


# Higher education teachers' and students' perceptions of open-book and proctored examinations in the COVID-19 pandemic

## *Percepción de docentes y estudiantes de educación superior de los exámenes a libro abierto y supervisados en la pandemia por COVID-19*

Beatriz Marcano <sup>1\*</sup> 

Beatriz Ortega-Ruipérez <sup>1</sup> 

Almudena Castellanos-Sánchez <sup>1</sup> 

<sup>1</sup> Universidad Internacional de La Rioja, Spain

\* Autor de correspondencia. E-mail: [beatriz.marcano@unir.net](mailto:beatriz.marcano@unir.net)

### How to reference this article/ Cómo referenciar este artículo:

Marcano, B., Ortega-Ruipérez, B. & Castellanos-Sánchez, A. (2023). Higher education teachers' and students' perceptions of open-book and proctored examinations in the COVID-19 pandemic. *Educación XX1*, 26(1), 207-228. <https://doi.org/10.5944/educxx1.33514>

**Fecha de recepción:** 15/03/2022  
**Fecha de aceptación:** 28/07/2022  
**Publicado online:** 02/01/2023

### ABSTRACT

In the early days of the COVID-19 pandemic, higher education was forced to review its assessment processes. Competency achievement and academic honesty should be ensured in online assessments. In the Master of Educational Technology and Digital Competences of a Spanish University, the open-book examination model was implemented to respond to the new situation considering the characteristics of authentic assessment (adapted to students,

intellectually challenging, related to practice, coherent with the didactic methodology, makes plagiarism difficult). We wanted to analyze the relevance of this change in evaluation. The main objective is to analyze the differences between traditional face-to-face exams from before the pandemic and open-book exams with and without proctoring according to the perception of teachers and students. The research is of an empirical nature and quantitative approach and is based on the responses of 66 teachers and 301 students to a questionnaire with sufficient validity (chi-2/GI: 2.453, RMSEA: .069, CFI: .99 and TLI: .99), and an Omega reliability coefficient of .882. Comparisons were made between model A: traditional face-to-face examination, model B: open-book examination with proctoring, model C: open book examination without proctoring. The results show that for teachers and students open-book exams with or without proctoring had no significant differences and are more in line with an authentic assessment than face-to-face exams. It is concluded that open-book exams with or without proctoring are suitable for authentic online assessment in higher education. It is recommended to contrast the results in other online university courses and to encourage authentic assessment in higher education institutions.

**Keywords:** assessment, higher education, open-book examination, proctored exams, authentic assessment

## RESUMEN

En los primeros días de la pandemia de COVID-19, la educación superior se vio obligada a revisar sus procesos de evaluación. Se debían asegurar el logro competencial y la honestidad académica en las evaluaciones online. En el Máster de Tecnología Educativa y Competencias Digitales de una universidad española se implantó el modelo de examen a libro abierto para responder a la nueva situación considerando las características de la evaluación auténtica (adaptada a los estudiantes, intelectualmente desafiante, relacionada con la práctica, coherente con la metodología didáctica, dificulta el plagio). Hemos querido analizar la pertinencia de este cambio en la evaluación. El objetivo principal es analizar las diferencias entre los exámenes presenciales tradicionales de antes de la pandemia y los exámenes de libro abierto con y sin supervisión según la percepción de profesores y alumnos. La investigación es de carácter empírico y enfoque cuantitativo y se basa en las respuestas de 66 profesores y 301 alumnos a un cuestionario con validez suficiente (chi-2/GI: 2.453, RMSEA:.069, CFI:.993 y TLI: .991), y un coeficiente de fiabilidad omega de .882. Se realizaron comparaciones entre modelo A: examen tradicional presencial, modelo B: examen a libro abierto con proctoring, modelo C: examen a libro abierto sin proctoring. Los resultados muestran que para profesores y alumnos los exámenes con libro abierto con o sin supervisión no presentan diferencias significativas y se ajustan más a una evaluación auténtica que los exámenes presenciales. Se concluye que los exámenes de libro abierto con o sin supervisión son adecuados para realizar una evaluación auténtica en línea en la educación superior. Se recomienda contrastar los resultados en otros cursos universitarios online e incentivar la evaluación auténtica en las instituciones de educación superior.

**Palabras clave:** evaluación, educación superior, examen a libro abierto, exámenes supervisados, evaluación auténtica

## INTRODUCTION

The closure of education, due to the global Covid19 pandemic, has led educational institutions to rely on virtuality to survive. Teachers have been forced to change their daily practices, including the way they assess students. Faced with the difficulties of testing the competencies of a confined student body, without the possibility of taking face-to-face exams, countries have adopted different strategies to adapt to the complicated situation (Fardoun, et al., 2020). These were oriented towards flexibility and the use of online means - oral exams, supervised writing, collaborative work, tests, open-book exams, etc. Recovering education, making it possible, must be a priority for countries to avoid a generational catastrophe (United Nations Educational, Scientific and Cultural Organization [UNESCO], 2021).

In this sense, some studies (García-Peñalvo et al., 2021; Guangul et al., 2020) point to the evaluation process as the greatest challenge that higher education has faced in this pandemic and make recommendations for such a unique process in which continuous assessment should be considered, guaranteeing student identification in final exams, using technology to which the student is accustomed, replacing traditional written tests with papers, exercises, infographics, etc. Specifically, it is suggested that in virtual exams students can face evaluation models where an online presentation made by themselves is required, in such a way that identity and authorship are verified, thus facing the great challenge of dishonesty (Guangul et al., 2020; Gudiño et al., 2021; Shaushenova et al., 2021).

Other research has focused on studying the experience of students and teachers with online assessment during confinement (Garcia-Alberti, 2021), finding important challenges such as the lack of preparation of teachers to deal with remote assessment and their distrust towards this type of tests. For education professionals, the possibility of cheating, the reliability of the technology, connectivity, problems with hardware, usability, ease of correcting and creating tests, are key concerns (Butler-Henderson & Crawford, 2020; Halak & El-Hajjar, 2019). Add to this the shortcomings detected in some teachers in terms of their training in assessment (Soodmand & Ranjbar, 2021), and the task of online measurement of university student competencies becomes more complicated. Recent studies point to the fact that teachers, to reformulate assessment during the pandemic, have preferentially opted to move their traditional exams to an online format, given the emergency (Slade et al., 2021).

