



Validación de una herramienta para evaluar las presentaciones teórico-prácticas en la Educación Superior

Validation of a tool to evaluate theoretical-practical presentations in Higher Education

Authors

Antonio Jesús Sánchez-Oliver ¹
 Francisco José Rodríguez Rojas ²
 Adrián Feria-Madueño ²
 Alejandro Muñoz-López ¹
 Ángel Carnero Díaz ²
 Antonio Muñoz-Llerena ²
 Borja Sañudo-Corrales ²
 Miguel Ángel Oviedo-Caro ²
 Moisés Grimaldi-Puyana ²
 Paula Bianchi ¹
 Raúl Domínguez ¹
 Salva Angosto ²

¹ Departamento de Motricidad Humana y Rendimiento Deportivo, Universidad de Sevilla, Sevilla (España)

² Departamento de Educación Física y Deporte, Universidad de Sevilla, Sevilla (España)

Autor de correspondencia:
 Moisés Grimaldi-Puyana
 mgrimaldi@us.es

How to cite in APA

Sanchez-Oliver, A. J., Rodríguez Rojas, F. J., Feria-Madueño, A., Muñoz-López, A., Carnero Díaz, Ángel, Muñoz-Llerena, A., Sañudo-Corrales, B., Oviedo-Caro, M., Ángel, Grimaldi-Puyana, M., Bianchi, P., Domínguez, R., & Angosto, S. (2025). Validation of a tool to evaluate theoretical-practical presentations in Higher Education. *Retos*, 67, 903-916. <https://doi.org/10.47197/retos.v67.112987>

Resumen

Introducción: En el contexto del Espacio Europeo de Educación Superior, la evaluación formativa y compartida se establece como un enfoque clave para fomentar el aprendizaje activo y las competencias transversales. Sin embargo, se requiere de herramientas validadas que apoyen este proceso en la educación superior.

Objetivo: Este estudio tuvo como propósito validar una rúbrica para evaluar presentaciones teórico-prácticas en los grados de Educación Primaria y Ciencias de la Actividad Física y del Deporte, integrando autoevaluación, heteroevaluación y coevaluación.

Metodología: La rúbrica fue diseñada basándose en criterios específicos para medir el desempeño estudiantil. Se utilizó una muestra de 1.257 evaluaciones realizadas por profesorado y alumnado. Se llevaron a cabo análisis factoriales exploratorios y confirmatorios para evaluar la estructura y fiabilidad del instrumento.

Resultados: Los resultados confirmaron la estructura y fiabilidad de la rúbrica, con índices de ajuste adecuados ($RMSEA < 0,08$; $CFI > 0,90$). La varianza explicada alcanzó el 57 % en las evaluaciones teóricas y el 59,85 % en las prácticas. Además, se observó una mejora significativa en la claridad y organización de las presentaciones estudiantiles.

Discusión: Los hallazgos son consistentes con estudios previos, destacando el impacto positivo de la rúbrica en el desarrollo de competencias como la autorregulación y el aprendizaje reflexivo. Se identificaron áreas de mejora en algunos ítems para optimizar su rendimiento.

Conclusiones: La rúbrica validada constituye un recurso efectivo para la evaluación formativa, promoviendo un aprendizaje significativo y la implicación activa del alumnado. Su replicabilidad y flexibilidad la posicionan como un modelo aplicable a diversos contextos educativos.

Palabras clave

Evaluación Formativa; Educación Superior; Rúbrica; Evaluación Compartida; Competencias Transversales

Abstract

Introduction: Within the framework of the European Higher Education Area, formative and shared evaluation is established as a key approach to foster active learning and transversal competencies. However, validated tools are needed to support this process in higher education.

Objective: This study aimed to validate rubric designed to evaluate theoretical-practical presentations in Primary Education and Sports Science degrees, integrating self-assessment, peer assessment, and teacher assessment.

Methodology: A rubric was developed based on specific criteria to measure student performance. A total of 1,257 evaluations conducted by faculty and students were analyzed using exploratory and confirmatory factor analyses to ensure the instrument's validity and reliability.

Results: The results confirmed the rubric's structure and reliability, with adequate fit indices ($RMSEA < 0.08$; $CFI > 0.90$). The explained variance reached 57% for theoretical evaluations and 59.85% for practical evaluations. Significant improvements were observed in the clarity and organization of student presentations.

Discussion: The findings align with previous studies, highlighting the positive impact of the rubric on developing competencies such as self-regulation and reflective learning. Areas for improvement were identified in some items to optimize performance.

Conclusions: The validated rubric provides an effective framework for formative evaluation, promoting meaningful learning and active student engagement. Its replicability and flexibility position it as a model applicable to various educational contexts.

Keywords

Formative Assessment; Higher Education; Rubric; Shared Assessment; Transversal Competencies



Introducción

En el contexto de la Espacio Europeo de Educación Superior (EEES), uno de los desafíos más persistentes radica en la distinción entre evaluación y calificación, términos que a menudo se confunden, lo que dificulta la adopción de modelos formativos (Gibbs, 2003). La excesiva preocupación por calificar, basada en exámenes parciales o finales, ha reducido la evaluación a un proceso centrado en notas y puntuaciones, generando un clima de tensión y ansiedad en el alumnado (López-Pastor & Palacios, 2012). Este fenómeno es evidente en las Ciencias de la Actividad Física y el Deporte (CCAFD), donde las evaluaciones se limitan con frecuencia a exámenes teóricos que, aunque pedagógicamente cuestionables, continúan empleándose. No obstante, un número creciente de docentes ha comenzado a implementar sistemas de evaluación formativa, promoviendo prácticas educativas más centradas en el aprendizaje integral y la participación del estudiante (Pascual-Arias et al., 2023).

La Evaluación Formativa y Compartida constituye un enfoque clave para alinear la evaluación con la programación por competencias en la EEES. Este modelo ha demostrado evidencias significativas sobre su impacto positivo en la calidad del aprendizaje, el rendimiento académico y el desarrollo de capacidades como la metacognición, la motivación, la implicación, la autonomía y la capacidad crítica del alumnado (Pascual-Arias et al., 2023). En este marco, se enfatiza la integración de la evaluación del aprendizaje y la enseñanza, la participación de los estudiantes, el uso continuo y compartido de instrumentos de evaluación, y un sistema de comunicación bidireccional y reiterado entre alumnado y profesorado, alineado con criterios competenciales (Navarro-Adelantado & Jiménez-Jiménez, 2021; Sánchez-Olivier et al., 2025).

La implicación del alumnado en los procesos evaluativos, especialmente a través de técnicas como la autoevaluación y la coevaluación, ha demostrado múltiples beneficios (López Guillén & Taveras Sandoval, 2022; Kambourova et al., 2021; Pérez-Pueyo, 2016). Estas estrategias no solo fomentan competencias clave como la autorregulación, la autocritica y la responsabilidad, sino que también mejoran el aprendizaje y favorecen un ambiente democrático en el aula (Bozkurt, 2020). En este sentido, la autoevaluación y la coevaluación destacan como herramientas fundamentales para promover un aprendizaje autónomo y auténtico, permitiendo al alumnado reflexionar sobre su desempeño y comprender mejor su propio proceso de aprendizaje (Carrizosa & Gallardo, 2012; Panadero et al., 2013). Paralelamente, estos enfoques brindan a los docentes información valiosa sobre el nivel de entendimiento del alumnado y áreas específicas de mejora (Sánchez Mesa, 2004). Su implementación, fomenta competencias esenciales, como la capacidad personal y social de aprender a aprender, consolidando su papel dentro de los enfoques educativos actuales (Gairín Sallán & Olmos Rueda, 2024; Lluch Molins & Portillo Vidiella, 2018).

