to or from the assessment (Respondus, 2020a). When used in a proctored testing center, equipped preferably with biometric identification verification and surveillance cameras, IBT can indeed effectively address security vulnerabilities during administration of exams.

All three types of testing, PPT, CBT, or IBT can be administered in either proctored or unproctored mode. And, all formats of course delivery, traditional in-person, distance education, or online can utilize any of the three types of testing. For example, Malaysia’s Wawasan Open University, established in August 2006, utilizes a course delivery model consisting of provision of the main self-learning course materials via print or CD-ROM, supplementary material and online discussion forums via Moodle LMS, tutorial sessions offered in-person at local learning centers equipped with tutorial rooms and computer labs, unproctored IBT interim assessments on Moodle, and a proctored three-hour PPT summative final examination (Wong & Liew, 2013).

Sufficient research has been conducted that establishes equivalence of PPT and computerized testing (Buchanan & Smith, 1999; Davis, 1999). However, there have been mixed findings regarding the question of proctored versus unproctored testing. Frein (2011) conducted an experiment with military cadets enrolled in an introductory psychology course. Students’ test performance on three 20-item, multiple-choice, closed-book exams administered, respectively, in proctored (i.e., in-class) PPT, proctored IBT, and unproctored (i.e., take-home/remote) IBT formats were compared against a comparison group of students who had taken the identical tests in a previous semester in proctored PPT format. With the percentage of students who had previously
correctly answered each question serving as the baseline data, a total of 20 scores (1 for each item) was compared for each exam. T-test results for all three exams revealed that there was no difference in the test scores between the two groups. Notwithstanding that fact, given that cheating should be a concern for unproctored exams, the results may not generalize to other schools where a strict Honor Code is not a central focus of student life (Frein, 2011).

At Embry-Riddle Aeronautical University, following the 2004 transition of its distance education program from proctored PPT to unproctored IBT, research revealed statistically significant increases in the mean scores of unproctored exams in each of the three courses (in accounting, management, and marketing) studied (Schultz, Schultz, & Gallogly, 2007). Carstairs and Myors (2009) conducted an experiment to compare performance of two cohorts of undergraduate students, in an industrial and organizational psychology course, on identical tests under differing conditions of testing and proctoring. Year 1 cohort took the 55 multiple-choice test items as a part of their final summative examination in PPT format under proctored conditions. The same test items were grouped into three separate tests and presented in year two: 20 items in an unproctored PPT take-home test; another 20 items in an unproctored IBT test using LMS quiz module; and the remaining 15 items as a part of the proctored PPT final examination for year 2 cohort. The results showed that unproctored performance was significantly better than proctored performance, both between-cohorts and within-cohorts. Unproctored performance did not significantly differ between PPT and IBT testing formats supporting that the effects appear to be due to proctoring, especially since both cohorts seemed to match in ability as they performed equally well on the final exam test items (Carstairs & Myors, 2009). In another research study of closed-book test performance under proctored and unproctored settings, Brallier and Palm (2015) utilized an introductory sociology course during four semesters. Each semester the same instructor taught a traditional in-person section and an online section. For both sections, proctored PPT format was used in two semesters, while in the other two semesters unproctored IBT was administered. The findings showed that students scored significantly higher on the unproctored tests than the proctored ones with no significant difference in test performance between students in the two different types of sections.

Proctored versus unproctored online testing has also been a subject of research interest as it pertains to recruiting and testing of prospective employees via the Internet. The Occupational Personality Questionnaire (OPQ) is a well-known psychometric evaluation used worldwide to assess workplace behavioral traits and to provide some insight into how a job candidate’s behavior might affect job performance (SHL, 2020). Since there are no correct or incorrect answers in a personality test, “cheating” takes the form of faking (i.e., a conscious attempt to represent oneself according to the situation), or socially desirable responding (i.e., tendency to give overly favorable self-descriptions), both of which can be affected by the presence of supervision (Bowen, Martin & Hunt, 2002). Although equivalence between modes of administration (i.e., proctored versus unproctored) is not fully established, research studies have found no differences between the scores of an individual who would take the OPQ test in a proctored environment as compared to a candidate who would take the test in an unproctored setting (Bartram & Brown, 2004; Gupta, 2007; Joubert & Kriek, 2009).

