

Toward Resilient Online Assessments: Lessons Learned from the COVID-19 Era



Walid Massoud and Mohamed Abdel-Latif

“During COVID-19: No managerial strategies, no teacher training, no debates on technological design or politics, no arguments about the pros and cons—we just do it”.

Michael Kerres

1 Motivation and Anticipations

At the beginning of COVID-19, educators and assessment practitioners all over the world had several hard questions. The most demanding questions were when this pandemic will end and how teaching/testing activities can be managed during this period. They found no choice but to expand the use of technology in all fields of education, which was taken with no debates or arguments (Kerres, 2020). They just did it! In fact, everyone gambled on the use of technology and its compliance with the requirements of education in these difficult circumstances.

Assessment is globally one of the toughest educational challenges during the COVID-19 era. According to UNESCO (2020), 58 out of 84 surveyed countries had postponed or rescheduled exams; 23 introduced alternative methods such as at-home testing; and 22 maintained exams. In comparison, exams in 11 countries were canceled altogether (UNESCO, 2020). It seems that there has been an increasing trend toward using online testing in summative and standardized exams. Despite the common challenges and concerns that have been noted by several communities, such as non-ready infrastructure and risk of malpractices, it seems that online assessments will most likely be one of the central characteristics of education in the future.

W. Massoud (✉)

Qatar University Testing Center, Doha, Qatar

M. Abdel-Latif

National Center of Examinations and Educational Evaluation, Cairo, Egypt

Since many researchers anticipate that a new educational ecosystem will be introduced in the coming era, it becomes a must to review/evaluate all existing online assessment options. This review can help different stakeholders by selecting assessment solutions that can satisfy their requirements and needs. Moreover, it can help in upgrading existing solutions or designing new ones that can adapt to the emerging characteristics of the new era, cope with the diversity of contexts, and provide high-quality assessments.

The design of this review has two phases. It has started with an exploratory online survey to explore assessment users' attitudes to online testing, major challenges, and the most commonly used platforms. In the second phase, a critical review of the online assessment models/systems has been conducted in terms of various aspects such as core features, security, practicality, fairness, maintaining test-takers' privacy, equality, and diversity of cultures.

2 An Exploratory Survey

The authors conducted an online survey of the test-takers' and teachers' attitudes of e-testing, challenges, and platforms used for online or e-testing during COVID-19. The survey consists of five multiple-choice questions (MCQs) and an open-ended question. The link of the survey on Google Forms was shared with faculty, teachers, and social media groups. Participation was voluntary and anonymous. Participants agreed on the survey consent before answering the questions.

The participants were 285 test-takers (164 females; 121 males) and 18 teachers (8 females; 10 males) who took or participated in electronic tests. Although the sample is not representative and may be biased (90% Egypt, 9% Qatar, and 1% KSA), we believe that the results can be considered an indication of reality due to the similar nature of challenges of infrastructure and culture in developing countries. The experience of e-testing seems successful, to some extent, from the views of both sub-samples, an average of 6.7 out of 10 (6.72 test-takers and 6.89 teachers). The preference of test types among test-takers is at-home e-tests (56%), paper-based tests (37%), and on-site (lab) e-tests (7%), respectively. Teachers' preference is the same (39%) for at-home and paper-based tests and then comes the on-site e-tests (22%). Fifty-three percent of the test-takers prefer using at-home e-tests for final exams. On the contrary, only 22% of the teachers trust at-home e-tests for final exams. While only 4% of the test-takers have not faced any challenges during at-home e-tests, the rest (96%) have faced challenges and issues; the most common of them are system glitches (45%) and technical issues (35%) such as Internet connections and sudden electricity outage. The testing time has not been enough for 9%, and 6% have the difficulty of having a place at home that complies with the requirements of the test environment. The most frequent platforms used for assessment by the sample can be seen under Learning Management Systems and included Blackboard, Google Classrooms, Microsoft Teams, Microsoft Forms, and Google Forms. Pure online assessment platforms used by the sample were Surpass and Assessment Gourmet.

3 Online Assessment Models: A Critical Review

To understand the nature and limitations of each online assessment model or approach, we first need to explore a taxonomy of the assessment-related systems. It may be hard to have one taxonomy that includes all systems. However, we propose a classification based on our analysis of the available platforms in the market and a review of known taxonomies in the field (Fig. 1). First of all, let us differentiate between two components of educational technologies: learning management systems (LMS) and assessment management systems (AMS). LM (and blended learning)¹ systems, on the one hand, include a testing component to support the learning process. Although the testing component may not fulfill some of the key features of standard online assessment, it has taken the spotlight in most educational institutions during COVID-19 because it is available with the e-learning platform in some organizations or because it is free (which was the case of many participants in the exploratory survey). On the other hand, AM systems are designed to provide high-quality assessments and exams and comply with the standards of testing.