Un paso avanzado en este ámbito es la evaluación triádica, que integra la participación del alumnado, sus compañeros/as y el docente en un proceso conjunto de reflexión, diálogo y consenso (Pérez Pueyo et al., 2024). Este modelo no solo facilita la identificación de logros y áreas de mejora, sino que también proporciona retroalimentación valiosa que enriquece el aprendizaje (Ortiz Ocaña, 2015). Para que sea efectiva, es fundamental que los tres agentes empleen instrumentos claros y objetivos, diseñados específicamente para mejorar la comprensión y el progreso del alumnado (Miranda Ullán & Caballero Blanco, 2024). La evaluación triádica, junto con modalidades diádicas como la autoevaluación y la heteroevaluación, demuestra que los procesos de evaluación formativa tienen un impacto significativo en el aprendizaje, fomentando habilidades como la justicia, la responsabilidad y la autorregulación (Padilla Carmona & Gil Flores, 2008).

En este contexto, los instrumentos de evaluación desempeñan un papel crucial, ofreciendo criterios claros y niveles de logro bien definidos que orientan al alumnado y garantizan la coherencia en las actividades evaluativas (Moreno, 2021). Entre estos, las rúbricas o escalas descriptivas han emergido como herramientas esenciales desde la aparición del enfoque competencial en la educación, proporcionando descripciones cualitativas que guían tanto al alumnado como al profesorado. El uso de rúbricas en la EEES ha crecido considerablemente en los últimos años, debido a su gran utilidad en la realización de evaluaciones (Velasco-Martínez et al., 2018). Para medir cualquier aspecto, es imprescindible comprenderlo (Spitzer, 2017). Las rúbricas son herramientas que promueven la transparencia en los procesos de evaluación, utilizando criterios claros, específicos y accesibles para todos. En esencia, las rúbricas se emplean para evaluar el desempeño (Brookhart, 2013). Esto se logra mediante procesos de monitoreo



y medición, donde se integran diversos indicadores o dimensiones que describen de manera precisa las características necesarias para alcanzar un nivel determinado de desarrollo. Así, permiten establecer niveles de logro en tareas concretas, programas de innovación o, por ejemplo, la certificación de competencias o habilidades específicas (López Pastor & Pérez Pueyo, 2017). En síntesis, las rúbricas facilitan la alineación entre estrategias y objetivos, definiendo previamente qué será evaluado, ya sea una competencia, el progreso estudiantil o el desempeño organizacional. Por tanto, representan una herramienta ideal para evaluar el proceso de aprendizaje, aportando retroalimentación valiosa e informativa (Andrade, 2000).

Este enfoque, basado en instrumentos claros y procesos de triangulación de información como la auto-evaluación, coevaluación y heteroevaluación, representa un cambio metodológico significativo que prioriza el aprendizaje reflexivo y continuo, transformando las aulas en espacios de desarrollo integral y significativo (Sánchez Mesa, 2004). En resumen, este enfoque evaluativo transforma la evaluación en un proceso continuo, compartido y bidireccional, donde instrumentos como las rúbricas permiten no solo registrar el nivel de logro, sino también potenciar el aprendizaje profundo y el desarrollo integral del alumnado (Ribera Vázquez, 2014).

Por otro lado, la comunicación oral es una competencia transversal fundamental para los estudiantes universitarios de los Grados en Educación Primaria (GEP) y GCCAFD, quienes, como futuros maestros y educadores físicos, desempeñarán un rol clave en contextos que requieren interacción constante con otras personas (Fernández García et al., 2008). Aunque el lenguaje es una facultad típicamente humana, la habilidad para expresarse con claridad, precisión y eficacia en situaciones formales y planificadas no es innata, sino que debe ser trabajada y perfeccionada (Avendaño & Miretti, 2007). Desarrollar esta competencia permite a los futuros profesionales convertirse en comunicadores eficaces, una habilidad esencial en su práctica profesional (Fernandez- López, 2007). Ya sea en exposiciones dentro o fuera del aula, reuniones de trabajo, presentaciones de informes y resultados, entrevistas con alumnos/as, padres, madres u otros profesionales, o en la gestión y coordinación de intervenciones orales, el éxito de su desempeño dependerá en gran medida de su capacidad para manejar estrategias comunicativas adaptadas a cada situación (Fernández García et al., 2008). Por ello, fomentar la comunicación oral en estos grados universitarios no solo enriquece su formación, sino que asegura que estén preparados para los desafíos comunicativos inherentes a su futura labor educativa.

Sin embargo, a pesar de los avances teóricos y prácticos en la implementación de modelos formativos y compartidos en la EEES, aún persiste la necesidad de contar con herramientas validadas que garanticen su eficacia en contextos específicos, como en los GEP y GCCAFD. La validación de instrumentos como las rúbricas o escalas descriptivas se presenta como un paso esencial para asegurar su capacidad de proporcionar una evaluación equitativa, transparente y centrada en el aprendizaje. Estos instrumentos no solo deben ofrecer criterios claros y niveles de logro definidos, sino también fomentar una interacción significativa entre alumnado y profesorado, promoviendo la reflexión, el diálogo y el aprendizaje profundo (Ruay Garcés, 2018). Por lo tanto, el objetivo de este artículo es la validación de una herramienta para evaluar las presentaciones teórico-prácticas en la EEES con participación de alumnado y profesorado siguiendo un proceso de evaluación basado en la autoevaluación, heteroevaluación y coevaluación.

Método

Participantes

La muestra total del estudio se compuso de 1.257 evaluaciones de profesorado y alumnado en los GEP y GCCAFD de la Universidad de Sevilla. La Tabla 1 muestra la distribución de los cuatro grupos considerados para la realización de la validación cruzada de las escalas de evaluación teórica y de evaluación práctica. Además, se especifican las asignaturas consideradas para la evaluación.

Tabla 1. Muestra de los grupos de validación del estudio.

Total	Profesor		Alumnado		Asignatura/s
	N	%	N	%	
Grupo 1. Teoría	349	63	18,1	286	81,9
					Iniciación Deportiva en la Escuela



Grupo 2. Teoría	141	56	39,7	85	60,3	Actividad física, deporte y recreación, Investigación en Actividad Físico-Deportiva como Hábito y Calidad de Vida en las Personas Adultas y Mayores
Grupo 3. Práctica	513	229	44,6	284	55,4	Acondicionamiento Físico en la Escuela, Iniciación Deportiva en la Escuela, Voleibol II: Fundamentos Avanzados y su Enseñanza, Preparación Física, Innovación Docente e Iniciación a la Investigación Educativa en Educación Física
Grupo 4. Práctica	254	98	38,6	156	61,4	Acondicionamiento Físico en la Escuela, Preparación física, Innovación Docente e Iniciación a la Investigación Educativa en Educación Física
Total	1,257	446	35,5	811	64,5	-

Instrumento

Siguiendo un proceso metódico basado en las indicaciones de cómo diseñar escalas descriptivas (rúbricas) de los autores López Pastor y Pérez Pueyo (2017), se diseñó una rúbrica para evaluar presentaciones teórico-prácticas en la EEES con la participación activa de alumnado y profesorado a través de autoevaluación, heteroevaluación y coevaluación. Este instrumento se creó definiendo objetivos y contenidos relacionados directamente con las tareas asignadas, identificando criterios claros que evalúan la calidad del trabajo, estableciendo niveles de logro descriptivos y diferenciados que reflejan el desempeño del alumnado, y determinando un enfoque cualitativo y cuantitativo adaptado tanto a la evaluación formativa como a la calificación. Además, se aplicó en modo de prueba para ajustarlo a partir de producciones previas del alumnado, comprobando su idoneidad y corrigiendo aspectos necesarios para garantizar su funcionalidad. Finalmente, la rúbrica se implementó en un contexto real, integrando la retroalimentación del alumnado para asegurar que el instrumento no solo sea útil en la evaluación, sino que también fomente el aprendizaje y el desarrollo de competencias.