Higher education has a primary role in identifying solutions to these challenges that online assessment presents to universities (Pagram et al., 2018). In that sense, with this study we want to analyze a proposed shift in online assessment towards open-book examinations as a mechanism to meet the challenge of maintaining quality education during the pandemic.

### **Theoretical background**

One of the strategies that higher education institutions have adopted to deal with online exams has been the use of different assessment modalities, among them, open book exams (Cheung, 2020). In these exams, consultation of class materials, notes, the web is allowed. It was an examination model proposed by some authors who associated this modality with the education of the future and high-level competencies (Feller, 1994; Herrington & Herrington, 1998; Koutselini, 1997). Prior to the Covid-19 pandemic, studies (Brown & Glasner, 1999; Williams & Wong, 2009) highlighted this type of examination for allowing students to creatively use knowledge acquired during the course, reflect deeply on problems, and apply critical thinking, rather than simply recalling information or using what they remember. At the same time, the open-book exam encourages students to conduct a deeper study of the content, assess their learning gaps during exam preparation, self-evaluate their preparation results, and reduce stress. On the other hand, generally the conditions of traditional exams are unrealistic, they do not represent real working world situations (Feller, 1994) so it is complicated to measure certain skills. To solve daily problems, professionals rely on all the resources at their disposal, including the Internet, which is incompatible with closed-book exams. Authentic assessment according to Boud (2020), Ibarra-Sáiz et al. (2020) and Herrington & Herrington (1998) requires that the context in which it is carried out reflects the conditions in which the professional will have to act, also, it should invite the student to transform his knowledge into creative actions or products even in collaboration with his peers. On the other hand, it implies solving complex challenges that require critical thinking, challenges that will be in line with what has been worked on during the course and finally, providing reliable learning indicators. Undoubtedly, this is an assessment model that involves students more as decision makers and applicators of newly constructed knowledge, which induces deeper learning (McArthur, 2020; Williams & Wong, 2009).

On the other hand, given that nowadays learning outcomes must be expressed in terms of competences, this open-book examination modality is ideal if what is required is to measure the entrepreneurship, flexibility, personal responsibility, innovation or creativity of students, skills that most countries value for their professionals (Organización para la Cooperación y el Desarrollo Económicos [OCDE],

2019). Knowledge is becoming increasingly complex and so are the problems, skills required of 21st century professionals such as finding and evaluating information or managing complex professional situations, are exercised in open-book examinations (Feller, 1994). However, for this to be possible, it is necessary to design tests consistent with competency-based models such as simulations or projects (Cano, 2008). The design of open-book exams is not an additional burden for teachers; however, it requires special preparation to design tests that are creative, meaningful, linked to the reality of each professional and that allow students to demonstrate the competencies required of them (Cano, 2008; Feller, 1994). In addition, they require a change in the pedagogical approach, so that teachers orient their sessions to the teaching of high-level skills. We cannot forget that assessment is not only intended to measure learning, but also to favour it.

Given these characteristics, open-book exams become a relevant modality in online tests whose objective is an authentic evaluation, in the sense of being realistic, capable of measuring competencies, considering the student as the protagonist of his or her learning. In this same sense, since the Covid-19 pandemic, studies are reappearing in which the results of experiences in higher education with this type of exams are presented, given their suitability to the emergency experienced. A relevant example is the experiment conducted by Vázquez et al. (2021), which analyzes the performance of students in supervised and unsupervised online open-book exams, or the study conducted by Prigoff et al. (2021), which evaluates the suitability of open-book exams in medical students.

The purpose of this study is to contribute to this line of research on open-book examinations and to know how well they adjust to the demands of learning today and to the new trends in evaluation in higher education. Having reliable and valid assessment instruments that facilitate research on this topic will allow progress in this field of study (Nunnally & Bernstein, 1994). Likewise, taking advantage of the experience of the main actors in this evaluation process, such as teachers and students, contributes to have a realistic assessment of open-book examinations. Therefore, the general objective of this study was to compare the examination models used before and during the pandemic according to the perception of teachers and students and to analyze the differences in relation to the characteristics of authentic assessment. To this end, the following specific objectives were proposed:

1. To know the teachers' perceptions of the traditional examination model and the open-book examination model with and without proctoring.
2. To know the students' perception of the traditional examination model and the open-book examination model with and without proctoring.
3. To compare the students' and teachers' assessment of traditional exams and open-book exams with and without proctoring.

## METHOD

### Design

In conducting this study, a quantitative methodology was applied, with a research design through an online survey (Callegaro et al.; 2015) to know the perception that teachers and students have of the different models of exams conducted (A, B, C). (Appendix 1). This methodology and design was considered the most relevant to access the information, because of the ease offered to students to respond, since they had finished their school term and their participation in the study was completely voluntary.

### Population and Sample

The population of this study consisted of all students taking the Master's Degree in Educational Technology and Digital Competences (MUTECD) at the International University of La Rioja (Spain) and the teachers of the courses from November 2019 to July 2020. The sampling procedure was non-probability purposive by proximity (Otzen & Manterola, 2017). After the exams were taken, students and teachers were asked to answer the questionnaire about the exam models. The sample consisted of 302 students, of whom 195 from the first semester responded to the questionnaire on model C and 107 from the second semester responded to the questionnaire on models A and B. The sample of teachers consisted of 37. All of them responded to the questionnaire on model A, and in addition 17 responded to the questionnaire on model B and 19 responded to the questionnaire on model C.

