



Evaluación formativa y compartida en Educación Física: efectos de una intervención pedagógica sobre las percepciones del estudiantado

Formative and shared assessment in Physical Education: effects of a pedagogical intervention on students' perceptions

Autores

Walter Fernando Chavarría Gatica¹
Michelle Acuña Guerra²
Ariane Alvarez Alvarez³
Juan Francisco Cabrera Ramos⁴

^{1, 2, 4} Universidad Católica de Temuco (Chile)

³ Universidad Mayor (Chile)

Autor de correspondencia:
Ariane Alvarez Alvarez
ariane.alvarez@umayor.cl

Recibido: 10-04-26
Aceptado: 30-05-26

Cómo citar en APA

Chavarría Gatica, W. F., Acuña-Guerra, M., Alvarez Alvarez, A., & Cabrera Ramos, J. F. (2026). Evaluación formativa y compartida en Educación Física: efectos de una intervención pedagógica sobre las percepciones del estudiantado. *Retos*, 82, 1055-1067.
<https://doi.org/10.47197/retos.v82.119088>

Resumen

Introducción: La persistencia de modelos tradicionales de calificación en Educación Física limita el aprendizaje competencial y la agencia del estudiantado.

Objetivo: Examinar los efectos de una intervención pedagógica basada en la Evaluación Formativa y Compartida (EFC) sobre las percepciones de estudiantes de enseñanza media en un contexto escolar vulnerable en Chile.

Metodología: Participaron 182 estudiantes en 15 sesiones, favoreciendo la transición progresiva desde la heteroevaluación a la autoevaluación y coevaluación democrática. Se utilizó un cuestionario validado con cuatro dimensiones: Concepciones, Actores, Momentos-Instrumentos y Contenidos. Se aplicó un diseño cuasiexperimental de grupo único con medición pre y post intervención. Se comprobó la normalidad mediante la prueba de Shapiro-Wilk y se analizaron las diferencias con la prueba t de Student para muestras relacionadas y se calculó el tamaño del efecto mediante d de Cohen.

Resultados: Se evidenciaron incrementos estadísticamente significativos ($p < 0.001$) en todas las dimensiones, especialmente en las dimensiones Concepciones ($d = 1.06$), y en Actores ($d = 0.83$), lo que sugiere una reconfiguración de las creencias del estudiantado respecto a la evaluación.

Discusión: La resistencia inicial a la coevaluación vinculada a dinámicas de poder, desconfianza o temor a la exposición, que advierte la literatura, pueden ser mitigadas cuando el profesorado crea un clima de aula seguro.

Conclusiones: los hallazgos sugieren un desplazamiento desde una cultura evaluativa centrada en la calificación hacia una orientada al aprendizaje, la reflexión y la participación, lo cual aporta evidencia a favor de la EFC como estrategia para promover la justicia evaluativa.

Palabras clave

Educación Física; evaluación del aprendizaje; evaluación formativa y compartida; justicia evaluativa; percepciones del estudiantado.

Abstract

Introduction: The persistence of traditional grading models in Physical Education limits competence-based learning and student agency. **Objective:** To examine the effects of a pedagogical intervention based on Formative and Shared Assessment (FSA) on the perceptions of high school students in a vulnerable school context in Chile. **Methodology:** 182 students participated in 15 sessions, promoting the gradual transition from teacher-controlled assessment to self-assessment and democratic co-assessment. A validated questionnaire with four dimensions was used: Conceptions, Actors, Moments-Instruments, and Contents. A single-group quasi-experimental design with pre- and post-intervention measurement was applied. Normality was checked using the Shapiro-Wilk test, and differences were analyzed with the paired samples Student's t-test, and the effect size was calculated using Cohen's d. **Results:** Statistically significant increases ($p < 0.001$) were observed in all dimensions, especially in the Conceptions dimension ($d = 1.06$), which suggests a reconfiguration of students' beliefs regarding assessment, and in the Actors dimension ($d = 0.83$).

Discussion: Initial resistance to peer assessment linked to power dynamics, mistrust, or fear of exposure, as noted in the literature, can be mitigated when teachers create a safe classroom climate.

Conclusions: The findings suggest a shift from an assessment culture focused on grading toward one oriented to learning, reflection, and participation, thereby supporting FSA as a strategy for promoting assessment justice.

Keywords

Physical Education; learning assessment; formative and shared assessment; assessment justice; student perceptions.

