

# Evaluando la evaluación. Validación mediante PLS-SEM de la escala ATAE para el análisis de las tareas de evaluación

*Evaluating Assessment. Validation with PLS-SEM of ATAE Scale for the Analysis of Assessment Tasks*

Ibarra-Sáiz, M.S. , & Rodríguez-Gómez, G. 

Cátedra UNESCO en Evaluación, Innovación y Excelencia en Educación. Universidad de Cádiz (España)

## Abstract

One of the essential functions of university teachers lies in the decision-making process regarding the various components included in assessment process design, where the quality of assessment tasks is a key aspect. This study presents both validation of an instrument for students to evaluate the assessment tasks and the model that upholds the relationships between constructs that characterise the assessment tasks. Working from a review of the literature, a theoretical model has been devised featuring the characteristics of the assessment tasks and the relationships between them. The *Analysis of the Assessment and Learning Tasks* questionnaire (ATAE) has been designed to check them, based on a formative measurement model. Using a cohort design, a total of 1,166 questionnaires were obtained, completed by students from the Business Administration and Management (BAM) and Finance and Accounting (F&A) degree courses. The measurement model and the structural model were evaluated by means of the *Partial Least Squares Structural Equation Modeling* (PLS-SEM) technique using SmartPLS\_3 software. The results show no collinearity problems plus high levels of absolute and relative importance for each questionnaire item. From the students' perception, it should be highlighted that the challenging aspect of an assessment task is related to transfer of learning, and that this is measured by use of communication strategies and demonstration of in-depth understanding.

**Keywords:** Assessment task; Assessment as learning; Empowerment; PLS-SEM; Partial Least Squares; Structural Equation Modeling, PLSpredict

## Resumen

Una de las funciones esenciales del profesorado universitario se concreta en el proceso de toma de decisiones sobre los diferentes componentes que constituyen el diseño de los procesos de evaluación, siendo uno de sus elementos clave la calidad de las tareas de evaluación. En este estudio se presenta tanto la validación de un instrumento para la valoración por el estudiantado de las tareas de evaluación como el modelo que sustenta las relaciones entre los constructos que caracterizan las tareas de evaluación. A partir de una revisión de la literatura se ha elaborado un modelo teórico de las características de las tareas de evaluación y las relaciones existentes entre ellas. Para su comprobación se ha diseñado, sobre la base de un modelo de medida de carácter formativo, el cuestionario *Análisis de las Tareas de Evaluación y Aprendizaje* (ATAE). Mediante un diseño de cohorte se han obtenido un total de 1.166 cuestionarios cumplimentados por estudiantes de los grados de Administración y Dirección de Empresas (ADE) y Finanzas y Contabilidad (FYCO). La evaluación del modelo de medida y del modelo estructural se ha realizado mediante la técnica *Partial Least Squares Structural Equation Modeling* (PLS-SEM) utilizando el software SmartPLS\_3. Los resultados muestran la no existencia de problemas de colinealidad y unos niveles elevados de importancia absoluta y relativa de cada uno de los ítems del cuestionario. Es de destacar, desde la percepción de los estudiantes, que el carácter retador de una tarea de evaluación se relaciona con la transferencia del aprendizaje, y cómo el uso de estrategias de comunicación y la demostración de una comprensión profunda son elementos mediadores de esta relación.

**Palabras clave:** Tarea de evaluación; Evaluación como aprendizaje; Empoderamiento; PLS-SEM; Mínimos cuadrados parciales; Modelo de ecuaciones estructurales; PLS predictivo

Received/Recibido 2020 May 15 | Approved/Aprobado 2020 June 10 | Published/Publicado 2020 June 23

Es frecuente oír comentarios de los estudiantes, en muchos casos poco positivos, sobre la oportunidad, utilidad o justicia de la evaluación. Pero ¿es realmente así cuando los estudiantes abordan tareas de evaluación retadoras? Este interrogante inicial ha servido para iniciar un proyecto que se centra en la calidad de las tareas de evaluación, la percepción que de las mismas tienen los estudiantes universitarios y cómo podría mejorar su diseño el profesorado universitario.

Una de las funciones esenciales del profesorado, desde la consideración de la planificación educativa, se concreta en el diseño de los procesos de evaluación, siendo numerosos los aspectos sobre los cuales se tendrán que tomar decisiones (Bearman et al., 2014, 2016, 2017). En este estudio se centra la atención sólo en uno de ellos, como es el de las características que debe reunir una tarea de evaluación para considerarse una tarea de evaluación de calidad, aspectos que han sido analizados previamente por autores como Ashford-Rowe et al. (2014), Gore et al. (2009) o Smith & Smith (2014). En definitiva, el centro de interés está en la naturaleza de las tareas de evaluación.

Encontramos numerosos estudios respecto a la percepción de los estudiantes sobre la evaluación en un sentido global (Wren et al., 2009; Cubero-Ibáñez & Ponce-González, 2020) o centrando la atención en aspectos concretos como la frecuencia de uso de diferentes medios de evaluación (Pereira et al., 2017), pero son escasas las investigaciones sobre la experiencia y percepción de los estudiantes sobre la naturaleza específica de las tareas de evaluación.

El estudio que aquí se presenta, contextualizado en un proyecto más global, tiene el propósito específico de aportar un modelo exploratorio/predictivo y un instrumento que permitan analizar y mejorar el diseño y práctica de las tareas de evaluación en el ámbito de la educación superior. En concreto, a través de este estudio se pretende:

- Aportar un modelo predictivo de la transferencia del aprendizaje de las tareas de

evaluación en el que se consideren las relaciones entre el carácter retador de estas tareas, la profundidad de las mismas y la comunicación.

- Ofrecer un instrumento que permita analizar y comprender la percepción de los estudiantes universitarios sobre la calidad de las tareas de evaluación que realizan.

En un primer momento se presentan los fundamentos teóricos y un modelo predictivo en el que se establecen las relaciones causales entre un conjunto de variables caracterizadoras de las tareas de evaluación y, posteriormente, se presentan los resultados obtenidos al contrastar este modelo a partir de la percepción de estudiantes de los grados de Administración y Dirección de Empresas (ADE) y Finanzas y Contabilidad (FYCO), aportando finalmente una serie de implicaciones teóricas y prácticas para la mejora de los procesos de evaluación.

## **Marco conceptual y desarrollo de hipótesis**

El diseño y la especificación de las tareas de evaluación es planteado por parte de Sadler (2016) como una de las tres reformas necesarias a llevar a cabo en el contexto de la evaluación del aprendizaje en la educación superior. Para este autor la evaluación y calificación de las ejecuciones de los estudiantes implican establecer una inferencia a partir de los productos y actuaciones del estudiante y, lógicamente, la calidad de esta inferencia viene determinada por la calidad de los datos (los productos y actuaciones del estudiante) y la habilidad del evaluador. Este trabajo se sitúa, esencialmente, en el segundo elemento planteado por Sadler (habilidad del evaluador), pero desde la consideración que el papel de evaluador puede ser ejercido tanto por parte del profesorado como por el estudiantado.

Desde la perspectiva del profesorado como evaluador nos referimos a la habilidad del mismo para diseñar e implementar tareas de evaluación y, desde el punto de vista del estudiante, su papel como ejecutor de estas tareas y la valoración que le merece las

mismas. Por lo tanto, obviamos en este momento la capacidad de los estudiantes para valorar su propia actividad, a través de la autoevaluación, o la actividad que desarrollan sus propios compañeros, mediante la evaluación entre iguales, aspectos éstos que son esenciales en el juicio evaluativo de los estudiantes (Ibarra-Sáiz et al., 2020; Tai et al., 2017) y centramos la atención en su capacidad para valorar no ya la actividad del profesorado en general, sino la calidad de un producto concreto de la actividad diseñadora del profesorado como son las tareas de evaluación. Por lo común, el profesorado diseña una tarea de evaluación y el estudiante ejecuta la misma intentando alcanzar el estándar que el profesorado ha determinado y, una vez finalizada la tarea, siente el alivio de la finalización y comienza a sentir la incertidumbre ante el resultado alcanzado. Cabría preguntarse en ese momento qué le ha supuesto al estudiante abordar esa tarea concreta, qué valoración le merece la tarea en sí misma, qué valor tiene para el estudiante haberla realizado, en definitiva, qué le ha supuesto como experiencia de aprendizaje. Nos situamos de esta forma en el reto planteado por Dawson et al. (2013) de pasar de estudiar las prácticas de evaluación desde una perspectiva abstracta a intentar comprender cómo los principios de la evaluación se pueden trasladar a la mejora de los resultados de aprendizaje y cómo son percibidos y valorados por los estudiantes.