La rúbrica o escala descriptiva diseñada se estableció en base a 4 diferentes grados o niveles para cada ítem o aspecto a evaluar: Excelente (4); Satisfactorio (3); Mejorable (2); e Insuficiente (1). Los ítems fueron 8 para las presentaciones o exposiciones orales teóricas y 10 para las sesiones prácticas (Tabla 2). En cada nivel aparece descripciones lo más precisas posibles de las características que puede tener una producción de un alumnado, de las posibles conductas del alumnado, o de las posibles respuestas a una pregunta o actividad de aprendizaje. El instrumento se diseñó para que fuera eficaz para la calificación, pero también para que formara parte del proceso de evaluación formativa, a pesar de transformarlos finalmente en valores numéricos. En el material suplementario se puede encontrar la escala descriptiva o rúbrica para ampliar la información sobre la herramienta.

Tabla 2. ítems o aspectos valorados en la escala descriptiva.

	Presentaciones o exposiciones orales teóricas	Presentaciones o exposiciones orales prácticas
Ítem 1	Contenido	Contenido
Ítem 2	Organización de la información	Exposición
Ítem 3	Exposición	Expresión oral/Lenguaje no verbal
Ítem 4	Expresión oral	Distribución y organización de grupos
Ítem 5	Lenguaje no verbal	Tiempo de compromiso motor
Ítem 6	Tiempo	Elección y organización de material
Ítem 7	Soporte	Tareas/actividades/juegos/ejercicios seleccionados
Ítem 8	Trabajo en equipo	Distribución del tiempo
Ítem 9		Instrucciones y control de la clase
Ítem 10		Trabajo en equipo

Procedimiento

En el marco de un proyecto de innovación educativa financiado por el IV Plan Propio de Docencia de la Universidad de Sevilla, convocatoria 2023, titulado *Evaluación Formativa: Autoevaluación, Coevaluación y Heteroevaluación en la Enseñanza Universitaria*, se llevó a cabo la recogida de respuestas necesarias para la validación de la rúbrica diseñada. Este proyecto involucró a profesorado y alumnado de los GEP y CCAFD de la Universidad de Sevilla, con un enfoque en la comunicación oral como competencia transversal y fundamental.

El procedimiento comenzó con reuniones entre el profesorado de diferentes asignaturas para elaborar la rúbrica mencionada. Posteriormente, esta fue presentada al alumnado con el objetivo de informarles sobre su contenido, promover su comprensión y, además, permitirles participar en su mejora mediante la revisión y modificación de algunos ítems. Una vez ajustada para optimizar su claridad y adecuación,



la rúbrica se implementó en la evaluación de tareas de presentación oral realizadas durante clases teóricas y prácticas.

La metodología utilizada incluyó tres modalidades de evaluación: (i) heteroevaluación: el profesorado evaluó al grupo que realizaba la exposición; (ii) autoevaluación: el alumnado evaluó su propio desempeño durante la exposición; y (iii) coevaluación: los compañeros y compañeras del grupo evaluaron las exposiciones de sus pares; asegurando así, la evaluación como formativo, compartido y bidireccional, en el que la rúbrica permitió registrar el nivel de logro y potenciar el aprendizaje y desarrollo integral del alumnado.

Todas las respuestas fueron recopiladas a través de la aplicación Forms de Microsoft, proporcionada por la Universidad de Sevilla. Finalmente, la base de datos fue depurada para garantizar la calidad y fiabilidad de los datos antes de proceder al análisis estadístico. Este procedimiento permitió validar la herramienta de evaluación y su eficacia en el desarrollo de la competencia comunicativa en el contexto universitario.

Análisis de los datos

En primer lugar, se comprobaron las propiedades psicométricas de los indicadores de la escala utilizando el software estadístico JASP 0.18.1 (JASP Team, 2023). Posteriormente, se comprobó la validez de la estructura factorial mediante el análisis factorial exploratorio (AFE) y el análisis factorial confirmatorio (AFC). Se realizó una validación cruzada tanto de la escala de evaluación para la teoría como para la práctica. El Grupo 1 y Grupo 3 se utilizaron para la realización del AFE, mientras que el Grupo 2 y el Grupo 4 para la realización del AFC considerando la validación de la escala teórica y práctica respectivamente.

El AFE se realizó con el programa FACTOR v.12 siguiendo las recomendaciones de Lloret-Segura et al. (2014). Este análisis se llevó a cabo mediante el método de extracción de Análisis factorial de rango mínimo (MRFA; ten Berge, & Kiers, 1991) y se utilizó el método de rotación Oblimin Directa. Para determinar el número de factores se utilizó el procedimiento de Implementación Óptima del Análisis Paralelo (Timmerman & Lorenzo-Seva, 2011), mientras que para comprobar el ajuste del modelo se analizaron los coeficientes de error cuadrático medio de aproximación (RMSEA) y el Índice de ajuste comparativo (CFI) propuesto por Tanaka y Huba (1989). Otros indicadores que se tuvieron en cuenta fueron el índice G-H generalizado para analizar la replicabilidad de los factores derivados del AFE. También se observaron las medidas de adecuación muestral de Kaiser Meyer Olkin (KMO), así como la prueba de esfericidad de Bartlett (BST). Por último, se comprobó la interpretabilidad teórica de la solución factorial extraída del AFE.

Se realizó un análisis factorial confirmatorio (AFC) a través del paquete MPLUS aplicando el método de Estimación de Máxima Verosimilitud (ML) con el objetivo de corregir la posible ausencia de normalidad multivariante. Así, para la evaluación del ajuste global se utilizaron diferentes índices de bondad de ajuste recomendados en la literatura (Kline, 2015), como la significación de la Chi-cuadrado. Además, se calcularon otros coeficientes que permitieron comprobar la adecuación de los modelos propuestos, como la relación entre χ^2 y sus grados de libertad (χ^2/df), siendo los valores aceptables inferiores a cinco (Byrne, 2013). Del mismo modo se comprobaron los coeficientes de los índices de bondad de ajuste robusto del modelo propuesto, el Índice de Ajuste No Normado (NNFI), el Índice de Ajuste Comparado (CFI) y el Índice de Ajuste Incremental (IFI). Para estos indicadores, se considera un buen ajuste con valores superiores a 0,90 (MacCallum & Austin, 2000). Para finalizar, se muestra el error cuadrático medio de aproximación (RMSEA), considerándose un buen ajuste una puntuación inferior a .08 (Kline, 2013). En segundo lugar, en la evaluación de la fiabilidad de las escalas se tuvieron en cuenta tres medidas: El Alfa de Cronbach (α), la Fiabilidad Compuesta (FC) y la Varianza Media Extraída (AVE) para cada factor (Hair et al., 2014).

Resultados

Análisis de las propiedades psicométricas de los ítems

Para comprobar las propiedades psicométricas de los ítems de la escala, se observaron los valores de correlación ítem-total, media, desviación típica, asimetría y curtosis. Esta información se muestra en la



Tabla 3, tanto para los ítems de la escala de evaluación teórica como la escala de evaluación práctica. Esta tabla muestra que la totalidad de los ítems tuvieron coeficientes de correlación ítem-total superiores al valor recomendado por la literatura ($>0,30$; Nunnally, 1995). Según Chou y Bentler (1995), el límite aceptable de asimetría y curtosis es un valor absoluto de 3,0, siendo superado únicamente la curtosis por el ítem de “distribución tiempo” de la escala teórica. Sin embargo, la eliminación del ítem no mejoraría significativamente la fiabilidad del mismo (Teoría: 0,766; Práctica: 0,769).

Tabla 3. Propiedades psicométricas de la escala.