Introducción

La evaluación es un elemento curricular estratégico que condiciona la forma en que los estudiantes se posicionan frente al saber y al aprendizaje. Este fenómeno la convierte en uno de los dispositivos de poder más influyentes en el entorno escolar por tratarse de una práctica social situada, con implicaciones simbólicas y éticas, que tiende a reproducir desigualdades educativas preexistentes y que, bajo determinadas condiciones pedagógicas, puede también contribuir a cuestionarlas. En consecuencia, evaluar implica siempre una toma de posición respecto de qué aprendizajes se consideran valiosos y quiénes son legitimados como aprendices competentes.

La promulgación del Decreto N° 67 (Ministerio de Educación de Chile [Mineduc], 2018), en el contexto chileno, ha impulsado un giro hacia la evaluación como una herramienta de aprendizaje, destinada a monitorear trayectorias y retroalimentar procesos más que a certificar resultados finales de manera punitiva. Sin embargo, la brecha entre las políticas y las prácticas docentes sigue siendo un desafío. Como advierte Tolgfors (2018), a menudo conviven distintas visiones de la evaluación para el aprendizaje, pero predomina una visión instrumental donde el docente controla el juicio evaluativo en detrimento de la agencia del alumnado. Cuando la evaluación se rige predominantemente por una racionalidad centrada en la estandarización y en la nota, como forma de control, tienden a invisibilizarse las trayectorias y metas individuales de aprendizaje y se restringe la participación democrática del estudiantado, especialmente en contextos de vulnerabilidad social (Pérez-Norambuena et al., 2025).

En la asignatura Educación Física y Salud, donde tradicionalmente ha predominado un enfoque centrado en la medición del rendimiento físico y la calificación sumativa (Fraile et al., 2018; López-Pastor, 2017; Velásquez Díaz, 2025), la cultura escolar centrada en la nota suele resistir los intentos de innovación y se tiende a privilegiar el uso de instrumentos tradicionales, que pueden generar experiencias de incompetencia, exclusión y vergüenza, especialmente en estudiantes con menor desarrollo de sus capacidades motrices (Gallardo-Fuentes et al., 2025). Superar esta situación requiere repensar la evaluación, poniendo el foco en las decisiones pedagógicas respecto a qué y cómo evaluar.

La transformación de las prácticas evaluativas requiere que el profesorado asuma un rol activo en la conducción pedagógica del aprendizaje. Desde enfoques de liderazgo transformacional, este rol implica crear condiciones que favorezcan la participación, la reflexión y la agencia del estudiantado, tensionando las lógicas tradicionales de enseñanza y evaluación centradas en el control y la calificación (Sliwka et al., 2024). En coherencia con Sanmartí (2020), el docente asume una responsabilidad ética de promover una cultura de la regulación, la colaboración y la corresponsabilidad por sobre la cultura de la nota. Este liderazgo pedagógico se expresa en la democratización del poder evaluativo y el tránsito hacia el diseño de escenarios de evaluación formativa y compartida (López-Pastor, 2017).

El rigor metodológico en la implementación de la evaluación del aprendizaje es una condición indispensable para asegurar que los procesos evaluativos en el contexto escolar sean justos y transparentes; sin embargo, no basta para transformar la cultura evaluativa (Ravela et al., 2017; Santos, 2019; Santos, 2024). La Evaluación Formativa y Compartida (EFC) se presenta como un cambio de paradigma particularmente relevante en la concepción de la evaluación escolar al entenderla como un proceso participativo orientado al aprendizaje y no a la certificación o medición de lo aprendido (Molina et al., 2023). La EFC implica una redistribución del poder evaluativo históricamente concentrado en el docente, habilitando formas más democráticas, dialogadas y transparentes de regulación del aprendizaje.

En este escenario, se considera la justicia evaluativa como un criterio de validez en el cual las prácticas de evaluación no pueden separarse de sus consecuencias sociales y educativas (Aquario, 2021; McArthur, 2023), pues una evaluación solo es plenamente válida cuando también es justa (Randall et al., 2024). Este planteamiento supera la imparcialidad y la inclusión procedimental, al exigir participación estudiantil y redistribución del poder en las decisiones evaluativas (Kakara et al., 2025; Nieminen, 2022). La noción de justicia evaluativa integra dimensiones procedimentales, distributivas e interactivas, visibles en quién define criterios, recoge evidencias e interpreta el aprendizaje (Banasadi et al., 2023), y se expresa relacionalmente cuando el estudiantado tiene voz, reconocimiento e influencia intencionada en los juicios y decisiones sobre su aprendizaje, lo que es conceptualizado por Rasooli y DeLuca (2024) como agencia evaluativa.