### Instrument

Data were collected through a questionnaire based on the one used by Williams and Wong (2009) designed to assess the effectiveness of open-book examinations. The survey consists of 9 items. *1. The time to take the exam (2 hours) was about right. The type of exam: 2. required the student to come in person to take the exam; 3. was in line with the pedagogy/methodology, 4. promoted more complex learning (creating, reflecting, etc.) and less rote learning, 5. is intellectually challenging, 6. is better suited to the learning style of each student, 7. is more related to professional practice, 8. The content of the exam was attractive, 9. The type of exam allowed cheating (plagiarism, copying, etc.).* The answers are indicated on a 5-point Likert scale (1=Strongly disagree, 2=Disagree, 3=Agree, 4=Strongly agree, 5=Don't know/no answer). The questionnaire was completed through the Forms application of Office 365 (Questionnaire). Data were collected in June and July 2021. The same questions were applied for the three examination models: model A: traditional

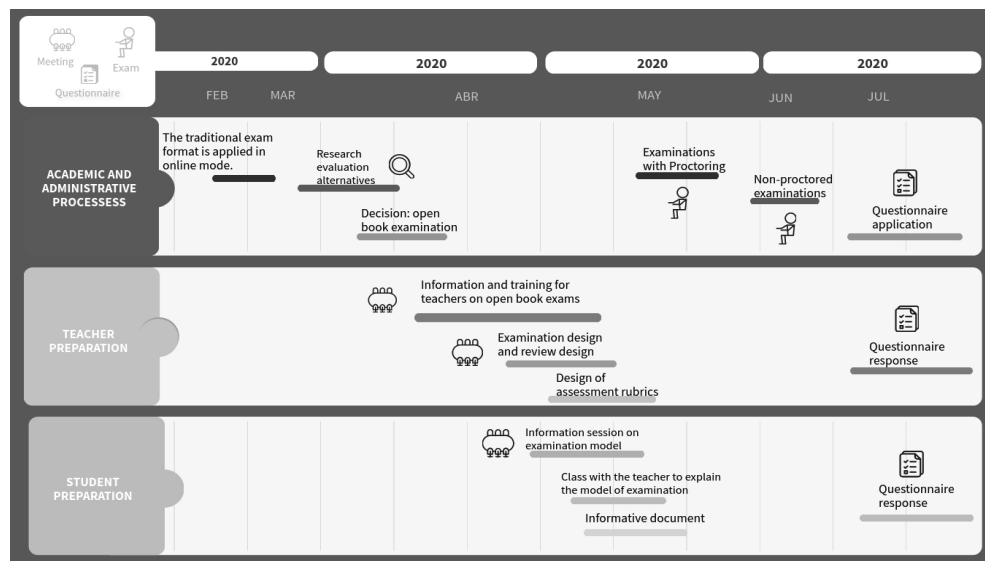
face-to-face examination, model B: open book examination with proctoring, model C: open book examination without proctoring.

## Procedure

We wanted to evaluate the different exam models because in the situation of the COVID-19 pandemic, online exams were implemented in the MUTECD. It was decided to make a change in the examination models that were traditionally applied in face-to-face mode (Fig.1) to an open-book examination model oriented to an authentic assessment that could be used as an opportunity for learning. These exams presented challenging approaches and real situations in which the student had to solve cases, simulations, assuming the role of an expert, proposing a solution to a problem in the classroom or an educational center, and so on. The answers would be presented in the form of a digital product such as a video, infographic, podcast, interactive presentation or other. Likewise, students had the option to connect to the web and use the technological applications they considered most relevant in each case, not only to show their knowledge but also to apply it, reflect, make decisions, create digital content, and share it. The design of this exam model was done collaboratively among the teachers of the same subject. They were also asked to create a common rubric for the evaluation of student competencies (Feller, 1994).

**Figure 1**

*Outline of procedure for changing the examination model*



Source. Created by authors using Genial.ly (online)

In addition, the students are informed about how to be evaluated; on the one hand, they receive information sessions from the Master’s management and professors; on the other hand, a document explaining the characteristics of the exam was published in the virtual classrooms of all subjects. Finally, the exams were conducted, the online survey was applied to students and teachers, and the collected data were analyzed.

### Statistical analysis

To guarantee the validity of the instrument, an ordinal confirmatory factor analysis was performed using the Mplus 8.0 tool. Previously, the normality analysis was performed with Mardia’s coefficient, it was found that the kurtosis and multivariate symmetry statistics showed significant results, so the normality of the distribution of the set of items of the questionnaire cannot be assumed (Table 1).

**Table 1**

*Multivariate skewness and kurtosis results according to Mardia’s coefficients*

	Coefficient	z	$\chi^2$	df	p
Skewness	24.1		1123	84	<.001
Kurtosis	114.8	38.6			<.001

Subsequently, from a matrix of polychoric correlations, the parameters are estimated with the robust weighted least squares method (WLSMV), suitable in these cases (Li, 2014 and Xia, 2016). In the model evaluation phase, the standardized indexes are used: normalized robust chi-square ( $\chi^2 /d.f.$ ), to assess the overall fit, where values between 3 and 5 are considered acceptable (Hair et al., 2014); RMSEA, to assess the residual matrix, which is acceptable with values below .08 (Byrne, 2009); and CFI and TLI (Tucker-Lewis Index), of comparative fit, which are acceptable from .90 (Hair et al., 2014). Following Hu and Bentler (1999) as evidence of validity, an acceptable fit in the combination of these indexes is sufficient. The results of the confirmatory factor analysis indicate that the unidimensional model is validated. The fit indices show acceptable values in all cases: chi-2/GI less than 3, RMSEA not significant (indicating, therefore, that it is equal to zero) and CFI and TLI greater than .95 (Table 2).



**Table 2**

*Results of the Chi-2 and RMSEA fit indices*

	No. Parameters	Chi-2				RMSEA			
		Chi-2	GL	P value	chi-2/GL	RMSEA	P value	CFI	TLI
No item 9	32	49.1	20	.0003	2.45	.069	.091	.993	.991
	31	56.4	21	.0000	2.69	.075	.04	.976	.968

A study of convergent validity is also carried out by calculating the average variance explained by the model (Average Variance Extracted, AVE),

$$AVE = \frac{\sum_{i=1}^N P_i^2}{n}$$

In addition, internal consistency is estimated by calculating the composite reliability, also called omega coefficient or internal consistency estimated by SEM (Viladrich et al.; 2017), the factor loadings (P) of the j items and the error variance (ei) are used, with the following formula:

$$\omega = \frac{(\sum_{j=1}^n P_j)^2}{(\sum_{j=1}^n P_j)^2 + (\sum_{j=1}^n e_j)}$$

The convergent validity results also point to a good fit of the dimension, reaching a total variance explained by the authentic assessment dimension of 0.607 (60%) and a composite reliability coefficient or omega of .882.

A comparison of the results for each item and the total scale score was also performed using the nonparametric Kruskal-Wallis test. And paired model comparisons (A vs B; A vs C and B vs C) are included. This analysis was performed comparatively for students and teachers. The statistical differences of the average ranges of each group were compared (Mann-Whitney U). We considered for the analyses the Bonferroni adjustment that gives a better significance when there are small groups. Cohen's D for the effect of differences is also included, where the intervals for r: 0.1 to 0.3: is considered small effect; 0.3 to 0.5: intermediate effect; 0.5 and above: strong effect, according to Fritz et al. (2012) based on Cohen.