	M	DT	r_{ix}	$\alpha.-x$	A	K
Escala evaluación teoría						
Ítem 1. Contenido	3,48	0,6	0,640	0,712	-0,46	-0,82
Ítem 2. Organización	3,68	0,5	0,493	0,738	-1,39	0,98
Ítem 3. Exposición	3,47	0,6	0,537	0,730	-0,50	-0,73
Ítem 4. Expresión oral	3,44	0,5	0,478	0,740	-0,18	-1,09
Ítem 5. Lenguaje no verbal	3,47	0,6	0,468	0,741	-0,75	0,01
Ítem 6. Distribución tiempo	3,65	0,6	0,312	0,768	-2,00	4,51
Ítem 7. Soporte o material	3,34	0,8	0,373	0,763	-0,84	-0,10
Ítem 8. Trabajo en equipo	3,62	0,6	0,483	0,738	-1,68	2,97
Escala evaluación práctica						
Ítem 1. Contenido	3,61	0,5	0,502	0,740	-0,76	-0,72
Ítem 2. Exposición	3,68	0,5	0,515	0,739	-1,00	-0,39
Ítem 3. Expresión	3,66	0,5	0,444	0,748	-1,01	-0,24
Ítem 4. Organización	3,70	0,5	0,379	0,757	-1,30	0,60
Ítem 5. Tiempo	3,53	0,6	0,403	0,755	-0,93	0,09
Ítem 6. Elección	3,75	0,5	0,384	0,756	-1,56	1,32
Ítem 7. Tareas	3,66	0,5	0,438	0,749	-1,21	0,52
Ítem 8. Distribución tareas	3,67	0,5	0,425	0,751	-1,27	0,61
Ítem 9. Instrucción	3,62	0,5	0,466	0,745	-0,95	-0,19
Ítem 10. Trabajo equipo	3,73	0,5	0,386	0,756	-1,48	1,11

Nota: M: media; DT: Desviación típica; r_{ix} : correlación ítem-total; $\alpha.-x$: alfa de Cronbach en caso de eliminar el ítem; A: asimetría; K: curtosis.

Análisis factorial exploratorio

Tras analizar las propiedades de los ítems, se evaluó la validez de constructo del instrumento mediante el AFE y el AFC. Estos análisis se realizaron con muestras independientes de diferentes asignaturas. De este modo, se ha podido realizar una validación cruzada y comprobar el ajuste de la estructura factorial en dos muestras independientes de la población de estudio (Hair et al., 2014).

Tras comprobar las propiedades de los ítems, se realizó un AFE para los ocho ítems de la escala teórica y para los diez ítems de la escala práctica. Se tuvieron en cuenta la medida KMO de adecuación muestral y la prueba Bartlett, que indican la proporción de la varianza en las variables de estudio que puede ser causada por cada dimensión subyacente. Los valores más altos indican que el análisis factorial puede ser útil con los datos, mientras que los valores más bajos indican lo contrario. En la muestra del estudio, el valor KMO (Teoría: 0,86; Práctica: 0,91) y la prueba de Bartlett (Teoría: 772,7(28); Práctica: 1862,3(45)) indicaba que la matriz de correlaciones de la muestra no era nula ($p < 0,001$) y los datos eran apropiados para el AFE.

El procedimiento de Aplicación Óptima del Análisis Paralelo sugirió agrupar los ítems de las dos escalas en un único factor. Para ello se procedió a realizar una evaluación de la unidimensionalidad siguiendo las indicaciones de Ferrando y Lorenzo-Seva (2018). Los resultados (Tabla 4) mostraron la existencia de valores de I-Unico superiores a 0,95 en todos los ítems excepto en ítem 6 de la escala teórica. En el caso de I-ECV todos los ítems mostraron valores superiores a 0,85 excepto los ítems 4 y 7 de la escala teórica y el ítem 3 de la escala práctica. Por último, el I-real, la gran mayoría de los ítems tuvieron unos valores inferiores a 0,30, excepto el ítem 7 de la escala teórica y el ítem 3 de la escala práctica. Considerando que la gran mayoría de los ítems de ambas escalas se encontraron en los valores indicados sugiere que los datos pueden tratarse como esencialmente unidimensionales.

Tabla 4. Evaluación de la unidimensionalidad de las escalas.

	I-UniCo	IC (95%)	I-ECV	IC (95%)	I-REAL	IC (95%)
Escala evaluación teoría						
Ítem 1. Contenido	1,00	(1,00-1,00)	1,00	(0,99-1,00)	0,02	(0,00-0,08)
Ítem 2. Organización	1,00	(0,99-1,00)	0,99	(0,87-1,00)	0,05	(0,00-0,24)
Ítem 3. Exposición	1,00	(1,00-1,00)	1,00	(0,96-1,00)	0,04	(0,00-0,13)
Ítem 4. Expresión oral	0,98	(0,78-1,00)	0,84	(0,55-1,00)	0,28	(0,03-0,65)
Ítem 5. Lenguaje no verbal	1,00	(1,00-1,00)	0,99	(0,92-1,00)	0,06	(0,00-0,17)
Ítem 6. Distribución tiempo	1,00	(0,74-1,00)	0,94	(0,52-1,00)	0,11	(0,00-0,40)
Ítem 7. Soporte o material	0,62	(0,32-0,99)	0,44	(0,25-0,87)	0,64	(0,18-0,86)
Ítem 8. Trabajo en equipo	1,00	(0,96-1,00)	0,95	(0,77-1,00)	0,13	(0,00-0,32)



Total	0,95	(0,90-0,99)	0,86	(0,79-0,93)	0,17	(0,15-0,17)
Escala evaluación práctica						
Ítem 1. Contenido	1,00	(0,96-1,00)	0,96	(0,78-1,00)	0,13	(0,01-0,33)
Ítem 2. Exposición	0,99	(0,97-1,00)	0,90	(0,81-0,98)	0,24	(0,11-0,35)
Ítem 3. Expresión	0,98	(0,86-1,00)	0,82	(0,63-0,95)	0,33	(0,15-0,52)
Ítem 4. Organización	1,00	(1,00-1,00)	1,00	(0,98-1,00)	0,03	(0,00-0,10)
Ítem 5. Tiempo	0,95	(0,68-1,00)	0,75	(0,48-0,91)	0,35	(0,19-0,70)
Ítem 6. Elección	1,00	(0,92-1,00)	0,91	(0,69-1,00)	0,17	(0,01-0,33)
Ítem 7. Tareas	0,99	(0,95-1,00)	0,90	(0,75-1,00)	0,23	(0,02-0,38)
Ítem 8. Distribución tareas	1,00	(0,85-1,00)	0,92	(0,62-1,00)	0,20	(0,04-0,55)
Ítem 9. Instrucción	1,00	(0,96-1,00)	0,91	(0,78-1,00)	0,20	(0,09-0,34)
Ítem 10. Trabajo equipo	1,00	(0,90-1,00)	0,91	(0,680-1,00)	0,19	(0,03-0,42)
Total	0,99	(0,98-0,99)	0,90	(0,88-0,92)	0,21	(0,16-0,24)

Nota: I-Unico: Congruencia unidimensional de los ítems; I-Ecv: varianza común explicada por ítems; I-Real: cargas absolutas individuales de los ítems.

Esta solución no presentó ningún ítem que tuviera que ser excluido según los criterios de exclusión indicados en la sección de análisis de datos (cargas factoriales inferiores a 0,40 o cargas cruzadas en dos o más factores superiores a 0,40). La Tabla 5 muestra los resultados de la solución factorial de único factor para las escalas de teoría y práctica respectivamente. Para comprobar el ajuste del modelo, se analizaron los coeficientes RMSEA y CFI, que mostraron valores dentro de los puntos de corte recomendados siendo menor de 0,10 en RMSEA (Teoría: 0,072; Práctica: 0,079) y superiores a 0,95 en CFI (Teoría: 0,97; Práctica: 0,97). Por otro lado, el Índice G-H Generalizado, mostró valores superiores a 0,80 en los factores detectados por el AFE para cada escala (Teoría: 0,88; Práctica: 0,91), indicando una posible buena replicabilidad de las dimensiones en otros estudios (Ferrando & Lorenzo-Seva, 2018). La varianza explicada por los ocho ítems de la escala de evaluación teórica fue del 57,0%, mientras que en la escala de evaluación práctica fue del 59,85%.

Tabla 5. Cargas factoriales análisis factorial exploratorio.