La EFC es especialmente relevante en contextos de vulnerabilidad social porque actúa como un dispositivo de justicia que permite validar procesos evaluativos que favorezcan diversas formas de ser competente (Nieminen, 2022; Oliver-Álvarez y Martos-García, 2023). Promueve prácticas evaluativas sensibles a las trayectorias individuales, fundamentadas en el reconocimiento y la equidad (Barrientos et al., 2023; Leão Pereira et al., 2025), y propicia la inclusión mediante la participación democrática, al transformar el acto evaluativo en un espacio de diálogo donde el estudiante ejerce su agencia, asumiendo un rol activo en su proceso de evaluación y, consecuentemente, en su proceso de aprendizaje (Hortigüela-Alcalá et al., 2019).

En Educación Física, la evidencia acumulada sobre la Evaluación Formativa y Compartida es consistente. Herrero-González et al. (2023) identifican que estas prácticas se asocian con mejoras consistentes en la implicación del alumnado, la comprensión de los criterios de evaluación y la percepción de equidad, destacando su potencial para incidir en la cultura evaluativa más allá de la mejora del rendimiento. Moura et al. (2024) y Treschman et al. (2025) coinciden en que la participación sustantiva del alumnado en procesos de autoevaluación, coevaluación y retroalimentación favorece la autorregulación, la autonomía y el aprendizaje profundo, desplazando la evaluación desde una lógica de control hacia una lógica de aprendizaje. No obstante, el desarrollo de la agencia evaluativa enfrenta resistencias culturales tanto del profesorado como del estudiantado, habituados a modelos de heteroevaluación que eximen, a estos últimos, de la implicación en la toma de decisiones y de la responsabilidad que ello implica (Ruz et al., 2024). La agencia estudiantil no emerge de forma espontánea, sino que requiere un andamiaje pedagógico que permita al estudiante apropiarse, progresivamente, de su proceso evaluativo (Slingerland et al., 2024).

La asignatura de Educación Física y Salud ofrece condiciones especialmente favorables para la innovación evaluativa. La visibilidad pública del desempeño motriz puede generar ansiedad y sentimientos de incompetencia cuando la evaluación se centra exclusivamente en estándares normativos, convirtiéndose en una barrera para el aprendizaje. Los modelos de evaluación formativa y compartida plantean el error como oportunidad de aprendizaje con efectos positivos en la experiencia emocional del estudiantado y en su adhesión a estilos de vida activos y saludables (Gallardo-Fuentes et al., 2025).

A pesar de la abundante literatura sobre EFC en la formación inicial del profesorado (Asun-Dieste y Guíu, 2023; Hortigüela-Alcalá et al., 2019; Ruiz et al., 2023), persiste una brecha de evidencia empírica en educación secundaria, especialmente en contextos escolares vulnerables. En estos escenarios las prácticas evaluativas cumplen, además de una función didáctica, una función de justicia educativa y de cohesión social (Rubiana et al., 2025), y pueden actuar como un factor de protección frente a la reproducción de desigualdades sociales.

El presente estudio tiene como objetivo examinar los efectos de una intervención pedagógica basada en la Evaluación Formativa y Compartida en las percepciones del estudiantado respecto a las dimensiones de Concepciones, Actores, Momentos-Instrumentos y Contenidos de la evaluación, en términos de agencia evaluativa y justicia evaluativa, tomando como escenario la asignatura de Educación Física y Salud en un contexto de vulnerabilidad social.

Método

La investigación adoptó un enfoque metodológico de carácter descriptivo-comparativo, con un diseño cuasiexperimental de grupo único con mediciones pre y post intervención (Hernández-Sampieri et al., 2014). Este diseño fue seleccionado porque la conformación aleatoria de grupos de control en contextos educativos reales implicaría privar a un grupo de estudiantes de una estrategia pedagógica que se presume beneficiosa.

Si bien el estudio fue predominantemente cuantitativo, se incorporaron registros cualitativos en bitácoras docentes que recogieron percepciones, reflexiones y ajustes realizados durante la implementación. En coherencia con la clasificación de diseños mixtos propuesta por Hernández-Sampieri y Mendoza (2018), esta integración corresponde a un diseño mixto con énfasis cuantitativo (QUAN-qual).



