

Evaluation systems in online environments

Inés Alegre¹, Jasmina Berbegal-Mirabent²

¹Managerial Decision Sciences Department, IESE Business School – University of Navarra, Barcelona, Spain, ²Department of Economy and Business Organization, Universitat Internacional de Catalunya, Barcelona, Spain.

Abstract

One of the biggest challenges of online teaching is student evaluation. With the students not being physically present, assessing their level of knowledge on a subject presents different challenges than those traditionally encountered in face-to-face teaching. In this paper we present an overview of different evaluation systems and reflect about its advantages and disadvantages when applying them in online environments.

The most common evaluation systems: multiple-choice quizzes, open question exams, essays, projects and oral exams, are ranked depending on several criteria. Criteria include items that any professor should take into consideration such as easiness of design and preparation or difficulty of student cheating. The advantages and downsides of each evaluation system are presented and several mechanisms to mitigate the disadvantages of each method are proposed.

This paper is helpful to professors and teachers, particularly in the current situation where the Covid-19 pandemic has moved most high-education teaching online.

Keywords: *Online; evaluation systems; exams; cheating; individual assessment.*

1. Introduction

The Covid-19 pandemic has forced schools, universities, business schools and all type of education institutions to either cancel courses or move to online teaching (García-Peñalvo et al., 2021). Moving online presents many challenges, from choosing the appropriate online platform to adapting materials and teaching styles (Barra et al., 2020). Within all these challenges, one of the most important ones is individual student evaluation.

Education institutions are required to guarantee that remote assessment is secure, reliable and fair, in particular, protecting against academic misconduct while also safeguarding a fair provision and treatment of students (Guangul et al., 2020). There is a wide array of evaluation systems that have traditionally been used in physical settings. Some professors chose to evaluate students through individual written exams, where students are gathered in a room and work on their own answers while an invigilator controls their behavior. Other professors propose some in-class activities while they go around the class solving doubts and gathering information on the attitude and knowledge of each of the students. None of these two examples of evaluation systems could be transferred to an online setting without losing much of its essence. The same is valid for many other evaluation methods.

Which evaluation systems are more adequate for online settings? What are their associated advantages and disadvantages? In this paper we aim to provide a clear overview of different evaluation systems that could be applied in online environments and rank them according to a set of diverse criteria. The criteria chosen capture the most relevant concerns of professors when designing an evaluation system: from easiness of grading to exam duration, including also difficulty of student cheating or objectivity in grading. Since no evaluation system is perfect, we discuss potential methods, called mitigation levers, that can help minimize the drawbacks of each evaluation system.

We believe our paper contributes to the current literature on higher education and particularly on evaluation systems by providing a general view of the evaluation methods available for online settings. Our overview is particularly relevant for professors. Although we acknowledge that each education level, subject and professor will differ on their preferences and needs, our framework is flexible enough to provide an answer to each professor who, depending on the relative weight he/she gives to each criteria, can find the most adequate evaluation tool to fulfill his/her purposes.

2. Literature review

Despite assessment practices in higher education institutions have been largely discussed in the literature, designing an appropriate assessment strategy is a continual challenge for instructors (Akimov & Malin, 2020). This situation is even more difficult in remote and

hybrid training modalities, where there is a lack of harmonized approaches of assessment (Kearns, 2012). Although online courses existed before the coronavirus spread, universities have been forced to readapt the way they are assessing students' performance and started using online assessment tools such as quizzes (either multiple-choice or in an open-question format), oral exams, evaluation through projects or written essays (Guangul et al., 2020).

This variety of approaches opens up many interrogates questioning which examination methods, particularly at the individual level, are the most appropriate ones (Barra et al., 2020) as distance modes of course delivery have brought new challenges.

In choosing the type of assessment, several considerations need to be taken into account (Hsiao & Watering, 2020). One of the critical issues is the validity and reliability of the assessment and if the method of delivery meets the intended purpose (Tuah & Naing, 2021). The assessment must be consistent, fairly applied, and must allow students to demonstrate the extent to which intended learning outcomes have been achieved (Shraim, 2019). The design of online exams must follow pedagogical principles, rather than merely embodying innovative technology, and the whole process must be carefully planned (Whitelock, 2006).