### ***Las tareas de evaluación como elemento esencial en el proceso de evaluación***

Si reducimos la evaluación a una especie de actividades de obligado cumplimiento o trabajos forzados para profesores y estudiantes, entonces estamos tergiversando el significado real del aprendizaje disciplinar en la educación superior. Estamos fallando al no reconocer la relevancia de la curiosidad, la importancia de preguntar y dar respuesta a cuestiones pertinentes y relevantes. La evaluación debería ser el punto de encuentro donde el conocimiento, las ideas, la divergencia, la crítica y la comprensión se

generan e intercambian y esta es la finalidad esencial de la educación superior (Sambell et al., 2013).

Esta importancia del papel que juega la evaluación ha sido destacada por parte de Biggs & Tang (2011) cuando desde su concepción del alineamiento constructivo plantean que las percepciones de la evaluación afectarán a la implicación de los estudiantes en el proceso de aprendizaje; o por Pereira et al. (2017) quienes, de una forma más específica, se refieren a la influencia que ejercen los diferentes medios de evaluación (presentaciones, informes, portafolio, proyectos ...) en la forma en que aprenden los estudiantes.

Diseñar la evaluación implica un proceso de toma de decisiones en torno a una serie de elementos que Bearman et al. (2014) llegan a concretar en un marco global en el que se contemplan los propósitos de la evaluación, el contexto de la evaluación, los procesos de retroalimentación, los resultados de aprendizaje, las interacciones y las tareas de evaluación. Debemos ser conscientes que todos y cada uno de estos elementos no son de naturaleza independiente, sino que se determinan unos a otros, de forma que ciertos propósitos o contextos de la evaluación implicarán la prevalencia de un tipo u otro de tarea de evaluación, pues como destacan Ibarra-Sáiz & Rodríguez-Gómez (2019, p. 192) “las tareas simples, memorísticas o repetitivas, no pueden captar la complejidad de realidades y escenarios que exigen soluciones múltiples y abiertas”.

En definitiva, como Sadler (2016, p. 1083) ha puesto de manifiesto, no debemos confundir una evidencia de baja calidad del rendimiento del estudiante con la evidencia de un bajo rendimiento. En esta misma línea, Boud (2020) nos advierte sobre la importancia de las tareas de evaluación, en tanto que una pobre elección de las mismas conducirá a un aprendizaje pobre y distorsionará lo que los estudiantes terminarán siendo capaces de realizar. Una tarea de evaluación de baja calidad nos aportará una información débil,

sesgada e injusta del rendimiento del estudiante.

### ***Caracterización de las tareas de evaluación***

Esencialmente podemos decir que una tarea de evaluación es una actividad diseñada con la finalidad de recabar información sobre la capacidad de los estudiantes para aplicar y utilizar sus competencias, conocimientos, habilidades y destrezas al abordar la resolución de problemas complejos y poder comprobar el grado de consecución de los resultados de aprendizaje esperados.

Tradicionalmente el aprendizaje de los estudiantes universitarios ha sido evaluado, sobre todo, por el grado de comprensión de algún dominio de conocimientos específicos centrado en la materia que estuvieran cursando, por lo que el foco de atención estaba ajustado a lo que los estudiantes conocían. Progresivamente, sobre todo a partir de la última década del pasado siglo, se ha cambiado el foco de atención y se está poniendo el énfasis en el valor de las habilidades transferibles, genéricas o competencias esenciales, aquéllas que todo estudiante debería desarrollar con independencia de la disciplina concreta en torno a la que desean desarrollar su actividad profesional futura (Boud, 2014; Strijbos et al., 2015).

Este cambio de rumbo ha implicado una renovación de los medios de evaluación que se utilizan, pasando de las clásicas pruebas, test o exámenes finales centrados en la reproducción del conocimiento, a una serie de nuevos medios de evaluación (portafolio, simulaciones, resolución de casos ...) mediante los cuales se intenta integrar y dar coherencia al aprendizaje que se pretende desarrollar a través del alineamiento de la enseñanza y la evaluación, lo que ha llevado a ciertos autores a hablar de una nueva cultura de la evaluación (Dochy, 2009). Estos nuevos medios de evaluación se centran en las ejecuciones del estudiante, en lo que es capaz de realizar y producir, utilizando el pensamiento crítico y la creatividad para la resolución de problemas complejos y actuales.

La calidad de las tareas de evaluación constituye un eje central de esta nueva cultura evaluadora, de suerte que, por ejemplo, Sambell et al. (2013) se refieren al énfasis en las tareas de evaluación auténticas y complejas como uno de los seis principios centrales del enfoque evaluativo conocido como evaluación para el aprendizaje. En esta misma línea, Rodríguez-Gómez & Ibarra-Sáiz (2015) consideran las tareas de evaluación como uno de los retos esenciales que deben abordarse desde el enfoque de la evaluación como aprendizaje y empoderamiento.

La calidad de las tareas de evaluación puede analizarse sobre la base de tres dimensiones concretas (Gore et al., 2009): la rigurosidad intelectual, la significación y el apoyo al estudiante que se ofrece. La rigurosidad intelectual se refiere a centrar las tareas de evaluación en producir una comprensión profunda de lo importante, de los conceptos habilidades e ideas esenciales. Requiere una construcción activa y la implicación de los estudiantes en un pensamiento de alto nivel, además de comunicar sustancialmente sobre aquello que han aprendido. Una tarea de evaluación será relevante en la medida que ayude a hacer que el aprendizaje sea más significativo e importante para los estudiantes conectándolo con las demandas intelectuales de su trabajo. De esta forma, las tareas de evaluación exigen una clara conexión con el conocimiento previo y con el conocimiento académico y extraacadémico. Por último, la tarea de evaluación apoya al estudiante en la medida que fija de manera explícita unas altas expectativas sobre el trabajo del estudiante.

Desde la consideración del diseño de las tareas de evaluación para estudiantes que inician su carrera universitaria, Thomas et al. (2019) consideran que estas tareas deben ser facilitadoras del aprendizaje de los estudiantes, han de favorecer la implicación de los estudiantes en el aprendizaje y deben aportar retroalimentación para un aprendizaje futuro, sirviendo así no sólo para analizar el grado de ejecución, más o menos acertado, sino para comprender potenciales futuras áreas de mejora.

Tras analizar diferentes aportaciones Dochy (2009) llega a la conclusión que los nuevos medios de evaluación mantienen cinco características esenciales. Como primera característica señala que una buena evaluación tiene que sustentarse en requerir de los estudiantes la construcción del conocimiento. No basta con que los estudiantes reproduzcan fielmente el conocimiento, es preciso que sean capaces de dominar la estructura y las interrelaciones existentes y dar coherencia al conocimiento. La segunda característica destaca la necesidad de valorar la capacidad de aplicar el conocimiento a casos de actuales, lo que exige analizar en qué medida el estudiante aplica el conocimiento a la resolución de problemas de la vida real y, además, toma decisiones apropiadas. La tercera característica la constituye la sensibilidad contextual y la multiplicidad de perspectivas. Para Dochy (2009) el estudiante no sólo necesita saber “qué”, sino que también precisa conocer “cuándo”, “dónde” y “por qué”. Para ello no son recomendables los medios de evaluación basados en afirmaciones y respuestas, es preciso que el estudiante domine e interiorice los mecanismos causales subyacentes. La

participación del estudiante es destacada por este autor como cuarta característica de estos nuevos medios de evaluación, donde el estudiante tiene un papel activo al debatir y participar en el diseño o elaboración de los criterios de evaluación, de los instrumentos de evaluación, e incluso actuando como evaluador. Por último, la evaluación no es algo finalista o independiente, sino que está integrada en el proceso de aprendizaje y es coherente con los métodos de enseñanza y el ambiente de aprendizaje.

Como hemos comprobado son múltiples y diversos los diferentes aspectos que pueden considerarse al diseñar una tarea de evaluación. Sobre la base de estas aportaciones previas y de otros estudios, en la Tabla 1 se presentan las cuatro características que hemos considerado en esta investigación como esenciales de una tarea de evaluación: ser un estímulo retador, la necesidad de demostrar una comprensión profunda, el uso de estrategias de comunicación y la capacidad de transferencia de lo aprendido durante la realización de la tarea.

