


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

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

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). Percepción de docentes y estudiantes de educación superior de los exámenes a libro abierto y supervisados en la pandemia por COVID-19 [Higher education faculty 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

### 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 a libro abierto con y sin proctoring según la percepción de docentes y estudiantes. La investigación es de carácter empírico y enfoque cuantitativo y se basa en las respuestas de 66 docentes y 301 estudiantes 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 docentes y estudiantes los exámenes a libro abierto con o sin proctoring no presentan diferencias significativas y se ajustan más a una evaluación auténtica que los exámenes presenciales. Se concluye que los exámenes a libro abierto con o sin proctoring 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

## 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 analyse the relevance of this change in evaluation. The main objective is to analyse 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

## INTRODUCCIÓN

El cierre de la educación, debido a la pandemia mundial de Covid19, ha llevado a las instituciones educativas a recurrir a la virtualidad para sobrevivir. Los profesores se han visto obligados a cambiar sus prácticas cotidianas, incluida la forma de evaluar a los alumnos. Ante las dificultades para evaluar las competencias de un alumnado confinado, sin posibilidad de realizar exámenes presenciales, los países adoptaron diferentes estrategias para adaptarse a la complicada situación (Fardoun, et al., 2020). Éstas se orientaron hacia la flexibilidad y el uso de medios en línea: exámenes orales, escritos supervisados, trabajos colaborativos, pruebas, exámenes a libro abierto, etc. Recuperar la educación, hacerla posible, debía ser una prioridad para los países para evitar una catástrofe generacional (Organización de las Naciones Unidas para la Educación, la Ciencia y la Cultura [UNESCO], 2021).

En este sentido, algunos estudios (García-Peñalvo et al., 2021; Guangul et al., 2020) señalan el proceso de evaluación como el mayor reto al que se ha enfrentado la educación superior en esta pandemia y hacen recomendaciones para dicho proceso único en el que se debe considerar la evaluación continua, garantizando la identificación del estudiante en los exámenes finales, utilizando la tecnología a la que el estudiante está acostumbrado, sustituyendo las pruebas escritas tradicionales por trabajos, ejercicios, infografías, etc. En concreto, se sugiere que en los exámenes virtuales los alumnos puedan enfrentarse a modelos de evaluación en los que se exija una presentación online realizada por ellos mismos, de forma que se verifique la identidad y la autoría, enfrentándose así al gran reto de la deshonestidad (Guangul et al., 2020; Gudiño et al., 2021; Shaushenova et al., 2021).

Otras investigaciones se han centrado en el estudio de la experiencia de alumnos y profesores con la evaluación en línea durante el encierro (García-Alberti, 2021), encontrando importantes retos como la falta de preparación de los profesores para enfrentarse a la evaluación a distancia y su desconfianza hacia este tipo de pruebas. Para los profesionales de la educación, la posibilidad de hacer trampas, la fiabilidad de la tecnología, la conectividad, los problemas con el hardware, la usabilidad, la facilidad para corregir y crear pruebas, son preocupaciones fundamentales (Butler-Henderson & Crawford, 2020; Halak & El-Hajjar, 2019). Si a esto se añaden las carencias detectadas en algunos profesores en cuanto a su formación en evaluación (Soodmand & Ranjbar, 2021), la tarea de medición online de las competencias de los estudiantes universitarios se complica. Estudios recientes señalan que los docentes, para reformular la evaluación durante la pandemia, han optado preferentemente por trasladar sus exámenes tradicionales a un formato en línea, dada la emergencia (Slade et al., 2021).

La educación superior tiene un papel primordial en la identificación de soluciones a estos retos que la evaluación online presenta a las universidades (Pagram et al., 2018). En ese sentido, con este estudio queremos analizar una propuesta de cambio

en la evaluación en línea hacia los exámenes a libro abierto como mecanismo para afrontar el reto de mantener la calidad de la educación durante la pandemia.

### **Antecedentes teóricos**

Una de las estrategias que las instituciones de educación superior han adoptado para hacer frente a los exámenes en línea ha sido el uso de diferentes modalidades de evaluación, entre ellas, los exámenes a libro abierto (Cheung, 2020). En estos exámenes se permite la consulta de materiales de clase, apuntes, la web. Fue un modelo de examen propuesto por algunos autores que asociaron esta modalidad con la educación del futuro y las competencias de alto nivel (Feller, 1994; Herrington & Herrington, 1998; Koutselini, 1997). Antes de la pandemia por Covid-19, los estudios (Brown & Glasner, 1999; Williams & Wong, 2009) destacaban este tipo de examen por permitir a los estudiantes utilizar de forma creativa los conocimientos adquiridos durante el curso, reflexionar profundamente sobre los problemas y aplicar el pensamiento crítico, en lugar de limitarse a recordar la información o utilizar lo que recuerdan. Al mismo tiempo, el examen a libro abierto anima a los estudiantes a realizar un estudio más profundo del contenido, a evaluar sus lagunas de aprendizaje durante la preparación del examen, a autoevaluar los resultados de su preparación y a reducir el estrés. Por otra parte, generalmente las condiciones de los exámenes tradicionales son poco realistas, no representan situaciones reales del mundo laboral (Feller, 1994), por lo que resulta complicado medir determinadas competencias. Para resolver los problemas cotidianos, los profesionales recurren a todos los recursos que tienen a su disposición, incluido Internet, lo que es incompatible con los exámenes a libro cerrado. La evaluación auténtica según Boud (2020), Ibarra-Sáiz et al. (2020) y Herrington & Herrington (1998) requiere que el contexto en el que se realiza refleje las condiciones en las que el profesional tendrá que actuar, además, debe invitar al alumno a transformar sus conocimientos en acciones o productos creativos incluso en colaboración con sus compañeros. Por otro lado, implica la resolución de retos complejos que requieran de un pensamiento crítico, retos que estarán en consonancia con lo trabajado durante el curso y, finalmente, proporcionar indicadores de aprendizaje fiables. Sin duda, se trata de un modelo de evaluación que involucra más a los estudiantes como tomadores de decisiones y aplicadores del conocimiento recién construido, lo que induce a un aprendizaje más profundo (McArthur, 2020; Williams & Wong, 2009).