Another key concern that has risen with the transition to online examinations is whether this will make cheating easier (Chirumamilla et al., 2020). The impossibility of sharing the physical space with students during examination situations and have them face-to-face has led to a number of cheating practices (e.g., impersonation, forbidden aids, peeking, peer collaboration, non-allowed outside assistance). Depending on the type of examining technique, different countermeasures can be implemented (e.g., proctors, biometry, randomizing questions, broadcasting, use of antiplagiarism software, etc.), yet, online assessments are still vulnerable to academic dishonesty.

The type of assessment practice chosen will have a major impact on students' learning and academic achievement. Therefore, when discussing which evaluation methods work best, it is necessary to self-reflect about the rationale behind the assessment.

3. Methodology

In order to investigate which assessment methods are more suitable to different situations, the study was organized in three stages. First, we conducted a review of the literature aimed at identifying the most commonly used online assessment methods in higher education and the pros and cons of each method. The search was conducted in Web of Science and Google Scholar. We retained articles in academic journals but also reports published by independent organizations (e.g., European Commission). Also discussions on public forums as a result of the Covid-19 outbreak and its impact in evaluation processes were considered. The keywords used in the searches combined relevant terms such as "online", "assessment" and

“evaluation”. In addition to the above searching terms, we filtered papers by year of publication, selecting only papers published after 2005, when we believe online education started to take off. Main journals in which the selected papers were published include *Higher Education*, *Studies in Higher Education*, *Assessment & Evaluation in Higher Education*. After analyzing the documents, a list of five online assessment methods was obtained. These methods can be defined as follows:

Multiple-choice quiz (MCQ). Online quiz that contains closed-answer type questions which allow assessing essential knowledge. Questions might include text, pictures, sound or other media and weight individual answers. In a quiz the grading is automatized and questions can be randomized. This assessment method requires a learning platform.

Open question exam. This is the conventional assessment method in which students are posed with open-answer type questions. It presents different questions (e.g. testing memory, testing knowledge about concepts, testing application of the key learnings, etc.) and usually requires an answer of a couple of paragraphs long (up to one page).

Essay. Students are challenged to come up with the key concepts and theories covered during the course and put them in their own words to interpret or discuss a given topic. This method allows evaluating students’ aptitudes to recall, organize and integrate different theories and viewpoints in the form of a written work.

Project. Students are asked to think beyond the boundaries of the classroom and are challenged to apply what they have learned to an in-depth exploration of a topic. The project can be evaluated by means of an oral presentation or with the written report. If the former, the assessment is based on a presentation prepared by the student and followed by a dialogue with the instructor on this piece of work. In the case of a written report, there is no face-to-face conversation between the student and the instructor and therefore, the evaluation occurs asynchronously.

Oral exam. Oral exams (also called *vivas*) test students’ ability to verbally communicate theories, ideas and key concepts covered in a course. The lecturer poses questions to the student in spoken form and the student has to respond to them. Depending on the answers, the lecturer has the opportunity to ask follow-up questions, and thus, make this evaluation tailored to the individual student.

In addition, we also distinguish between closed-/open- book assessment situations (applicable to MCQ and open question exams). In a closed-book assessment students are confronted with the exam by solely relying on their own memory. On the contrary, in open-book exams they are allowed to refer to any material they want to consult while carrying out the exam. This later form of examination tests for more than just rote-learning.

Another outcome was the identification of the main aspects lecturers look at when choosing an assessment method: the workload required (easiness of design and preparation, and easiness of grading), the type of knowledge to evaluate (covering a module/the entire course, and deepness of the learning), the extent to which the assessment method prevents students from dishonesty (cheating with peers, and access to other sources of information), the reliability of the instrument (grading objectivity) and its feasibility (duration of the exam), and whether it is possible to maintain visual contact with student.

Next, in a second stage we organized two focus groups with professors aimed at discussing the challenges and effective practices in online assessment. Professors were selected from different disciplines and met the requirement of being actively involved in teaching innovation practices. Two focus groups of 8 professors each were performed, one with professors whose main teaching experience is with undergraduate students and another one with professors whose main experience is with graduate students. The professors came from different programs, schools and universities including medical school, architecture, management, journalism and engineering. The participants were also diverse in gender, age and career level, including lecturers, assistant professors, associate and full professors. The focus groups allowed the researchers to contrast the findings from the literature review, make sure that no relevant evaluation methodology was left out and that all the critical characteristics that an evaluation method should possess were taken into account.

