



## Instrumentos y estrategias de evaluación auténtica en Educación Física escolar: una revisión sistemática

*Instruments and strategies for authentic assessment in school Physical Education: a systematic review*

### Autores

Macarena Hurtado-Guerrero <sup>1</sup>  
 Sebastián Espoz-Lazo <sup>2</sup>  
 Rodrigo Yáñez-Sepúlveda <sup>3,7</sup>  
 Mario Soto-González <sup>4</sup>  
 Andrea Castro-Dávila <sup>5</sup>  
 Dilan Galeano-Rojas <sup>6</sup>  
 Pedro Valdivia-Moral <sup>6</sup>  
 Claudio Hinojosa-Torres <sup>3</sup>

<sup>1</sup> Colegio SSCC Padres Franceses (Chile)

<sup>2</sup> Pontificia Universidad Católica de Chile (Chile)

<sup>3</sup> Universidad Andres Bello (Chile)

<sup>4</sup> Colegio Arrayán de Casablanca (Chile)

<sup>5</sup> Universidad Pedagógica Experimental Libertador (Venezuela)

<sup>6</sup> Universidad Granada (España)

<sup>7</sup> Universidad Espíritu Santo (Ecuador)

Autor de correspondencia:  
 Claudio Hinojosa-Torres  
[claudio.hinojosa@unab.cl](mailto:claudio.hinojosa@unab.cl)

Recibido: 24-08-25  
 Aceptado: 26-09-25

### Cómo citar en APA

Hurtado-Guerrero, M., Espoz-Lazo, S., Yáñez-Sepúlveda, R., Soto-González, M., Castro-Dávila, A., Galeano-Rojas, D., Valdivia-Moral, P., & Hinojosa-Torres, C. (2025). Instrumentos y estrategias de evaluación auténtica en Educación Física escolar: una revisión sistemática. *Retos*, 73, 784-797.  
<https://doi.org/10.47197/retos.v73.117462>

### Resumen

**Introducción:** La evaluación auténtica se ha consolidado como una estrategia clave al proponer instrumentos y situaciones que reflejan fielmente las demandas, contextos y complejidades de la práctica real. Sin embargo, su implementación en la práctica docente presenta tensiones vinculadas a la calidad de los instrumentos, la formación del profesorado y las condiciones institucionales.

**Objetivo:** Analizar cómo se materializa la evaluación auténtica en la Educación Física escolar, con énfasis en los instrumentos y estrategias empleadas por los docentes, así como en los factores que condicionan su aplicación.

**Metodología:** La búsqueda se realizó en Web of Science, Scopus, SPORTDiscus y ERIC, siguiendo la guía PRISMA 2020, lo que permitió seleccionar 11 estudios empíricos. Se aplicaron criterios de inclusión con enfoque SPIDER. La calidad metodológica de los estudios seleccionados se desarrolló mediante la herramienta QATSDD.

**Resultados:** Los resultados muestran que la evaluación auténtica se expresa principalmente mediante rúbricas, observación situada en el juego, portafolios y proyectos integradores. No obstante, persisten limitaciones en la validez y fiabilidad de los instrumentos, así como una cobertura insuficiente de las dimensiones socioemocionales y ciudadanas.

**Conclusiones:** Esta revisión aporta evidencia inédita al identificar la brecha entre el discurso favorable a la evaluación auténtica y su implementación real en contextos escolares, subrayando la necesidad de validar instrumentos, fortalecer la alfabetización evaluativa docente y generar apoyos institucionales que aseguren la sostenibilidad de estas prácticas.

### Palabras clave

Aprendizaje; educación; instrumentos de evaluación; práctica docente; retroalimentación.

### Abstract

**Introduction:** Authentic assessment has become a key strategy by proposing instruments and situations that faithfully reflect real practice's demands, contexts, and complexities. However, its implementation in teaching practice reveals tensions related to the quality of the instruments, teachers' assessment literacy, and institutional conditions.

**Objective:** To analyse how authentic assessment is enacted in school Physical Education, with particular emphasis on the instruments and strategies employed by teachers and the contextual factors that shape its application.

**Methodology:** A systematic search was conducted in Web of Science, Scopus, SPORTDiscus, and ERIC, following the PRISMA 2020 guidelines, which led to the selection of 11 empirical studies. The methodological quality of the included studies was appraised using the QATSDD tool.

**Results:** The findings indicate that authentic assessment is primarily expressed through rubrics, game-based observation, portfolios, and integrative projects. Nevertheless, limitations remain concerning the validity and reliability of the instruments, alongside insufficient coverage of socio-emotional and civic dimensions.

**Conclusions:** This review offers novel evidence by identifying the gap between the favourable discourse surrounding authentic assessment and its actual implementation in school contexts. It highlights the need to validate instruments, strengthen teachers' assessment literacy, and provide institutional support to ensure the sustainability of such practices.

### Keywords

Learning; education; assessment instruments; teaching practice; feedback.

## Introducción

La evaluación auténtica (EA) se ha consolidado como un enfoque que desplaza el énfasis de la reproducción de contenidos hacia la demostración situada de competencias mediante tareas de desempeño con alta validez ecológica (Darling-Hammond & Snyder, 2000; Fawns et al., 2024; Villarroel et al., 2017; Wiggins, 1998). En lugar de centrarse únicamente en “lo que el estudiante recuerda”, este enfoque se focaliza en “lo que el estudiante puede hacer” frente a demandas complejas, en contextos que emulan situaciones reales de práctica. Sus fundamentos se remontan a la crítica de Wiggins (1990) a las pruebas descontextualizadas y al giro performativo de la evaluación a finales del siglo XX, ampliamente documentado en la literatura sobre autenticidad en evaluación contemporánea (Ashford-Rowe et al., 2013). Entre las contribuciones más influyentes se encuentra el modelo de cinco dimensiones de Gulikers et al. (2004)—tarea, contexto físico, contexto social, forma de evaluación y criterios/estándares—, que ofrece un marco sólido para diseñar y auditar evaluaciones alineadas con prácticas de referencia y con saberes transferibles.

En el ámbito de la Educación Física (EF) escolar, la EA adquiere un significado particular debido a la naturaleza motriz, situada y social del aprendizaje (López-Pastor & Pérez-Pueyo, 2017). Evaluar en contexto —ya sea durante el juego, en tareas cooperativas o en proyectos de diseño de actividades físicas— permite valorar de manera simultánea componentes motrices, cognitivos y socioemocionales (toma de decisiones, comunicación, autorregulación, juego limpio), que difícilmente emergen a través de pruebas convencionales (Hay & Penney, 2012).

La literatura internacional ha documentado tanto el potencial de estas prácticas como los riesgos de su simplificación excesiva cuando se aplican sin criterios claros, sin evidencias o desvinculadas del currículo (MacPhail & Halbert, 2010; Oliver-Álvarez & Martos-García, 2023). En la literatura anglosajona se ha destacado también la presencia de tensiones en la implementación de la EA en EF (Hu et al., 2025; Ravi & Besharat, 2025; Treschman et al., 2025). Autores como Casey & Goodyear (2015) han mostrado cómo los enfoques pedagógicos basados en el aprendizaje cooperativo y el uso de tecnologías digitales plantean oportunidades, pero también desafíos en términos de criterios y fiabilidad. En esta misma línea, Casey et al. (2017) han documentado experiencias pedagógicas que muestran el potencial y las limitaciones de la incorporación de tecnologías digitales para sostener prácticas auténticas en EF. Kirk (2009) subraya que la evaluación en EF no puede desligarse de los debates sobre la finalidad y el futuro de la asignatura, ya que constituye un elemento central en la definición de qué aprendizajes se valoran y con qué legitimidad.

Desde una perspectiva sociocultural, Hay & Penney (2010) argumentan que evaluar implica también asignar valor y regular oportunidades de aprendizaje. En consecuencia, el diseño y uso de la evaluación (qué se evalúa, cómo se puntúa, quién participa) influyen en la distribución de logros, en la configuración de identidades y en las trayectorias educativas del alumnado. Así, la autenticidad no debe concebirse como un recurso metodológico superficial, sino como una condición esencial para la justicia curricular y la alfabetización evaluativa del profesorado (Boud & Falchikov, 2007).