	Carga Factorial	Comunalidad
Escala evaluación teoría		
Ítem 1. Contenido	0,789	0,723
Ítem 2. Organización	0,635	0,522
Ítem 3. Exposición	0,668	0,546
Ítem 4. Expresión oral	0,653	0,574
Ítem 5. Lenguaje no verbal	0,619	0,503
Ítem 6. Distribución tiempo	0,548	0,379
Ítem 7. Soporte o material	0,524	0,514
Ítem 8. Trabajo en equipo	0,617	0,476
Escala evaluación práctica		
Ítem 1. Contenido	0,644	0,555
Ítem 2. Exposición	0,724	0,605
Ítem 3. Expresión	0,688	0,635
Ítem 4. Organización	0,715	0,605
Ítem 5. Tiempo	0,597	0,542
Ítem 6. Elección	0,527	0,377
Ítem 7. Tareas	0,687	0,584
Ítem 8. Distribución tareas	0,673	0,608
Ítem 9. Instrucción	0,661	0,523
Ítem 10. Trabajo equipo	0,630	0,552

Análisis factorial confirmatorio

A continuación, se realizó un AFC utilizando la estimación ML (máxima verosimilitud) y el paquete MPlus. Los modelos teóricos, tanto de la escala teórica como la escala práctica, mostraron estimaciones adecuadas con cargas superiores a 0,30 exceptuando el ítem 6 de la escala teórica, aunque su valor fue significativo ($p < 0,001$). El modelo final presentó índices de ajuste adecuados, de acuerdo con los parámetros recomendados por la literatura (Tabla 6), exceptuando el valor de RMSEA de la escala teórica con un valor superior a 0,10 según los límites establecidos por Kline (2013) y el valor NNFI ligeramente inferior a 0,90. Sin embargo, dado que el resto de los índices son adecuados se mantuvo el ítem 6 de la escala.

Tabla 6. Índices de bondad de ajuste de las soluciones factoriales.

Escala	χ^2 (gl)	χ^2/gl	RMSEA (IC)	NNFI	CFI	IFI
Teoría	75,03 (20)	3,75	0,14 (0,11-0,17)	0,86	0,90	0,90
Práctica	94,77 (27)	3,51	0,10 (0,08-0,13)	0,91	0,93	0,93
Valores recomendados	<5		<0,10		>0,90	



Nota: gl: Grados de libertad; IC: Intervalo de confianza; NNFI: Índice de ajuste no normalizado; CFI: Índice de ajuste comparativo; IFI: Índice de ajuste incremental.

Por último, para analizar la fiabilidad, se observaron las medidas de alfa de Cronbach, FC y AVE. Los valores se ajustaron a los parámetros recomendados en la literatura tanto para la escala de evaluación de teoría (α : 0,88; FC: 0,63; AVE: 0,50) como la escala de evaluación de prácticas (α : 0,90; FC: 0,77; AVE: 0,51). Aunque el valor mínimo de FC, alfa de Cronbach considerado adecuado es 0,70 (Nunnally, 1995), y se recomiendan valores superiores a 0,50 para el AVE (Bagozzi & Yi, 1988; Hair et al., 2014), sin embargo, Fornell y Larcker (1981) indican que valores de FC superiores a 0,60 también se pueden considerar adecuados. Todos los factores cumplieron los valores mínimos según los criterios indicados (Tabla 7).

Tabla 7. Análisis factorial confirmatorio.

	Estimación	R ²	EE	valor z	Sig.
Escala evaluación teórica (α : 0,88; FC: 0,63; AVE: 0,50)					
Ítem 1. Contenido	0,490	0,594	0,047	10,44	<0,001
Ítem 2. Organización	0,391	0,527	0,041	9,56	<0,001
Ítem 3. Exposición	0,424	0,564	0,042	10,00	<0,001
Ítem 4. Expresión oral	0,435	0,617	0,041	10,64	<0,001
Ítem 5. Lenguaje no verbal	0,505	0,555	0,051	9,95	<0,001
Ítem 6. Distribución tiempo	0,296	0,206	0,055	5,38	<0,001
Ítem 7. Soporte o material	0,463	0,493	0,050	9,19	<0,001
Ítem 8. Trabajo en equipo	0,365	0,483	0,041	9,00	<0,001
Escala evaluación práctica (α : 0,90; FC: 0,77; AVE: 0,51)					
Ítem 1. Contenido	0,603	0,575	0,046	13,18	<0,001
Ítem 2. Exposición	0,424	0,509	0,035	12,07	<0,001
Ítem 3. Expresión	0,393	0,442	0,036	11,00	<0,001
Ítem 4. Organización	0,517	0,583	0,039	13,32	<0,001
Ítem 5. Tiempo	0,499	0,458	0,044	11,28	<0,001
Ítem 6. Elección	0,539	0,544	0,043	12,68	<0,001
Ítem 7. Tareas	0,632	0,556	0,049	12,86	<0,001
Ítem 8. Distribución tareas	0,527	0,491	0,045	11,82	<0,001
Ítem 9. Instrucción	0,481	0,499	0,040	11,91	<0,001
Ítem 10. Trabajo equipo	0,469	0,462	0,041	11,36	<0,001

Nota: EE: Error estándar; Sig: significación; α : Alfa de Cronbach; FC: Fiabilidad compuesta; AVE: Varianza Media Extraída.

Discusión

El objetivo de este estudio fue validar una escala breve para evaluar las presentaciones teórico-prácticas realizadas por estudiantes de Educación Superior. Esta herramienta ha mostrado índices adecuados dentro de los parámetros marcados dentro de la literatura para confirmar que la escala permite que, no solo el profesor realice una heteroevaluación de los estudiantes, sino que los propios estudiantes utilicen la misma escala para realizar su propia autoevaluación y evaluar a los compañeros (coevaluación). Además, los resultados obtenidos en este estudio no solo han confirmado la validez, sino también la fiabilidad y reproducibilidad de la herramienta. De esta forma, este estudio aporta evidencias sólidas para su uso como instrumento evaluativo, destacándose como una herramienta robusta y alineada con los estándares internacionales de evaluación educativa.

Este estudio ha abordado la evaluación de una herramienta educativa breve y rápida de realizar. Dado el bajo número de ítems de las escalas, el AFE respaldó la estructura unidimensional de cada una de las dos escalas. Esta estructura unidimensional fue respaldada posteriormente por el AFC que ratificó mantener tanto el ítem 6 de la escala teórica como el ítem 3 de la escala práctica. Aunque en estos dos ítems en algunos análisis no obtuvieron valores idóneos dentro de los estándares de la literatura, los investigadores decidieron mantener dichos ítems ya que la variación de los indicadores de validez y fiabilidad no era significativa. Hair et al. (2014) exponen que, en los procesos de validación, las decisiones no deben ir tomadas únicamente por los resultados y datos de los análisis. Estos autores reafirmaron la potestad del investigador de tomar decisiones con base a los resultados y a los propios intereses de la investigación. Además, tanto los valores de cargas factoriales, como los valores de propiedades psicométricas y fiabilidad, no indicaban la necesidad de eliminar los ítems. Por ello, los investigadores tomaron la decisión de mantener ambos ítems en cada una de sus escalas.

Asimismo, la varianza explicada en ambas escalas, situada entre el 55% y el 60% reforzó la capacidad del instrumento para capturar los aspectos fundamentales del desempeño académico por parte de los



estudiantes de Educación Superior. En términos de fiabilidad, todos los valores de los índices analizados se situaron por encima de los umbrales recomendados por la literatura (Hair et al., 2014; Nunnally, 1995). A pesar de que el índice de error RMSEA en la escala teórica, que fue ligeramente superior al valor indicado por Kline (2013), futuros ajustes podrían optimizar su rendimiento.

El diseño de esta herramienta responde a las demandas del EEEs, que prioriza el aprendizaje activo y el desarrollo de competencias transversales y específicas (Hortigüela-Alcalá et al., 2015). Al ofrecer criterios claros y descriptores detallados, la rúbrica permite a los estudiantes identificar áreas de mejora y logros alcanzados, reduciendo la incertidumbre asociada a la evaluación (López-Pastor & Pérez-Pueyo, 2017). En este caso, los estudiantes se hacen responsables también de la propia evaluación al autoevaluarse y evaluar a sus compañeros. Asimismo, estas herramientas también proporcionan al profesorado un marco para realizar retroalimentaciones precisas y constructivas, facilitando un enfoque centrado en el aprendizaje procesual y no únicamente en la calificación (Ortiz Ocaña, 2015).