Finally, in the third stage, we created a survey targeted to students in order to capture their opinions and preferences about the different forms of assessment. Although students are not the most adequate to estimate the difficulty of design and preparation of a certain type of evaluation method, they are probably the most competent to appraise the easiness of cheating. With that in mind, we designed a questionnaire where the different assessment methods were listed. Opinions were asked about difficulty of cheating, either interacting with other peers or accessing not-allowed information and about deepness of the learning that the evaluation method was able to test. General questions about evaluation method preferences as well as open questions about what type of online assessment methods have they experienced were also included. We collected responses from a variety of students enrolled in different disciplines (e.g., engineering, management, law, economics, nursing, journalism). Students represented different nationalities (Spanish, Italian, English, French, Netherlands, USA) and were diverse in gender and age (from 18 to 30 years old). Survey responses were in line to what had previously been observed in the literature and confirmed in the professors' focus group, reassuring that we did not forget any relevant point.

With the information gathered (i.e., relevant literature, and professors and students' points of view) we developed the double entry matrix presented in the following section. The cells in the matrix were filled out separately by two researchers. Later, the matrices were compared. In case of disagreement a third researcher was consulted to reach an agreement.

4. Results and discussion

Our results are summarized in Table 1. Assessment methods are displayed in columns while the selection criteria appear in rows. A green cell means that a certain evaluation system perfectly fulfills a particular criterion, while yellow should be interpreted as a partial fulfillment and red as a poor performance. In some instances a criteria is not applicable. The bottom part of the table presents the mitigation levers (Stack et al., 2020) that have been identified and that, if applied properly, can help overcome some of the shortfalls of each evaluation method.

Table 1. Evaluation systems overview.

Criteria	Multiple choice quiz		Open question exam		Essay	Project		Oral exam
	Open book	Closed book	Open book	Closed book		Written	Oral	
Easiness of design and preparation	Red	Yellow	Yellow	Green	Green	Yellow	Yellow	Red
Easiness of grading	Green	Green	Yellow	Yellow	Red	Red	Yellow	Green
Prevents interaction with peers	Red	Red	Yellow	Yellow	Red	Red	Green	Green
Prevents access to other information	N/A	Red	N/A	Red	N/A	N/A	N/A	Green
Course coverage	Green	Green	Green	Green	Red	Green	Green	Yellow
Deepness of learning	Red	Red	Yellow	Yellow	Green	Green	Green	Green
Grading objectivity	Green	Green	Yellow	Yellow	Red	Red	Red	Red
Exam duration	Green	Green	Green	Green	N/A	N/A	Red	Red
Visual contact with student	Red	Red	Red	Red	Red	Red	Green	Green
Mitigation levers								
Synchronous	•	•	•	•				•
Rubric					•	•	•	•
Question parameters	•	•	•	•				
Code of honor	•	•	•	•	•	•	•	
Camera on	•	•	•	•				•
Strict time limitation	•	•	•	•				•
Plagiarism detection			•	•	•	•	•	
Browser lockdown	•	•	•	•				

Source: Authors. Note: N/A means Not Applicable.

Multiple choice quizzes, either open-book or closed-book (columns 2 and 3 of Table 1), are a popular online evaluation system. They offer some great advantages like the easiness and objectivity of grading, sometimes even done by a computer, limited exam duration and, if well designed, can cover much of the course content. In contrast, designing them properly is not easy as questions should be mutually exclusive and present no ambiguity. In addition, especially if the software used does not allow for permanent visual contact with the students, this type of evaluation method makes cheating, either by interacting with peers or by accessing forbidden information, relatively easy. To avoid these shortfalls, some mitigation mechanisms can be put in place. We suggest that the exam is done in a synchronous manner, so that all students are doing the quiz at the same time and with a computer software that permits seeing student faces. Additionally, putting some effort in changing question parameters such as randomizing the order of the questions and answers or not allowing the student to go back and review a question already answered, are usually good practices to avoid student dishonesty. Finally, there are software applications such as Respondus¹© that have a browser lockdown system that only permits students to have open the exam window during all the exam duration.