Este planteamiento converge con un campo consolidado: la evaluación formativa. Black & Wiliam (2009) demostraron que los procesos orientados a recoger evidencias del aprendizaje y transformarlas en retroalimentación accionable —tanto para estudiantes como para docentes— producen mejoras significativas y consistentes en diversos contextos educativos. Hattie & Timperley (2007) precisaron que el impacto de la retroalimentación depende de su alineación con la meta, de la información sobre la brecha y de la guía para cerrarla. Sadler (1989) enfatizó que la competencia evaluativa del estudiante exige comprender los criterios, contrastar su desempeño con los estándares y regularse en consecuencia.

Una EA bien diseñada tiende a ser formativa por construcción, ya que transforma el desempeño observado en información útil para la intervención docente y la autorregulación del alumnado (Castillo-Retamal et al., 2024). No obstante, trasladar estos principios a la EF escolar presenta desafíos significativos. Primero, la calidad de las tareas y los criterios: no toda tarea con apariencia de realismo es auténtica. Sin criterios explícitos y compartidos, la evaluación puede volverse opaca, difícil de replicar y susceptible al sesgo (Gulikers et al., 2006). El marco de Gulikers resulta fundamental para evitar este “realismo ingenuo”, diferenciando entre la simple apariencia de realidad y la correspondencia funcional con la

práctica profesional. Segundo, la alfabetización evaluativa docente: el diseño de rúbricas válidas, la observación de desempeños en acción y la negociación de estándares requieren competencias específicas que no siempre se desarrollan en la formación inicial ni en los procesos de inducción profesional (López-Pastor et al., 2020). Tercero, las condiciones de implementación: factores como el tamaño del grupo, el tiempo disponible, la presión por calificar y los regímenes de accountability tienden a favorecer soluciones rápidas —como las pruebas de selección— en detrimento de tareas complejas que demandan observación, diálogo y coevaluación (Hay, 2006). En este sentido, la literatura en EF ha documentado claramente la brecha entre el discurso favorable a la EA y su aplicación plena en contextos escolares reales (Castillo-Retamal et al., 2024).

En los últimos años, diversas síntesis y posicionamientos internacionales han reforzado la necesidad de profesionalizar la evaluación en EF, explicitando propósitos, estándares y criterios de calidad. El Position Statement on Assessment in Physical Education de la AIESEP (2020) sostiene que la evaluación debe generar evidencias creíbles de aprendizaje, alineadas con metas curriculares y respaldadas por políticas que aseguren la equidad y la validez de las decisiones evaluativas. Sin embargo, este marco normativo coexiste con la persistencia de prácticas heterogéneas y con una insuficiente inversión en alfabetización evaluativa docente (Atienza et al., 2018).

En paralelo, la investigación empírica ha documentado experiencias que muestran distintos grados de sistematicidad en la aplicación de estrategias auténticas. Estas incluyen el uso de rúbricas analíticas y criterios compartidos (Hall-López & Ochoa-Martínez, 2020); observación situada del juego para evaluar la toma de decisiones y los principios tácticos (Harvey & Jarrett, 2013); portafolios que articulan evidencias de progreso (Bailey & Pickard, 2010); coevaluación y autoevaluación guiadas por criterios (Lorenzo-Catalán & Kirk, 2015); y proyectos integradores de conocimiento disciplinar y didáctico (López-Pastor et al., 2020).

Una parte significativa de estas prácticas se enmarca en la tradición hispano-europea de evaluación formativa y compartida (EFyC), que, cuando se implementa rigurosamente, incrementa la participación del alumnado, mejora la transparencia y fortalece la coherencia entre enseñanza, aprendizaje y calificación (López-Pastor & Pérez-Pueyo, 2017). Estudios recientes confirman la viabilidad de estos dispositivos, aunque señalan la necesidad de reforzar su fiabilidad y documentar su impacto en aprendizajes de orden superior, más allá de la ejecución técnica aislada (Molina et al., 2023).

A pesar de estos avances, persisten vacíos que requieren atención. Primero, la falta de una síntesis sistemática que describa cómo se manifiesta la EA en la EF escolar: ¿qué tipos de tareas predominan?, ¿qué criterios y evidencias se utilizan?, ¿en qué medida se alinean estas prácticas con los principios de autenticidad (tarea, contexto, forma y criterios)? Segundo, la literatura evidencia un sesgo hacia la evaluación del dominio psicomotor, con mayores dificultades para valorar de manera válida y fiable las dimensiones afectivas y sociales (Hay & Penney, 2012), lo que debilita la promesa integral de la autenticidad. Tercero, la variabilidad contextual —determinada por políticas de evaluación, culturas escolares y recursos disponibles— condiciona la transferencia y sostenibilidad de las prácticas (MacPhail & Halbert, 2010). Estas limitaciones subrayan la necesidad de revisiones que trasciendan el inventario de instrumentos y que analicen las finalidades, estrategias e instrumentos empleados, así como los factores condicionantes de su implementación.

En este contexto, el objetivo de la presente revisión fue explorar cómo se materializa la EA en la EF escolar, con énfasis en los instrumentos y estrategias utilizadas por los docentes. Asimismo, busca establecer el grado de alineación de dichas prácticas con los principios fundamentales de la EA, con el propósito de orientar acciones de mejora tanto en la investigación como en la práctica docente, y contribuir a la discusión internacional sobre evaluación educativa en contextos curriculares diversos.

Esta revisión examina de manera explícita cómo se manifiesta la EA en la EF escolar, destacando los instrumentos, estrategias y condiciones contextuales que influyen en su implementación. Este aporte ofrece una perspectiva novedosa al conectar el discurso internacional sobre la EA con su materialización en los contextos escolares, aportando insumos tanto para la investigación como para la práctica docente a nivel global.

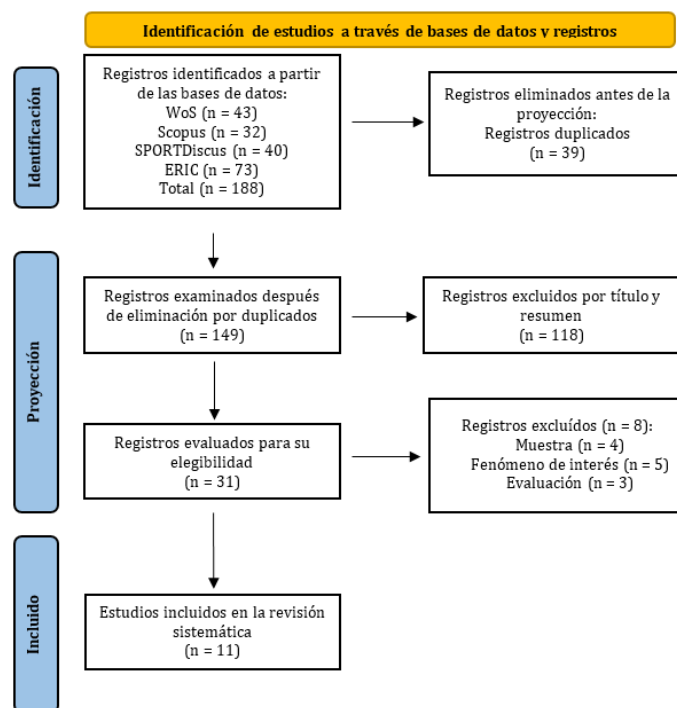
## Método

La siguiente revisión sistemática fue elaborada conforme a los lineamientos establecidos por la guía PRISMA 2020 (Page et al., 2021). Del mismo modo, se consideraron cada uno de los principios éticos de la Declaración Universal sobre Bioética y Derechos Humanos de la UNESCO, garantizando así el uso responsable de la evidencia científica publicada. Asimismo, se resguardaron principios de transparencia e integridad propios de las revisiones sistemáticas, asegurando la trazabilidad de las decisiones metodológicas y la coherencia en todo el proceso de análisis.