En este sentido, las rúbricas o escalas descriptivas ofrecen un marco tangible que facilita la precisión en la retroalimentación. Estas herramientas incluyen indicadores claramente definidos, niveles de logro organizados en progresión y descriptores que ayudan a los estudiantes a identificar áreas de mejora y logros alcanzados (López-Pastor & Pérez-Pueyo, 2017). Cuando las rúbricas están bien diseñadas, tienen el potencial de transformar la evaluación universitaria. Estas herramientas pueden ofrecer una solución coherente y replicable para otros contextos educativos, centrada en el proceso de aprendizaje y promoviendo la participación activa de todos los agentes involucrados. Así, la herramienta teórico-práctica de este estudio refuerza la evaluación como un medio de mejora continua y desarrollo integral tanto para el alumnado como para el profesorado. Al fomentar un enfoque formativo, las rúbricas permiten superar el paradigma de la calificación tradicional, favoreciendo un aprendizaje centrado en el proceso más que en la obtención de una nota (Gibbs, 2003).

El uso de modalidades como la autoevaluación y la coevaluación promueve un aprendizaje activo y reflexivo, permitiendo a los estudiantes identificar sus logros y áreas de mejora (Maldonado-Fuentes et al., 2024). Estos enfoques han sido identificados como esenciales para el desarrollo de competencias críticas y transversales en la Educación Superior (Biggs, 2005). Además, la participación activa del alumnado en la revisión inicial del instrumento refuerza su percepción de equidad y fomenta un diálogo constructivo con el profesorado. Estas dinámicas contribuyen a desarrollar competencias transversales esenciales, como la capacidad de emitir juicios razonados y participar en procesos colaborativos (Flecha, 1998; Biggs, 2005).

Estudios previos que han desarrollado herramientas de evaluación basadas en rúbricas, como la desarrollada en este estudio, han demostrado ser instrumentos eficaces para medir competencias transversales en la Educación Superior (Andrade & Du, 2005; Cañadas-Martín, 2018; Fernández, 2016; Jonsson & Svingby, 2007). En comparación con otros estudios, este instrumento no solo permite la medición del desempeño, sino que también incorpora elementos que fomentan la autoevaluación, heteroevaluación y coevaluación, promoviendo un enfoque de evaluación formativa y compartida (Andrade & Du, 2005; Cañadas-Martín, 2018; Jonsson & Svingby, 2007; López-Pastor, 2006).

La implementación de esta herramienta permitió observar mejoras en la organización y claridad de las presentaciones, según lo reflejado en las puntuaciones medias obtenidas con puntuaciones superiores a 3.68 sobre cuatro puntos totales tanto en la herramienta teórica como la práctica. De esta forma se demuestra que el diseño de la rúbrica asegura que los criterios de evaluación sean claros, consistentes y comprensibles para el alumnado, reduciendo la incertidumbre asociada a los procesos evaluativos (López-Pastor & Pérez-Pueyo, 2017). Otro aspecto destacable en los resultados descriptivos es la alta puntuación que han otorgado los estudiantes al trabajo en equipo. Estos resultados pueden ser indicativos de la concienciación de los estudiantes que al final no solo la preparación de las presentaciones es cosa de un grupo, sino también la propia evaluación. Ya que no solo deben saber valorar su propio esfuerzo, sino que también deben ser evaluados por sus propios compañeros. Al introducir la coevaluación el estudiante debe ser consciente que debe esforzarse y colaborar para obtener el reconocimiento de sus iguales (Basurto-Mendoza et al., 2021; Rodríguez Gómez et al., 2013). En este contexto, nuestra herramienta validada da un paso más allá, incorporando al alumnado no solo como parte activa del proceso de evaluación, sino también como agente clave en su diseño e implementación. Este enfoque dialogado y democrático no solo refuerza la transparencia y el compromiso de los estudiantes, sino que también potencia su aprendizaje y desarrollo integral, al involucrarlos directamente en la construcción de



instrumentos como la rúbrica o escala descriptiva utilizada (López Pastor et al., 2005). Además, siguiendo la reflexión de López Pastor y colaboradores (2005), defendemos la importancia de avanzar hacia procesos de evaluación compartida, superando la visión parcial que ofrecen únicamente la coevaluación y/o la autoevaluación, aunque estas representen un notable progreso frente a la heteroevaluación tradicional del profesorado (López Pastor et al., 2005).

A pesar de los buenos resultados en general obtenidos en la validez y fiabilidad de las herramientas teórico-prácticas, este estudio presenta algunas limitaciones. Por ejemplo, algún aspecto de mejora se trata del índice RMSEA de la escala teórica. Este valor sugiere que algunos ítems, como el ítem 6 o 7 de la escala, podrían ajustarse para mejorar el ajuste global del modelo. Además, aunque los datos de este estudio son representativos en el contexto de los grados evaluados, no pueden generalizarse los resultados a toda la población estudiantil o de las propias titulaciones de la facultad. Por ello sería beneficioso replicar la validación en otros contextos y disciplinas académicas.

Otra limitación a tener en cuenta y es importante destacar ha sido la falta de análisis del criterio de evaluación de estudiantes sobre sí mismos o los compañeros. La falta de experiencia en estos futuros profesionales en formación podría producir alteraciones en los resultados. Por tanto, la implementación de sesiones de formación específicas sobre el uso de rúbricas podría maximizar su impacto en el aprendizaje y la evaluación. Además, el profesorado podría establecer rangos de diferencias de medias entre su evaluación y la del estudiante para poder identificar las evaluaciones interesadas que pueden realizar los estudiantes para mejorar sus evaluaciones o la de compañeros o, al contrario, perjudicarlos. Una futura línea de investigación podría incluir estudios cualitativos para profundizar en la percepción de los estudiantes y docentes sobre la utilidad y aplicabilidad de la herramienta.

Conclusiones

Este estudio valida una herramienta robusta para la evaluación de presentaciones teórico-prácticas en la Educación Superior, particularmente en los grados de Educación Primaria y Ciencias de la Actividad Física y del Deporte. La fiabilidad y validez del instrumento, combinadas con su enfoque en la evaluación formativa y compartida, lo posicionan como un recurso valioso para medir competencias transversales y promover un aprendizaje significativo. Su replicabilidad y flexibilidad lo convierten en un modelo a seguir para el desarrollo de herramientas evaluativas en otros contextos educativos.

Agradecimientos

Agradecemos sinceramente al profesorado y alumnado de los grados de Educación Primaria y Ciencias de la Actividad Física y del Deporte de la Universidad de Sevilla por su valiosa colaboración en la validación de la rúbrica diseñada. Su participación activa fue esencial para el desarrollo de la competencia comunicativa y la mejora de la evaluación formativa. También agradecemos a la Universidad de Sevilla por su apoyo a través del IV Plan Propio de Docencia, que hizo posible este proyecto.

Financiación

El presente artículo deriva de un proyecto financiado por el IV Plan Propio de docencia de la Universidad de Sevilla, en la convocatoria del año 2023, concretamente en actividades del profesorado para la innovación educativa (L.2.2. del IV Plan Propio de Docencia), dentro de esta actuación la acción 221., que se centra en el Apoyo a la innovación docente. La publicación forma parte de la subvención JDC2022-048886-I, financiada por MCIN/AEI/10.13039/501100011033 y por la Unión Europea «NextGenerationEU»/PRTR. Este artículo forma parte de la tesis doctoral de Francisco José Rodríguez Rojas.