Por otro lado, dado que hoy en día los resultados de aprendizaje deben expresarse en términos de competencias, esta modalidad de examen a libro abierto es ideal si lo que se requiere es medir el espíritu emprendedor, la flexibilidad, la responsabilidad personal, la innovación o la creatividad de los estudiantes,

habilidades que la mayoría de los países valoran para sus profesionales (Organización para la Cooperación y el Desarrollo Económicos [OCDE], 2019). El conocimiento es cada vez más complejo y los problemas también, las habilidades requeridas a los profesionales del siglo XXI como la búsqueda y evaluación de información o la gestión de situaciones profesionales complejas, se ejercitan en los exámenes a libro abierto (Feller, 1994). Sin embargo, para que esto sea posible, es necesario diseñar exámenes coherentes con modelos basados en competencias como simulaciones o proyectos (Cano, 2008). El diseño de exámenes a libro abierto no es una carga adicional para los docentes; sin embargo, requiere una preparación especial para diseñar pruebas que sean creativas, significativas, vinculadas a la realidad de cada profesional y que permitan a los estudiantes demostrar las competencias que se les exigen (Cano, 2008; Feller, 1994). Además, requieren un cambio en el enfoque pedagógico, para que los profesores orienten sus sesiones a la enseñanza de competencias de alto nivel. No podemos olvidar que la evaluación no sólo pretende medir el aprendizaje, sino también favorecerlo.

Dadas estas características, los exámenes a libro abierto se convierten en una modalidad relevante en las pruebas online cuyo objetivo es una evaluación auténtica, en el sentido de ser realista, capaz de medir competencias, considerando al estudiante como protagonista de su aprendizaje. En este mismo sentido, desde la pandemia por Covid-19, están reapareciendo estudios en los que se presentan los resultados de experiencias en educación superior con este tipo de exámenes, dada su adecuación a la emergencia vivida. Un ejemplo relevante es el experimento realizado por Vázquez et al. (2021), que analiza el rendimiento de los estudiantes en exámenes online a libro abierto con proctoring y sin proctoring, o el estudio realizado por Prigoff et al. (2021), que evalúa la idoneidad de los exámenes a libro abierto en estudiantes de medicina.

El propósito de este estudio es contribuir a esta línea de investigación sobre los exámenes a libro abierto y conocer su ajuste a las demandas de aprendizaje actuales y a las nuevas tendencias de evaluación en la educación superior. Contar con instrumentos de evaluación fiables y válidos que faciliten la investigación sobre este tema permitirá avanzar en este campo de estudio (Nunnally y Bernstein, 1994). Asimismo, aprovechar la experiencia de los principales actores de este proceso de evaluación, como son docentes y estudiantes, contribuye a tener una valoración realista de los exámenes a libro abierto. Por lo tanto, el objetivo general de este estudio fue comparar los modelos de examen utilizados antes y durante la pandemia según la percepción de docentes y estudiantes y analizar las diferencias en relación con las características de la evaluación auténtica. Para ello, se propusieron los siguientes objetivos específicos:

1. Conocer la percepción de docentes sobre el modelo de examen tradicional y el modelo de examen a libro abierto con y sin proctoring.

2. Conocer la percepción de estudiantes sobre el modelo de examen tradicional y el modelo de examen a libro abierto con y sin proctoring.
3. Comparar la valoración de estudiantes y profesores de los exámenes tradicionales y de los exámenes a libro abierto con y sin proctoring.

## MÉTODO

### Diseño

En la realización de este estudio se aplicó una metodología cuantitativa, con un diseño de investigación a través de una encuesta online (Callegaro et al.; 2015) para conocer la percepción que tienen los docentes y estudiantes de los diferentes modelos de exámenes realizados (A, B, C). (Anexo 1). Esta metodología y diseño se consideró la más pertinente para acceder a la información, por la facilidad que se ofreció a los estudiantes para responder, ya que habían terminado su periodo escolar y su participación en el estudio fue totalmente voluntaria.

### Población y muestra

La población de este estudio estuvo formada por todos los estudiantes que cursan el Máster Universitario en Tecnología Educativa y Competencias Digitales (MUTECD) de la Universidad Internacional de La Rioja (España) y los docentes de las asignaturas desde noviembre de 2019 hasta julio de 2020. El procedimiento de muestreo fue no probabilístico intencional por proximidad (Otzen & Manterola, 2017). Tras la realización de los exámenes, se pidió a alumnos y profesores que respondieran al cuestionario sobre los modelos de examen. La muestra consistió en 302 estudiantes, de los cuales 195 del primer semestre respondieron al cuestionario sobre el modelo C y 107 del segundo semestre respondieron al cuestionario sobre los modelos A y B. La muestra de profesores fue de 37. Todos ellos respondieron al cuestionario sobre el modelo A, y además 17 respondieron al cuestionario sobre el modelo B y 19 respondieron al cuestionario sobre el modelo C.