### Proceso de búsqueda y selección

La búsqueda bibliográfica se llevó a cabo en las bases de datos Web of Science, Scopus, SPORTDiscus y ERIC. El periodo de cobertura se fijó hasta el 31 de mayo de 2025, sin restricción inferior de fecha, a fin de capturar el desarrollo histórico y actual del campo. Se incluyeron estudios en inglés, español y portugués, dado que concentran la mayor parte de la producción indexada en el área y permiten cubrir tanto literatura anglosajona como hispano-lusófona relevante para el currículo de EF escolar. Para estructurar la búsqueda, se definieron dos núcleos: a) para evaluación auténtica se emplearon los términos: “authentic assessment” OR “performance-based assessment” OR “authentic tasks” OR “real-world assessment” OR “competency-based assessment” OR “authentic learning” OR “simulated assessment”; y b) para EF: “physical education” OR “physical education teacher education” OR “PE teacher education” OR “physical education pedagogy” OR “PE teaching”. La cadena final fue: (núcleo 1) AND (núcleo 2), con búsqueda en título, resumen y palabras clave.

Figura 1. Flujograma PRISMA de estudios identificados



### Criterios de inclusión y exclusión

Para esta revisión sistemática se utilizaron criterios de inclusión basados en el enfoque SPIDER, que es adecuado para sintetizar evidencia cualitativa, mixta o interpretativa en estudios educativos. Para Sample (S) se incluyeron estudios en el ámbito de la EF; en Phenomenon of Interest (PI), aquellos que abordaron la EA; en Design (D), investigaciones de carácter cuantitativo, cualitativo o mixto; en Evaluation (E), estudios que consideraron la EA como herramienta docente; y en Research Type (R), únicamente trabajos con evidencia empírica. Se excluyeron: (a) ensayos teóricos, editoriales y revisiones; (b) trabajos en educación superior, clubes o alto rendimiento; (c) investigaciones donde la “evaluación” no correspondía a EA o no reportaba instrumentos/estrategias evaluativas; y (d) artículos sin texto completo

o no disponibles en los idiomas definidos. La selección inicial fue llevada a cabo por dos revisores de manera independiente (H-G, M, y Y-S, R). Ante cualquier discrepancia en la inclusión de estudios, se recurrió a la evaluación de un tercer investigador (H-T, C) para resolver el desacuerdo mediante consenso.

### **Evaluación de la calidad metodológica**

La calidad metodológica de los estudios incluidos se evaluó mediante la herramienta QATSDD (Sirriyeh et al., 2012), que permite valorar de manera uniforme diseños cuantitativos, cualitativos y mixtos (lo que se ajusta a la heterogeneidad del corpus analizado) y puntúa 16 ítems en una escala de 0 a 3 (mínimo 0 y máximo 48) (tabla 1). No se calculó un índice estadístico de concordancia (Cohen's Kappa), ya que las discrepancias entre revisores fueron mínimas y se resolvieron íntegramente mediante discusión y consenso. Los puntajes totales se expresaron como porcentaje del máximo posible, facilitando la comparación entre estudios. Finalmente, los estudios fueron clasificados según el porcentaje alcanzado respecto del puntaje máximo: de  $\geq 75\%$  fueron considerados de alta calidad (6 estudios), entre 50% y 74% se catalogaron como de calidad media (5 estudios) y menos de 50% fueron clasificados como de baja calidad (0 estudios).

Tabla 1. Resultados de la calidad metodológica por estudio

Autor/año	Diseño	Puntaje total/máx.	% de calidad	Observaciones
Barrientos et al., 2023	Cualitativo	34 / 48	70,8 %	Objetivos y diseño claros, fundamentación teórica sólida, análisis adecuado; limitada información sobre validación de datos.
Fuentes-Nieto et al., 2022	Cuantitativo	36/48	75,0 %	Objetivos claros; contexto y muestra descritos; análisis descriptivo y Kruskal-Wallis sin diferencias por grupo/sexo. Limitaciones: Sin reporte de fiabilidad/validez de instrumentos, sin detalles de ética.
Gréhaigne et al., 1997	Cuantitativo	37/48	77,1 %	Metodología innovadora, índices objetivos, evidencia sólida de validez y fiabilidad. Limitaciones: Falta de caracterización detallada de la muestra, escasa referencia ética.
Gutiérrez et al., 2014	Observacional	35/48	72,9 %	Propósito claro, metodología coherente con TGfU, resultados bien presentados. Limitaciones: Muestra reducida y poco diversa, limitada evidencia de fiabilidad/validez en este estudio.
Hariadi et al., 2025	Mixto	34/48	70,8 %	Objetivos claros, integración de métodos y resultados bien estructurados. Limitaciones: Descripción ética superficial, poca evidencia de fiabilidad/validez en el contexto comparado.
Hariadi, 2017	I+D	35/48	72,9 %	Objetivos claros, metodología de desarrollo coherente, análisis sólido de validez y fiabilidad. Limitaciones: Escaso planteamiento ético, limitada representatividad de la muestra.
Mintah, 2003	Cuantitativo	36/48	75,0 %	Objetivos claros, relevancia temática, resultados bien comunicados. Limitaciones: Poca profundización analítica, sin ética reportada, ausencia de reflexividad.
Monfette & Grenier, 2015	Cuantitativo	38/48	79,2 %	Aplicación rigurosa, resultados claros y útiles. Limitaciones: escasa reflexividad de los autores, ética poco discutida, validación de codificación limitada.
Syafruddin et al., 2023	Cualitativo	33/48	68,8 %	Objetivos claros, metodología cualitativa sólida, análisis temático riguroso. Limitaciones: Muestra pequeña, débil reporte ético, escasa reflexividad.
Tu et al., 2025	Cuantitativo	37/48	77,1 %	Objetivo claro y relevante, reconoce debilidades, lo que aporta transparencia. Limitaciones: Muestra reducida y poco diversa. Débil descripción del resguardo ético.
Williams, 2018	Cualitativo	36/48	75,0 %	Diseño robusto, aporta la voz docente; debilidades en ética, reflexividad y diversidad muestral.

De los seis estudios que fueron clasificados de alta calidad (54,5% del corpus), donde solo uno de ellos alcanzó una puntuación elevada (38 de 48), lo que sugiere que, si bien este grupo evidencia un nivel aceptable de rigor metodológico, sus resultados se sitúan en el umbral inferior de la categoría "alta". Los cinco estudios restantes (45,45% del corpus) fueron evaluados como de calidad media. Este hallazgo obliga a ser cautelosos al momento de extrapolar o generalizar las conclusiones de la presente revisión, ya que la consistencia metodológica global de los estudios incluidos resulta limitada.

## **Resultados**

Los datos recopilados en los estudios seleccionados fueron organizados en una matriz de análisis diseñada para garantizar una lectura sistemática y comparativa de la evidencia. Esta matriz incluyó: la referencia de cada estudio, sus objetivos, el diseño metodológico y la muestra considerada, los instrumentos



de investigación utilizados, los instrumentos específicos de evaluación en el ámbito de la EF, los principales hallazgos y, finalmente, el nivel de alineación de dichos resultados con la EA. De esta manera, se buscó no solo sintetizar la información de forma ordenada, sino también ofrecer una visión integral que permitiera valorar la coherencia y pertinencia de los aportes en relación con el estudio (Tabla 2).