Tabla 1. Definición de constructos

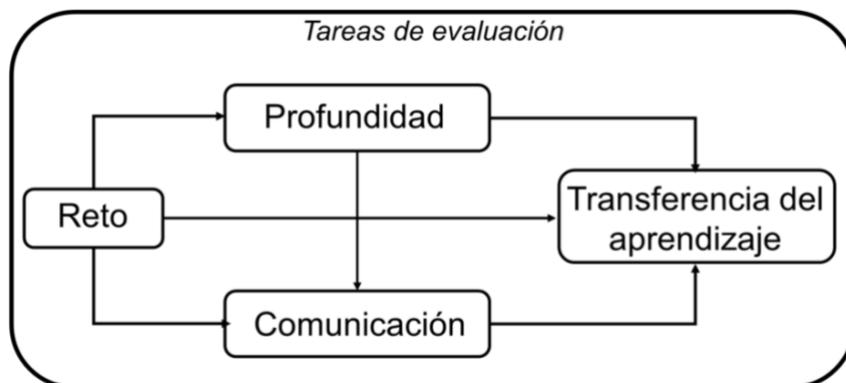
Constructo	Definición	Referencias
Reto	Abordar problemas abiertos y complejos que requieren pensamiento divergente, creatividad y establecer relaciones y conexiones significativas	Ashford-Rowe et al., 2014; Dochy & Gijbels, 2006; Gore et al., 2009; Sambell et al., 2013
Profundidad	Demostrar una comprensión profunda mediante la utilización de métodos de indagación y pensamiento reflexivo y crítico	Dochy, 2009; Entwistle & Karagiannopoulou, 2014; Herrington & Herrington, 2006; O'Donovan, 2016
Comunicación	Utilizar estrategias de comunicación, oral, escrita o simbólica, mediante presentaciones, realizaciones o productos basados en la argumentación fundamentada	Gore et al., 2009; Gulikers et. al, 2004; Smith & Smith, 2014
Transferencia	Relacionar el conocimiento y la experiencia con otras asignaturas y con la realidad social y profesional	Ashwin et al. 2015; Glofcheski, 2017; Gulikers et al., 2004, 2006; Ibarra-Sáiz et al., 2020; Strijbos et al. 2015

### **Modelo de investigación e hipótesis**

El modelo de partida en este trabajo propone que la percepción que los estudiantes tienen sobre la capacidad de transferencia del aprendizaje de las tareas de evaluación que

abordan está determinada por su profundidad y la comunicación exigida, siendo éstas a su vez determinadas por el carácter retador de la tarea de evaluación. En la Figura 1 se presenta este modelo básico indicando las relaciones que se establecen entre estos diferentes constructos.

Figura 1. Modelo para comprobar las relaciones entre elementos caracterizadores de las tareas de evaluación



A partir de este modelo teórico y tomando como base las aportaciones analizadas previamente en este estudio, se proponen las siguientes hipótesis:

- H1. La transferencia se relaciona positivamente con el reto (H1a), la profundidad (H1b) y la comunicación (H1c).
- H2. El reto se relaciona positivamente con la profundidad (H2a) y la comunicación (H2b).
- H3. La profundidad se relaciona directamente con la comunicación.
- H4. La relación entre el reto y la transferencia está mediada por la profundidad (H4a) y la comunicación (H4b).

## Método

Para la realización de este estudio se ha seguido una metodología de encuesta utilizando un diseño de cohortes, ya que se ha recabado la percepción de los estudiantes durante cuatro cursos académicos sucesivos, iniciando en el curso 2016/17 y finalizando en el curso 2019/20, por lo que no han sido los mismos sujetos cada año, sino que han respondido sujetos diferentes en cada curso académico.

Se diseñaron un conjunto de cuatro tareas de evaluación cuyas características aparecen descritas en el trabajo de Ibarra-Sáiz et al. (2020). A medida que los estudiantes finalizaban cada tarea de evaluación

respondían el cuestionario *Análisis de las tareas de evaluación y aprendizaje (ATAE)* expresando su valoración y experiencia en cada caso.

### Participantes

Se han recabado un total de 1.166 cuestionarios ATAE cumplimentados por estudiantes de la Facultad de Ciencias Económicas y Empresariales de la Universidad de Cádiz (Tabla 2). Estos estudiantes cursaban la asignatura Gestión de Proyectos, impartida en el último curso de los grados de Administración y Dirección de Empresas (ADE) y Finanzas y Contabilidad (FYCO).

Tabla 2. Características demográficas de la muestra

		Frecuencia	Porcentaje
<b>Género</b>	Hombre	555	47.6
	Mujer	611	52.4
<b>Grado</b>	FYCO	142	12.2
	ADE	1024	87.8
<b>Cohorte</b>	2017	361	31.0
	2018	369	31.6
	2019	240	20.6
	2020	196	16.8

En la Tabla 3 se presenta la distribución de los cuestionarios que cumplimentaron estos estudiantes durante los cuatro años y para cada una de las cuatro tareas de evaluación que abordaron.

Tabla 3. Distribución de cuestionarios por anualidad y tarea

Tarea	Anualidad				Total
	2017	2018	2019	2020	
1	93	110	75	64	342
2	89	50	67	55	261
3	98	104	51	41	294
4	81	105	47	36	269
Total	361	369	240	196	1166

### Instrumento

Los constructos e indicadores de medida del cuestionario ATAE se desarrollaron sobre la base de una revisión de la literatura y, posteriormente, se procedió a un proceso de validación mediante jueces (Figura 2). Se revisaron diferentes métodos utilizados para la validación de contenido mediante jueces expertos (Johnson y Morgan, 2016) y se optó por el método de consenso grupal, ya que mediante el mismo se evitan los sistemas de votación. La definición de los constructos fue revisada al finalizar cada uno de los ciclos y se concretaron los indicadores durante el proceso

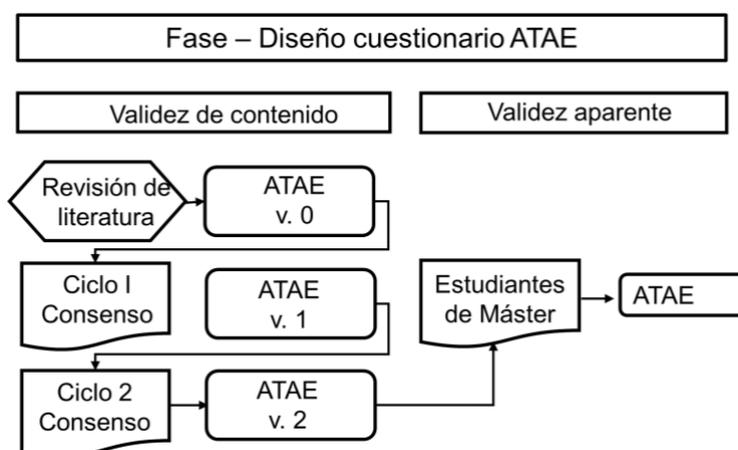
de discusión. Por último, para analizar la validez aparente se presentó el cuestionario a un grupo de estudiantes de máster pudiendo mejorar el cuestionario en cuanto a su claridad y la facilidad de comprensión del mismo.

El cuestionario ATAE (Anexo I) está estructurado en cuatro dimensiones (Tabla 4) y formado por 16 ítems en formato de escala tipo Likert (0-10) distribuidos en cada una de las dimensiones, y cuatro cuestiones abiertas. Aunque el debate sobre el número de opciones y si debe existir una opción intermedia no está cerrado (Horst & Pyburn, 2018), en este caso se ha optado por las recomendaciones de la OCDE (2013) de mantener una escala numérica de 0 a 10. La cumplimentación del cuestionario requería unos 10 minutos.

Tabla 4. Estructura del cuestionario ATAE

Dimensiones	# Ítems
Profundidad	4
Comunicación	4
Reto	3
Transferencia	5

Figura 2. Proceso de diseño del cuestionario ATAE (Análisis de Tareas de Evaluación)



### Análisis de datos

Para la estimación del modelo se ha utilizado el método PLS-SEM (Hair et al., 2017) y PLSpredict (Shmueli et al., 2016), utilizando para los cálculos el software SmartPLS 3 (Ringle et. al., 2015). PLS-SEM es un enfoque de análisis multivariante que se utiliza para estimar modelos con variables latentes, siendo una técnica recomendada cuando, como

sucede en este estudio, el objetivo es la predicción de un constructo objetivo o se pretende identificar constructos relevantes, el modelo de investigación es complejo según el tipo de relaciones que se hipotetizan (directa y mediación), los constructos que forman parte del modelo estructural se han diseñado siguiendo un modelo de medida formativo, el modelo estructural es complejo y los datos no se atienen a una distribución de normalidad

(Roldán & Sánchez-Franco, 2012; Hair et al., 2016; Jiménez-Cortés, 2019).