### Instrumento

Los datos se recogieron mediante un cuestionario basado en el utilizado por Williams y Wong (2009), diseñado para evaluar la eficacia de los exámenes a libro abierto. La encuesta consta de 9 ítems. *1.El tiempo para realizar el examen (2 horas) era más o menos correcto. 2.El tipo de examen: 2. requería que el alumno acudiera en persona a realizar el examen; 3. estaba en línea con la pedagogía/metodología, 4. promovía un aprendizaje más complejo (crear, reflexionar, etc.) y*

*menos memorístico, 5.es intelectualmente desafiante, 6.se adapta mejor al estilo de aprendizaje de cada alumno, 7. está más relacionado con la práctica profesional, 8. El contenido del examen era atractivo, 9. El tipo de examen permitía hacer trampas (plagio, copia, etc.).* Las respuestas se indican en una escala Likert de 5 puntos (1=Muy en desacuerdo, 2=Desacuerdo, 3=De acuerdo, 4=Muy de acuerdo, 5=No sabe/no contesta). El cuestionario se completó a través de la aplicación *Forms de Office 365* (Cuestionario). Los datos se recogieron en junio y julio de 2021. Se aplicaron las mismas preguntas para los tres modelos de examen: modelo A: examen presencial tradicional, modelo B: examen a libro abierto con proctoring, modelo C: examen a libro abierto sin proctoring.

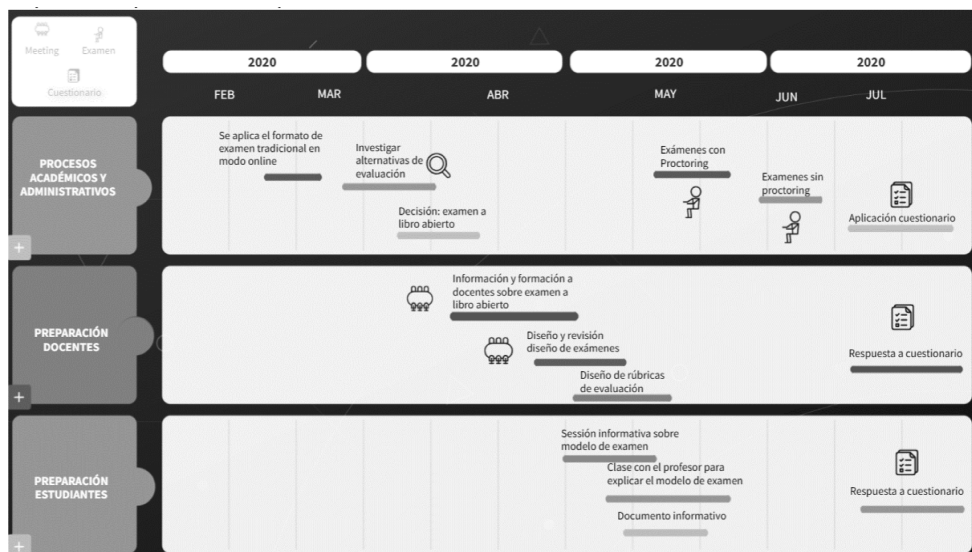
## **Procedimiento**

Se querían evaluar los diferentes modelos de examen porque en la situación de la pandemia por COVID-19 se implementaron los exámenes online en el MUTECD. Se decidió hacer un cambio en los modelos de examen que tradicionalmente se aplicaban en modalidad presencial (Fig.1) a un modelo de examen a libro abierto orientado a una evaluación auténtica que pudiera ser utilizada como una oportunidad de aprendizaje. Estos exámenes presentaban planteamientos desafiantes y situaciones reales en las que el alumno debía resolver casos, simulaciones, asumiendo el papel de experto, proponiendo una solución a un problema en el aula o en un centro educativo, etc. Las respuestas se presentaban en forma de producto digital como un vídeo, infografía, podcast, presentación interactiva u otros. Asimismo, los alumnos tenían la opción de conectarse a la red y utilizar las aplicaciones tecnológicas que considerasen más relevantes en cada caso, no sólo para mostrar sus conocimientos sino también para aplicarlos, reflexionar, tomar decisiones, crear contenidos digitales y compartirlos. El diseño de este modelo de examen se hizo de forma colaborativa entre los profesores de la misma asignatura. También se les pidió que crearan una rúbrica común para la evaluación de las competencias de los alumnos (Feller, 1994).

Además, se informó a los alumnos sobre cómo se iba a evaluar; por un lado, recibieron sesiones informativas por parte de la dirección del máster y de los profesores; por otro, se publicó en las aulas virtuales de todas las asignaturas un documento que explicaba las características del examen. Finalmente, se realizaron los exámenes, se aplicó la encuesta online a alumnos y profesores y se analizaron los datos recogidos.

**Figura 1**

*Esquema del procedimiento para cambiar el modelo de examen*



Fuente. Elaborado por las autoras con Genial.ly (online)

## Análisis estadístico

Para garantizar la validez del instrumento, se realizó un análisis factorial confirmatorio ordinal con la herramienta Mplus 8.0. Previamente, se realizó el análisis de normalidad con el coeficiente de Mardia, se encontró que los estadísticos de curtosis y simetría multivariante mostraron resultados significativos, por lo que no se puede asumir la normalidad de la distribución del conjunto de ítems del cuestionario (Tabla 1).