Referencias

Andrade, H. G. (2000). Using rubrics to promote thinking and learning. *Educational Leadership*, 57(5), 13-18. <https://eric.ed.gov/?id=EJ609600>



- Andrade, H., & Du, Y. (2005). Student perspectives on rubric-referenced assessment. *Practical Assessment, Research & Evaluation*, 10(3). https://scholarsarchive.library.albany.edu/edpsych_fac_scholar/2
- Avendaño, F., & Miretti, M. L. (2007). *El desarrollo de la lengua oral en el aula: Estrategias para enseñar a escuchar y a hablar*. Sevilla: Eduforma.
- Bagozzi, R. P., & Yi, Y. (1988). On the evaluation of structural equation models. *Journal of the Academy of Marketing Science*, 16(1), 74-94. <https://doi.org/10.1007/BF02723327>
- Basurto-Mendoza, S. T., Moreira-Cedeño, J. A., Velásquez-Espinales, A. N., & Rodríguez-Gámez, M. (2021). Autoevaluación, Coevaluación y Heteroevaluación como enfoque innovador en la práctica pedagógica y su efecto en el proceso de enseñanza-aprendizaje. *Polo del Conocimiento*, 6(3), 828-845. <https://doi.org/10.23857/pc.v6i3.2408>
- Bozkurt, F. (2020). Teacher candidates' views on self and peer assessment as a tool for student development. *Australian Journal of Teacher Education*, 45 (1). <https://doi.org/10.14221/ajte.2020v45n1.4>
- Brookhart, S. M. (2013). *How to create and use rubrics for formative assessment and grading*. Ascd. Alexandria, VA, USA.
- Byrne, B. M. (2013). *Structural Equation Modeling with Mplus: Basic Concepts, Applications, and Programming*. Routledge. <https://doi.org/10.4324/9780203807644>
- Cañadas-Martín, L. (2018). *La evaluación formativa en la adquisición de competencias docentes en la formación inicial del profesorado de Educación Física*. Tesis doctoral, Universidad Autónoma de Madrid, Madrid. <https://repositorio.uam.es/handle/10486/684192>
- Carrizosa, E., & Gallardo, J. (2012). Autoevaluación, coevaluación y evaluación de los aprendizajes. *Universitat Oberta de Catalunya*. Recuperado de https://www.uoc.edu/pdf/symposia/dret_tic2012/pdf/4.6.carrizosa-esther-y-gallardo-jose.pdf
- Chou, C.-P., & Bentler, P. M. (1995). Estimates and tests in structural equation modeling. In R. H. Hoyle (Ed.), *Structural equation modeling: Concepts, issues, and applications* (pp. 37-55). Sage Publications, Inc.
- Fernández, A.M. (2016). La rúbrica como instrumento de evaluación y de aprendizaje. En C. Hamodi. (Coord). *Formar mediante la evaluación en la Universidad. Propuestas prácticas útiles para docentes* (pp. 53-63). Valladolid: Ediciones de la Universidad de Valladolid.
- Fernández García, C. M., Torío López, S., Viñuela Hernández, M. P., Molina Martín, S., y Bermúdez Rey, T. (2008). La comunicación oral como competencia transversal de los estudiantes de pedagogía y magisterio: Presentación del diseño metodológico de una innovación para su trabajo en el aula. *Revista de Enseñanza Universitaria*, 31, 26-38. <http://hdl.handle.net/10651/48628>
- Fernández López, S. (2007). *Cómo gestionar la comunicación en organizaciones públicas y no lucrativas*. Madrid: Narcea.
- Ferrando, P. J., & Lorenzo-Seva U. (2018). Assessing the quality and appropriateness of factor solutions and factor score estimates in exploratory item factor analysis. *Educational and Psychological Measurement*, 78, 762-780. <https://doi.org/10.1177/0013164417719308>
- Fornell, C., & Larcker, D. F. (1981). Evaluating Structural Equation Models with Unobservable Variables and Measurement Error. *Journal of Marketing Research*, 18(1), 39-50. <https://doi.org/10.1177/002224378101800104>
- Gairín Sallán, J., & Olmos Rueda, P. (2024). Desarrollo y evaluación de la competencia 'Aprender a Aprender'. *Aula Abierta*, 53(3), 285-293. <https://doi.org/10.17811/rifie.21274>
- Gibbs, G. (2003). *Uso estratégico de la evaluación en el aprendizaje*. En S. G. Brown & A. Glasner (Eds.), *Evaluuar en la universidad. Problemas y nuevos enfoques* (pp. 61-75). Madrid: Narcea.
- Hair, J. F., Black, W. C., Babin, B. J., & Anderson, R. E. (2014). *Multivariate Data Analysis* (7th edition). Pearson.
- Hortigüela-Alcalá, D., Pérez-Pueyo, A., & López-Pastor, V. (2015). Implicación y regulación del trabajo del alumnado en los sistemas de evaluación formativa en educación superior. *Relieve- Revista Electrónica De Investigación Y Evaluación Educativa*, 21(1), ME6. <https://doi.org/10.7203/relieve.21.1.5171>
- JASP Team (2023). *JASP (Version 0.18.1)* [Computer software]. Disponible en: <https://jasp-stats.org/>
- Jiménez, F.; Navarro, V., & Souto, R. (coord.) (2021). *Explorando colaborativamente alternativas de evaluación formativa en la universidad*. La Laguna (Tenerife): Universidad de La Laguna. <http://riull.ull.es/xmlui/handle/915/22232>



- Jonsson, A., & Svingby, G. (2007). The use of scoring rubrics: reliability, validity and educational consequences. *Educational Research Review*, 2, 130-144. DOI: 10.1016/j.edurev.2007.05.002
- Kambourova, M., González-Agudelo, E. M., & Grisales-Franco, L. M. (2021). La autoevaluación del estudiante universitario: revisión de la literatura. *Teoría de la Educación: Revista Interuniversitaria*: 33, 2, 2021, 217-264. <https://doi.org/10.14201/teri.23672>
- Kline, R. (2013). Exploratory and confirmatory factor analysis. En Y. Petscher, C. Schatschneider, D. L. Compton (Eds.), *Applied quantitative analysis in education and the social sciences* (pp. 171-207). Routledge.
- Kline, R. B. (2015). *Principles and Practice of Structural Equation Modeling* (Fourth Edition). Guilford Publications.
- Lloret-Segura, S., Ferreres-Traver, A., Hernández-Baeza, A., & Tomás-Marco, I. (2014). El análisis factorial exploratorio de los ítems: una guía práctica, revisada y actualizada. *Annals of Psychology*, 30(3), 1151-1169. <https://doi.org/10.6018/analesps.30.3.199361>
- Lluch Molins, L., & Portillo Vidiella, M. C. (2018). La competencia de aprender a aprender en el marco de la educación superior. *Revista Iberoamericana De Educación*, 78(2), 59-76. <https://doi.org/10.35362/rie7823183>
- López Guillén, R. G., y Taveras Sandoval, J. (2022). Uso del aprendizaje cooperativo en educación física y su relación con la responsabilidad individual en estudiantes del nivel secundario. *Retos*, 43, 1-9. <https://doi.org/10.47197/retos.v43i0.82607>
- López Pastor, V. M., González Pascual, M., & Barba Martín, J. J. (2005). La participación del alumnado en la evaluación: la autoevaluación, la coevaluación y la evaluación compartida. *Tándem: Didáctica de la educación física*, 17, 21-37. <https://www.grao.com/revistas/la-evaluacion-en-educacion-fisica-2584?contenido=354124>
- López Pastor, V. M. y Palacios Picos, A. (2012). Percepción de los futuros docentes sobre los sistemas de evaluación de sus aprendizajes. *Revista Teoría de la Educación: Educación y Cultura en la Sociedad de la Información*. 13(3), 317-340. http://campus.usal.es/~revistas_trabajo/index.php/re-vistatesi/article/view/9143/9376
- López-Pastor, V.M. & Pérez-Pueyo, A. (2017). Introducción a la creación y uso de escalas descriptivas y rúbricas. En V. M., López-Pastor, & A. Pérez-Pueyo (Eds.), Buenas prácticas docentes. *Evaluación formativa y compartida en educación: Experiencias de éxito en todas las etapas educativas* (pp. 92-115). León: Universidad de León.
- MacCallum, R. C., & Austin, J. T. (2000). Applications of Structural Equation Modeling in Psychological Research. *Annual Review of Psychology*, 51(1), 201-226. <https://doi.org/10.1146/annurev.psych.51.1.201>
- Maldonado-Fuentes, A. C., Peña Vargas, C. S., & Meza Romero, J. C. (2024). Connotaciones atribuidas a la evaluación por futuros docentes de Educación Física en su primer año universitario presencial. *Retos: Nuevas Perspectivas de Educación Física, Deporte y Recreación*, 51, 75-85. <https://doi.org/10.47197/retos.v51.99916>
- Maldonado Fuentes, A. C., Peña Vargas, C. S., y Meza-Romero, J. C. (2024). Connotaciones atribuidas a la evaluación por futuros docentes de Educación Física en su primer año universitario presencial. *Retos*, 51, 75-85. <https://doi.org/10.47197/retos.v51.99916>
- Miranda Ullán, R., & Caballero Blanco, P. J. (2024). La evaluación triádica en el Grado Medio de Técnico en Guía en el Medio Natural y de Tiempo Libre: Una propuesta en el módulo de Maniobras con cuerdas. En Á. Pérez Pueyo, D. Hortigüela Alcalá, C. Gutiérrez García, & R. Barba Martín (Coords.), *Evaluación formativa y compartida en Educación Física: Fundamentos y experiencias prácticas en todas las etapas educativas* (pp. 735-755). Universidad de León.
- Moreno, A. (2021). *Instrumentos de evaluación LOMLOE: Tipos y ejemplos*. Vírgula Blog. Recuperado de <https://virgulablog.es/programacion-didactica/elementos-de-la-programacion-didactica/evaluacion/que-son-los-instrumentos-de-evaluacion/>
- Nunnally, J. C. (1995). *Teoría psicométrica*. McGraw-Hill.
- Ortiz Ocaña, A. (2015). Enseñanza, aprendizaje y evaluación: una configuración triádica. *Tecné, Episteme y Didaxis: TED*, (37), 87-108. http://www.scielo.org.co/scielo.php?script=sci_arttext&pid=S0121-38142015000100006&lng=en&tlang=es.