The AM systems and the testing components can also be classified based on delivery mode into two models (Fig. 2): center-based testing (CBT) and home-based testing (HBT), which has recently been called at-home testing. The CBT

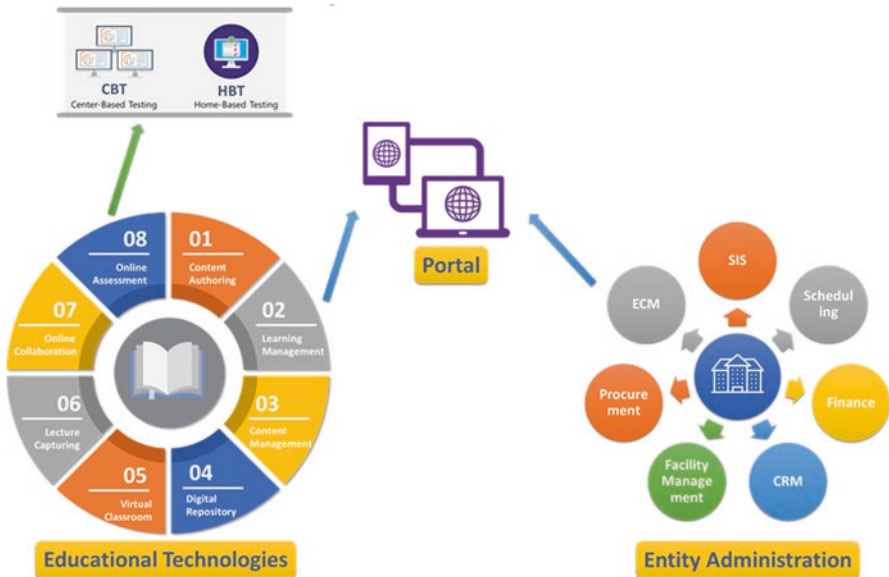


Fig. 1 Digital transformation for educational institutions

¹ Blended learning (BL) system is a style of education in which students learn via electronic and online media (e-learning) as well as traditional face-to-face teaching, such as Microsoft Teams and Google Classroom.

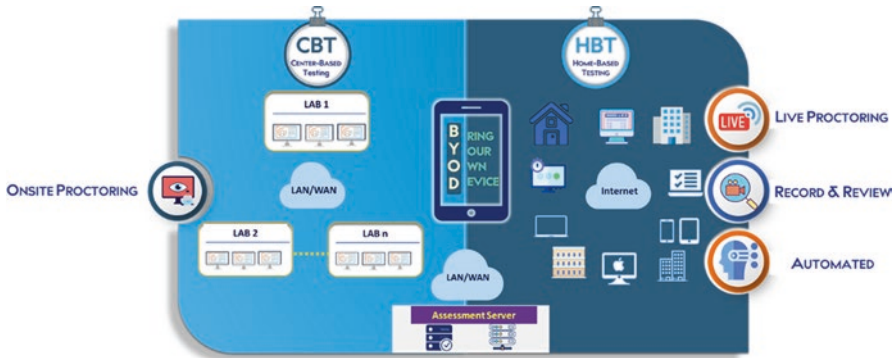


Fig. 2 Classification of assessment management systems

model usually takes place in computer labs by using desktops or laptops under the direct supervision of human proctors. Due to a limited number of labs in some educational institutions, a new sub-model has emerged, which is bring-your-own-device (BYOD). Institutions using the BYOD model ask test-takers to bring their laptops or tablets to have their tests in classrooms under the supervision of a teacher or a proctor. The HBT model usually takes place at home or any other locally suitable test environment by using the Internet on desktops, laptops, tablets, or mobiles. Mobiles with screen size less than 10 inches are not recommended for testing. The requirements of a test environment may include that the test-taker is alone in the room and the testing area is free of outside materials.

Within the HBT model, some institutions have utilized the learning management or blended learning systems to offer their summative exams (course quizzes and final exam) by integrating secure browsers. Secure browsers are applications that have control over the test-takers' devices to prevent test-takers from using the Internet or materials that are saved on their devices. They also prevent test-takers from copying-and-pasting the questions or taking screenshots of the test content.