Although the benefits of unproctored online testing – in terms of cost savings and streamlined administration for the testing organization, and myriad conveniences for the test-takers from choosing the time and the place of testing – are clear, there remains justifiable concern about malpractice in unproctored IBT. Research has shown that for noncognitive assessments, such as personality measures, comparable validities can be obtained in proctored and unproctored settings.
And, even though faking and socially desirable responding may be prevalent, the job candidate can still be screened out in subsequent job assessments/interviews that would normally follow the online personality test. On the other hand, for cognitive tests, such as a summative final examination, because of the finality and the high stakes implied it is safe to assume that some test-takers will try to exploit an unproctored environment to improve their test performance. Therefore, to ensure that evaluation of students enrolled in a fully online course is fair, valid, and educationally beneficial, university administration should provide the option for proctored IBT taken anywhere the student chooses.

ALTERNATIVE SOLUTIONS FOR PROCTORING ONLINE EXAMINATIONS

There are currently four categories of solutions for proctoring online examinations: proctoring in-person; real-time proctoring of the exam taken anywhere the student chooses by using a remote human proctor; proctoring review of the recorded exam by a human proctor; and automated proctoring.

Proctoring In-Person

This approach requires that the students take the test in a classroom or at a testing center under supervision of human proctors. Typically, student identification is verified by the proctor(s) and the students log on and take the exam using the LMS. For added proctoring ease, a customized browser such as LockDown (Respondus, 2020a) that turns the computer temporarily into a secure workstation can be utilized. In examinations that require external application software, such as Excel or R, to be used for solving problems and answering questions on the exam, a software solution such as Safe Exam Browser (SEB, 2020) can be employed that locks down the examination computer, interfaces with the quiz module of the LMS, and allows starting, and switching to, allowed external applications during the exam.

Real-Time/Live Remote Proctoring

In real-time/live remotely proctored IBT, a proctor remotely monitors the student during the examination. The test is taken anywhere the student chooses, but at an appointed time. Audio, video, and screen sharing from the student’s computer (including web cam and microphone feeds) are transmitted to the proctor’s in real time. Proctoring service providers such as Examity (2020), ProctorU (2020a), and Loyalist Exam Services (LES, 2020) will have professionally-trained live proctors sitting in a remote location who ensure student authentication, perform room scanning, and prevent/red flag any form of cheating. Active intervention into cheating behaviors practically removes the need for subsequent review of the recorded session by faculty. Although a proctor can monitor from 2 up to 16 test-takers at a time, this solution is not very scalable and is the most expensive of all alternative options.

Record and Review Proctoring

In this approach, the test can be taken on-demand, that is, anywhere and at any time the student chooses. Audio, video, and screen sharing from the student’s computer (including web cam and microphone feeds) are transmitted in encrypted fashion to the proctoring service server and
recorded. Later, a trained proctor plays back these recordings in accelerated speeds and red flags any suspicious activity through annotations for faculty consideration. RPNow is a proctoring platform that utilizes this approach and allows instructors to quickly receive and verify student’s exam results from within the LMS once proctors have completed their proctoring review process (PSI, 2020). Honorlock (2020) is also a record and review proctoring system that supports additional features including delivering proctoring services within non-LMS, third-party exam systems such as Pearson Vue, Pearson’s MyLab, and McGraw Hill’s Connect.

Kryterion (2020) utilizes several technologies in its online proctoring platform. Facial recognition algorithms and keystroke biometrics are employed for test candidate authentication. A lockdown browser, blacklisted URLs, and restricted keystroke combinations secure candidate’s computer. Specified thresholds for abnormal test-taking behavior, configurable by the test sponsor, when reached can automatically suspend an exam or enforce corrective action. Certified proctors provide a comprehensive session review for test sponsors with recorded video, audio, photos, and transcripts.

ProctorU’s Review+ is another record and review proctoring system that combines automated identification verification and launch process with artificial intelligence behavior monitoring and professional review. The software recording the session flags suspicious behavior such as lighting changes, unusual noises, and looking off screen consistently, which are subsequently investigated by trained proctors and, once confirmed, reported to the instructor as breach of integrity (ProctorU, 2020b).

Automated Proctoring

To reduce, even eliminate, the labor costs associated with proctoring, machine learning and advanced artificial intelligence techniques are utilized to develop automated proctoring systems. Proctortrack (2020) application uses student’s baseline biometric profile (face and knuckle scans) to verify student identity, delivers the LMS-based exam, transfers session data (audio, video, and screen shares) to secure servers while allowing for the video of the proctored session to be captured even if the Internet connection is temporarily lost. All proctoring data is then processed using proprietary algorithms that perform second-by-second analysis of the exam session and any suspicious behavior or deviations from exam guidelines are red flagged and delivered to the instructor for review. All data will be purged in accordance to the testing institution’s data retention policy. Proctorio (2020) is another fully automated proctoring platform that integrates with the LMS students are already using and employs analytics to capture suspicious behavior during the assessment to produce integrity reports for the instructor immediately upon exam submission.