Participantes

La investigación se realizó en un establecimiento educacional particular subvencionado ubicado en el sur de Chile. Esta institución educativa se encuentra inserta en un contexto socioeducativo vulnerable, conformado principalmente por familias de clase media baja. En este contexto, la vulnerabilidad social se entiende como una condición contextual caracterizada por desventajas socioeconómicas y educativas acumuladas que limitan el acceso a recursos culturales y de apoyo escolar, lo cual tiene incidencia en las trayectorias educativas (Junta Nacional de Auxilio Escolar y Becas, 2024).

El establecimiento cuenta con una matrícula de 700 estudiantes distribuidos desde educación parvularia hasta cuarto año de enseñanza media (Grado 12). Su proyecto educativo institucional se orienta hacia el desarrollo integral del estudiante mediante metodologías activas centradas en las habilidades, fomentando la creatividad y el pensamiento crítico, con valores activos en los estudiantes. La institución adhiere a los principios de educación inclusiva, entregando herramientas concretas para formar hombres y mujeres que sean un aporte a la sociedad, mientras promueve la participación democrática de todos los miembros de la comunidad educativa.

La muestra inicial estuvo compuesta por 200 estudiantes de enseñanza media seleccionados mediante un muestreo no probabilístico por conveniencia, con edades que oscilan entre los 11 y los 19 años. La muestra final quedó conformada por los 182 estudiantes que completaron tanto el pretest como el postest. Se incluyeron estudiantes de 7° básico a IV medio (grado 12), donde se consideraron como criterios de inclusión una alta asistencia a clases y la aprobación del consentimiento informado de los padres/apoderados y asentimiento firmado por el estudiante. Se excluyeron del estudio (aunque no de las acciones formativas) aquellos estudiantes con ausencias reiteradas o que no firmaron los documentos correspondientes. La distribución por curso y género, según cada momento, se presenta a continuación en la tabla 1.

Tabla 1. Participantes en el estudio según momento, género y curso

Nivel	Entrada			Re-test			Salida		
	M	H	T	M	H	T	M	H	T
7° Básico	24	24	48	14	9	23	23	24	47
8° Básico	20	17	37	9	6	15	19	13	32
I° Medio	18	25	43	7	8	15	16	23	39
II° Medio	13	15	28	5	7	12	12	13	25
III° Medio	10	8	18	3	3	6	8	8	16
IV° Medio	10	16	26	4	7	11	9	14	23
Total	95	105	200	42	40	82	87	95	182

Nota. M=Mujer, H=Hombre, T=Total. Los niveles de I a IV medio en Chile corresponden a los grados de 9 a 12 en otros contextos.

Instrumentos

Para la recogida de datos cuantitativos se utilizó el instrumento denominado <<Cuestionario: sobre concepciones y percepción de los estudiantes de enseñanza media sobre la evaluación en la clase de educación física>>, diseñado por González-Palacio et al. (2022). Para facilitar su mención a lo largo del manuscrito, en el contexto de este estudio se le denominará cuestionario PEEMEEF. Este instrumento presenta una alta confiabilidad, evidenciada por un coeficiente alfa de Cronbach de 0.804 en el contexto en que se validó inicialmente. Consta de un total de 26 preguntas. Considera una escala de cinco elementos (1=Muy en desacuerdo, 2=En desacuerdo, 3=Indiferente, 4=De acuerdo y 5=Muy de acuerdo). El instrumento posee 4 dimensiones: concepciones, actores, instrumentos-momentos y contenido.

Previo a su aplicación, el cuestionario fue revisado por un panel de especialistas en evaluación, quienes coincidieron en que la redacción de los ítems era pertinente para el contexto de enseñanza media en Educación Física en que se aplicó.

Para la recolección de datos cualitativos se utilizaron bitácoras de reflexión semiestructuradas, completadas por el docente implementador al término de cada sesión. Las bitácoras contemplaban tres categorías predefinidas de registro: (a) participación estudiantil en las actividades de coevaluación; (b) cambios notables en el discurso del alumnado sobre la evaluación; y (c) desafíos y logros de la sesión.

Procedimientos



El estudio se desarrolló en tres fases consecutivas. En una primera etapa, se aplicó un cuestionario a los 200 estudiantes participantes del estudio con el objetivo de identificar sus perspectivas respecto a los procesos de evaluación en la asignatura de Educación Física y Salud. La aplicación se realizó mediante un formulario digital y los resultados fueron procesados utilizando las recomendaciones metodológicas del instrumento.