Although many institutions use this approach, it lacks very important testing characteristics, which include having full control of the test situation and preventing impersonation. Failure to visually observe test-takers during the exam allows anyone to help the examinee in answering the questions or even answer the test on their behalf. Although before the start of the test all test-takers must accept the testing policies that impede these violations, the use of such an approach will lead to unfair situations as the results will be biased to test-takers who will crack the system over those who will follow the policies. Furthermore, it represents a high risk to the quality of the test (American Educational Research Association et al., 2014) and the credibility of the test results (Association of Test Publishers & National College Testing Association, 2020). Nevertheless, some institutions have resorted to this approach due to several reasons, such as simplicity, the large number of test-takers to be tested, and limitations of budget and resources.

Far from non-proctored tests, we can propose three models for online proctoring that can be recognized from the various platforms on the market. The proctoring component is available either as a sub-module of an LM or AM platforms or as a standalone software that integrates with other platforms. These proctoring models are live, record-and-review, and automated proctoring.

- A. **The live proctoring model** refers to a session in which one or more human proctors use technology to manage, monitor, and supervise remotely one or more test-takers, ID verifications, test authentications, test environment, and test device(s). It ensures that test integrity is maintained through the use of secure browsers, web cameras, and screen-sharing software. It can use machine learning and artificial intelligence (AI) algorithms to flag irregularities and suspicious instances, alerting the proctor to take action or decision.
- B. **The record-and-review model** refers to recording test-takers’ behaviors, data, and screens during the test to be reviewed later by human proctors. It also uses the same technologies of the Live Proctoring and flags or adds time-stamped remarks on potential suspicious instances to ensure test integrity. However, the irregularities and suspicious instances are reported for the post-test review. Some systems provide a review of the recording by professional proctors as a service to educational institutions.
- C. **Fully integrating AI** technology is the third model to monitor test-takers during exams and ensure test integrity by providing reports of potential suspicious instances just after the test session ends. Table 1 summarizes the most important features of proctoring models.

Table 1 Key features of proctoring models

Key features	Live proctoring	Record-and-review	Automated
Test-takers’ ID verification	Manual/automatic	Automatic	Automatic
Test-takers’ Authentication	Manual/automatic	Automatic	Automatic
Session launch with a live proctor	Yes	Yes	Yes
Real-time security checks (test-takers’ environment)	Yes	-	-
Real-time proctor, monitoring, and flagging	Yes	-	-
Proctor intervention	Real-time	Post-test	Post-test
Post-test review of the session to catch misconduct	-	Yes	Optional
AI flags any irregularities and suspicious instances	Optional	Yes	Yes
Video recording of test-takers’ behaviors and screens	No	Yes	Yes
Time-stamped video events for review	Optional	Yes	Yes

Moreover, the proctored online exams, especially using the live model, have been adopted, for standardized and high-stakes exams that include school leaving and university entrance exams as well as gateways for jobs. Based on research, the developers and users of this approach claim the robustness of the measures they use. The main purpose of this approach is to have full control of the test situation and prevent impersonation.

A pure live proctored exam refers to live proctored exams implemented without the supported of AI. The “pure” live proctored exams have considerable challenges, such as the limited number of test-takers to be tested in each session (max. 6), availability of experienced proctors, dividing proctors’ attention, the complexity of the systems, and inconvenient scheduling as based on available proctors and time zone differences. The utilization of AI has empowered live proctoring through flexible scheduling and an increased number of test-takers to be tested at the same time (Harmon & Lambrinos, 2008; Hylton et al., 2016; Milone et al., 2017).

One major plus of the live proctored exams is the real-time intervention as the proctor can take action like canceling the test if there is an attempt of cheating or malpractice, especially if there is an impact on the security of the test content. Both record-and-review and automated models have a post-test intervention, which may be sometimes late.

Exams using the record-and-review and automated online assessment proctoring and monitoring models have been reported to negatively affect test-takers’ academic performance (Crişan & Copaci, 2015; D’Souza & Siegfeldt, 2017; Dawood, 2016). Allowing human proctors to start the test to decrease test-takers’ test anxiety is one way to overcome this challenge. Another method can be incorporating materials within test-takers’ e-learning courses and sample online exams that help test-takers become calmer while taking their online exams (Vitasari et al., 2010). Furthermore, the systems using the automated model should allow test-takers to provide an appeal to a human reviewer to evaluate the fairness and appropriateness of the decision taken (Association of Test Publishers & National College Testing Association, 2020). Hence, some of these systems added a layer to automated-proctoring features called audit for human reviewers.