Talview’s online assessment platform uses advanced video and audio analytics to monitor proctoring session data (i.e., real-time image, video, and audio capturing of user’s actions) for any suspicious activity. It ensures that the test-taker is focusing on test screen, triggers notification of absence during the proctored session, and using the webcam and through the microphone checks for suspect objects in video and background voice activity to red flag the test (Talview, 2020). Combining the benefits of automation with the triggered notification of a live proctor in case of potential violation detected by the monitoring software is also a feature of the Honorlock platform. This allows the student to get back on track if there was no violation.

Respondus Monitor utilizes artificial intelligence and predictive analytics to provide a fully automated proctoring system in a multi-layered approach. The first layer uses advanced algorithms for facial recognition, motion, and lighting to analyze the student and examination environment.
The next layer uses data such as keyboard activity, mouse movements, and hardware changes to identify patterns and anomalies associated with cheating. Finally, question-by-question comparisons with other students who took the same exam are considered by the analytics. The resulting analysis provides a review priority ranking score to help the instructor quickly evaluate the proctoring results (Respondus, 2020b).

**SUMMARY AND CONCLUSIONS**

The Covid-19 shutdown of in-person university classrooms in Spring 2020 forced many instructors to move closed-book final examinations, intended to be proctored in the classroom, to the online environment. That, in turn, brought into focus the proctoring dilemma associated with online examinations. The surge in demand for online proctoring solutions has been significant as universities scramble to prepare themselves for the expected new reality of a continually increasing ratio of online delivery formats in their class schedules (Jose, 2020).

As it has been presented above, there are many alternative solutions for proctoring online examinations to ensure that evaluation of students is fair, valid, and educationally beneficial. Since the acquisition of an online proctoring platform/service is an institutional decision, careful evaluation and selection of alternatives must be undertaken. There are many features to compare including LMS and third-party exam system integration, management approach for proctoring the session from student authentication through monitoring to reporting results, computer lockdown and support for permissible external applications during the exam, incident logging and intervention, session recording and retention, fault tolerance and reliability, ease of access by instructors to review session and flagged violations, usability in configuring proctoring parameters for an exam, proctoring effectiveness research studies, data privacy and security standards, customer feedback, vendor reputation and longevity, and, of course, licensing and total cost of operation. Feature comparison studies (Foster & Layman, 2013) provide a useful resource for the proctoring platform selection process, even though the continuing advances in machine learning and artificial intelligence that are being exploited in upgrading proctoring software quickly date such studies. Indeed, research has already started to make use of smart watches and fitness monitors that may be worn by test-takers for authentication, and using detected changes in pulse and temperature along with tone of voice and facial expressions for additional inputs to predictive modeling of test malpractice (Jose, 2020).

None of the solutions is perfect or foolproof (Chase, 2018). Control of examination malpractice is a never-ending battle. If high stakes/rewards are associated with success in an exam, it is safe to assume that some test-takers will try to exploit vulnerabilities in the testing environment to improve their test performance. As this paper’s overview of published pedagogical research on proctored versus unproctored testing has shown, proctoring remains the foremost weapon in the fight against examination malpractice. And, online courses have indeed different options for proctoring their examinations.

It is important to conclude with a more sobering question. As we are forced to move our in-person classes to an online delivery mode, should our syllabi remain the same in their grading components and weighted emphasis on summative mid-term and final examinations? Today’s learning management systems with their amalgam of features resoundingly answer in the negative. An LMS is not only the classroom for the online course, but also a teaching assistant. It dutifully records student attendance, participation in activities, replies to instructor’s classroom questions, and contributions to group project deliverables and team discussions. It helps the student to follow...
the instructor’s prescribed paths for learning such as completing the viewing of a lecture video, answering a question from the instructor, starting a discussion for the class, before being able to take a weekly quiz. As such, the LMS offers the instructor a variety of ways to promote student learning and engagement throughout the course while maintaining the audit trail of student performance that can contribute to grading components specified in the syllabus. These components together with individual and group assignments allow a lesser weight to be associated with mid-term and final examinations. It may then become pedagogically sound to allow time-limited, open book/open note examinations that include more complex questions to assess student learning – making the question of proctoring or not proctoring the exam less of a dilemma.

REFERENCES