SPSS version 27 was used to carry out the analysis described above. The validation process was carried out with the students' database because it was much larger than that of the teachers.

## RESULTS

Descriptive and comparative analyses of teachers' and students' perceptions of the examination models are presented.

Table 3 shows the descriptive statistics for each of the items in each exam model and according to the results of the teachers and students. The Kruskal-Wallis bilateral asymptotic significance index and Cohen's index are also included.

**Table 3**

*Descriptive statistics of the variables according to teachers, students, and exam models, Kruskal Wallis bilateral asymptotic significance index and D Cohen index*

		Model evaluation			Kruskal-Wallis			D. Cohen
		A Mean (SD)	C Mean (SD)	B Mean (SD)	N	Test statistic	Asymptotic sig. (bilateral test)	
I_1	Teacher	3.41 (0.837)	3.76 (0.437)	3.71 (0.849)	66	3.635a.b	0.162	0.326
	Student	2.79 (0.930)	2.75 (0.951)	3.01 (0.914)	302	4.607a.b	0.100	0.188
I_2	Teacher	2.22 (1.128)	1.31 (0.793)	1.00 (0.000)				
	Student	1.62 (0.885)	1.45 (0.794)	1.61 (0.921)				
I_3	Teacher	2.72 (1.143)	3.88 (0.332)	3.88 (0.342)	65	20.606a	0.000	1.31
	Student	3.25 (0.952)	3.71 (0.559)	3.49 (0.745)	295	14.079a	0.001	0.415
I_4	Teacher	2.75 (0.984)	3.82 (0.529)	3.94 (0.250)	65	24.874a	0.000	1.529
	Student	3.15 (1.049)	3.75 (0.539)	3.58 (0.740)	294	20.674a	0.000	0.524
I_5	Teacher	2.72 (0.888)	3.53 (0.624)	3.88 (0.342)	65	22.298a	0.000	1.395
	Student	3.22 (0.967)	3.65 (0.702)	3.54 (0.706)	291	12.108a	0.002	0.381
I_6	Teacher	2.53 (1.047)	3.41 (0.712)	3.81 (0.403)	65	19.412a	0.000	1.25
	Student	2.95 (1.031)	3.59 (0.692)	3.38 (0.793)	289	22.782a	0.000	0.56
I_7	Teacher	2.50 (1.191)	3.94 (0.243)	3.81 (0.403)	65	25.145a	0.000	1.544
	Student	3.00 (1.116)	3.79 (0.497)	3.57 (0.728)	289	32.701a	0.000	0.694
I_8	Teacher	2.53 (1.016)	3.56 (0.512)	3.88 (0.342)	64	24.302a	0.000	1.518
	Student	2.95 (0.974)	3.58 (0.703)	3.42 (0.811)	291	23.959a	0.000	0.575
I_9	Teacher	2.45 (1.179)	2.35 (1.169)	2.44 (1.031)	64	0.090a.b	0.956	0.36
	Student	1.87 (0.911)	1.50 (0.823)	1.55 (0.942)	283	11.315a	0.003	0.371
Total_mean	Teacher	2.71 (0.799)	3.57 (0.283)	3.70 (0.234)				
	Student	3.05 (0.760)	3.53 (0.438)	3.43 (0.537)				

*Note.* No data for Kruskal Wallis on item two because there was no variability.

In teachers and students there are statistically significant differences in some respects. In the following, these differences are specified with the comparison between the examination models.

In some items there were no differences in the results between teachers and students as, for example, in item 1 (teachers  $p=.162$ ; students  $p= .100$ ), so it can be indicated that for both teachers and students the time to take the exam was correct in all models.

When comparing the exam models, it was found that for teachers the open-book exams were more in line with the pedagogical approach of the course than the traditional model. For the students, both the exam that asked them to respond to the questions textually (traditional model) and the models that asked them to create a digital product were consistent with the pedagogical approach during the course. Although a greater difference stands out in the open book model without proctoring (Table 4).

**Table 4**

*Comparisons between assessment models for: The type of exam was in line with the pedagogy/methodology used during the course*

Source	Sample 1-Sample 2	Test statistic	Dev. Error	Dev. test statistic	Sig.	Sig. adjusted <sup>a</sup>	D Cohen
Teacher	A-B	-18.250	4.987	-3.660	.000	0.001	1.121
	A-C	-18.430	4.888	-3.770	.000	0.000	1.131
	B-C	-0.180	5.673	-0.032	.975	1.000	0.011
Student	A-B	-16.012	12.057	-1.328	.184	0.553	0.226
	A-C	-37.556	10.471	-3.587	.000	0.001	0.53
	B-C	-21.544	10.011	-2.152	.031	0.094	0.304

Each row tests the null hypothesis that the distributions of Sample 1 and Sample 2 are equal. Asymptotic significances (bilateral tests) are displayed. The significance level is .05.

<sup>a</sup> Significance values have been adjusted by Bonferroni correction for various tests.

Teachers believe that open-book exams with or without proctoring promote more complex learning (creating, reflecting) and less rote learning than traditional exams. However, students believe that both traditional and open-book exams promote more complex learning and less rote learning, although they perceive a greater difference between open-book exams without proctoring compared to traditional (Table 5).

**Table 5**

*Comparisons between assessment models for: The type of exam promotes complex learning (creation, reflection, etc.) and less rote learning*

Source	Sample 1-Sample 2	Test statistic	Dev. Error	Dev. test statistic	Sig.	Sig. adjusted <sup>a</sup>	D Cohen
Teacher	A-C	-19.281	4.924	-3.916	.000	0.000	1.175
	A-B	-21.281	5.024	-4.236	.000	0.000	1.297
	C-B	2.000	5.715	0.350	.726	1.000	0.122
Student	A-C	-45.194	9.964	-4.536	.000	0.000	0.67
	A-B	-28.988	11.462	-2.529	.011	0.034	0.43
	B-C	-16.206	9.527	-1.701	.089	0.267	0.24

Each row tests the null hypothesis that the distributions of Sample 1 and Sample 2 are equal. Asymptotic significances (bilateral tests) are displayed. The significance level is .050.