**Tabla 1**

*Resultados de curtosis y asimetría multivariada y según los coeficientes de Mardia*

	Coefficientes	z	$\chi^2$	df	p
Asimetría	24.1		1123	84	<.001
Curtosis	114.8	38.6			<.001



Posteriormente, a partir de una matriz de correlaciones policóricas, se estiman los parámetros con el método de mínimos cuadrados ponderados robustos (WLSMV), adecuado en estos casos (Li, 2014 y Xia, 2016). En la fase de evaluación del modelo se utilizan los índices normalizados: chi-cuadrado robusto normalizado ( $\chi^2/d.f.$ ), para evaluar el ajuste global, donde se consideran aceptables valores entre 3 y 5 (Hair et al., 2014); RMSEA, para evaluar la matriz residual, que es aceptable con valores inferiores a .08 (Byrne, 2009); y CFI y TLI (Tucker-Lewis Index), de ajuste comparativo, que son aceptables a partir de .90 (Hair et al., 2014). Siguiendo a Hu y Bentler (1999) como evidencia de validez, un ajuste aceptable en la combinación de estos índices es suficiente. Los resultados del análisis factorial confirmatorio indican que el modelo unidimensional está validado. Los índices de ajuste muestran valores aceptables en todos los casos: chi-2/Gr inferior a 3, RMSEA no significativo (indicando, por tanto, que es igual a cero) y CFI y TLI superiores a .95 (Tabla 2).

**Tabla 2**  
*Resultados de los índices de ajuste Chi-2 y RMSEA*

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

También se realiza un estudio de la validez convergente calculando la varianza media explicada por el modelo (Average Variance Extracted, AVE),

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

Además, la consistencia interna se estima calculando la fiabilidad compuesta, también llamada coeficiente omega o consistencia interna estimada por SEM (Viladrich et al.; 2017), se utilizan las cargas factoriales (P) de los j ítems y la varianza del error (ei), con la siguiente fórmula:

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

Los resultados de la validez convergente también apuntan a un buen ajuste de la dimensión, alcanzando una varianza total explicada por la dimensión de evaluación auténtica de 0,607 (60%) y un coeficiente de fiabilidad compuesto u omega de 0,882.

También se realizó una comparación de los resultados de cada ítem y de la puntuación total de la escala mediante la prueba no paramétrica de Kruskal-Wallis. Y se incluyen comparaciones de modelos emparejados (A vs B; A vs C y B vs C). Este análisis se realizó de forma comparativa para alumnos y profesores. Se compararon las diferencias estadísticas de los rangos promedio de cada grupo (U de Mann-Whitney). Se consideró para los análisis el ajuste de Bonferroni que da una mejor significación cuando hay grupos pequeños. También se incluye la D de Cohen para el efecto de las diferencias, donde los intervalos para r: 0,1 a 0,3: se considera efecto pequeño; 0,3 a 0,5: efecto intermedio; 0,5 y más: efecto fuerte, según Fritz et al. (2012) basado en Cohen.

Para realizar el análisis descrito se utilizó el SPSS versión 27. El proceso de validación se llevó a cabo con la base de datos de los alumnos porque era mucho mayor que la de los profesores.

## RESULTADOS

A continuación, se presentan análisis descriptivos y comparativos de las percepciones de docentes y estudiantes sobre los modelos de examen.

La tabla 3 muestra los estadísticos descriptivos para cada uno de los ítems de cada modelo de examen y según los resultados de los docentes y estudiantes. También se incluyen el índice de significación asintótica bilateral de Kruskal-Wallis y el índice de Cohen.

**Tabla 3**

*Estadísticas descriptivas de las variables en los diferentes modelos de examen según los docentes y estudiantes, índice de significación asintótica bilateral de Kruskal Wallis e índice D de Cohen*

		Modelo de evaluación			Kruskal-Wallis			D. Cohen
		A Media (DT)	C Media (DT)	B Media (DT)	N	Estadísticos de prueba	Significación asintótica. (prueba bilateral)	
I_1	Docentes	3.41 (0.837)	3.76 (0.437)	3.71 (0.849)	66	3.635a.b	0.162	0.326
	Estudiantes	2.79 (0.930)	2.75 (0.951)	3.01 (0.914)	302	4.607a.b	0.100	0.188
I_2	Docentes	2.22 (1.128)	1.31 (0.793)	1.00 (0.000)				
	Estudiantes	1.62 (0.885)	1.45 (0.794)	1.61 (0.921)				

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

		Modelo de evaluación			Kruskal-Wallis			
		A	C	B	N	Estadísticos de prueba	Significación asintótica. (prueba bilateral)	D. Cohen
		Media (DT)	Media (DT)	Media (DT)				
I_3	Docentes	2.72 (1.143)	3.88 (0.332)	3.88 (0.342)	65	20.606a	0.000	1.31
	Estudiantes	3.25 (0.952)	3.71 (0.559)	3.49 (0.745)	295	14.079a	0.001	0.415
I_4	Docentes	2.75 (0.984)	3.82 (0.529)	3.94 (0.250)	65	24.874a	0.000	1.529
	Estudiantes	3.15 (1.049)	3.75 (0.539)	3.58 (0.740)	294	20.674a	0.000	0.524
I_5	Docentes	2.72 (0.888)	3.53 (0.624)	3.88 (0.342)	65	22.298a	0.000	1.395
	Estudiantes	3.22 (0.967)	3.65 (0.702)	3.54 (0.706)	291	12.108a	0.002	0.381
I_6	Docentes	2.53 (1.047)	3.41 (0.712)	3.81 (0.403)	65	19.412a	0.000	1.25
	Estudiantes	2.95 (1.031)	3.59 (0.692)	3.38 (0.793)	289	22.782a	0.000	0.56
I_7	Docentes	2.50 (1.191)	3.94 (0.243)	3.81 (0.403)	65	25.145a	0.000	1.544
	Estudiantes	3.00 (1.116)	3.79 (0.497)	3.57 (0.728)	289	32.701a	0.000	0.694
I_8	Docentes	2.53 (1.016)	3.56 (0.512)	3.88 (0.342)	64	24.302a	0.000	1.518
	Estudiantes	2.95 (0.974)	3.58 (0.703)	3.42 (0.811)	291	23.959a	0.000	0.575
I_9	Docentes	2.45 (1.179)	2.35 (1.169)	2.44 (1.031)	64	0.090a.b	0.956	0.36
	Estudiantes	1.87 (0.911)	1.50 (0.823)	1.55 (0.942)	283	11.315a	0.003	0.371
Total_mean	Docentes	2.71 (0.799)	3.57 (0.283)	3.70 (0.234)				
	Estudiantes	3.05 (0.760)	3.53 (0.438)	3.43 (0.537)				

Nota. No hay datos para Kruskal Wallis en el ítem dos porque no hubo variabilidad.