Para comprobar la adecuación del modelo de medida se ha utilizado el Análisis Confirmatorio Tetrádico (CTA-PLS). Utilizando esta técnica se puede contrastar la hipótesis nula de que los indicadores de un modelo son reflectivos (Garson, 2016), de forma que se puede confirmar el carácter reflectivo o formativo de las variables latentes (Hair et al., 2018).

La evaluación del modelo se ha efectuado de acuerdo al carácter formativo del mismo, para lo que se ha realizado un análisis de multicolinealidad y un análisis de los pesos. Posteriormente, se ha procedido a analizar la capacidad predictiva del modelo y las relaciones entre los constructos. Para ello se han realizado los siguientes análisis: a) evaluación de la colinealidad (VIF); b) coeficientes path del modelo estructural; c) coeficiente de determinación ( $R^2$ ); d) tamaño del efecto ( $f^2$ ); e) relevancia predictiva ( $Q^2$ ), f) tamaño del efecto ( $q^2$ ) y g) análisis del poder predictivo mediante PLSpredict (Shmueli et al., 2016, 2019).

## Resultados

### *Evaluación del modelo de medida*

Cuando se analizan indicadores reflectivos lo usual es analizar la consistencia interna (alfa de Cronbach), la validez convergente o la validez discriminante (Muñoz-Cantero et al., 2019), pero cuando se utilizan indicadores formativos, la evaluación del modelo de medida se basa en el análisis de la colinealidad, importancia relativa (pesos externos) y la importancia absoluta (cargas externas) (Hair et al., 2019). En nuestro caso (Tabla 5), con valores de colinealidad (VIF) entre 1.18 y 2.11 podemos concluir que, tomando como valor de referencia 5, la colinealidad no alcanza niveles críticos en ninguno de los constructos formativos y, por lo tanto, no hay dificultad para la estimación del modelo. Se encontró un indicador (COM\_3) cuyo peso no era estadísticamente significativo pero en cambio tenía carga próxima a 0.5, por lo que de acuerdo con las normas de decisión manifestadas por Hair et al. (2016) y el propio contenido del indicador, se optó por mantener todos los indicadores formativos.

Tabla 5. Pesos, cargas y valores VIF de los constructos formativos

Constructos	Indicadores	Pesos	Cargas	VIF
RETO (RET)	RET_1	0.373	0.832	1.700
	RET_2	0.450	0.842	1.590
	RET_3	0.090	0.598	1.447
	RET_4	0.343	0.749	1.518
PROFUNDIDAD (PRO)	PRO_1	0.269	0.738	1.485
	PRO_2	0.350	0.834	1.793
	PRO_3	0.311	0.814	1.770
	PRO_4	0.323	0.795	1.607
COMUNICACIÓN (COM)	COM_1	0.416	0.793	1.436
	COM_2	0.669	0.909	1.335
	COM_3	0.133	0.477	1.183
TRANSFERENCIA (TRA)	TRA_1	0.312	0.827	2.013
	TRA_2	0.303	0.834	2.110
	TRA_3	0.129	0.666	1.712
	TRA_4	0.239	0.712	1.667
	TRA_5	0.320	0.728	1.363

Complementariamente, siguiendo las orientaciones de Hair et al. (2019) para comprobar la robustez del modelo de medida, se procedió a realizar un Análisis Tetrádico Confirmatorio (CTA-PLS), que ha permitido comprobar empíricamente el carácter formativo de los constructos RET y TRA, al encontrarse tétradas significativamente diferentes de cero.

### Evaluación del modelo estructural

Para proceder a la evaluación del modelo estructural se realizaron los siguientes análisis: a) colinealidad; b) significación y relevancia de las relaciones estructurales; c) poder y relevancia predictiva; d) tamaño del efecto; y e) poder predictivo.

Tabla 6. Valores VIF del modelo estructural

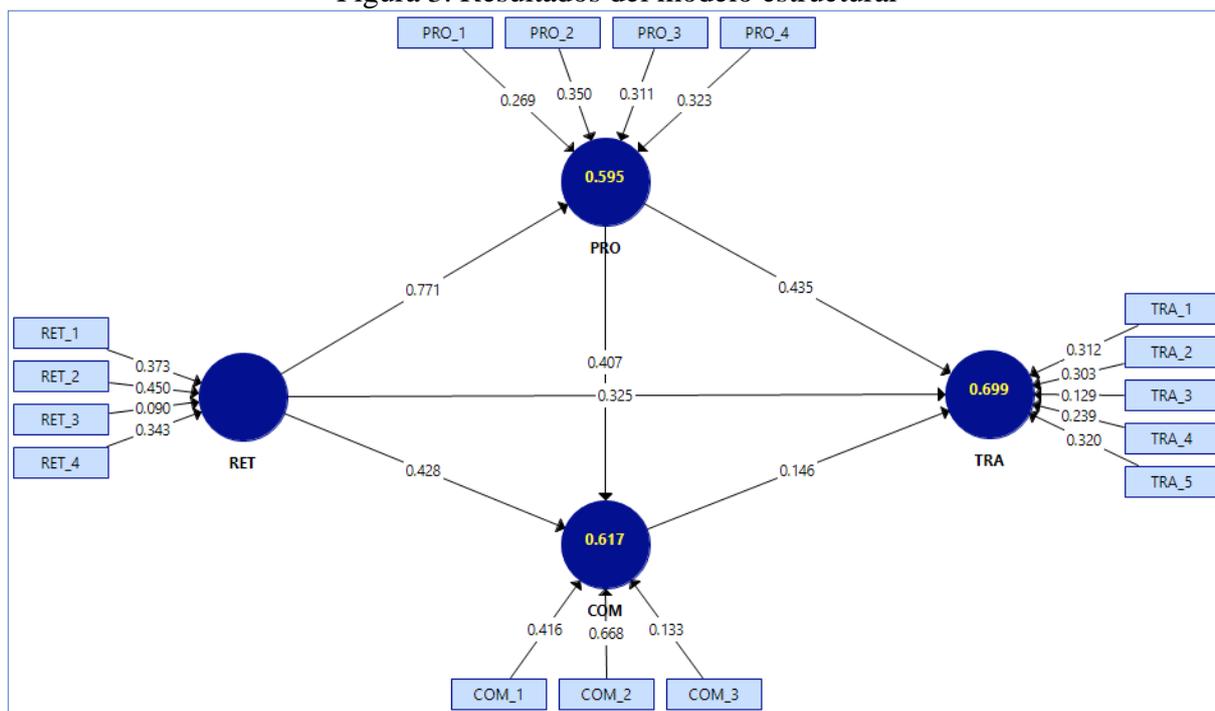
	COM	PRO	RET	TRA
COM				2.610
PRO	2.469			2.900
RET	2.469	1.000		2.947
TRA				

Con respecto a los problemas de colinealidad, en los resultados presentados en la Tabla 6

podemos comprobar que todos los valores VIF están claramente por debajo del límite de 5, por lo que podemos concluir que no existen problemas de colinealidad.

El valor predictivo del modelo se ha analizado a través del coeficiente de determinación ( $R^2$ ). Así, en la Figura 3 se evidencia cómo casi el 70% de la varianza ( $R^2$ ) del constructo transferencia (TRA) es explicado por los otros tres constructos. Según los criterios establecidos por Chin (1998) y Hair et al. (2017) podemos considerarlo como sustancial. El efecto más fuerte sobre la transferencia (TRA) es ejercido por el constructo profundidad (PRO, 0.435), seguido del constructo reto (RET, 0.325) y la comunicación (COM, 0.146). Así mismo, se evidencia que los valores de  $R^2$  para los constructos PRO (0.595) y COM (0.617) alcanzan niveles que pueden considerarse como moderados ( $R^2 > 0.50$ ). El modelo presenta un SRMR de 0.03, lo que indica un adecuado nivel de ajuste tomando como referencia el criterio usual de situarlo por debajo de 0.08.

Figura 3. Resultados del modelo estructural



Para establecer la significación estadística de los coeficientes path, de acuerdo con Hair et al. (2017), se realizó un remuestreo (*bootstrapping*) con 5.000 submuestras para generar los estadísticos *t* y los intervalos de confianza (Tabla 7). Observamos tamaños del

efecto grandes en el caso de la relación entre RET->PRO, siendo medio en las relaciones PRO->TRA, RET->COM y PRO->COM y pequeño en los casos RET->TRA y COM->TRA.