The proctored online exams, in general, have financial, technical, and sociocultural challenges. The financial challenge is mainly the high cost of proctored online platforms that prevent institutions and test-takers from trying it. The technical challenges may also be related to a limited budget, such as the absence of suitable Internet bandwidth for using the camera and video recording. Furthermore, cracking the security of the used technology is always a potential threat. The financial and technical challenges may increase educational inequalities and limit access to education at the end. The sociocultural concerns include the refusal of some test-takers, especially females, to allow photographing and video recording of themselves and their rooms during the test. This concern is considered a possible breach of test-takers’ privacy and a limitation in some cultural/religious contexts. This concern has highly increased due to data leaks of many technology firms and governmental servers in previous years. The European data protection law (European Union, 2016) and the Privacy Guidance When Using Video In The Testing Industry

(ATP Security Committee, 2020) are useful resources that can help to govern such issues, especially the collection and processing of a test-taker's personal information, the nature of data, and the purposes for which it will be used.

Further to the review of the online assessment models and to help the readers learn more about the features of various platforms, we came up with a comparison between the commonly used platforms. The comparison is based on the key features that can be impacted by the purpose of the test (high, medium, or low stakes) and accordingly can impact the quality of delivered assessments. The features are item-banking capabilities, item development, test construction methods, delivery modes, proctoring options, scoring methods, and supported statistics. We believe that this comparison can guide readers to identify their needs and select a suitable solution/platform. This comparison can be accessed via the following link: <https://sites.google.com/view/assessment-platforms/home>, and we will keep updating it by adding new platforms.

4 Conclusion

It may be hard to draw a conclusion about the future of online assessments as the pandemic has not ended yet, and there is an opportunity for emerging factors that may add to the future of using technology in assessment. However, we can draft some lines about the overall image based on our review of the available online assessment models.

It is noticeable that at the beginning of the COVID-19 era, the HBT model has dominated most of the exams in many institutions. However, due to challenges of the proctoring models and with the re-opening of some testing centers, the CBT model has started slowly to re-operate with social distancing and precaution measures. It seems there is little trust in the usage of AI and a lack of well-developed standards for using video in the testing industry. Accordingly, we can conclude that in the post COVID-19 era, the CBT will be the best model for high-stakes exams, and the HBT can be used widely in medium and low stakes. This conclusion is consistent with teachers' fears about using HBT for final exams reported in the survey. We expect that the non-proctored secure exams will stop soon due to the weaknesses related to its security.

The gap between the learning management systems (with an assessment component) and the assessment management systems will gradually disappear due to the high demand from educational institutions to use one solution for both purposes, which is the case of many platforms according to our online survey. The vast development of AI and educational technologies will assist in achieving the fusion between the two systems. Nevertheless, the need for pure, highly equipped online assessment systems will continue for high-stakes standardized exams only.

It is also obvious that educational entities, especially in developing countries, have to upgrade their infrastructures, redesign their organizational schema, and

develop their human resources to adapt to a new era. All these changes have to comply and achieve 21st century skills in education. Ministries of education and organizations that fail to cope with these changes will have hard times and may not be able to qualify their graduates for the market.

5 Recommendations

Based on our analysis and evaluation of the current approaches and common practices that have been adopted during COVID-19, as well as the lessons learned from the COVID-19 era, we recommend these guidelines for each group of assessment users.

For educational policymakers, we recommend they review the current testing standards and policies to create new policies and adapt the current ones to ensure their appropriateness for the post COVID-19 era. They also have to establish new policies that highly consider test-takers' privacy and cultural/religious concerns.

As for decision-makers in educational institutions, they need to be open-minded to modern methodologies and innovations that can cope with the shift happening to test-takers while transferring from a traditional testing environment to the new conditions. They also have to choose the most fitting methodologies for the context of their institution. Then, they have to ensure the selected methodologies fit with the context and standards of their educational institution, as there is no single methodology that fits for all educational settings.

For test designers, it is recommended to consider the new conditions and methodologies while designing assessments by shifting from assessment of knowledge to critical thinking and problem-solving skills, which should be reflected in the selection of item types and design of scoring rubrics. Estimation of answer time and test-takers' experience of using technology should be considered as well to ensure testing quality and fairness.