<sup>a</sup> Significance values have been adjusted by Bonferroni correction for various tests.

Teachers find open-book exams with and without proctoring more intellectually challenging than traditional exams, but students find all models of exams intellectually challenging, although open-book exams without proctoring more so. The other two models consider them equally challenging (Table 6).

**Table 6**

*Comparisons between assessment models for: The type of exam is intellectually challenging*

Source	Sample 1-Sample 2	Test statistic	Dev. Error	Dev. test statistic	Sig.	Sig. adjusted <sup>a</sup>	D Cohen
Teacher	A-C	-15.765	5.259	-2.998	.003	0.008	0.9
	A-B	-23.875	5.365	-4.450	.000	0.000	1.392
	C-B	8.110	6.103	1.329	.184	0.552	0.463
Student	A-C	-35.926	10.362	-3.467	.001	0.002	0.516
	A-B	-22.557	11.882	-1.898	.058	0.173	0.324
	B-C	-13.368	9.856	-1.356	.175	0.525	0.192

Each row tests the null hypothesis that the distributions of Sample 1 and Sample 2 are equal. Asymptotic significances (bilateral tests) are displayed. The significance level is .050.

<sup>a</sup> Significance values have been adjusted by Bonferroni correction for various tests.

Both teachers and students feel that open-book examination models are better suited to each student’s learning style, especially the unmonitored model, as opposed to the traditional model (Table 7).

**Table 7**

*Comparisons between assessment models for: The type of exam is better adapted to the learning style of each student*

Source	Sample 1-Sample 2	Test statistic	Dev. Error	Dev. test statistic	Sig.	Sig. adjusted <sup>a</sup>	D Cohen
Teacher	A-C	-14.522	5.314	-2.733	.006	0.019	0.82
	A-B	-22.656	5.421	-4.179	.000	0.000	1.28
	C-B	8.134	6.167	1.319	.187	0.562	0.459
Student	A-C	-52.459	11.087	-4.732	.000	0.000	0.712
	A-B	-31.151	12.730	-2.447	.014	0.043	0.423
	B-C	-21.308	10.474	-2.034	.042	0.126	0.289

Each row tests the null hypothesis that the distributions of Sample 1 and Sample 2 are equal. Asymptotic significances (bilateral tests) are displayed. The significance level is .050.

<sup>a</sup> Significance values have been adjusted by Bonferroni correction for various tests.

Likewise, teachers and students agree that the open-book examination models are more related to professional practice than the traditional model (Table 8). Likewise, they agree that the open-book examination models were more attractive than the traditional model (Table 9).

**Table 8**

*Comparisons between assessment models for: The type of exam is more related to professional practice*

Source	Sample 1-Sample 2	Test statistic	Dev. Error	Dev. test statistic	Sig.	Sig. adjusted <sup>a</sup>	D Cohen
Teacher	A-B	-19.063	5.102	-3.736	.000	0.001	1.144
	A-C	-22.086	5.001	-4.416	.000	0.000	1.325
	B-C	-3.024	5.805	-0.521	.602	1.000	0.181
Student	A-B	-36.668	11.510	-3.186	.001	0.004	0.552
	A-C	-56.916	9.983	-5.701	.000	0.000	0.857
	B-C	-20.248	9.475	-2.137	.033	0.098	0.305

Each row tests the null hypothesis that the distributions of Sample 1 and Sample 2 are equal. Asymptotic significances (bilateral tests) are displayed. The significance level is .050.

<sup>a</sup> Significance values have been adjusted by Bonferroni correction for various tests.

**Table 9**

*Comparisons between assessment models for: The content of the exam was attractive*

Source	Sample 1-Sample 2	Test statistic	Dev. Error	Dev. test statistic	Sig.	Sig. adjusted <sup>a</sup>	D Cohen
Teacher	A-C	-17.109	5.322	-3.215	.001	0.004	0.984
	A-B	-24.453	5.322	-4.595	.000	0.000	1.407
	C-B	7.344	6.145	1.195	.232	0.696	0.422
Student	A-C	-54.406	11.116	-4.895	.000	0.000	0.732
	A-B	-39.018	12.741	-3.062	.002	0.007	0.525
	B-C	-15.388	10.513	-1.464	.143	0.430	0.207

Each row tests the null hypothesis that the distributions of Sample 1 and Sample 2 are equal. Asymptotic significances (bilateral tests) are displayed. The significance level is .050.

<sup>a</sup> Significance values have been adjusted by Bonferroni correction for various tests.

According to the teachers’ perception, there is no difference between the exam models in terms of the possibility of committing plagiarism; however, according to the students, the traditional exam allows plagiarism as opposed to the open-book models, regardless of whether they were with or without proctoring (Table 10).

**Table 10**

*Comparisons between assessment models for: The type of exam allowed cheating (plagiarism)*

Source	Sample 1-Sample 2	Test statistic	Dev. Error	Dev. test statistic	Sig.	Sig. adjusted <sup>a</sup>	D Cohen
Student	C-B	0.157	10.401	0.015	.988	1.000	0.002
	C-A	34.883	10.817	3.225	.001	0.004	0.49
	B-A	34.726	12.657	2.744	.006	0.018	0.488

Each row tests the null hypothesis that the distributions of Sample 1 and Sample 2 are equal. Asymptotic significances (bilateral tests) are displayed. The significance level is .050.

<sup>a</sup> Significance values have been adjusted by Bonferroni correction for various tests.

**Comparisons of evaluation model according to teachers and students**

In the Kruskal-Wallis test the results are similar between teachers and students, that is, both perceive significant differences between models A and C, and models A and B and similarity between models C and B (Table 11).

**Table 11**

*Comparison of traditional face-to-face and open-book evaluation models with and without proctoring for teachers and students*

Source	Sample 1-Sample 2	Test statistic	Dev. Error	Dev. test statistic	Sig.	Sig. adjusted <sup>a</sup>	D Cohen
Teacher	A-C	-17.621	5.739	-3.070	.002	0.006	0.921
	A-B	-23.386	5.739	-4.075	.000	0.000	1.223
	C-B	5.765	6.559	0.879	.379	1.000	0.301
Student	A-C	-54.260	12.629	-4.296	.000	0.000	0.625
	A-B	-41.233	14.649	-2.815	.005	0.015	0.475
	B-C	-13.027	12.200	-1.068	.286	0.857	0.150

Each row tests the null hypothesis that the distributions of Sample 1 and Sample 2 are equal. Asymptotic significances (bilateral tests) are displayed. The significance level is .050.