Tabla 2. Síntesis de los estudios revisados

Referencia	Objetivo	Metodología-Muestra	Instrumento de investigación	Instrumento de evaluación en EF	Principales hallazgos	Relación
Barrientos, 2022	Analizar los desafíos que experimentan profesores de EF al utilizar EA.	Estudio cualitativo. 4 docentes de EF: 3 primaria y 1 secundaria.	Entrevista semiestructurada	Pruebas físicas estandarizadas y observación de actitudes en clase.	La evaluación es percibida como centrada en lo físico; poca atención a lo actitudinal y procedimental.	Parcial
Fuentes-Nieto et al., 2022	Analizar resultados del uso de EA y ET con TIC.	Cuantitativo con diseño descriptivo. 38 estudiantes de secundaria.	Rúbrica descriptiva auto/coevaluar video. Escala Likert para evaluar experiencia.	Cuestionario específico de percepción de evaluación en EF.	Videotutorial mejora aprendizaje/reflexión. La aplicación de Pickers facilita y ahorra tiempo.	Alta
Gréhaigne et al., 1997	Proponer un procedimiento de EA en deportes de equipo en el contexto escolar.	Estudio cuantitativo. 302 estudiantes de secundaria.	Planilla de observación en juego que registra aspectos del juego.	Procedimiento juego-basado (tipo TSAP)	El procedimiento integra técnica y táctica en contexto real de juego, ofrece puntaje único comparable y muestra validez y fiabilidad.	Alta
Gutiérrez et al., 2014	Diseñar/implementar un enfoque "game context" para evaluar desempeño en juego.	Estudio observacional. 19 estudiantes (11 niñas, 8 niños) de EF secundaria.	Videograbación (2 cámaras) y análisis observacional.	Evaluación mediante juego de invasión 5x5 (balonmano).	Evaluación en contexto real de juego. Se observa al alumnado jugando (5v5), registrando decisiones y ejecuciones según problemas tácticos; esto mantiene la autenticidad.	Alta
Hariadi et al., 2025	Evaluar la viabilidad, obstáculos y adherencia en la implementación del IPOPJ en EF.	Estudio mixto. 70 escuelas primarias públicas y 55 escuelas nacionales.	Cuestionarios, hojas de observación, guías para entrevistas.	Instrumento para la EA en EF (IPOPJ).	La EA permitió valorar competencias de manera integral; la efectividad depende de la preparación docente y recursos disponibles.	Alta
Hariadi, 2017	Desarrollar un instrumento de EA para habilidades motoras y actitudinales en EF.	Estudio I+D. 102 estudiantes (prueba piloto) y 962 (prueba de campo).	IPOPJ (actitudes, habilidades y conocimiento).	Rúbricas de desempeño físico (habilidades motrices), listas de cotejo de actitudes y pruebas de observación directa.	Los instrumentos desarrollados permitieron evaluar tanto aspectos motores como actitudinales, integrando dimensiones de desempeño real en contextos auténticos.	Alta
Mintah, 2003	Examinar el grado de utilización de la EA en EF y deportes.	Estudio cuantitativo descriptivo. Profesores de EF y entrenadores de secundaria y universitarios.	Cuestionario estructurado con ítems de frecuencia y percepciones sobre efectos de la EA.	Portafolios de desempeño, diarios reflexivos, proyectos prácticos, rúbricas de observación, autoevaluación y coevaluación.	Se constató una presencia creciente de prácticas de EA, especialmente portafolios y autoevaluación. Los docentes percibieron efectos positivos en la motivación, autoconcepto y en habilidades.	Alta
Monfette & Grenier, 2015	Describir las prácticas evaluativas declaradas por docentes de EF en primaria.	Estudio cuantitativo. 327 docentes de EF (230 hombres y 97 mujeres)	Cuestionario de 27 ítems Likert (1-5) con las seis características de EA.	Sin antecedentes	Las prácticas cumplen en gran medida con características de EA, pero requieren ajustes para ser más auténticas.	Parcial
Syafruddin et al., 2023	Desarrollar y validar una rúbrica auténtica para evaluar el rendimiento en EF.	Estudio cualitativo. Muestra 10 docentes y estudiantes de secundaria.	Observación directa. Entrevistas. Análisis documental.	Cognitivo: pruebas escritas y orales. Psicomotor: observación con rúbrica. Afectivo: listas de cotejo actitudinales.	Rúbrica válida y confiable; promueve evaluación justa y realista.	Alta
Tu et al., 2025	Analizar las percepciones del profesorado de EF de secundaria respecto a la implementación de evaluación por competencias.	Estudio cuantitativo descriptivo. 50 docentes de EF (35 hombres y 15 mujeres).	Cuestionario Likert de evaluación por competencias en EF.	Evaluación basada en competencias en EF (CAHPE).	> 90% de los docentes entiende el concepto y propósito de la CAHPE; 56% la implementa regularmente y 44% rara vez o nunca. El 90% reporta dificultades para desarrollar criterios adecuados de evaluación.	Parcial
Williams & Penney, 2011	Explorar el potencial de tecnologías digitales para diseñar evaluaciones auténticas en cursos de secundaria	Estudio cualitativo descriptivo. 12 docentes de 6 escuelas; 148 estudiantes.	Entrevistas a profesores y estudiantes. Observaciones de clases.	Examen digital: a) dos secciones teóricas; b) dos ejecuciones prácticas grabadas; y c) una reflexión escrita final.	Las tareas integraron de forma efectiva dimensiones teóricas y prácticas, potenciando la autenticidad y relevancia de la evaluación. Se necesita replantear prácticas evaluativas.	Alta

Los once estudios incluidos presentan una diversidad metodológica y temática que exige una lectura integradora. En términos de diseño, el corpus combina investigaciones cualitativas, cuantitativas (observacionales, encuestas y desarrollo/validación de instrumentos) y un solo estudio de método mixto. Entre los cualitativos, se sitúan trabajos centrados en la comprensión de prácticas y desafíos docentes vinculados a la evaluación en EF, a partir de entrevistas semiestructuradas y observación (por ejemplo, Barrientos, 2022; Syafruddin et al., 2023; Williams & Penney, 2011). En el polo cuantitativo, emergen tres subgrupos: (i) estudios observacionales de rendimiento en deportes de colaboración/oposición, con construcción de índices e indicadores de validez/fiabilidad (Grehaigine et al., 1997; Gutiérrez et al., 2014); (ii) encuestas que cartografían prácticas y percepciones docentes/estudiantiles mediante cuestionarios estructurados (Mintah, 2003; Monfette & Grenier, 2015; Tu et al., 2025); y (iii) desarrollo/validación de instrumentos específicos para la evaluación en EF, con análisis psicométricos (Hariadi, 2017). A esta diversidad se suma un estudio aplicado con análisis no paramétrico de percepciones estudiantiles sobre experiencias evaluativas mediadas por TIC (Fuentes-Nieto et al., 2022), y un único estudio mixto que combina cuestionarios con observaciones/entrevistas para analizar la implementación comparada de un instrumento de EA en dos sistemas educativos (Hariadi et al., 2025).

A nivel de instrumentos y procedimientos, los estudios observacionales (Grehaigine et al., 1997; Gutiérrez et al., 2014) emplean sistemas de codificación del juego e indicadores de eficiencia/decisión que permiten estimar validez y fiabilidad en contextos reales de práctica. Las encuestas (Mintah, 2003; Monfette & Grenier, 2015; Tu et al., 2025) utilizan cuestionarios estructurados (escala Likert o ítems abiertos) para describir frecuencia de uso, condiciones de implementación y creencias sobre la evaluación; su fortaleza es la cobertura muestral, mientras que su limitación radica en el autoinforme y la menor capacidad explicativa sobre el “cómo” se implementan las prácticas. En la línea aplicada, Fuentes-Nieto et al. (2022) integra auto/coevaluación y Plickers para valorar la experiencia de evaluación transformativa en una unidad de danza, con análisis descriptivo e inferencias no paramétricas (Kruskal-Wallis). En el plano instrumental, Hariadi (2017) aporta un instrumento para habilidades motrices manipulativas con evidencia psicométrica básica y Hariadi et al. (2025) analiza la implementación de un instrumento de EA en primaria con triangulación de fuentes y contraste estadístico. Los cualitativos (Barrientos, 2022; Syafruddin et al., 2023; Williams & Penney, 2011) profundizan en la coherencia pedagógica de la evaluación, la organización de la enseñanza con la evaluación en mente y las condiciones de posibilidad (tiempo, recursos, acompañamiento), ofreciendo comprensión contextual que complementa la evidencia cuantitativa.