Tabla 7. Resultados del modelo estructural utilizando los valores *t* y percentiles con un intervalo de confianza del 95% (n=5.000 submuestras)

Relaciones	Coeficientes Path				Tamaño del efecto				Hipótesis
	Path(*)	95% CI	<i>t</i>	<i>p</i>	<i>f</i> <sup>2</sup> (+)	95% CI	<i>t</i>	<i>p</i>	
RET -> TRA	0.325	[0.252, 0.390]	9.006	0.000	0.119	[0.071, 0.180]	4.347	0.000	H1a
PRO -> TRA	0.435	[0.365, 0.515]	11.660	0.000	0.217	[0.142, 0.312]	4.949	0.000	H1b
COM -> TRA	0.146	[0.077, 0.217]	4.092	0.000	0.027	[0.008, 0.061]	1.977	0.048	H1c
RET -> PRO	0.771	[0.734, 0.808]	41.047	0.000	1.469	[1.179, 1.862]	8.311	0.000	H2a
RET -> COM	0.428	[0.703, 0.779]	37.986	0.000	0.194	[0.130, 0.270]	5.414	0.000	H2b
PRO -> COM	0.407	[0.338, 0.475]	11.763	0.000	0.175	[0.117, 0.247]	5.308	0.000	H3

Notas: (\*) 0.75 sustancial/0.50 moderado/0.25 débil / (+) 0.35 grande/0.15 medio /0.02 pequeño

Para comprobar la relevancia predictiva del modelo se han calculado los valores  $Q_2$  mediante el procedimiento de *blindfolding* (Tabla 8). Podemos observar que todos los valores para los constructos endógenos están por encima de cero. Más específicamente, el mayor valor es presentado por TRA (0.396), seguido de PRO (0.373) y COM (0.341). Estos resultados apoyan la relevancia del modelo predictivo a partir de las variables latentes endógenas.

Tabla 8. Redundancia de constructo validada de forma cruzada (valores  $Q_2$ )

	SSO	SSE	$Q^2 (=1 - SSE/SSO)$
COM	3498.000	2306.506	0.341
PRO	4664.000	2922.958	0.373
RET	4664.000	4664.000	
TRA	5830.000	3518.702	0.396

Los tamaños del efecto ( $q_2$ ) permiten evaluar cómo un constructo exógeno contribuye a un

constructo latente endógeno. En nuestro caso constatamos que se alcanzan valores pequeños en el tamaño del efecto (Tabla 9).

Tabla 9. Tamaños del efecto ( $q_2$ )

	COM	PRO	TRA
COM		-0.001	0.002
PRO	0.051		0.069
RET	0.066		0.034
TRA	-0.001	0.000	

Notas: \* 0.02 pequeño/0.15 medio/0.35 grande

Por último, para comprobar el poder predictivo del modelo se llevó a cabo el procedimiento PLSpredict (Sharma et al., 2018) obteniendo los resultados que se presentan en la Tabla 10. Se evidencia que, en todos los casos, los valores  $Q_{2predict}$  son mayor de cero y en la mitad de los indicadores se obtienen valores de RMSE mayores utilizando PLS frente a LM, lo que indica que el modelo tiene un poder predictivo medio (Shmueli et al., 2019; Hair et al., 2019).

Tabla 10. Resumen de predicción de variables manifiestas (indicadores)

Indicadores	PLS		LM		PLS-LM
	RMSE	$Q^2_{\text{predict}}$	RMSE	$Q^2_{\text{predict}}$	(RMSE)
COM_1	1.275	0.375	1.274	0.376	0.001
COM_2	1.061	0.434	1.057	0.438	0.004
COM_3	1.993	0.109	2.000	0.103	-0.007
PRO_4	1.289	0.369	1.287	0.370	0.002
PRO_1	1.361	0.355	1.363	0.353	-0.002
PRO_2	1.176	0.409	1.177	0.407	-0.001
PRO_3	1.242	0.365	1.243	0.363	-0.002
TRA_1	1.135	0.392	1.137	0.388	-0.002
TRA_2	1.152	0.426	1.155	0.423	-0.003
TRA_3	1.582	0.252	1.567	0.267	0.015
TRA_4	1.496	0.314	1.494	0.317	0.002
TRA_5	1.329	0.295	1.320	0.303	0.009

### Análisis de la mediación

Para el análisis de la mediación se han seguido las orientaciones planteadas por Zhao et al. (2010) y Nitzl et al. (2016) quienes proponen un proceso de dos pasos. En un primer momento se trata de determinar la significación de los efectos indirectos mediante un procedimiento de remuestreo (*bootstrapping*) y, en segundo lugar, establecer el tipo de mediación siguiendo el árbol de decisión planteado por Zhao et al. (2010) y actualizado por Hair et al. (2017).

a) La profundidad y la comunicación como variables mediadoras

En el caso del modelo que hemos presentado (Fig. 3), la profundidad y la comunicación operan como variables mediadoras entre el carácter retador de las tareas de evaluación y la transferencia del aprendizaje, por lo que podemos decir que se trata de un modelo de mediación múltiple. En la Tabla 11 se presentan los resultados obtenidos al comprobar el efecto de esta mediación. Así, comprobamos que el reto tiene un efecto directo (0.325) significativo ( $t=9.006, p<.05$ ) y que el efecto total indirecto de las relaciones entre reto y transferencia (0.444) también es

significativo ( $t=15.272, p<0.05$ ) y, en ambos casos, el intervalo de confianza no incluye el cero. Observamos cómo la profundidad (0.336) presenta un efecto específico indirecto significativo ( $t=11.378, p<.05$ ) y, aunque en menor medida, el efecto de la comunicación (0.063) también es significativo ( $t=3.914, p<.05$ ), siendo así mismo significativo el efecto indirecto múltiple de la profundidad y la comunicación (0.046).

Para analizar la magnitud de la mediación se ha calculado el índice de varianza explicada (VAF) siguiendo las orientaciones de Cepeda-Carrión et al. (2017), concluyendo que el mayor poder de mediación (75.6%) lo ejerce la profundidad, seguido de la comunicación (14.1%) y la interacción entre profundidad y comunicación (10.3%).

Al ser en todos los casos el producto del efecto directo y el indirecto específico positivo podemos concluir que se trata de una mediación complementaria por lo que, como señalan Zhao et al. (2010) estas variables mediadoras son consistentes con el modelo teórico hipotetizado, aunque podrían existir otros mediadores no contemplados que pudieran completar este modelo.

Tabla 11. Resumen de la comprobación del efecto de mediación de RET sobre TRA

<i>Efecto total</i>	<b>Efecto</b>	<b>95% CI</b>	<i>t</i>	<i>p</i>	<b>Tipo de mediación</b>
RET -> TRA	0.769	[0.733, 0.805]	42.047	0.000	
<i>Efecto directo</i>					
RET -> TRA	0.325	[.252, 0.390]	9.006	0.000	
<i>Total efecto indirecto</i>					
RET -> TRA	0.444	[0.389, 0.503]	15.272	0.000	
<i>Efectos indirectos específicos</i>					
RET -> PRO -> COM -> TRA	0.046	[0.023, 0.072]	3.662	0.000	Complementaria
RET -> PRO -> TRA	0.336	[0.277, 0.393]	11.378	0.000	Complementaria
RET -> COM -> TRA	0.063	[0.033, 0.095]	3.914	0.000	Complementaria

## Discusión

A través de este estudio se pretendía, en primer lugar, aportar un modelo predictivo de la transferencia del aprendizaje basado en las variables que caracterizan la naturaleza de las tareas de evaluación y, en segundo lugar, ofrecer un instrumento útil para analizar y comprender la percepción de los estudiantes universitarios sobre la calidad de las tareas de evaluación a las que se enfrentan en su proceso de aprendizaje. Los resultados que se han presentado sugieren implicaciones tanto desde una perspectiva teórica como práctica para comprender el diseño e implementación de las tareas de evaluación en las aulas universitarias, al tiempo que permite vislumbrar líneas futuras de investigación.

### *Implicaciones teóricas*

Como hemos señalado, el primer objetivo consistía en aportar un modelo predictivo sobre los elementos constitutivos de la naturaleza de las tareas de evaluación, situando este estudio en el contexto de las investigaciones en torno a la caracterización y naturaleza de las tareas de evaluación.

Una de las principales aportaciones de este trabajo radica en la confirmación de un modelo que integra la relación entre un conjunto de variables caracterizadoras de la calidad de las tareas de evaluación. Los resultados que se han alcanzado evidencian que el modelo hipotetizado puede predecir en gran medida las relaciones establecidas entre las variables destacadas y ponen de manifiesto, en primer

lugar, que el carácter retador de las tareas de evaluación se relaciona directamente con la transferencia del aprendizaje, la profundidad y la comunicación y, por otra parte, el carácter de mediación que ejercen la profundidad y la comunicación.