<sup>a</sup> Significance values have been adjusted by Bonferroni correction for various tests.

## DISCUSSION AND CONCLUSIONS

The challenges that higher education had to assume to carry out summative assessment processes in times of pandemic led it to innovate and implement assessment modalities different from those applied before the pandemic. At that time, having to assume the development of courses and assessment in online modality, it was necessary to take advantage of the moment to make the improvements that technologies and digital formats could allow. And at the same time assume an appropriate assessment modality for higher education in which this is assumed as an opportunity for critical, reflective, and empowering learning for the student as recommended by Boud (2020), Brown (2015), Ibarra-Sáiz et al. (2020) and Ibarra-Sáiz & Rodríguez-Gómez (2020). In this sense, in the master's degree in educational technology and digital competencies, it was proposed to make a change in the examination models from face-to-face to online open-book exams with proctoring and without proctoring. The objective of this study was to compare traditional face-to-face examinations and open-book examinations with and without supervision as perceived by teachers and students considering the characteristics of authentic assessment.

Statistical tests confirm the validity of the instrument used to analyze authentic assessment as conceived by Boud (2020), Ibarra-Sáiz et al. (2020), Herrington and Herrington (1998) and Williams and Wong, (2009), who consider that authentic assessment must have certain characteristics. Among these characteristics are precisely that it be challenging, that it be linked to professional practice, that it be

adapted to the learning style of students, that it promotes complex learning, and that it be consistent with the pedagogical approach.

Next, we will reflect on the differences and similarities found according to the perspective of the main parties involved in the process: teachers and students. We would like to begin by emphasizing that one of the major criticisms made of summative evaluation in the different courses of higher education is that it often does not correspond to the didactic methodology or the pedagogical approach, so it is considered that there is incoherence in relation to the evaluative model (Cano, 2008). This was evident in many cases during the COVID-19 pandemic, in which professors only transformed their face-to-face exams to digital format to carry out their course evaluations, as pointed out by Slade et al. (2021). In this study, the results indicate that the open-book exam model with or without proctoring as perceived by the professors was more in line with the pedagogical approach of the course than traditional face-to-face exams. This may represent an indication of success in the change of the assessment model that was adopted due to the need to take the exams online because of the pandemic. However, for students the traditional face-to-face exams were also consistent with the pedagogical model of the course, as were the open-book exams. This difference may be because the face-to-face exams also had a competency-based assessment approach, with case studies or simulations of educational problems that students had to solve in textual form. From the students' perspective, this coincided with the practical and applied approach of the subjects taken.

Open-book exams promote more complex learning that involves synthesizing, creating, reflecting, and making decisions, as pointed out by Feller (1994), Vázquez et al. (2021), Williams & Wong, (2009), and encourage less rote learning. In this aspect, both teachers and students agreed. Likewise, both considered that open-book exams represented an intellectual challenge, as highlighted by Herrington & Herrington, (1998) and Koutselini (1997). In this case, they had to create a digital product, such as an infographic, video, interactive presentation to answer the exam question. In this sense, teachers consider traditional exams to be less demanding, but students perceive that they are also intellectually challenging, although less so than open-book exams. This may be due to the characteristics of the face-to-face exam, although it had to be solved textually, practical situations had to be solved. The demands posed in the open-book exams with or without proctoring were more related to the professional practice of the future teachers and were more attractive than the face-to-face exams, and this was agreed upon by teachers and students. This is consistent with the results of Williams & Wong, (2009) and reaffirms that this type of exams is more suitable for authentic assessment (Boud, 2020; Ibarra-Sáiz et al., 2020; Herrington & Herrington, 1998).



The issue of plagiarism in assessments in higher education courses became one of the most contentious in the early days of the pandemic, many teachers lacked confidence that students would comply with academic honesty in online assessments (Cheung, 2020; Guangul et al., 2020; Gudiño et al., 2021; Shaushenova et al., 2021). This study confirms that teachers felt that open-book exams with or without proctoring offered the same potential for plagiarism as face-to-face exams, which converges with the position of Butler-Henderson & Crawford (2020). The vulnerability of academic honesty in these types of assessments with exams is evident regardless of their modality: face-to-face or online with or without proctoring. Perhaps it would be more effective to avoid plagiarism to train students to perform honest practices in their evaluations, as proposed by Halak & El-Hajjar (2019). Or implement another type of assessment more focused on project development, oral presentations, or collaborative work with comprehensive peer to peer assessment, focused on solving real educational situations as proposed by Boud (2020), Ibarra-Sáiz et al. (2020), McArthur (2020) and Williams & Wong, (2009). However, results in other studies have determined that there is no plagiarism in either open-book (Cheung, 2020) or proctored (Gudiño et al., 2021; Shaushenova et al., 2021) examinations.

Contrary to the perception of professors, for students, open-book exams with or without supervision do not allow plagiarism. This can be explained by the demands of this exam model in which students had to apply knowledge, solve challenging situations, create digital content, decide which web applications to use, test them and give an answer in a maximum time of two hours. This demands complex learning from students as highlighted by Brown and Glasner, (1999), Koutselini, (1997) and Williams & Wong, (2009) in relation to open-book exams.

For the teachers, the open-book exams made a big difference in relation to the face-to-face exams, although for the students they did not. This may have been due to the challenge for the teachers to reconfigure the exam approaches so that they had different questions in the same subject and in that subject the answer was required to be given through the same digital product, e.g., a video or an infographic. They had to work collaboratively among the three or four teachers of the same subject and agree on the evaluation rubrics so that the exams were as equivalent as possible. We felt that it was mainly the novelty in the type of exam that made the difference compared to traditional exams. Open-book exams were being implemented for the first time, which demanded more thought when designing the exams (Soodmand & Ranjbar, 2021). Additionally, as the proctoring system penalizes web queries, permissions had to be requested from the university to allow students to query the web, get up from their seats, and record with the smartphone camera in cases where videos were requested or record the operation of a programmed robot during the exam.