En cuanto a relaciones y convergencias, el conjunto coincide en: (a) la centralidad de la retroalimentación formativa y la participación activa del alumnado (auto/coevaluación, evidencias auténticas), destacadas tanto en estudios empíricos como en aproximaciones reflexivas; (b) la importancia de emplear instrumentos pertinentes al contexto de la tarea (observación sistemática en juego, rúbricas, listas de cotejo); y (c) la necesidad de apoyos institucionales y tiempos pedagógicos que permitan sostener prácticas de calidad. Como tensiones, emergen la fragmentación entre el discurso favorable y la implementación efectiva; la variabilidad en la calidad metodológica, especialmente en aspectos de validación psicométrica, consistencia analítica y reporte ético; y la heterogeneidad de contextos que limita la generalización. En conjunto, la evidencia sugiere que la EA en EF progresa cuando se alinea con la enseñanza (tareas significativas, criterios explícitos, evidencias observables), se instrumenta con procedimientos válidos y fiables, y cuenta con soportes organizacionales que la hagan sostenible. Esta convergencia refuerza la pertinencia de la EA como marco común para orientar la investigación y la práctica en el área.

## Discusión

La presente revisión sistemática tiene como objetivo explorar cómo se materializa la EA en la EF escolar, poniendo énfasis en los instrumentos y estrategias empleadas por los docentes, así como en los factores que condicionan su desarrollo. Este enfoque resulta pertinente dado que, a pesar del consenso en torno al valor de la EA como vía para valorar desempeños de manera situada y significativa, persisten desafíos que dificultan su implementación en contextos reales.

Los resultados de esta revisión confirman y amplían lo señalado por investigaciones anteriores en el campo de la EF. En particular, López-Pastor et al. (2012) y Molina et al. (2023) ya habían advertido que la EA se caracteriza por su potencial para integrar dimensiones motrices, cognitivas y socioemocionales,



pero que su implementación requiere de una alfabetización evaluativa sólida y de un compromiso institucional que no siempre está garantizado. En línea con estos antecedentes Hinojosa-Torres et al. (2025), resaltan cómo la falta de formación sistemática en competencias evaluativas y la ausencia de políticas y recursos dificultan la consolidación de la EA en la formación inicial docente de EF.

En línea con estos antecedentes, los estudios incluidos muestran que la retroalimentación formativa y la participación activa del alumnado mediante auto y coevaluación se han consolidado como pilares de la EA, lo cual coincide con el marco teórico de Black & Wiliam (2009) y con la perspectiva de Sadler (1989) sobre la necesidad de que los estudiantes comprendan criterios, contrasten su desempeño y regulen sus aprendizajes. Estas prácticas son sistematizadas por Cañadas (2022), quien destaca su papel clave para motivar al alumnado y favorecer un aprendizaje más auténtico en la EF escolar, lo que coincide con lo reportado por Treschman et al. (2025), quienes muestran que la activación del alumnado en procesos de retroalimentación y evaluación fortalece la calidad del aprendizaje en EF. Además, Hu et al. (2025) destacan que la integración de la tecnología amplifica el potencial de la retroalimentación auténtica, mientras que Ravi & Besharat (2025) muestran que la percepción estudiantil sobre la flexibilidad y creatividad en el diseño de la evaluación influye en la motivación y el compromiso. Sin embargo, a diferencia de lo que reportaban trabajos más generales en educación superior (Ashford-Rowe et al., 2013; Villarroel et al., 2017), esta revisión revela que en la EF escolar los instrumentos disponibles muestran escasa validación psicométrica y una limitada capacidad para valorar aprendizajes socioemocionales y ciudadanos, estando en línea con los hallazgos de Vlachopoulos & Makri (2024) quienes plantean que la EA encuentra dificultades en la construcción de rúbricas confiables y transferibles entre contextos.

Por otra parte, Fuentes-Nieto et al. (2022) subrayan el potencial de las tecnologías digitales para ofrecer una retroalimentación más inmediata y eficaz, mejorando la aceptación y el compromiso del alumnado con las tareas auténticas, aunque la evidencia sobre su integración aún es incipiente y heterogénea. Por su parte, George & González-Moreno (2020) destacan la importancia de diseñar evaluaciones auténticas que integren dimensiones teóricas y prácticas, poniendo énfasis en la relevancia pedagógica y en la participación activa de los estudiantes para potenciar la autenticidad y la significatividad del aprendizaje.

En lo referido a las finalidades de la EA, los estudios revisados reflejan una orientación predominante hacia el desarrollo de competencias motrices y técnico-tácticas en contextos situados (Gréhaigne et al., 1997; Gutiérrez et al., 2014). No obstante, la promesa integral de la autenticidad, que debería incluir dimensiones cognitivas y socioemocionales, aparece solo parcialmente materializada. La tendencia a privilegiar lo observable y medible en el juego genera un sesgo psicomotor que ya había sido advertido por Hay & Penney (2012) y que limita la posibilidad de captar aspectos como la cooperación, la autorregulación o la ciudadanía. No obstante, estudios como los de Mintah (2003) y Monfette & Grenier (2015) evidencian que, aun con limitaciones, la incorporación de portafolios, diarios reflexivos y proyectos amplía el foco hacia dimensiones motivacionales y actitudinales, mostrando una transición hacia concepciones más integrales de la EA. Del mismo modo, investigaciones recientes como la de Fuentes-Nieto et al. (2022) evidencian que la integración de tecnologías digitales puede ampliar la autenticidad de la evaluación, permitiendo generar retroalimentación más inmediata y ahorrar tiempo pedagógico. No obstante, la evidencia es aún incipiente y heterogénea, lo que evidencia la distancia entre el discurso de la EA como práctica transformadora y la realidad de su implementación en escuelas con recursos limitados.

En relación con los instrumentos y estrategias empleados, la revisión evidencia una diversidad de enfoques. Los procedimientos de observación situados en el juego (Gréhaigne et al., 1997; Gutiérrez et al., 2014) ofrecen indicadores fiables de toma de decisiones y ejecución táctica, alineados con el principio de autenticidad de evaluar en contexto. Los cuestionarios estructurados (Monfette & Grenier, 2015; Tu et al., 2025) proporcionan una panorámica amplia de prácticas y percepciones, aunque con la limitación de depender del autoinforme y de no capturar la complejidad real de las interacciones. La utilización de rúbricas emerge como una estrategia recurrente (Syafuruddin et al., 2023), con potencial para mejorar la transparencia y la equidad, pero la falta de criterios compartidos y de procesos de validación rigurosos reduce su impacto. Estudios como los de Hariadi (2017) y Hariadi et al. (2025) aportan instrumentos con evidencia psicométrica básica, aunque la revisión muestra que estas iniciativas son excepcionales más que la norma. En general, los instrumentos parecen diseñados de manera puntual y localizada, sin

un marco estandarizado que facilite su transferencia entre contextos escolares. En este sentido, Welsandt et al. (2024) destacan que la construcción de instrumentos auténticos requiere escenarios de evaluación que integren criterios claros y contextos realistas, apoyándose en recursos digitales que mejoran la validez y la transferibilidad. Del mismo modo, Grant et al. (2025) plantean que la autenticidad no radica solo en la fidelidad al entorno real, sino también en la capacidad del instrumento para promover juicios evaluativos complejos y socialmente situados, lo que aporta un marco sólido para repensar las rúbricas y herramientas digitales en EF.

Un hallazgo adicional es el crecimiento de experiencias que combinan autoevaluación, coevaluación y retroalimentación docente. Estas prácticas, en línea con la tradición hispano-europea de la Evaluación Formativa y Compartida (López-Pastor & Pérez-Pueyo, 2017), muestran efectos positivos en la participación y en la coherencia entre enseñanza y aprendizaje. Sin embargo, la literatura sigue reportando debilidades en términos de fiabilidad interobservador, consistencia analítica y documentación ética (Castillo-Retamal et al., 2024). Esta constatación coincide con lo advertido por Fawns et al. (2024), quienes plantean que la EA no puede entenderse como una “panacea” aplicable en cualquier contexto, sino que requiere una mirada crítica que reconozca sus límites, los riesgos de banalización y la necesidad de condiciones estructurales que aseguren su sostenibilidad.