Se ha contrastado la hipótesis en la que se afirma la relación directa de la transferencia con el carácter retador de la tarea de evaluación (H1a), así como con la profundidad (H1b) y la comunicación (H1c). Así mismo, se ha contrastado la hipótesis que relaciona el carácter retador de las tareas con la profundidad (H2a) y la comunicación (H2b) y la que relaciona la relación entre profundidad y comunicación (H3). Por último, se ha contrastado la hipótesis en la que se establece la mediación de la profundidad (H4a) y la comunicación (H4b) en la relación entre el carácter retador de las tareas y la transferencia del aprendizaje.

El carácter retador de las tareas de evaluación, la necesidad de utilizar un pensamiento de alto nivel en su resolución mediante la demostración de una comprensión profunda o utilizar la comunicación son todas características que determinan la calidad de las tareas de evaluación y su capacidad para transferir el aprendizaje realizado a otros contextos o situaciones (Ashford-Rowe et al., 2014; Sambell et al., 2013; Smith & Smith, 2014). Este estudio evidencia el papel relevante del carácter retador de las tareas de evaluación y la necesidad de enfrentar a los estudiantes a abordar problemas complejos que

le motiven en su resolución y cómo ello repercute en la transferencia del aprendizaje que realizan durante la resolución de estas tareas.

### ***Implicaciones prácticas***

Un segundo objetivo que ha orientado este estudio ha sido el de ofrecer un instrumento para analizar y comprender la percepción de los estudiantes universitarios sobre la calidad de las tareas de evaluación. La estimación del modelo de medida apoya la validez del cuestionario ATAE utilizado para operativizar las variables latentes, ya que los ítems son relevantes y cargan en el constructo correcto. En consecuencia, disponemos de un instrumento de fácil aplicación tras la finalización de una tarea de evaluación a través del cual se puede recabar de una forma rápida la valoración de los estudiantes sobre la misma y, a su vez, sirve como elemento de reflexión crítica de los propios estudiantes en torno a la utilidad de la propia tarea de evaluación.

### ***Limitaciones y prospectiva***

Desde una perspectiva metodológica esta investigación adolece de una serie de limitaciones que, a su vez, se convierten en posibilidades de mejora y nuevas líneas de investigación. En primer lugar, esta investigación se ha realizado en el contexto de los estudios de grado de Ciencias Económicas y Empresariales y, más concretamente, con estudiantes que cursaban su último curso. Además, las tareas de evaluación que tuvieron que abordar estos estudiantes fueron diseñadas desde un enfoque evaluativo basado en la evaluación como aprendizaje y empoderamiento, razones todas ellas que nos conducen a una cierta debilidad en las posibilidades de generalizar los hallazgos alcanzados. Por lo tanto, una vía futura de investigación radica en realizar otros estudios que permitan, por un lado, la extrapolación a otros contextos diferentes tanto en el campo de las ciencias sociales como en otras áreas de conocimiento y, por otro lado, confrontar las valoraciones realizadas ante tareas de evaluación diseñadas por el profesorado desde

diferentes enfoques evaluativos, lo que permitiría contrastar las valoraciones realizadas por los estudiantes entre tareas de evaluación de naturaleza bien diferente.

En segundo lugar, el instrumento de medida que se ha presentado se basa en la valoración de los estudiantes desde una perspectiva totalmente subjetiva, por lo que podría mejorarse a través del uso de instrumentos que faciliten indicadores alternativos (Panadero et al., 2018), o incorporando otras fuentes de información diferentes al propio estudiante, combinando la medición y la intervención (Panadero et al., 2016).

### **Conclusión**

En esta investigación hemos presentado resultados que evidencian cómo el carácter retador de las tareas de evaluación, o su exigencia para poner en práctica un conocimiento de alto nivel utilizando estrategias de comunicación, son características que favorecen la transferencia del aprendizaje en los procesos de evaluación. Se ha aportado un instrumento que pueden ser adaptado o replicado en otros contextos diferentes y se han presentado nuevas líneas de trabajo que permitirán una mejor comprensión de la naturaleza de las tareas de evaluación.

Este estudio afianza la importancia del papel del profesorado universitario como diseñador de experiencias de aprendizaje y, concretamente, como diseñador de tareas de evaluación con un alto nivel de calidad que exijan poner en juego toda la potencialidad del estudiante. Pero este no es un papel fácil, como han evidenciado los trabajos de Bearman et al. (2016, 2017), exige pensar en torno a la racionalidad y justificación de la evaluación, sobre las actividades que van a servir para ser revisadas o calificadas, los criterios que van a servir para valorar si se han alcanzado los resultados de aprendizaje deseados, la contribución y aportaciones que puede realizar el propio estudiante en el proceso de evaluación o cuál será el esquema temporal que seguirá la realización de las diferentes tareas de evaluación a lo largo del curso. Es un

proceso de toma de decisiones del profesorado que, en gran medida, viene dificultado por limitaciones contextuales que hacen necesaria una mayor formación específica del profesorado en evaluación y el desarrollo de políticas universitarias favorecedoras de nuevos medios de evaluación.

Diseñar tareas de evaluación retadoras, que exijan el uso de un conocimiento profundo de las materias, el dominio de estrategias de comunicación y, en consecuencia, que el aprendizaje que se realiza durante su resolución se pueda transferir a otros contextos, exige de un profesorado que domine los nuevos medios de evaluación y su uso contextualizado. Medios de evaluación que habrán de ser valorados no ya sobre los antiguos criterios de fiabilidad y validez, sino sobre nuevos criterios alternativos (Dochy, 2009) como la transparencia, la justicia, la apreciación directa (no inferida) o su autenticidad y complejidad cognitiva que, evidentemente, exigen del profesorado y de los estudiantes una nueva visión y perspectiva en torno a los procesos de evaluación.

## Reconocimiento

Este trabajo ha sido posible gracias al proyecto TransEval (Ref. I+D+i 2017/01) financiado por la Universidad de Cádiz y al apoyo de la Cátedra UNESCO Evaluación, Innovación y Excelencia en Educación.

## Referencias

- Ashford-Rowe, K., Herrington, J., & Brown, C. (2014). Establishing the critical elements that determine authentic assessment. *Assessment & Evaluation in Higher Education*, 39(2), 205–222. <https://doi.org/10.1080/02602938.2013.819566>
- Ashwin, P., Boud, D., Coate, K., Hallet, F., Keane, E., Krause, K.-L., ... Tooher, M. (2015). *Reflective teaching in higher education*. Bloomsbury. <https://doi.org/10.1142/S0129183114500405>
- Bearman, M., Dawson, P., Bennett, S., Hall, M., Molloy, E., Boud, D., & Joughin, G. (2017). How university teachers design assessments: a cross-disciplinary study. *Higher Education*, 74(1), 49–64. <https://doi.org/10.1007/s10734-016-0027-7>
- Bearman, M., Dawson, P., Boud, D., Bennett, S., Hall, M., & Molloy, E. (2016). Support for assessment practice: developing the Assessment Design Decisions Framework. *Teaching in Higher Education*, 21(5), 545–556. <https://doi.org/10.1080/13562517.2016.1160217>
- Bearman, M., Dawson, P., Boud, D., Hall, M., Bennett, S., Molloy, E., & Joughin, G. (2014). *Guide to the Assessment Design Decisions Framework*. Retrieved from <http://www.assessmentdecisions.org/guide/>
- Biggs, J., & Tang, C. (2011). *Teaching for quality learning at university. What the students does* (4th ed.). McGraw-Hill-SRHE & Open University Press.
- Boud, D. (2014). Shifting views of assessment: from secret teachers' business to sustaining learning. In C. Kreber, C. Anderson, N. Entwistle, & J. McArthur (Eds.), *Advances and innovations in university assessment and feedback* (pp. 13–31). Edinburgh University Press Ltd. <https://doi.org/10.3366/edinburgh/9780748694549.003.0002>
- Boud, D. (2020, en prensa). Challenges in reforming higher education assessment: a perspective from afar. *RELIEVE - Electronic Journal of Educational Research, Assessment and Evaluation*.
- Cepeda-Carrión, G., Nitzl, C., & Roldán, J. L. (2017). Mediation Analyses in Partial Least Squares Structural Equation Modeling: Guidelines and Empirical Examples. In H. Latan & R. Noonan (Eds.), *Partial Least Squares Path Modeling: Basic concepts, methodological issues and applications* (pp. 173–195). Springer. <https://doi.org/10.1007/978-3-319-64069-3>