Con el propósito de evaluar la fiabilidad temporal del instrumento, se aplicó nuevamente el cuestionario a un subgrupo de 82 estudiantes transcurridas tres semanas desde la primera medición y antes del inicio de la estrategia de intervención. Este intervalo fue definido en concordancia con la literatura psicométrica, que recomienda periodos de entre 2 y 6 semanas para estudios de estabilidad para minimizar posibles sesgos de memoria de respuesta y, al mismo tiempo, reducir la probabilidad de cambios sustantivos en las percepciones evaluadas en ausencia de intervención (Koo y Li, 2016; Streiner y Norman, 2008).

A partir de los resultados del diagnóstico se diseñó y aplicó un plan de mejora a la totalidad del grupo participante. Una vez finalizada la intervención, se replicó el mismo cuestionario a los participantes con el fin de identificar posibles cambios en sus percepciones.

Consideraciones sobre la Validez Interna

Se implementaron acciones para controlar las principales amenazas a la validez interna descritas por Hernández-Sampieri et al. (2014). En relación con la maduración, si bien el desarrollo cognitivo y socio-emocional del estudiantado puede influir progresivamente en sus concepciones, la duración acotada de la intervención (tres meses), junto con la magnitud y especificidad de los cambios observados en variables de carácter perceptivo, hacen poco probable que estos se expliquen exclusivamente por procesos de desarrollo natural.

Respecto a la amenaza de historia, no se registraron durante el periodo de intervención eventos institucionales, curriculares o evaluativos externos que pudieran explicar de forma alternativa los cambios detectados. La amenaza de instrumentación se controló rigurosamente mediante la aplicación del mismo cuestionario digital bajo condiciones idénticas (González-Palacio et al., 2022). Asimismo, el riesgo de regresión a la media se minimizó al incluir a la totalidad de la población estudiantil disponible sin seleccionar grupos con valores extremos. Se reconoce el posible efecto de la prueba como una limitación inherente al diseño, la cual se asume como parte del proceso.

De manera complementaria, la interpretación de los resultados cuantitativos se trianguló con registros cualitativos provenientes de bitácoras docentes, evidenciando que los cambios en las percepciones del estudiantado coinciden temporalmente con la implementación de actividades específicas de autoevaluación y coevaluación. Esta convergencia entre fuentes de datos aporta mayor consistencia en la interpretación de los efectos observados.

Plan de acción

La intervención se integró orgánicamente en una unidad didáctica de 15 sesiones de 90 minutos cada una, como un continuo pedagógico en las siguientes cuatro fases:

Fase 1: Deconstrucción y Contrato Didáctico (Sesiones 1-3). Esta fase se orientó a dismantelar la visión punitiva de la evaluación. En la primera sesión, tras aplicar el pretest, se realizó un plenario titulado ¿Para qué evaluamos?, donde emergieron conceptos asociados al miedo y al control. El docente introdujo explícitamente la distinción entre calificación (nota) y evaluación (mejora). Se negoció un Contrato de Evaluación donde los estudiantes, por votación, decidieron ponderar un 40% de la nota final a procesos de autoevaluación y coevaluación, reservando el 60% a la heteroevaluación docente.

Fase 2: Alfabetización Evaluativa y Modelaje (Sesiones 4-6). Antes de solicitar a los estudiantes que se evaluaran, se dedicaron sesiones al análisis de videos de desempeño motriz. Los estudiantes evaluaron ejecuciones de terceros a través de rúbricas, discutiendo sus juicios en grupos pequeños para calibrar criterios. Se introdujo la técnica del semáforo para la retroalimentación entre pares: Verde (lo que hiciste bien), Amarillo (lo que puedes mejorar) y Rojo (lo que te impidió avanzar).

Fase 3: Práctica con Coevaluación (Sesiones 7-13). Durante la práctica motriz, se establecieron pausas evaluativas cada 20 minutos. En parejas, un estudiante ejecutó mientras el otro observó con una lista de cotejo simplificada, entregando retroalimentación inmediata. El docente asumió un rol de consultor e



intervino solo cuando las parejas no lograron consenso. Se priorizó una transición progresiva, donde la retroalimentación evolucionó de comentarios superficiales (bien hecho) a indicaciones técnicas precisas.

Fase 4: Calificación Dialogada y Cierre (Sesiones 14-15). La unidad culminó con charlas de autoevaluación, donde cada estudiante presentó su portafolio de evidencias y propuso una calificación. El docente contrastó esta propuesta con sus propios registros.