Los factores contextuales emergen como determinantes en la viabilidad de la EA. La falta de tiempo, la presión por calificar de manera rápida, el tamaño de los grupos y la ausencia de apoyos institucionales se mencionan de forma recurrente en los estudios revisados (Barrientos et al., 2022; Williams & Penney, 2011). Estas condiciones refuerzan la tendencia a recurrir a pruebas tradicionales o simplificadas en detrimento de tareas auténticas que requieren observación detallada, diálogo y procesos de retroalimentación sostenidos. Desde una perspectiva sociocultural, Hay & Penney (2010) ya advertían que evaluar implica asignar valor y distribuir oportunidades de aprendizaje. En este sentido, la insuficiente inversión en alfabetización evaluativa docente (Atienza et al., 2018) y la ausencia de políticas que respalden la autenticidad como principio curricular limitan la posibilidad de que estas prácticas se consoliden en los sistemas educativos. A este panorama se suma la falta de preparación específica del profesorado en evaluación auténtica. Levy-Feldman (2025) subraya que muchos docentes no han recibido formación formal en enfoques alternativos de evaluación, lo que repercute directamente en su capacidad para diseñar, implementar y sostener prácticas auténticas. Esta carencia de alfabetización evaluativa constituye un obstáculo estructural que limita la adopción de la EA, incluso cuando existen recursos o intenciones pedagógicas favorables.

La diversidad de contextos analizados también permite establecer comparaciones internacionales. En Europa, la tradición de la Evaluación Formativa y Compartida ha permitido desarrollar dispositivos sistemáticos como portafolios, rúbricas y proyectos integradores (López-Pastor & Pérez-Pueyo, 2017), que muestran un grado mayor de institucionalización y respaldo normativo. En contraste, los estudios de Asia (Hariadi, 2017; Hariadi et al., 2025; Tran et al., 2025; Tu et al., 2025) reflejan procesos más incipientes, vinculados a reformas curriculares que promueven la evaluación por competencias, pero con dificultades para establecer criterios claros y consistentes. En Latinoamérica, la evidencia apunta a una adopción gradual de tecnologías digitales y de estrategias de coevaluación, aunque todavía con fuerte condicionamiento por los recursos disponibles y la cultura escolar (Castillo-Retamal et al., 2024; Fuentes-Nieto et al., 2022). Esta diversidad refuerza la idea de que la EA no debe entenderse como un modelo único, sino como un conjunto de principios adaptables a cada contexto. Al mismo tiempo, muestra la necesidad de articular aprendizajes internacionales que eviten la fragmentación y promuevan criterios comunes de validez y fiabilidad.

A partir de estos hallazgos, se proyecta la necesidad de avanzar hacia estudios más rigurosos, particularmente en la validación psicométrica de instrumentos que integren dimensiones cognitivas, socioemocionales y ciudadanas. Sin embargo, estas tensiones no solo reflejan limitaciones metodológicas, sino también una disyuntiva conceptual: la diferencia entre autenticidad y un realismo superficial en la evaluación. Simular contextos reales no basta si no se acompaña de un cambio epistemológico en la enseñanza, entendiendo el conocimiento como construcción situada (Grecic et al., 2024). En la formación docente, se vuelve prioritario desarrollar la alfabetización evaluativa mediante experiencias sistemáticas de diseño e implementación de evaluaciones auténticas. Finalmente, en el ámbito de las políticas

educativas, se requiere alinear marcos curriculares y normativos con los principios de autenticidad, asegurando tiempos pedagógicos y apoyos institucionales que permitan superar el realismo superficial y consolidar un cambio epistemológico en la enseñanza.

Los hallazgos de esta revisión sugieren que la consolidación de la EA en la EF escolar requiere acciones en tres planos. En el ámbito docente, se debe fortalecer la alfabetización evaluativa mediante la formación en diseño de rúbricas, observación situada y uso de tecnologías digitales que favorezcan la retroalimentación. En la formación inicial, resulta clave integrar experiencias sistemáticas de aplicación de EA en la práctica supervisada, de modo que los futuros profesores desarrollen competencias para implementarla en contextos reales. Finalmente, en el nivel de políticas educativas, es necesario promover marcos normativos y apoyos institucionales que aseguren tiempo pedagógico y recursos, evitando que las presiones por la calificación reduzcan la autenticidad a un mero formalismo. Estas implicaciones ofrecen orientaciones concretas para avanzar hacia una cultura evaluativa más coherente con las demandas contemporáneas de la EF escolar.

## Conclusiones

Los hallazgos de esta revisión sistemática permiten concluir que la EA en la EF escolar constituye un enfoque con enorme potencial para transformar las prácticas evaluativas, al desplazar la atención desde la mera reproducción de contenidos hacia la valoración situada de competencias en contextos que emulan la práctica real. La evidencia muestra que este tipo de evaluación favorece aprendizajes significativos al integrar dimensiones motrices, cognitivas y socioemocionales, reforzando la retroalimentación formativa y la participación activa del alumnado como principios centrales. Sin embargo, también se constata que esta promesa se materializa de manera desigual y fragmentada, con fuertes tensiones entre el discurso pedagógico favorable y las condiciones efectivas de su implementación en las escuelas.

En términos de instrumentos, la revisión confirma que las rúbricas, la observación situada en el juego, los portafolios y los proyectos integradores representan los dispositivos más utilizados para dar forma a la EA. No obstante, persisten limitaciones vinculadas a la fiabilidad y la validez de los instrumentos empleados, así como una cobertura insuficiente de las dimensiones socioemocionales y ciudadanas del aprendizaje. Esto revela que, aunque existe un avance en la incorporación de estrategias contextualizadas y participativas, aún se requiere un esfuerzo sostenido para garantizar evaluaciones más coherentes con los principios de autenticidad.

La revisión también permite concluir que los factores contextuales ejercen un peso decisivo en la viabilidad de la EA. La presión institucional por calificar de manera rápida, la falta de tiempo y de recursos, los grupos numerosos y la ausencia de apoyos sistemáticos emergen como obstáculos recurrentes que condicionan la sostenibilidad de estas prácticas. De este modo, la efectividad de la EA no depende únicamente de la voluntad del profesorado, sino de la existencia de marcos curriculares, normativos y organizacionales que respalden su implementación estructural.

En conjunto, esta revisión aporta una síntesis crítica que no solo describe el estado actual del campo, sino que también señala los desafíos y oportunidades para su desarrollo futuro. La necesidad de validar instrumentos, ampliar el foco hacia las dimensiones socioemocionales, fortalecer la alfabetización evaluativa docente e integrar apoyos institucionales constituye un horizonte imprescindible para consolidar la EA en la EF escolar. Se trata, en definitiva, de avanzar hacia prácticas evaluativas más justas, situadas y coherentes con los desafíos educativos contemporáneos, contribuyendo al desarrollo integral de los estudiantes y a la consolidación de una cultura de evaluación que valore el aprendizaje en toda su complejidad.

## Limitaciones

Las principales limitaciones de esta revisión se relacionan con el bajo número de estudios considerados, la diversidad de enfoques metodológicos y la calidad desigual de la evidencia disponible. A ello se suma la falta de un metaanálisis cuantitativo, lo que obliga a interpretar los resultados con cautela y restringe la posibilidad de generalizar las conclusiones.

## Futuras líneas de investigación

A futuro, resulta clave avanzar en la validación psicométrica de los instrumentos de evaluación en EF, integrando no solo el componente motriz, sino también aspectos socioemocionales y ciudadanos. También es necesario desarrollar estudios longitudinales que permitan observar los efectos a largo plazo. En este escenario, el uso de tecnologías digitales e inteligencia artificial ofrece nuevas oportunidades, mientras que las comparaciones interculturales pueden aportar una comprensión más amplia sobre las condiciones que favorecen o dificultan su implementación.