At this point we would like to highlight that the application of the proctor, did not affect the perception of the open-book exams, which may be a point in favour of the supervision of online exams to ensure academic honesty as verified by Gudiño et al. (2021) and Shaushenova et al. (2021). Likewise, the implementation of monitoring systems in assessment has an impact on the quality of distance education outcomes and in this we agree with Shaushenova et al. (2021). We also agree with these authors that it will be necessary to improve supervision systems from the technological point of view and to impact on innovation in educational evaluation. This would solve a need expressed by several higher education institutions, especially during the context of the COVID-19 pandemic, but which can be installed in evaluation practices in the post-pandemic.

## **CONCLUSION**

Finally, we can conclude that there are no major differences between the perception of open-book exams with or without proctoring between teachers and students, they only differ in that for students the open-book exams without proctoring are much more challenging than the others and they consider that it is more difficult to commit plagiarism in open-book exams with or without proctoring. Likewise, both teachers and students perceive a significant difference between traditional face-to-face exams and open-book exams. According to this study, it can also be concluded that open-book exams with or without proctoring constituted a relevant alternative for the final evaluation in online courses. However, it would be necessary to verify whether these results are confirmed in other studies within the same institution or others in which the same examination modality is applied under the same conditions. Nevertheless, the evaluation proposal according to the empirical background and theoretical contributions is pertinent to assume evaluation as authentic, self-reflective learning, adapted to real professional demands.

It would be necessary to overcome some methodological limitations, regarding the control of certain variables that can enrich this study, for example, to know the previous level of digital competences of the students. A qualitative-quantitative research design could be implemented, complemented with interviews or a focus group to follow up on the perception of the exam and confirm whether it was a learning and competency consolidation experience. It is also recommended to address other aspects such as the levels of anxiety associated with proctored examination models and the satisfaction of teachers and students with this type of evaluation.

Based on these results, it could be recommended to apply this model of examination and assessment in higher education institutions with online courses that wish to join the current trends in assessment.

## REFERENCES

- Boud, D. (2020). Retos en la reforma de la evaluación en educación superior: una mirada desde la lejanía. *RELIEVE-Revista Electrónica de Investigación y Evaluación Educativa*, 26(1). <https://doi.org/10.7203/relieve.26.1.17088>
- Brown, S. (2015). La evaluación auténtica: el uso de la evaluación para ayudar a los estudiantes a aprender. *RELIEVE-Revista Electrónica de Investigación y Evaluación Educativa*, 21(2). <https://doi.org/10.7203/relieve.21.2.7674>
- Brown, S., & Glasner, A. (1999). *Assessment matters in higher education*. Open University Press.
- Byrne, B. (2009). *Structural Equation Modeling with AMOS (2ª Ed.)*. Routledge.
- Butler-Henderson, K., & Crawford, J. (2020). A systematic review of online examinations: A pedagogical innovation for scalable authentication and integrity. *Computers & Education*, 159. <https://doi.org/10.1016/j.compedu.2020.104024>
- Callegaro, M., Manfreda, K. L., & Vehovar, V. (2015). *Web survey methodology*. Sage.
- Cano, E. (2008). La evaluación por competencias en la educación superior. *Profesorado. Revista de Currículum y Formación de Profesorado*, 12(3), 1-16. <https://www.redalyc.org/articulo.oa?id=56712875011>
- Cheung, C. (2020). Evaluation of academic integrity of online open book assessments implemented in an undergraduate medical radiation science course during COVID-19 pandemic. *Journal of Medical Imaging and Radiation Sciences*, 51(4), 610-616. <https://doi.org/10.1016/j.jmir.2020.09.009>
- Fardoun, H., González-González, C., Collazos, C., & Yousef, M. (2020). Estudio exploratorio en iberoamérica sobre procesos de enseñanza-aprendizaje y propuesta de evaluación en tiempos de pandemia. *Education in the Knowledge Society*, 21, 17. <https://doi.org/10.14201/eks.23537>
- Feller, M. (1994). Open-book testing and education for the future. *Studies in Educational Evaluation*, 20(2), 235-238. [https://doi.org/10.1016/0191-491X\(94\)90010-8](https://doi.org/10.1016/0191-491X(94)90010-8)
- Fritz, C., Morris, P., & Richler, J. (2012). Effect size estimates: current use, calculations, and interpretation. *Journal of Experimental Psychology: General*, 141(1), 2-18. <https://psycnet.apa.org/doi/10.1037/a0024338>
- García-Alberti, M., Suárez, F., Chiyón, I., & Mosquera, J. (2021). Challenges and Experiences of Online Evaluation in Courses of Civil Engineering during the Lockdown Learning Due to the COVID-19 Pandemic. *Education Sciences*, 11(2), 59. <https://doi.org/10.3390/educsci11020059>
- García-Peñalvo, F., 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 & A. Tabacco (Eds) *Radical Solutions*

- for Education in a Crisis Context. *Lecture Notes in Educational Technology*. Springer. [https://doi.org/10.1007/978-981-15-7869-4\\_6](https://doi.org/10.1007/978-981-15-7869-4_6)
- Guangul, F., Suhail, A., Khalit, M., & Khidhir B. (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*, 32, 519–535. <https://doi.org/10.1007/s11092-020-09340-w>
- Gudiño, S., Jasso, F., & de La Fuente, J. (2021). Remote proctored exams: Integrity assurance in online education? *Distance Education*, 42(2), 200-218. <https://doi.org/10.1080/01587919.2021.1910495>
- Hair, J., Black, W., Babin, B., & Anderson, R. (2014). *Multivariate Data Analysis* (7<sup>a</sup> Ed.). Pearson.
- Halak, B., & El-Hajjar, M. (2019). Design and evaluation of plagiarism prevention and detection techniques in engineering education. *Higher Education Pedagogies*, 4(1), 197-208. <https://doi.org/10.1080/23752696.2018.1563757>
- Herrington, J., & Herrington, A. (1998). Authentic assessment and multimedia: how university students respond to a model of authentic assessment. *Higher Education Research and Development*, 17, 3, 305–322.
- Hu, L., & Bentler, P. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling*, 6, 1–55. <http://dx.doi.org/10.1080/10705519909540118>
- Ibarra-Sáiz, M., & Rodríguez-Gómez, G. (2020). Aprendiendo a evaluar para aprender en la educación superior. *Revista Iberoamericana de Evaluación Educativa*, 13(1), 5–8. <https://revistas.uam.es/riee/article/view/12070>
- Ibarra-Sáiz, M., Rodríguez-Gómez, G., Boud, D., Rotsaert, T., Brown, S., Salinas Salazar, M. L., & Rodríguez Gómez, H. M. (2020). El futuro de la evaluación en la educación superior. *RELIEVE-Revista Electrónica de Investigación y Evaluación Educativa*, 26(1), 1-6. <https://doi.org/10.7203/relieve.26.1.17323>
- Koutselini, M. (1997). Testing and life-long learning: Open-book and closed-book examination in a university course. *Studies in Educational Evaluation*, 23, 2, 131-139. [https://doi.org/10.1016/S0191-491X\(97\)00008-4](https://doi.org/10.1016/S0191-491X(97)00008-4)
- Li, C. H. (2014). *The performance of MLR, USLMV, and WLSMV estimation in structural regression models with ordinal variables*. Michigan State University. <https://tinyurl.com/3fypvhos>
- McArthur, J. (2020). Participación e implicación del estudiante en la evaluación: implicar a todo el estudiante en la búsqueda de la justicia y el bien social. *RELIEVE-Revista Electrónica de Investigación y Evaluación Educativa*, 26(1). <https://doi.org/10.7203/relieve.26.1.17089>
- Nunnally, J. C. & Bernstein, I. H. (1994) *Psychometric theory*. McGraw-Hill.