Análisis de datos

Para constatar la confiabilidad del instrumento se calcularon el alfa de Cronbach y el omega de McDonald, a nivel global y por dimensión (Dunn et al., 2014), y se estimó la estabilidad temporal test-retest mediante correlaciones de Pearson y coeficientes de correlación intraclase (Koo y Li, 2016). En segundo lugar, para descartar un sesgo de atrición, los participantes que completaron ambas mediciones ($n = 182$) se compararon con quienes no lo hicieron ($n = 18$) en las medidas del pretest mediante pruebas t de Welch para muestras independientes. En tercer lugar, se verificó el supuesto de normalidad de las diferencias pre-post mediante la prueba de Shapiro-Wilk, complementada con la inspección visual de los histogramas de distribución; las diferencias entre mediciones se contrastaron con la prueba t de Student para muestras relacionadas y, como análisis de robustez, con la prueba no paramétrica de rangos con signo de Wilcoxon. Finalmente, se calcularon los tamaños del efecto mediante la d de Cohen con sus intervalos de confianza al 95%. La totalidad de los procedimientos estadísticos se realizó con JASP (versión 0.95.3).

Los registros de las bitácoras docentes se examinaron mediante análisis temático, siguiendo las fases propuestas por Braun y Clarke (2006), que incluyen la lectura reiterada de las entradas, codificación de los segmentos vinculados a las categorías de registro y agrupación en temas. Como criterios de rigor se resguardó la trazabilidad de cada extracto a su bitácora. Los hallazgos se integraron con los resultados cuantitativos mediante una estrategia de complementariedad, triangulándolos por dimensión en el apartado de Resultados.

Consideraciones éticas

La ética de la investigación se resguardó mediante la participación voluntaria del estudiantado, quienes otorgaron su consentimiento explícito para formar parte del estudio. Se aseguró en todo momento la confidencialidad y el resguardo de la información recolectada. Con este fin, se implementó un sistema de codificación basado en una matriz asociativa, lo que permitió anonimizar las muestras sin perder la trazabilidad necesaria para vincular las mediciones pre, re y postest correspondientes a cada participante.

Resultados

Confiabilidad del cuestionario en el contexto estudiado

Se evaluó la consistencia interna del cuestionario PEEMEEF a la primera recogida de datos, tanto por dimensión como de manera global. Los coeficientes se presentan en la Tabla 2.

Tabla 2. Consistencia interna de la escala por dimensión y globalmente

Dimensión	Ítems	α	ω
Concepciones	5	0.61	0.66
Actores	6	0.55	0.56
Momentos-instrumentos	6	0.71	0.76
Contenidos	9	0.66	0.63
Global	26	0.85	0.83

Aunque la consistencia interna varió entre dimensiones ($\alpha = 0.55$ – 0.76), se decidió mantener el instrumento completo debido a su coherencia teórica, su validación en trabajos previos y los niveles adecua-

dos de fiabilidad global ($\alpha = 0.85$; $\omega = 0.83$). En particular, la menor consistencia observada en la dimensión Actores se interpretó considerando la heterogeneidad de los roles implicados en los procesos evaluativos, más que como una limitación del instrumento.

Los análisis de correlaciones test-retest fueron significativas en todas las dimensiones, con coeficientes de Pearson entre 0.55 y 0.75 ($p < 0.001$). De manera complementaria, los coeficientes de correlación intraclase (ICC) oscilaron entre 0.55 y 0.74 en las dimensiones específicas, ubicándose todas en el rango de moderada fiabilidad (Koo y Li, 2016). Estos resultados son especialmente relevantes en el caso de la dimensión 'Actores', que, a pesar de presentar una menor consistencia interna, muestra niveles de estabilidad temporal comparables al resto de las dimensiones, lo que permite justificar su mantenimiento en la escala.

El valor de ICC = 0.74 para la dimensión *Final* se encuentra muy próximo al umbral de 0.75, lo que lo sitúa en el límite superior de la categoría moderada y cercano a ser considerado bueno. En conjunto, estos resultados permiten interpretar que el cuestionario presenta una estabilidad temporal aceptable en todas sus dimensiones (Tabla 3).