En el caso de docentes y estudiantes existen diferencias estadísticamente significativas en algunos aspectos. A continuación, estas diferencias se especifican con la comparación entre los modelos de examen. En algunos ítems no hubo diferencias en los resultados entre docentes y estudiantes como, por ejemplo, en el ítem 1 (profesores  $p=.162$ ; estudiantes  $p=.100$ ), por lo que se puede indicar que tanto para docentes como para estudiantes el tiempo de realización del examen fue correcto en todos los modelos.

Al comparar los modelos de examen, se encontró que para los docentes los exámenes a libro abierto estaban más en línea con el enfoque pedagógico del curso que el modelo tradicional. Para los estudiantes, tanto el examen que les pedía responder a las preguntas de forma textual (modelo tradicional) como los modelos que les pedían crear un producto digital eran coherentes con el enfoque pedagógico del curso. Aunque destaca una mayor diferencia en el modelo a libro abierto sin proctoring (Tabla 4).

**Tabla 4**

*Comparaciones entre los modelos de evaluación para: El tipo de examen se ajustaba a la pedagogía/metodología utilizada durante el curso*

Fuente	Muestra 1-Muestra 2	Estadístico de prueba	Desv. Error	Desv. Estadístico de prueba	Sig.	Sig. ajustada <sup>a</sup>	D Cohen
Docentes	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
Estudiantes	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

Cada fila prueba la hipótesis nula de que las distribuciones de la Muestra 1 y la Muestra 2 son iguales. Se visualizan las significaciones asintóticas (pruebas bilaterales). El nivel de significación es .05.

<sup>a</sup> Los valores de significación se han ajustado mediante la corrección de Bonferroni para varias pruebas.

Los docentes creen que los exámenes a libro abierto con o sin proctoring promueven un aprendizaje más complejo (crear, reflexionar) y menos memorístico que los exámenes tradicionales. Sin embargo, los estudiantes creen que tanto los exámenes tradicionales como los a libro abierto promueven un aprendizaje más complejo y menos memorístico, aunque perciben una mayor diferencia entre los exámenes a libro abierto sin proctoring en comparación con los tradicionales (Tabla 5).

**Tabla 5**

*Comparaciones entre los modelos de evaluación para: El tipo de examen promueve el aprendizaje complejo (creación, reflexión, etc.) y menos memorístico*

Fuente	Muestra 1-Muestra 2	Estadístico de prueba	Desv. Error	Desv. Estadístico de prueba	Sig.	Sig. ajustada <sup>a</sup>	D Cohen
Docentes	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
Estudiantes	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

Cada fila prueba la hipótesis nula de que las distribuciones de la Muestra 1 y la Muestra 2 son iguales. Se visualizan las significaciones asintóticas (pruebas bilaterales). El nivel de significación es .05.

<sup>a</sup> Los valores de significación se han ajustado mediante la corrección de Bonferroni para varias pruebas.

Los docentes consideran que los exámenes a libro abierto con y sin proctoring son más exigentes intelectualmente que los exámenes tradicionales, pero los estudiantes consideran que todos los modelos de exámenes son exigentes intelectualmente, aunque los exámenes a libro abierto sin proctoring lo son más. Los otros dos modelos los consideran igualmente desafiantes (Tabla 6).

**Tabla 6**

*Comparaciones entre los modelos de evaluación para: El tipo de examen es un reto intelectual*

Fuente	Muestra 1-Muestra 2	Estadístico de prueba	Desv. Error	Desv. Estadístico de prueba	Sig.	Sig. ajustada <sup>a</sup>	D Cohen
Docentes	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
Estudiantes	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

Cada fila prueba la hipótesis nula de que las distribuciones de la Muestra 1 y la Muestra 2 son iguales. Se visualizan las significaciones asintóticas (pruebas bilaterales). El nivel de significación es .05.

<sup>a</sup> Los valores de significación se han ajustado mediante la corrección de Bonferroni para varias pruebas.

Tanto los docentes como los estudiantes consideran que los modelos de examen a libro abierto se adaptan mejor al estilo de aprendizaje de cada alumno, especialmente el modelo sin proctoring, frente al modelo tradicional (Tabla 7).

**Tabla 7**

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

Fuente	Muestra 1-Muestra 2	Estadístico de prueba	Desv. Error	Desv. Estadístico de prueba	Sig.	Sig. ajustada <sup>a</sup>	D Cohen
Docentes	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
Estudiantes	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

Cada fila prueba la hipótesis nula de que las distribuciones de la Muestra 1 y la Muestra 2 son iguales. Se visualizan las significaciones asintóticas (pruebas bilaterales). El nivel de significación es .05.

<sup>a</sup> Los valores de significación se han ajustado mediante la corrección de Bonferroni para varias pruebas.

Asimismo, docentes y estudiantes coinciden en que los modelos de examen a libro abierto están más relacionados con la práctica profesional que el modelo tradicional (Tabla 8). También, están de acuerdo en que los modelos de examen a libro abierto eran más atractivos que el modelo tradicional (Tabla 9).