For educators, administrators, and IT professionals, it is recommended to choose the most appropriate tools and practices that standardize and facilitate the use of the selected methodologies, whether in test administration or invigilation. Appropriate training should be provided to those who are involved in testing situations to enforce the new policies and avoid exposing test-takers to obstacles that may affect their performance in the test.

Finally, for researchers in the field of educational technology, we recommend reviewing our proposed classification and conducting research that can use the new computer science technologies and innovations to empower online assessment systems.

References

- American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (2014). *Standards for educational and psychological testing* (2014th ed.). American Educational Research Association. <https://www.apa.org/science/programs/testing/standards>
- Association of Test Publishers, & National College Testing Association. (2020). *Online Proctoring and Monitoring Standards (Draft Version)*.
- ATP Security Committee. (2020). *Privacy guidance when using video in the testing industry*. Amazon Digital Services LLC - KDP Print US. <https://books.google.com.qa/books?id=JbSVzQEACAAJ>
- Crîșan, C., & Copaci, I. (2015). The relationship between primary school childrens' test anxiety and academic performance. *Procedia – Social and Behavioral Sciences*. <https://doi.org/10.1016/j.sbspro.2015.02.311>
- D'Souza, K. A., & Siegfeldt, D. V. (2017). A conceptual framework for detecting cheating in Online and take-home exams. *Decision Sciences Journal of Innovative Education*. <https://doi.org/10.1111/dsji.12140>
- Dawood, E. (2016). Relationship between test anxiety and academic achievement among undergraduate nursing students. *Journal of Education and Practice*, 7, 57–65.
- European Union. (2016). *General Data Protection Regulation (GDPR) – Official Legal Text*. General Data Protection Regulation. <https://gdpr-info.eu/>
- Harmon, O. R., & Lambrinos, J. (2008). Are online exams an invitation to cheat? *Journal of Economic Education*. <https://doi.org/10.3200/JECE.39.2.116-125>
- Hylton, K., Levy, Y., & Dringus, L. P. (2016). Utilizing webcam-based proctoring to deter misconduct in online exams. *Computers and Education*. <https://doi.org/10.1016/j.compedu.2015.10.002>
- Kerres, M. (2020). Against All Odds: Education in Germany Coping with Covid-19. *Postdigital Science and Education*. <https://doi.org/10.1007/s42438-020-00130-7>
- Milone, A. S., Cortese, A. M., Balestrieri, R. L., & Pittenger, A. L. (2017). The impact of proctored online exams on the educational experience. *Currents in Pharmacy Teaching and Learning*. <https://doi.org/10.1016/j.cptl.2016.08.037>
- UNESCO. (2020). Exams and assessments in COVID-19 crisis: Fairness at the centre. *United Nations Educational, Scientific and Cultural Organization*, 1. <https://en.unesco.org/news/exams-and-assessments-covid-19-crisis-fairness-centre>
- Vitasari, P., Wahab, M. N. A., Othman, A., Herawan, T., & Sinnadurai, S. K. (2010). The relationship between study anxiety and academic performance among engineering students. *Procedia – Social and Behavioral Sciences*. <https://doi.org/10.1016/j.sbspro.2010.12.067>



Dr. Walid Massoud is a senior psychometrician/researcher who is recognized as one of the Arab assessment experts in national assessments, item banking, large-scale studies, performance standards, data management, and IRT analysis. He has worked in different institutions such as the National Center of Examinations and Educational Evaluation (Egypt), American University in Cairo, Development Associates Inc., and Qatar University. He also worked as a consultant for EU, USAID, UNESCO, and WB. With 20 years of comprehensive corporate and academic experience, he participated in Egyptian national projects, such as the Education Enhancement Project (2000), National Standardized Test (2010), and School Team Excellence Awarded Program (STEAP-2007).



Dr. Mohamed Abdel-Latif has been a researcher in the National Center of Examinations and Educational Evaluation, Egypt, since 2004. He has more than 16 years of experience in the field of Computer Science and Educational Technology. He is highly renowned as a computer science expert for developing assessment and educational solutions identifying customer needs, providing technical assistance in the area of test development, item banking, setting performance standards, program evaluation, and building capacity. Mohamed has participated in several nationwide assessment reform projects in Egypt, such as CAPS for G4, 8, and 10 (2005–2009), the National Standardized Test (2010), and Teacher Cadre Tests statistical analysis committee (2008–2010).