- Chin, W. W. (1998). The partial least squares approach to structural equation modeling. In G. A. Marcoulides (Ed.), *Modern methods for business research* (pp. 295–398). Lawrence Erlbaum.
- Cubero-Ibáñez, J., & Ponce-González, N. (2020). Aprendiendo a través de tareas de evaluación auténticas: Percepción de estudiantes de Grado en Educación Infantil. *Revista Iberoamericana de Evaluación Educativa*, 13(1), 41–69. <https://doi.org/10.15366/riee2020.13.1.002>
- Dawson, P., Bearman, M., Boud, D. J., Hall, M., Molloy, E. K., Bennett, S., & Joughin, G. (2013). Assessment Might Dictate the Curriculum, But What Dictates Assessment? *Teaching & Learning Inquiry: The ISSOTL Journal*, 1(1), 107–111. <https://doi.org/10.2979/teachlearninqu.1.1.107>
- Dochy, F. (2009). The Edumetric Quality of New Modes of Assessment: Some Issues and Prospect. In G. Joughin (Ed.), *Assessment, Learning and Judgement in Higher Education* (pp. 85–114). Springer. [https://doi.org/10.1007/978-1-4020-8905-3\\_6](https://doi.org/10.1007/978-1-4020-8905-3_6)
- Dochy, F., & Gijbels, D. (2006). New learning, assessment engineering and edumetrics. In L. Verschaffel, F. Dochy, M. Boekaerts, & S. Vosniadou (Eds.), *Instructional Psychology: Past, present and future trends. Sixteen essays in honour of Erik De Corte*. Elsevier.
- Entwistle, N., & Karagiannopoulou, E. (2014). Perceptions of Assessment and their Influences on Learning. In C. Kreber, C. Anderson, N. Entwistle, & J. McArthur (Eds.), *Advances and innovations in university assessment and feedback* (pp. 75–98). Edinburgh University Press Ltd. <https://doi.org/10.3366/edinburgh/9780748694549.003.0005>
- Garson, G. D. (2016). *Partial Least Squares: Regression & Structural Equation Models*. Statistical Publishing Associates.
- Glofcheski, R. (2017). Making Assessment for Learning Happen Through Assessment Task Design in the Law Curriculum. In D. Carless, S. M. Bridges, C. K. Y. Chan, & R. Glofcheski (Eds.), *Scaling up Assessment for Learning in Higher Education* (pp. 67–80). Singapore. [https://doi.org/10.1007/978-981-10-3045-1\\_5](https://doi.org/10.1007/978-981-10-3045-1_5)
- Gore, J., Ladwig, J., Eslworth, W., & Ellis, H. (2009). *Quality assessment framework: A guide for assessment practice in higher education*. Callaghan, NSW Australia: The Australian Learning and Teaching Council. The University of Newcastle. Retrieved from [http://www.olt.gov.au/system/files/resources/QAF\\_FINAL\\_doc\\_for\\_print.pdf](http://www.olt.gov.au/system/files/resources/QAF_FINAL_doc_for_print.pdf)
- Gulikers, J. T. M., Bastiaens, T. J., Kischner, P. A., & Kester, L. (2006). Relations between student perception of assessment authenticity, study approaches and learning outcomes. *Studies in Educational Evaluation*, 32, 381–400. <https://doi.org/10.1016/j.stueduc.2006.10.003>
- Gulikers, J. T. M., Bastiaens, T., & Kirschner, P. A. (2004). A five-dimensional framework for authentic assessment. *Educational Technology Research and Development*, 52(3), 67–85. <https://doi.org/10.1007/BF02504676>
- Hair, J. F., Hult, G. T. M., Ringle, C. M., & Sarstedt, M. (2017). *A Primer on Partial Least Squares Structural Equation Modeling (PLS-SEM)* (2nd ed.). Sage.
- Hair, J. F., Risher, J. J., Sarstedt, M., & Ringle, C. M. (2019). When to use and how to report the results of PLS-SEM. *European Business Review*, 31(1), 2–24. <https://doi.org/10.1108/EBR-11-2018-0203>
- Hair, J. F., Sarstedt, M., Ringle, C. M., & Gudergan, S. P. (2018). *Advanced Issues in Partial Least Squares Structural Equation Modeling*. Sage.
- Herrington, J., & Herrington, A. (2006). Authentic conditions for authentic assessment: Aligning task and assessment. *Proceedings of the 29th HERDSA Annual Conference*, 146–151.
- Horst, S. J., & Pyburn, E. M. (2018). Likert scaling. In B. B. Frey (Ed.), *The SAGE encyclopedia of educational research*,

- Ibarra-Sáiz, M.S. y Rodríguez-Gómez, G. (2020). Evaluando la evaluación. Validación mediante PLS-SEM de la escala ATAE para el análisis de las tareas de evaluación. *RELIEVE*, 26(1), art. M4. <http://doi.org/10.7203/relieve.26.1.17403>
- 
- measurement, and evaluation* (pp. 974–976). Sage.
- Ibarra-Sáiz, M. S. y Rodríguez-Gómez, G. (2019). Una evaluación como aprendizaje. En J. Paricio, A. Fernández e I. Fernández (Eds.), *Cartografía de la buena docencia. Un marco para el desarrollo del profesorado basado en la investigación* (pp. 175–196). Narcea.
- Ibarra-Sáiz, M. S., Rodríguez-Gómez, G., & Boud, D. (2020). Developing student competence through peer assessment: the role of feedback, self-regulation and evaluative judgement. *Higher Education*. <https://doi.org/10.1007/s10734-019-00469-2>
- Jiménez-Cortés, R. (2019). Aprendizaje de las mujeres en las redes sociales: Validación de la escala MAIA con PLS. *Revista de Investigación Educativa*, 37(2), 431–449. <https://doi.org/http://dx.doi.org/10.6018/rie.37.2.325721>
- Johnson, R. L., & Morgan, G. B. (2016). *Survey scales. A guide to development, analysis, and reporting*. The Guilford Press.
- Muñoz-Cantero, J.-M., Rebollo-Quintela, N., Mosteiro-García, J., & Ocampo-Gómez, C.-I. (2019). Validación del cuestionario de atribuciones para la detección de coincidencias en trabajos académicos. *RELIEVE - Revista Electrónica de Investigación y Evaluación Educativa*, 25(1), 1–16. <https://doi.org/10.7203/relieve.25.1.13599>
- Nitzl, C., Roldán, J. L., & Cepeda, G. (2016). Mediation Analysis in Partial Least Squares Path Modeling: Helping researchers discuss more sophisticated models. *Industrial Management & Data Systems*, 116(9), 1849–1864. <https://doi.org/10.1108/IMDS-07-2015-0302>
- O'Donovan, B. (2016). How student beliefs about knowledge and knowing influence their satisfaction with assessment and feedback. *Higher Education*, 74(4), 617–633. <https://doi.org/10.1007/s10734-016-0068-y>
- OECD. (2013). Methodological considerations in the measurement of subjective well-being. In *OECD Guidelines on measuring subjective well-being* (pp. 61–138). Paris: OECD Publishing. <https://doi.org/10.1787/9789264191655-en>
- Panadero, E., Andrade, H., & Brookhart, S. (2018). Fusing self-regulated learning and formative assessment: a roadmap of where we are, how we got here, and where we are going. *The Australian Educational Researcher*, 45(1), 13–31. <https://doi.org/10.1007/s13384-018-0258-y>
- Panadero, E., Klug, J., & Järvelä, S. (2016). Third wave of measurement in the self-regulated learning field: when measurement and intervention come hand in hand. *Scandinavian Journal of Educational Research*, 60(6), 723–735. <https://doi.org/10.1080/00313831.2015.1066436>
- Pereira, D., Niklasson, L., & Flores, M. A. (2017). Students' perceptions of assessment: a comparative analysis between Portugal and Sweden. *Higher Education*, 73(1), 153–173. <https://doi.org/10.1007/s10734-016-0005-0>
- Ringle, C. M., Wende, S., & Becker, J.-M. (2015). *SmartPLS 3*. Bönningstedt: SmartPLS. Retrieved from <http://www.smartpls.com>
- Rodríguez-Gómez, G., & Ibarra-Sáiz, M. S. (2015). Assessment as learning and empowerment: Towards sustainable learning in higher education. In M. Peris-Ortiz & J. M. Merigó Lindahl (Eds.), *Sustainable learning in higher education. Developing competencies for the global marketplace* (pp. 1–20). Springer. [https://doi.org/10.1007/978-3-319-10804-9\\_1](https://doi.org/10.1007/978-3-319-10804-9_1)
- Roldán, J. L., & Sánchez-Franco, M. J. (2012). Variance-Based Structural Equation Modeling: Guidelines for Using Partial Least Squares in Information Systems Research. In *Research Methodologies, Innovations and Philosophies in Software Systems Engineering and Information Systems* (pp. 193–221). IGI Global. <https://doi.org/10.4018/978-1-4666-0179-6.ch010>
- Sadler, D. R. (2016). Three in-course assessment reforms to improve higher