**Tabla 8**

*Comparaciones entre los modelos de evaluación para: El tipo de examen está más relacionado con la práctica profesional*

Fuente	Muestra 1-Muestra 2	Estadístico de prueba	Desv. Error	Desv. Estadístico de prueba	Sig.	Sig. ajustada <sup>a</sup>	D Cohen
Docentes	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
Estudiantes	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

Cada fila prueba la hipótesis nula de que las distribuciones de la Muestra 1 y la Muestra 2 son iguales. Se visualizan las significaciones asintóticas (pruebas bilaterales). El nivel de significación es .05.

<sup>a</sup> Los valores de significación se han ajustado mediante la corrección de Bonferroni para varias pruebas.

**Tabla 9**

*Comparaciones entre los modelos de evaluación para: El contenido del examen era atractivo*

Fuente	Muestra 1-Muestra 2	Estadístico de prueba	Desv. Error	Desv. Estadístico de prueba	Sig.	Sig. ajustada <sup>a</sup>	D Cohen
Docentes	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
Estudiantes	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

Cada fila prueba la hipótesis nula de que las distribuciones de la Muestra 1 y la Muestra 2 son iguales. Se visualizan las significaciones asintóticas (pruebas bilaterales). El nivel de significación es .05.

<sup>a</sup> Los valores de significación se han ajustado mediante la corrección de Bonferroni para varias pruebas.

Según la percepción de los docentes, no hay diferencia entre los modelos de examen en cuanto a la posibilidad de cometer plagio; sin embargo, según los estudiantes, el examen tradicional permite el plagio frente a los modelos a libro abierto, independientemente de que sean con o sin proctoring (Tabla 10).

**Tabla 10**

*Comparaciones entre modelos de evaluación para: El tipo de examen permitía hacer trampas (plagio)*

Fuente	Muestra 1-Muestra 2	Estadístico de prueba	Desv. Error	Desv. Estadístico de prueba	Sig.	Sig. ajustada <sup>a</sup>	D Cohen
Estudiantes	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

Cada fila prueba la hipótesis nula de que las distribuciones de la Muestra 1 y la Muestra 2 son iguales. Se visualizan las significaciones asintóticas (pruebas bilaterales). El nivel de significación es .05.

<sup>a</sup> Los valores de significación se han ajustado mediante la corrección de Bonferroni para varias pruebas.

### Comparaciones del modelo de evaluación según docentes y estudiantes

En la prueba de Kruskal-Wallis los resultados son similares entre docentes y estudiantes, es decir, ambos perciben diferencias significativas entre los modelos A y C, y los modelos A y B y similitud entre los modelos C y B (Tabla 11).

**Tabla 11**

*Comparación de los modelos de evaluación tradicional presencial y a libro abierto con y sin proctoring para docentes y estudiantes*

Fuente	Muestra 1-Muestra 2	Estadístico de prueba	Desv. Error	Desv. Estadístico de prueba	Sig.	Sig. ajustada <sup>a</sup>	D Cohen
Docentes	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
Estudiantes	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

Cada fila prueba la hipótesis nula de que las distribuciones de la Muestra 1 y la Muestra 2 son iguales. Se visualizan las significaciones asintóticas (pruebas bilaterales). El nivel de significación es .05.

<sup>a</sup> Los valores de significación se han ajustado mediante la corrección de Bonferroni para varias pruebas.

## DISCUSIÓN Y CONCLUSIONES

Los retos que la educación superior tuvo que asumir para llevar a cabo procesos de evaluación sumativa en tiempos de pandemia la llevaron a innovar e implementar modalidades de evaluación diferentes a las aplicadas antes de la pandemia. En ese momento, al tener que asumir el desarrollo de los cursos y la evaluación en la modalidad en línea, era necesario aprovechar el momento para realizar las mejoras que las tecnologías y los formatos digitales podían permitir. Y al mismo tiempo asumir una modalidad de evaluación adecuada para la educación superior en la que ésta se asuma como una oportunidad de aprendizaje crítico, reflexivo y potenciador para el estudiante como recomiendan Boud (2020); Brown (2015); Ibarra-Sáiz et al. (2020); Ibarra-Sáiz & Rodríguez-Gómez (2020). En este sentido, en el máster de tecnología educativa y competencias digitales, se propuso realizar un cambio en los modelos de examen, pasando de exámenes presenciales a exámenes online a libro abierto con proctoring y sin proctoring. El objetivo de este estudio fue comparar los exámenes presenciales tradicionales y los exámenes a libro abierto con y sin proctoring, tal y como los perciben docentes y estudiantes, considerando las características de la evaluación auténtica.

Las pruebas estadísticas confirman la validez del instrumento utilizado para analizar la evaluación auténtica tal y como lo conciben Boud (2020), Ibarra-Sáiz et al. (2020), Herrington y Herrington (1998) y Williams y Wong, (2009), quienes consideran que la evaluación auténtica debe tener ciertas características. Entre estas características están precisamente que sea desafiante, que esté vinculada a la práctica profesional, que se adapte al estilo de aprendizaje de los estudiantes, que promueva el aprendizaje complejo y que sea coherente con el enfoque pedagógico.

A continuación, reflexionaremos sobre las diferencias y similitudes encontradas según la perspectiva de los principales implicados en el proceso: docentes y estudiantes. Queremos comenzar destacando que una de las mayores críticas que se hacen a la evaluación sumativa en los diferentes cursos de la educación superior es que muchas veces no se corresponde con la metodología didáctica ni con el enfoque pedagógico, por lo que se considera que hay incoherencia en relación con el modelo evaluativo (Cano, 2008). Esto se evidenció en muchos casos durante la pandemia por COVID-19, en los que los profesores sólo transformaron sus exámenes presenciales a formato digital para realizar las evaluaciones de sus cursos, como señalan Slade et al. (2021). En este estudio, los resultados indican que el modelo de examen a libro abierto con o sin proctoring, tal como lo perciben los docentes, estaba más en consonancia con el enfoque pedagógico del curso que los exámenes presenciales tradicionales. Esto puede representar una indicación de éxito en el cambio del modelo de evaluación que se adoptó debido a la necesidad de realizar los exámenes en línea a causa de la pandemia. Sin embargo, para los alumnos los exámenes presenciales tradicionales también fueron coherentes con



el modelo pedagógico del curso, al igual que los exámenes a libro abierto. Esta diferencia puede deberse a que los exámenes presenciales también tenían un enfoque de evaluación por competencias, con estudios de casos o simulaciones de problemas educativos que los alumnos debían resolver en forma de texto. Desde la perspectiva de los estudiantes, esto coincidía con el enfoque práctico y aplicado de las asignaturas cursadas.