- Padilla Carmona, M. T., & Gil Flores, J. (2008). La evaluación orientada al aprendizaje en la Educación Superior: Condiciones y estrategias para su aplicación a la docencia universitaria. *Revista Española de Pedagogía*, 66(241), 467-486. <https://www.jstor.org/stable/23766196>
- Panadero, E., Tapia, J. A., & Huertas, J. A. (2013). Autoevaluación: Connotaciones teóricas y prácticas. *Electronic Journal of Research in Educational Psychology*, 11(2), 551-576. Disponible en: <http://www.redalyc.org/articulo.oa?id=293128257012>
- Pascual-Arias, C., López-Pastor, V. M., & Hortigüela-Alcalá, D. (2023). Estudio longitudinal sobre los efectos del desarrollo de la Evaluación Formativa y Compartida en la Formación Inicial del Profesorado. *Cultura, Ciencia y Deporte*, 18(55), 5-31. <https://doi.org/10.12800/ccd.v18i55.1904>
- Pérez-Pueyo, Á., Hortigüela Alcalá, D., & Gutiérrez García, C. (2024). La evaluación triádica en los aprendizajes grupales: El proceso de evaluación de los montajes finales con combas en el Estilo actitudinal. En Á. Pérez Pueyo, D. Hortigüela Alcalá, C. Gutiérrez García, & R. Barba Martín (Coords.), *Evaluación formativa y compartida en Educación Física: Fundamentos y experiencias prácticas en todas las etapas educativas* (pp. 523-557). Universidad de León
- Pérez Pueyo, Ángel. (2016). El Estilo Actitudinal en Educación Física: Evolución en los últimos 20 años. *Retos*, 29, 207-215. <https://doi.org/10.47197/retos.v0i29.38720>
- Ribera Vázquez, L. (2014). *La evaluación como fuente de aprendizaje: Las rúbricas* [Trabajo Fin de Máster, Universidad Internacional de La Rioja]. <https://reunir.unir.net/handle/123456789/3062>
- Rodríguez Gómez, G., Ibarra Saiz, M., & García Jiménez, E. (2013). Autoevaluación, evaluación entre iguales y coevaluación: conceptualización y práctica en las universidades españolas. *Revista de Investigación en Educación*, 2 (11), 198-210. <http://agora.edu.es/servlet/articulo?codigo=4734976>
- Ruay Garcés, R. (2018). La evaluación: una estrategia para desarrollar aprendizajes profundos en el estudiante. *Boletín Virtual*, 7(8), 46-51. <https://revista.redipe.org/index.php/1/article/view/541>
- Sánchez Mesa, S. J. (2004). La heteroevaluación, coevaluación y autoevaluación como dinamizadores del proceso enseñanza-aprendizaje. *Universidad Industrial de Santander*. Recuperado de <https://noesis.uis.edu.co/bitstreams/21edf8a9-4a76-44c6-bc0b-06a306208664/download>
- Sánchez Mesa, S. J. (2004). *La heteroevaluación, coevaluación y autoevaluación como dinamizadores del proceso enseñanza-aprendizaje* [Tesis de especialización, Universidad Industrial de Santander]. <https://noesis.uis.edu.co/bitstreams/21edf8a9-4a76-44c6-bc0b-06a306208664/download>
- Sánchez-Oliver, A. J., Carnero-Díaz, Á., Muñoz-Llerena, A., & Bianchi, P. (2025). Proyecto de innovación docente para la implementación de la evaluación formativa en la universidad. In *Acciones educativas innovadoras en el ámbito universitario* (pp. 85-103). Dykinson.
- Spitzer, D. R. (2007). *Transforming performance measurement: Rethinking the way we measure and drive organizational success*. AMACOM Div American Mgmt Assn. New York, NY, USA.
- Tanaka, J. S., & Huba, G. J. (1989). A general coefficient of determination for covariance structure models under arbitrary GLS estimation. *British Journal of Mathematical and Statistical Psychology*, 42(2), 233-239. <https://doi.org/10.1111/j.2044-8317.1989.tb00912.x>
- Ten Berge, J.M.F., & Kiers, H.A.L. (1991). A numerical approach to the exact and the approximate minimum rank of a covariance matrix. *Psychometrika*, 56, 309-315. <https://doi.org/10.1007/bf02294464>
- Timmerman, M. E., & Lorenzo-Seva, U. (2011). Dimensionality assessment of ordered polytomous items with parallel analysis. *Psychological Methods*, 16(2), 209-220. <https://doi.org/10.1037/a0023353>
- Velasco-Martínez, L. C., & Tójar Hurtado, J. C. (2017). Uso de rúbricas en educación superior y evaluación de competencias. *Revista de currículum y formación del profesorado*, 22 (3), 184-208. <https://doi.org/10.30827/profesorado.v22i3.7998>

Datos de los/as autores/as y traductor/a:

Antonio Jesús Sánchez-Oliver	sanchezoliver@us.es	Autor
Francisco José Rodríguez Rojas	fran.rr.frr@gmail.com	Autor
Adrián Feria-Madueño	aferia1@us.es	Autor
Alejandro Muñoz-López	amunoz26@us.es	Autor
Ángel Carnero Díaz	acarnero1@us.es	Autor
Antonio Muñoz-Llerena	amllerena@us.es	Autor
Borja Sañudo-Corrales	bsancor@us.es	Autor
Miguel Ángel Oviedo-Caro	moviedo@us.es	Autor
Moisés Grimaldi-Puyana	mgrimaldi@us.es	Autor
Paula Bianchi	pbianchi@us.es	Autora
Raúl Domínguez	rdherrera@us.es	Autor
Salva Angosto	angosto@us.es	Autor