Tabla 3. Estabilidad temporal del cuestionario

Dimensión	n	r (Pearson)	p	ICC	Interpretación
Concepciones	82	0.56	< 0.001	0.55	Moderada
Actores	82	0.55	< 0.001	0.55	Moderada
Momentos-instrumentos	82	0.69	< 0.001	0.69	Moderada
Contenidos	82	0.68	< 0.001	0.66	Moderada
Final	82	0.75	< 0.001	0.74	Moderada

Nota. Interpretación basada en los criterios de Koo y Li (2016): ICC < 0.50 = pobre; 0.50-0.75 = moderada; 0.75-0.90 = buena; > 0.90 = excelente.

Aplicación inicial del cuestionario

La aplicación inicial revela una percepción ambivalente sobre la evaluación. Si bien una amplia mayoría reconoce la importancia del rol docente (90%) y el aporte de la evaluación al aprendizaje (88%), se obtuvo solo un 41% de acuerdo en su utilidad y un 29% del alumnado (sumando las categorías a veces y casi nunca) consideró que la calificación no reflejaba adecuadamente su aprendizaje real.

La Tabla 4 muestra que la media más alta corresponde a la dimensión Contenidos y la más baja a Actores. Las dimensiones Concepciones y Momentos-instrumentos se ubican en valores intermedios. Las desviaciones estándar se mantienen estables en torno a 0.40 en todas las dimensiones, lo que indica una dispersión relativamente homogénea de las respuestas.

Tabla 4. Resultados iniciales del cuestionario PEEMEEF

Dimensión	Media	DE
Concepciones	3.03	0.37
Actores	2.74	0.43
Momentos-instrumentos	3.41	0.43
Contenidos	3.83	0.39

Nota. N = 200 casos válidos.

En el análisis de los ítems con menor valoración se revelan focos específicos de tensión en la cultura evaluativa. Los estudiantes muestran escepticismo sobre si la calificación refleja realmente el aprendizaje (ítem 6, M = 1.96) y una desconfianza inicial hacia la participación de sus pares en la evaluación (ítem 5, M = 2.41). Se observan percepciones críticas respecto a la influencia de factores no académicos en la nota, tales como el uso del uniforme escolar o el comportamiento, y una baja percepción de justicia evaluativa por parte del docente (M = 3.33). Estos hallazgos de entrada confirman la necesidad de enfatizar la democratización y transparencia de la evaluación durante la intervención.

Resultados de la aplicación de la estrategia de intervención

Las pruebas t de Welch no evidenciaron diferencias estadísticamente significativas entre los participantes que completaron ambas mediciones (n = 182) y quienes solo respondieron el pretest (n = 18) en ninguna de las dimensiones del cuestionario ni en la puntuación total ($|t| \leq 0.94$, gl de Welch entre 20.77



y 26.92, $p \geq .355$), lo que sugiere que la atrición no introdujo un sesgo relevante en las estimaciones. Una vez completada la intervención, se verificó el supuesto de normalidad de las diferencias pre-post mediante la prueba de Shapiro-Wilk. Los resultados no evidencian desviación significativa respecto de la normalidad ($W = 0.992$; $p = 0.426$; n pareadas = 182), lo que, junto con la inspección del histograma de diferencias, justifica el uso de la prueba t de Student para muestras relacionadas como análisis principal. En aras de robustez, se complementa la inferencia con la prueba no paramétrica de rangos con signo de Wilcoxon.

Al comparar el pretest con el posttest se observa un incremento estadísticamente significativo en todas las dimensiones evaluadas (Tabla 5 y figura 1). Las puntuaciones medias aumentan entre 0.34 y 0.71 puntos, y tanto la prueba t de Student como la prueba de rangos con signo de Wilcoxon confirman que las diferencias fueron significativas ($p < 0.001$).

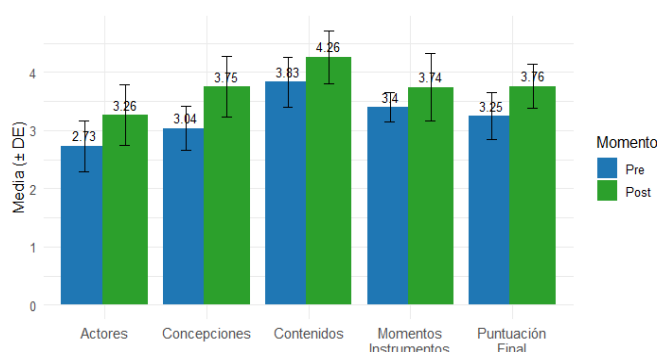
Los tamaños del efecto observados son de moderados a grandes en todas las dimensiones, destacando la dimensión Concepciones ($d = 1.06$; IC 95%: 0.88-1.24) y la puntuación final ($d = 1.18$; IC 95%: 0.99-1.37), lo que indica cambios sustantivos entre las mediciones pre y post intervención.