Los exámenes a libro abierto promueven un aprendizaje más complejo que implica sintetizar, crear, reflexionar y tomar decisiones, como señalan Feller (1994), Vázquez et al. (2021), Williams & Wong, (2009), y fomentan un aprendizaje menos memorístico. En este aspecto, tanto los profesores como los alumnos estuvieron de acuerdo. Asimismo, ambos consideraron que los exámenes a libro abierto representaban un reto intelectual, como destacan Herrington & Herrington, (1998) y Koutselini (1997). En este caso, tenían que crear un producto digital, como una infografía, un vídeo o una presentación interactiva para responder a la pregunta del examen. En este sentido, los profesores consideran que los exámenes tradicionales son menos exigentes, pero los alumnos perciben que también son un reto intelectual, aunque menos que los exámenes a libro abierto. Esto puede deberse a las características del examen presencial, ya que, aunque había que resolverlo textualmente, había que resolver situaciones prácticas. Las exigencias planteadas en los exámenes a libro abierto con o sin proctoring estaban más relacionadas con la práctica profesional de los futuros profesores y resultaban más atractivas que las de los exámenes presenciales, y en esto coincidieron docentes y estudiantes. Esto converge con los resultados de Williams & Wong, (2009) y reafirma que este tipo de exámenes es más adecuado para la evaluación auténtica (Boud, 2020; Ibarra-Sáiz et al., 2020; Herrington & Herrington, 1998).

El tema del plagio en las evaluaciones en los cursos de educación superior se convirtió en uno de los más polémicos en los primeros días de la pandemia, muchos profesores no confiaban en que los estudiantes cumplirían con la honestidad académica en las evaluaciones en línea (Cheung, 2020; Guangul et al., 2020; Gudiño et al., 2021; Shaushenova et al., 2021). Este estudio confirma que los profesores consideraban que los exámenes a libro abierto con o sin proctoring ofrecían el mismo potencial de plagio que los exámenes presenciales, lo que converge con la posición de Butler-Henderson & Crawford (2020). La vulnerabilidad de la honestidad académica en este tipo de evaluaciones con exámenes es evidente independientemente de su modalidad: presencial o en línea con o sin proctoring. Quizás sería más efectivo para evitar el plagio entrenar a los estudiantes para que realicen prácticas honestas en sus evaluaciones, como proponen Halak & El-Hajjar (2019). O implementar otro tipo de evaluación más centrada en el desarrollo de proyectos, presentaciones orales o trabajo colaborativo con evaluación integral entre pares, enfocada a la resolución de situaciones educativas reales como proponen Boud (2020), Ibarra-Sáiz et al. (2020), McArthur (2020) y Williams & Wong, (2009).

Sin embargo, los resultados de otros estudios han determinado que no hay plagio ni en los exámenes a libro abierto (Cheung, 2020) ni en los proctorados (Gudiño et al., 2021; Shaushenova et al., 2021).

En contra de la percepción de los docentes, para los estudiantes, los exámenes a libro abierto con o sin proctoring no permiten el plagio. Esto puede explicarse por las exigencias de este modelo de examen, en el que los estudiantes debían aplicar los conocimientos, resolver situaciones desafiantes, crear contenidos digitales, decidir qué aplicaciones web utilizar, probarlas y dar una respuesta en un tiempo máximo de dos horas. Esto exige un aprendizaje complejo por parte de los estudiantes, como destacan Brown y Glasner, (1999), Koutselini, (1997) y Williams y Wong, (2009) en relación con los exámenes a libro abierto.

Para los docentes, los exámenes a libro abierto supusieron una gran diferencia con respecto a los exámenes presenciales, aunque para los estudiantes no. Esto puede deberse al reto que supuso para los docentes reconfigurar los planteamientos de los exámenes de forma que tuvieran diferentes preguntas en la misma asignatura y en la que se exigía que la respuesta se diera a través de un mismo producto digital, por ejemplo, un vídeo o una infografía. Tuvieron que trabajar en colaboración entre los tres o cuatro docentes de la misma asignatura y acordar las rúbricas de evaluación para que los exámenes fueran lo más equivalentes posible. Consideramos que la diferencia respecto a los exámenes tradicionales era sobre todo la novedad en el tipo de examen. Era la primera vez que se aplicaban los exámenes a libro abierto, lo que exigía una mayor reflexión a la hora de diseñarlos (Soodmand y Ranjbar, 2021). Además, como el sistema de proctoring penaliza las consultas en la web, había que solicitar permisos a la universidad para que los estudiantes pudieran consultar la web, levantarse de sus asientos y grabar con la cámara del smartphone en los casos en los que se solicitaban vídeos o grabar el funcionamiento de un robot programado durante el examen.