## Agradecimientos

Agradecemos al Observatorio Chileno de la Educación Física y Deporte Escolar (OCHEFID) de la Universidad Andrés Bello por su apoyo y contribución al trabajo desarrollado en esta investigación.

## Financiación

No se recibió financiación para la realización de esta investigación.

## Referencias

- AIESEP (2020). *Position Statement on Assessment in Physical Education*. <https://aiesep.org/scientific-meetings/position-statements/>
- Ashford-Rowe, K., Herrington, J., & Brown, C. (2013). 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>
- Atienza, R., Valencia, A., & Devís, J. (2018). Experiencias de evaluación en educación física. Una aproximación desde la formación inicial del profesorado. *Estudios Pedagógicos (Valdivia)*, 44(2), 127–147. <https://dx.doi.org/10.4067/S0718-07052018000200127>
- Bailey, R., & Pickard, A. (2010). Body learning: examining the processes of skill learning in dance. *Sport, Education and Society*, 15(3), 367–382. <https://doi.org/10.1080/13573322.2010.493317>
- Barrientos, E., López-Pastor, V. M., Lorente-Catalán, E., & Kirk, D. (2022). Challenges with using formative and authentic assessment in physical education teaching from experienced teachers' perspectives. *Curriculum Studies in Health and Physical Education*, 14(2), 109–126. <https://doi.org/10.1080/25742981.2022.2060118>
- Black, P., & Wiliam, D. (2009). Developing the theory of formative assessment. *Educational Assessment Evaluation and Accountability*, 21(1), 5–31. <https://doi.org/10.1007/s11092-008-9068-5>
- Boud, D., & Falchikov, N. (2007). *Rethinking assessment in higher education: Learning for the longer term*. Routledge.
- Cañadas, L. (2022). Procesos de auto-evaluación y co-evaluación en educación física. Una revisión sistemática. *Revista Iberoamericana de Evaluación Educativa*, 15(1). <https://doi.org/10.15366/riee2022.15.1.009>
- Casey, A., & Goodyear, V. (2015). Can cooperative learning achieve the four learning outcomes of physical education? A review of literature. *Quest*, 67(1), 56–72. <https://doi.org/10.1080/00336297.2014.984733>
- Casey, A., Goodyear, V., & Armour, K. (2017). *Digital technologies and learning in physical education: Pedagogical cases*. Routledge.
- Castillo-Retamal, F., Sánchez-Quiroz, A., Montepil-Núñez, R., Quintana-Moya, L., Rojas-Román, M., & Villanueva-Carrasco, E. (2024). Evaluación en educación física: percepción del profesorado desde el territorio escolar (Assessment in physical education: perception of teachers from the school territory). *Retos*, 52, 610–622. <https://doi.org/10.47197/retos.v52.101527>
- Darling-Hammond, L., & Snyder, J. (2000). Authentic assessment of teaching in context. *Teaching and Teacher Education*, 16(5–6), 523–545. [https://doi.org/10.1016/S0742-051X\(00\)00015-9](https://doi.org/10.1016/S0742-051X(00)00015-9)
- Fawns, T., Bearman, M., Dawson, P., Nieminen, J., Ashford-Rowe, K., Willey, K., ... Press, N. (2024). Authentic assessment: from panacea to criticality. *Assessment & Evaluation in Higher Education*, 50(3), 396–408. <https://doi.org/10.1080/02602938.2024.2404634>



- Fuentes-Nieto, T., López-Pastor, V. M., & Palacios-Picos, A. (2022). Combinando una evaluación auténtica y transformativa a través de las TIC en educación física (A combination of transformative and authentic assessment through ICT in Physical Education). *Retos*, 44, 728-738. <https://doi.org/10.47197/retos.v44i0.91459>
- George, E., & González-Moreno, P. (2020). Evaluación auténtica como alternativa para la mejora en el proceso de enseñanza-aprendizaje de educación física en escuelas primarias. *Revista Ciencias de la Actividad Física UCM*, 21(2), 1-19. <https://doi.org/10.29035/rcaf.21.2.3>
- Grant, K., Fedoruk, L., & Nowell, L. (2021). Conversations and reflections on authentic assessment. *Imagining SoTL*, 1, 146-162. <https://doi.org/10.29173/isotl532>
- Grecic, D., Sprake, A., Thomson, A., Christodoulides, E., & Palmer, C. (2024). The epistemic judgement framework: a reflexive tool for physical education teachers' professional development to support quality physical education. *Frontiers in Education*, 9(1480690). <https://doi.org/10.3389/educ.2024.1480690>
- Gréhaigne, J., Godbout, P., & Bouthier, D. (1997). Performance assessment in team sports. *Journal of Teaching in Physical Education*, 16(4), 500-516. <https://doi.org/10.1123/jtpe.16.4.500>
- Gulikers, J., Bastiaens, T., & Kirschner, P. (2004). A five-dimensional framework for authentic assessment. *Educational Technology Research and Development*, 52, 67-86. <https://doi.org/10.1007/BF02504676>
- Gulikers, J., Bastiaens, T., & Kirschner, P. (2006). Authentic assessment, student and teacher perceptions: the practical value of the five-dimensional framework. *Journal of Vocational Education & Training*, 58(3), 337-357. <https://doi.org/10.1080/13636820600955443>
- Gutiérrez, D., Fisette, J., García-López, L. M., & Contreras, O. (2014). Assessment of secondary school students' game performance related to tactical contexts. *Journal of Human Kinetics*, 42, 223-234. <https://doi.org/10.2478/hukin-2014-0076>
- Hall-López, J., & Ochoa-Martínez, P. (2020). Enseñanza virtual en educación física en primaria en México y la pandemia por COVID-19. *Revista Ciencias de la Actividad Física UCM*, 21(2), 1-7. <https://doi.org/10.29035/rcaf.21.2.4>
- Hariadi, H., Valianto, B., Akhmad, I., & Syed Ali, S. K. (2025). Implementation study of authentic assessment in physical education in Indonesia and Malaysia. *Jurnal Cakrawala Pendidikan*, 44(1), 102-115. <https://doi.org/10.21831/cp.v44i1.71485>
- Hariadi. (2017). Instrument development of authentic assessment for manipulative fundamental motor skill at elementary school. *IOP Conference Series: Materials Science and Engineering*, 180(1), 12164. <https://doi.org/10.1088/1757-899X/180/1/012164>
- Harvey, S., & Jarrett, K. (2013). A review of the game-centred approaches to teaching and coaching literature since 2006. *Physical Education and Sport Pedagogy*, 19(3), 278-300. <https://doi.org/10.1080/17408989.2012.754005>
- Hattie, J., & Timperley, H. (2007). The power of feedback. *Review of Educational Research*, 77(1), 81-112. <https://doi.org/10.3102/003465430298487>
- Hay, P., & Penney, D. (2010). Proposing conditions for assessment efficacy in physical education. *European Physical Education Review*, 15(3), 389-405. <https://doi.org/10.1177/1356336X09364294>
- Hay, P., & Penney, D. (2012). *Assessment in physical education: A sociocultural perspective*. Routledge. <https://doi.org/10.4324/9780203133163>
- Hay, P. (2006). Assessment for learning in physical education. In D. Kirk, D. Macdonald, M. O'Sullivan (Eds.) *Assessment for learning in physical education* (pp. 312-325). SAGE Publications. <https://doi.org/10.4135/9781848608009.n18>
- Hinojosa-Torres, C., Blázquez-Sánchez, D., Barahona-Fuentes, G., Espoz-Lazo, S., Zavala-Crichton, J. P., Hurtado-Guerrero, M., Cerda-Miranda, L., Valdivia-Moral, P., & Yáñez-Sepúlveda, R. (2025). Evaluación de competencias en la formación inicial docente en educación física: revisión sistemática. *Revista Digital de Investigación en Docencia Universitaria*, 19(1), e2047. <https://doi.org/10.19083/ridu.2025.2047>
- Hu, A., Liu, Q., & Daniel, B. (2025). El papel de la tecnología digital en la evaluación auténtica: perspectivas de los educadores universitarios. *Assessment & Evaluation in Higher Education*, 1-14. <https://doi.org/10.1080/02602938.2025.2524095>
- Kirk, D. (2009). *Physical education futures* (1st ed.). Routledge. <https://doi.org/10.4324/9780203874622>