- Organización para la Cooperación y el Desarrollo Económicos. (2019). *OECD Skills Strategy 2019. Skills to Shape a Better Future*. <https://www.oecd.org/skills/oecd-skills-strategy-2019-9789264313835-en.htm>
- Otzen, T., & Manterola, C. (2017). Técnicas de Muestreo sobre una Población a Estudio. *International journal of morphology*, 35(1), 227-232. <http://dx.doi.org/10.4067/S0717-95022017000100037>
- Pagram, J., Cooper, M., Jin, H., & Campbell. (2018). Tales from the Exam Room: Trialing an E-Exam System for Computer Education and Design and Technology Students. *Education Sciences*, 8, 4. 8. <https://doi.org/10.3390/educsci8040188>
- Prigoff, J., Hunter, M., & Nowygrod, R. (2021). Medical student assessment in the time of COVID-19. *Journal of Surgical Education*, 78(2), 370-374. <https://doi.org/10.1016/j.jsurg.2020.07.040>
- Shaushenova, A., Zulpykhar, Zh., Zhumasseitova, S., Ongarbayeva, M., Akhmetzhanova, Sh., Mutalova, Zh., Niyazbekova, Sh., & Zueva, A. (2021). The influence of the proctoring system on the results of online tests in the conditions of distance learning. *AD ALTA: Journal of Interdisciplinary Research*, 11(2) 250-256.
- Slade, C., Lawrie, G., Taptamat, N., Browne, E., Sheppard, K. & Matthews, K. (2021). Insights into how academics reframed their assessment during a pandemic: disciplinary variation and assessment as afterthought. *Assessment & Evaluation in Higher Education*, 47(4), 588-605. <https://doi.org/10.1080/02602938.2021.1933379>
- Soodmand, H. & Ranjbar, N. (2021). EAP teachers' assessment literacy: From theory to practice. *Studies in Educational Evaluation*, 70. <https://doi.org/10.1016/j.stueduc.2021.101042>
- United Nations Educational, Scientific and Cultural Organization (2021). *One year into COVID: prioritizing education recovery to avoid a generational catastrophe*. <https://unesdoc.unesco.org/ark:/48223/pf0000376984>
- Vázquez, J., Chiang, E., & Sarmiento, I. (2021). Can we stay one step ahead of cheaters? A field experiment in proctoring online open book exams. *Journal of Behavioral and Experimental Economics*, 90. <https://doi.org/10.1016/j.socec.2020.101653>
- Williams, J., & Wong, A. (2009). The efficacy of final examinations: A comparative study of closed-book, invigilated exams and open-book, open-web exams. *British Journal of Educational Technology*, 40(2), 227–236. <https://doi.org/10.1111/j.1467-8535.2008.00929.x>
- Xia, Y. (2016). *Investigating the chi-square-based model-fit indexes for WLSMV and ULSMV estimators* [Doctoral Dissertation]. Florida State University Libraries. <https://bit.ly/3lISnc6>

## Appendix 1

### Description of model exams

**Model A** is an exam that has two parts, one with comprehension questions and the other with development questions or case studies. The comprehension questions propose to relate concepts, analyze a situation, or evaluate a strategy or methodology. In the development questions, practical situations are proposed, such as the design of a methodological strategy in which active methodologies are applied and technological tools are incorporated. The answers are presented in writing and the exam is taken in face-to-face mode, without the support of any kind of materials. It is an exam model more focused on an evaluation for learning (Carless, 2015) in which specific and transversal competencies and skills are demanded (capacity for analysis, application and decision making, among others).

**Model B** is an exam in which two approaches are presented and the student chooses one. Two situations are proposed, for example, a problem in the classroom or the development of an educational activity with active methodologies and technology support, or an approach to improve the educational center. In all cases, technological tools must be used. The answer is elaborated as a digital product that can be a video, an infographic, a virtual classroom, an interactive presentation, etc. The students send the url of the digital product as an answer to the exam. It is presented in online mode with a proctoring system, and it is an open book exam (sources can be consulted on the internet). It is an exam that presents a quality evaluative activity, which is challenging for the student who must solve the problem in two hours, think of solutions, create educational digital content, use only material with creative commons licenses, consider the characteristics of their students or their educational center to which the solution or educational proposal supported by technologies will be directed. From this perspective, it is a test focused on evaluation as learning (Ibarra-Sáiz & Rodríguez-Gómez, 2020), in which the student constructs his own learning and strengthens his competence development. It also meets the characteristics of an authentic assessment in the terms of Herrington and Herrington (1998); Williams and Wong, (2009), Boud (2020) and Ibarra-Sáiz et al. (2020) in terms of being challenging, linked to professional practice, adapting to the learning style of students, promoting complex learning and being coherent with the pedagogical approach.

**Model C.** has the same characteristics as model B only in that it is performed without a proctoring system. So, the student can record his own videos to explain for example how he has done the programming in Scratch or show the operation of the SpheroMini robot he has programmed during the exam and send the url of the video and the programmed project.