En este punto queremos destacar que la aplicación del proctoring no afectó la percepción de los exámenes a libro abierto, lo cual puede ser un punto a favor de la supervisión de los exámenes en línea para asegurar la honestidad académica como lo verifican Gudiño et al. (2021) y Shaushenova et al. (2021). Asimismo, la implementación de sistemas de proctoring en la evaluación incide en la calidad de los resultados de la educación a distancia y en esto coincidimos con Shaushenova et al. (2021). También coincidimos con estos autores en que será necesario mejorar los sistemas de proctoring desde el punto de vista tecnológico e incidir en la innovación de la evaluación educativa. Esto resolvería una necesidad expresada por varias instituciones de educación superior, especialmente durante el contexto de la pandemia por COVID-19, pero que puede instalarse en las prácticas de evaluación en la post-pandemia.

## CONCLUSIÓN

Por último, podemos concluir que no existen grandes diferencias entre la percepción de los exámenes a libro abierto con o sin proctoring entre docentes y estudiantes, sólo difieren en que para los estudiantes los exámenes a libro abierto sin proctoring son mucho más desafiantes que los otros y consideran que es más difícil cometer plagio en los exámenes a libro abierto con o sin proctoring. Asimismo, tanto los docentes como los alumnos perciben una diferencia significativa entre los exámenes presenciales tradicionales y los a libro abierto. Según este estudio, también se puede concluir que los exámenes a libro abierto con o sin proctoring constituyen una alternativa relevante para la evaluación final en los cursos online. Sin embargo, sería necesario verificar si estos resultados se confirman en otros estudios dentro de la misma institución u otros en los que se aplique la misma modalidad de examen en las mismas condiciones. No obstante, la propuesta de evaluación según los antecedentes empíricos y los aportes teóricos es pertinente para asumir la evaluación como un aprendizaje auténtico, autorreflexivo y adaptado a las exigencias profesionales reales.

Sería necesario superar algunas limitaciones metodológicas, en cuanto al control de ciertas variables que pueden enriquecer este estudio, por ejemplo, conocer el nivel previo de competencias digitales de los estudiantes. Se podría implementar un diseño de investigación cualitativo-cuantitativo, complementado con entrevistas o un grupo focal para hacer un seguimiento de la percepción del examen y confirmar si fue una experiencia de aprendizaje y consolidación de competencias. También se recomienda abordar otros aspectos como los niveles de ansiedad asociados a los modelos de examen proctorado y la satisfacción de docentes y estudiantes con este tipo de evaluación.

A partir de estos resultados, se podría recomendar la aplicación de este modelo de examen y evaluación en instituciones de educación superior con cursos en línea que deseen sumarse a las tendencias actuales de evaluación.

## REFERENCIAS BIBLIOGRÁFICAS

- 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.). New York: 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. 10.14201/eks.23437
- 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. 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. <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: Burgos, D., Tlili, A., Tabacco, A. (eds) *Radical Solutions for Education in a Crisis Context. Lecture Notes in Educational Technology*. Springer, Singapore. [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* (7ª Ed.). Essex: 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. Recuperado a partir de <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*. <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*. Retrieved september 9, 2021, from <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/3IISnc6>

## Anexo 1

### Descripción de modelos de exámenes

**Modelo A**, es un examen que tiene dos partes, una de preguntas de comprensión y otra de preguntas de desarrollo o casos prácticos. En las preguntas de comprensión se propone relacionar conceptos, analizar alguna situación o valorar una estrategia o metodología. En las preguntas de desarrollo, se plantean situaciones prácticas como el diseño de una estrategia metodológica en la que se apliquen metodologías activas y se incorporen herramientas tecnológicas. Las respuestas se presentan por escrito y el examen se realiza en modo presencial, sin apoyo de ningún tipo de materiales. Es un modelo de examen más enfocado a una evaluación para el aprendizaje (Carles, 2015) en el que se demandan competencias y habilidades específicas y transversales (capacidad de análisis, de aplicación y toma de decisiones, entre otras).

**Modelo B** es un examen en el que se presentan dos planteamientos, el estudiantado elige uno. Se proponen dos situaciones, por ejemplo, alguna problemática en el aula o desarrollo de actividad educativa con metodologías activas y apoyo en tecnología, o planteamiento para mejorar el centro educativo. En todos los casos se debe hacer uso de las herramientas tecnológicas. La respuesta se elabora como un producto digital que puede ser un vídeo, una infografía, un aula virtual, presentación interactiva, etc. El estudiantado envía la url del producto digital como respuesta del examen. Se presenta en modo online con sistema de vigilancia y es un examen a libro abierto (se pueden consultar fuentes por internet). Es un examen que presenta una actividad evaluativa de calidad, que resulta retadora para el estudiante quien tiene que resolver en dos horas la problemática, pensar en las soluciones, crear contenido digital educativo, usar solo material con Licencias Creative Commons, considerar las características de sus estudiantes o su centro educativo al que va a dirigir la solución o propuesta educativa apoyada en las tecnologías. Desde esta perspectiva es un examen enfocado en la evaluación como aprendizaje (Ibarra-Sáiz & Rodríguez-Gómez, 2020), en el que el estudiante construye su propio aprendizaje y fortalece su desarrollo competencial. Así mismo cumple con las características de una evaluación auténtica en los términos de Herrington and Herrington (1998); Williams and Wong, (2009), Boud (2020) e Ibarra-Sáiz et al. (2020) en cuanto a ser desafiante, que esté vinculada a la práctica profesional, que se adapte al estilo de aprendizaje de los estudiantes, que promueva el aprendizaje complejo y que sea coherente con el enfoque pedagógico.

**Modelo C** tiene las mismas características que el modelo B solo que se realiza sin sistema de vigilancia. Por lo que el estudiante puede grabar sus propios vídeos para explicar por ejemplo cómo ha realizado la programación en Scratch o mostrar el funcionamiento del robot SpheroMini que ha programado durante el examen y enviar la url del vídeo y del proyecto programado.