- education learning outcomes. *Assessment & Evaluation in Higher Education*, 41(7), 1081–1099. <https://doi.org/10.1080/02602938.2015.1064858>
- Sambell, K., McDowell, L., & Montgomery, C. (2013). *Assessment for Learning in Higher Education*. Routledge.
- Sharma, P. N., Shmueli, G., Sarstedt, M., Danks, N., & Ray, S. (2018). Prediction-Oriented Model Selection in Partial Least Squares Path Modeling. *Decision Sciences*. <https://doi.org/10.1111/deci.12329>
- Shmueli, G., Ray, S., Velasquez Estrada, J. M., & Chatla, S. B. (2016). The elephant in the room: Predictive performance of PLS models. *Journal of Business Research*, 69(10), 4552–4564. <https://doi.org/10.1016/j.jbusres.2016.03.049>
- Shmueli, G., Sarstedt, M., Hair, J. F., Cheah, J. H., Ting, H., Vaithilingam, S., & Ringle, C. M. (2019). Predictive model assessment in PLS-SEM: guidelines for using PLSpredict. *European Journal of Marketing*, 53(11), 2322–2347. <https://doi.org/10.1108/EJM-02-2019-0189>
- Smith, J. K., & Smith, L. F. (2014). Developing Assessment Tasks. In C. Wyatt-Smith, V. Klenowski, & P. Colbert (Eds.), *Designing Assessment for Quality Learning* (pp. 123–136). Springer. <https://doi.org/10.1007/978-94-007-5902-2>
- Srijbos, J., Engels, N., & Struyven, K. (2015). Criteria and standards of generic competences at bachelor degree level: A review study. *Educational Research Review*, 14, 18–32. <https://doi.org/10.1016/j.edurev.2015.01.001>
- Tai, J., Ajjawi, R., Boud, D., Dawson, P., & Panadero, E. (2018). Developing evaluative judgement: enabling students to make decisions about the quality of work. *Higher Education*, 76(3), 467–481. <https://doi.org/10.1007/s10734-017-0220-3>
- Thomas, T., Jacobs, D., Hurley, L., Martin, J., Maslyuk, S., Lyall, M., & Ryan, M. (2019). Students' perspectives of early assessment tasks in their first-year at university. *Assessment & Evaluation in Higher Education*, 44(3), 398–414. <https://doi.org/10.1080/02602938.2018.1513992>
- Wren, J., Sparrow, H., Northcote, M., & Sharp, S. (2009). Higher Education Students' Perceptions of Effective Assessment. *International Journal of Learning*, 15(12), 11–23. <https://doi.org/10.18848/1447-9494/CGP/v15i12/46037>
- Zhao, X., Lynch, J. G., & Chen, Q. (2010). Reconsidering Baron and Kenny: Myths and truths about mediation analysis. *Journal of Consumer Research*, 37(2), 197–206. <https://doi.org/10.1086/651257>

### Anexo I: Estructura dimensional del cuestionario ATAE (Análisis de tareas de evaluación y aprendizaje)

Reflexiona sobre la tarea de evaluación y valora, en una escala de 0 (nada) a 10 (total), *en qué grado o medida has puesto en práctica o desarrollado los siguientes aspectos* que se detallan

Dimensión	ETIQUETA	ITEM N°	Items
<b>PROFUNDIDAD (PRO)</b>	<b>PRO_1</b>	I1	Utilizar métodos de indagación e investigación
	<b>PRO_2</b>	I6	Demostrar una comprensión profunda de conceptos e ideas fundamentales
	<b>PRO_3</b>	I11	Identificar, articular y relacionar los conceptos y tópicos fundamentales de la asignatura
	<b>PRO_4</b>	I12	Desarrollar el pensamiento reflexivo y crítico
<b>COMUNICACIÓN (COM)</b>	<b>COM_1</b>	I2	Hacer uso de estrategias de comunicación oral o escrita
	<b>COM_2</b>	I8	Argumentar de forma razonada y fundamentada
	<b>COM_3</b>	I16	Presentar productos ante audiencias internas o externas
<b>RETADORA (RET)</b>	<b>RET_1</b>	I3	Establecer relaciones y conexiones significativas
	<b>RET_2</b>	I4	Coordinar el proceso y las actuaciones para dar respuesta a lo requerido en la tarea
	<b>RET_3</b>	I5	Asumir riesgos optando por soluciones que implican creatividad, mayor dificultad o incertidumbre
	<b>RET_4</b>	I7	Buscar soluciones o perspectivas alternativas
<b>TRANSFERENCIA (TRA)</b>	<b>TRA_1</b>	I9	Integrar y relacionar los conocimientos, habilidades y experiencias previas con otros nuevos, estableciendo conexiones significativas y relevantes
	<b>TRA_2</b>	I10	Relacionar conocimientos, habilidades y experiencias previas con otros nuevos
	<b>TRA_3</b>	I13	Relacionar los conocimientos y experiencias con otras materias
	<b>TRA_4</b>	I14	Relacionar los conocimientos y experiencias con la realidad social
	<b>TRA_5</b>	I15	Realizar productos concretos (proyectos, ensayos, presentaciones, debates, ejecuciones, etc.)
<b>PREGUNTAS ABIERTAS</b>		P1	¿Consideras que esta tarea de evaluación ha sido retadora? ¿Por qué? Justifica razonadamente tu respuesta
		P2	¿Consideras que esta tarea de evaluación te ha exigido rigurosidad intelectual? ¿Por qué? Justifica razonadamente tu respuesta
		P3	¿Consideras que esta tarea de evaluación tiene relación con el ámbito profesional? ¿Por qué? Justifica razonadamente tu respuesta
		P4	Globalmente, ¿qué te ha supuesto realizar esta tarea?, ¿qué competencias consideras que has puesto en práctica? ¿cómo consideras que ha sido tu desempeño en esta tarea?

#### Authors / Autores

**Ibarra-Sáiz, M.S.** ([marisol.ibarra@uca.es](mailto:marisol.ibarra@uca.es))  0000-0003-4513-702X

Profesora Titular de Universidad en el Área de Métodos de Investigación y Diagnóstico en Educación de la Universidad de Cádiz. Directora de la Cátedra UNESCO en *Evaluación, Innovación y Excelencia en Educación*. Directora del Grupo de Investigación EVALfor-SEJ509 – *Evaluación en contextos formativos*, del Plan Andaluz de Investigación, Desarrollo e Innovación (PAIDI). Desarrolla su investigación principalmente en el ámbito de la evaluación en educación superior. Ha sido investigadora principal de más de 10 proyectos europeos, internacionales y nacionales cuyos resultados han sido publicados en diversos artículos, capítulos de libros y contribuciones a congresos internacionales. Actualmente es coinvestigadora principal del Proyecto FLOASS - *Resultados y analíticas de aprendizaje en la educación superior: Un marco de acción desde la evaluación sostenible* (RTI2018-093630-B-I00) en el que participan 6 universidades españolas.

**Rodríguez-Gómez, G.** ([gregorio.rodriguez@uca.es](mailto:gregorio.rodriguez@uca.es))  0000-0001-9337-1270

Catedrático de Universidad de Métodos de Investigación y Diagnóstico en Educación de la Universidad de Cádiz. Es coordinador del área estratégica "Estudios e investigaciones en evaluación" de la Cátedra UNESCO en *Evaluación, Innovación y Excelencia en Educación*. Miembro fundador del Grupo de Investigación EVALfor-SEJ509 - *Evaluación en contextos formativos*. Sus intereses de investigación se centran en la metodología de investigación y en la evaluación en educación superior. Actualmente es coinvestigador principal del Proyecto FLOASS - *Resultados y analíticas de aprendizaje en la educación superior: Un marco de acción desde la evaluación sostenible* (RTI2018-093630-B-I00). Autor de artículos, capítulos de libro y aportaciones en conferencias internacionales. Ha sido Presidente de la Asociación Interuniversitaria de Investigación Pedagógica (AIDIPE). Actualmente es Presidente de la Red Estatal de Docencia Universitaria RED-U.



**Revista ELectrónica de Investigación y EValuación Educativa**  
*E-Journal of Educational Research, Assessment and Evaluation*

[ISSN: 1134-4032]



Esta obra tiene [licencia de Creative Commons Reconocimiento-NoComercial 4.0 Internacional](https://creativecommons.org/licenses/by-nc/4.0/).  
This work is under a [Creative Commons Attribution 4.0 International license](https://creativecommons.org/licenses/by-nc/4.0/).