Open question exams (columns 4 and 5 of table 1) are easier to prepare than multiple-choice quizzes but are usually more time consuming to grade and allow for a higher degree of subjectivity. By giving students the opportunity to express themselves with more freedom, it is easier to assess the deepness of the knowledge acquired. Cheating by interacting with peers or accessing information is less straightforward than in quizzes but can potentially still be an issue. In addition to the mitigation strategies already discussed for quizzes, the use of anti-plagiarism systems can be adequate in this setting.

Essays and projects (column 6, 7 and 8 of Table 1) are by nature asynchronous and require the student to think deeper. Plagiarism is one of the biggest risks in this context that can be mitigated by using a plagiarism detection system. Grading is lengthier and more subjective but this drawback can be softened by using a rubric that clearly establishes the grading criteria and to which students must have adhered on.

Finally, oral exams (column 9 of Table 1) allow a one to one interaction between the professor and the student. This evaluation system has many advantages. As soon as the exam is finished, the professor has a clear idea of the knowledge of the student on the subject and can grade accordingly. In addition, since it is synchronous and individualized, cheating becomes difficult. In contrast, being impossible to run several oral exams in parallel, doing an oral exam is very time consuming for the professor, something that can only be minimized by having a strict time control.

¹ <https://web.respondus.com/he/>

5. Conclusion

Our study is a practical compendium on the advantages and disadvantages of online individual evaluation methods. By reviewing the literature and gathering the point of view of professors and students, we have developed a framework that professors can use to assess which evaluation method is more adequate in their context, taking into account the subject, the technology and the time available.

References

- Akimov, A., & Malin, M. (2020). When old becomes new: a case study of oral examination as an online assessment tool. *Assessment & Evaluation in Higher Education*, 45(8), 1205-1221. doi: 10.1080/02602938.2020.1730301.
- Barra, E., López-Pernas, S., Alonso, Á., Sánchez-Rada, J. F., Gordillo, A., & Quemada, J. (2020). Automated Assessment in Programming Courses: A Case Study during the COVID-19 Era. *Sustainability*, 12(18), 7451. doi: 10.3390/su12187451.
- Chirumamilla, A., Sindre, G., & Nguyen-Duc, A. (2020). Cheating in e-exams and paper exams: the perceptions of engineering students and teachers in Norway. *Assessment & Evaluation in Higher Education*, 45(7), 940-957. doi: 10.1080/02602938.2020.1719975.
- García-Peñalvo, F. J., Corell, A., Abella-García, V., & Grande-de-Prado, M. (2021). Recommendations for mandatory online assessment in higher education during the covid-19 pandemic. In: D. Burgos, A. Tlili and A. Tabacco (Eds.), *Radical Solutions for Education in a Crisis Context* (pp. 85-98). Singapore: Springer.
- Guangul, F. M., Suhail, A. H., Khalit, M. I., & Khidhir, B. A. (2020). Challenges of remote assessment in higher education in the context of COVID-19: a case study of Middle East College. *Educational Assessment, Evaluation and Accountability*, 1-17. doi: 10.1007/s11092-020-09340-w.
- Hsiao, Y. P., & Watering, G. A.v. d. (2020). *Guide for choosing a suitable method for remote assessment considerations and options*. The Netherlands: University of Twente.
- Kearns, L. R. (2012). Student assessment in online learning: Challenges and effective practices. *Journal of Online Learning and Teaching*, 8(3), 198.
- Shraim, K. (2019). Online examination practices in higher education institutions: learners' perspectives. *Turkish Online Journal of Distance Education*, 20(4), 185-196. doi: 10.17718/tojde.640588.
- Stack, A., Boitshwarelo, B., Reedy, A., Billany, T., Reedy, H., Sharma, R., & Vemuri, J. (2020). Investigating online tests practices of university staff using data from a learning management system. *Australasian Journal of Educational Technology*, 36(4), 72-81. doi: 10.14742/ajet.4975.
- Tuah, N. A. A., & Naing, L. (2021). Is online assessment in higher education institutions during COVID-19 pandemic reliable? *Siriraj Medical Journal*, 73(1), 61-68. doi: 10.33192/Smj.2021.09.