- Levy-Feldman, I. (2025). The role of assessment in improving education and promoting educational equity. *Education Sciences*, 15(2), 224. <https://doi.org/10.3390/educsci15020224>
- López-Pastor, V. M., Kirk, D., Lorente-Catalán, E., MacPhail, A., & MacDonald, D. (2012). Alternative assessment in physical education: a review of international literature. *Sport, Education and Society*, 18(1), 57–76. <https://doi.org/10.1080/13573322.2012.713860>
- López-Pastor, V. M., & Pérez-Pueyo, Á. (coords) (2017). *Evaluación formativa y compartida en educación: Experiencias de éxito en todas las etapas educativas*. Universidad de León
- López-Pastor, V. M., Molina, M., Pascual, C., & Manrique, J. C. (2020). La importancia de utilizar la evaluación formativa y compartida en la formación inicial del profesorado de educación física: los proyectos de aprendizaje tutorado como ejemplo de buena práctica. *Retos*, 37, 620–627. <https://doi.org/10.47197/retos.v37i37.74193>
- Lorente-Catalán, E., & Kirk, D. (2015). Student teachers' understanding and application of assessment for learning during a physical education teacher education course. *European Physical Education Review*, 22(1), 65-81. <https://doi.org/10.1177/1356336X15590352>
- MacPhail, A., & Halbert, J. (2010). 'We had to do intelligent thinking during recent PE': students' and teachers' experiences of assessment for learning in post-primary physical education. *Assessment in Education: Principles, Policy & Practice*, 17(1), 23–39. <https://doi.org/10.1080/09695940903565412>
- Mintah, J. (2003). Authentic assessment in physical education: prevalence of use and perceived impact on students' self-concept, motivation, and skill achievement. *Measurement in Physical Education and Exercise Science*, 7(3), 161–174. [https://doi.org/10.1207/S15327841MPEE0703\\_03](https://doi.org/10.1207/S15327841MPEE0703_03)
- Molina, M., López-Pastor, V. M., Hortigüela-Alcalá, D., Pascual-Arias, C., & Fernández-Garcimartín, C. (2023). Formative and shared assessment and feedback: an example of good practice in physical education in pre-service teacher education. *Cultura, Ciencia y Deporte*, 18(55), 157-169. <https://doi.org/10.12800/ccd.v18i55.1986>
- Monfette, O., & Grenier, J. (2015). Portrait des pratiques évaluatives déclarées par des enseignants d'éducation physique et à la santé au primaire. *Canadian Journal of Education Revue Canadienne De l'éducation*, 38(2), 1–28. <https://cje.journals.publicknowledgeproject.org/cje/index.php/cje-rce/article/view/1820>
- Oliver-Álvarez, M., & Martos-García, D. (2023). Traditional assessment and its consequences. A case in the hegemonic physical education. *Cultura, Ciencia y Deporte*, 18(58), 27-49. <https://doi.org/10.12800/ccd.v18i58.1976>
- Page, M., McKenzie, J., Bossuyt, P., Boutron, I., Hoffmann, T., Mulrow, C., Shamseer, L., Tetzlaff, J., Akl, E., Brennan, S., Chou, R., Glanville, J., Grimshaw, J., Hróbjartsson, A., Lalu, M., Li, T., Loder, E., Mayo-Wilson, E., McDonald, S., McGuinness, L., ... Moher, D. (2021). The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. *BMJ (Clinical research ed.)*, 372, n71. <https://doi.org/10.1136/bmj.n71>
- Ravi, M., & Besharat, M. (2025). A holistic consideration of authentic assessments: student perception of assessment design, delivery, flexibility and creativity. *European Journal of Engineering Education*, 1–18. <https://doi.org/10.1080/03043797.2025.2480116>
- Sadler, D. (1989). Formative assessment and the design of instructional systems. *Instructional Science*, 18(2), 119–144. <https://doi.org/10.1007/BF00117714>
- Sirriyeh, R., Lawton, R., Gardner, P., & Armitage, G. (2012). Reviewing studies with divers designs: the development and evaluation of a new tool. *Journal of Evaluation in Clinical Practice*, 18(4), 746–752. <https://doi.org/10.1111/j.1365-2753.2011.01662.x>
- Syafuruddin, A., Sufitriyono, S., & Sutriawan, A. (2023). Authentic assessment in physical education. *JOURNAL RESPECS (Research Physical Education and Sports)*, 5(1), 103–108. <https://doi.org/10.31949/respecs.v5i1.4149>
- Tran, T., Tuan, M., & Quyen, D. (2025). Competency-based assessment of high school students in physical education in Vietnam: Perspectives and practices. *International Journal of Learning Teaching and Educational Research*, 24(4), 501–524. <https://doi.org/10.26803/ijlter.24.4.23>
- Treschman, P., Stylianou, M., Brooks, C., & Sperka, L. (2025). Activating students in feedback and assessment processes in physical education. *Sport, Education and Society*, 1–14. <https://doi.org/10.1080/13573322.2025.2510427>

- Villarroel, V., Bloxham, S., Bruna, D., Bruna, C., & Herrera-Seda, C. (2017). Authentic assessment: creating a blueprint for course design. *Assessment & Evaluation in Higher Education*, 43(5), 840–854. <https://doi.org/10.1080/02602938.2017.1412396>
- Vlachopoulos, D., & Makri, A. (2024). A systematic literature review on authentic assessment in higher education: Best practices for the development of 21st century skills, and policy considerations. *Studies in Educational Evaluation*, 83(101425), 101425. <https://doi.org/10.1016/j.stueduc.2024.101425>
- Welsandt, N., Fortunati, F., Winther, E., & Abs, H. (2024). Constructing and validating authentic assessments: the case of a new technology-based assessment of economic literacy. *Empirical Research in Vocational Education and Training*, 16(1). <https://doi.org/10.1186/s40461-024-00158-0>
- Wiggins, G. (1990). The case for authentic assessment. *Practical Assessment, Research, and Evaluation*, 2(2), 1–3. <https://doi.org/10.7275/ffb1-mm19>
- Wiggins, G. (1998). *Educative assessment: Designing assessments to inform and improve student performance*. Jossey-Bass.
- Williams, J., & Penney, D. (2011). Authentic assessment in performance based subjects. *Teachers and Curriculum*, 12(1). <https://doi.org/10.15663/tandc.v12i1.28>

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

Macarena Hurtado-Guerrero	<a href="mailto:hurtado.macarena@gmail.com">hurtado.macarena@gmail.com</a>	Autora
Sebastián Espoz-Lazo	<a href="mailto:sebastian.espoz@uc.cl">sebastian.espoz@uc.cl</a>	Autor
Rodrigo Yáñez-Sepúlveda	<a href="mailto:rodrigo.yanez.s@unab.cl">rodrigo.yanez.s@unab.cl</a>	Autor
Mario Soto-González	<a href="mailto:mariosotog82@gmail.com">mariosotog82@gmail.com</a>	Autor
Andrea Castro-Dávila	<a href="mailto:ancasdavila@outlook.es">ancasdavila@outlook.es</a>	Autora
Dilan Galeano-Rojas	<a href="mailto:dagaleanor@correo.ugr.es">dagaleanor@correo.ugr.es</a>	Autor
Pedro Valdivia-Moral	<a href="mailto:pvaldivia@ugr.es">pvaldivia@ugr.es</a>	Autor
Claudio Hinojosa-Torres	<a href="mailto:claudio.hinojosa@unab.cl">claudio.hinojosa@unab.cl</a>	Autor/ Traductor