Tabla 5. Comparativa entre resultados al inicio y final del plan de acción

Dimensión	M Pre $\pm$ DE	M Post $\pm$ DE	ΔM (Post-Pre)	t (gl=181)	d de Cohen	Magnitud
Concepciones	3.04 $\pm$ 0.37	3.75 $\pm$ 0.58	+0.71	-14.26***	1.06 [0.88-1.24]	Grande
Actores	2.73 $\pm$ 0.43	3.26 $\pm$ 0.52	+0.52	-11.25***	0.83 [0.66-1.00]	Grande
Momentos-Instrumentos	3.40 $\pm$ 0.43	3.74 $\pm$ 0.46	+0.34	-7.48***	0.56 [0.40-0.71]	Moderado
Contenidos	3.83 $\pm$ 0.40	4.26 $\pm$ 0.52	+0.43	-10.18***	0.76 [0.59-0.92]	Moderado
Puntuación Final	3.25 $\pm$ 0.25	3.76 $\pm$ 0.38	+0.51	-15.94***	1.18 [0.99-1.37]	Grande

Nota. Los tamaños del efecto se reportan en valor absoluto. Los contrastes no paramétricos de Wilcoxon confirmaron el mismo patrón de diferencias significativas ($p < 0.001$). $N = 182$. *** = $p < .001$

Figura 1. Comparativa visual entre resultados al inicio y final del plan de acción



Fuente: Elaboración propia.

Concepciones registró el mayor incremento absoluto, al pasar de 3.04 a 3.75 (+0.71; $d = 1.06$); Actores, que partía de la media más baja del pretest (2.73), aumentó 0.52 puntos ($d = 0.83$); Contenidos se incrementó 0.43 puntos y alcanzó la media más alta del posttest (4.26; $d = 0.76$); y Momentos-Instrumentos presentó el cambio más discreto (+0.34; $d = 0.56$). La puntuación final aumentó 0.51 puntos ($d = 1.18$).

El incremento significativo en la dimensión Actores ($\Delta M = +0.52$), que parte con la media más baja en el pretest, es respaldado por los registros cualitativos donde se identifica el tema de la <<Transición de la pasividad a la agencia evaluativa>>. Las bitácoras docentes documentan que, tras una resistencia inicial, los estudiantes asumieron un rol autocrítico y constructivo. Una entrada de bitácora ilustra este cambio: <<los alumnos asumieron un rol autocrítico en los ejercicios de coevaluación, manteniéndose atentos a lo que ocurría en la clase y mostrando capacidad para describir situaciones con observaciones de carácter constructivo>> (Bitácora Docente 2, Sesión 10).

La dimensión Concepciones resulta ser la de mayor incremento ($\Delta M = +0.71$), lo que sugiere avances en diferenciar evaluación de calificación. Las jornadas de reflexión con el profesorado revelan que este es un punto central del trabajo, donde se valora la crítica amiga para desmontar creencias arraigadas.

Como se señala en los registros, se trabaja activamente en <<desmontar la creencia de que ambos términos son equivalentes, explicando que muchas veces la calificación no refleja el verdadero aprendizaje ni el esfuerzo del estudiante>> (Registro de Jornada de Reflexión 3), lo que da pruebas de que se avanza en una nueva visión pedagógica en el equipo docente.

Discusión

Los resultados de este estudio muestran que una intervención pedagógica basada en la Evaluación Formativa y Compartida (EFC) puede generar cambios positivos y estadísticamente significativos en las percepciones del estudiantado sobre la evaluación en Educación Física, fortaleciendo su orientación hacia el aprendizaje y desplazando la evaluación desde una lógica de calificación hacia su función reguladora del aprendizaje, en coherencia con los postulados de Canabal y Margalef (2017), López-Pastor (2017), Sanmartí (2020) y Molina et al. (2023).

El desplazamiento se expresó, además, en una resignificación de los contenidos evaluados, donde el estudiantado comenzó a valorar la evaluación no solo como verificación del rendimiento motor, sino como una oportunidad para profundizar aprendizajes de carácter conceptual y actitudinal, en línea con una evaluación orientada al aprendizaje (Canabal y Margalef, 2017) y especialmente relevante en el contexto escolar en que se desarrolla el estudio (Velásquez, 2025). Estos cambios pueden ser interpretados desde una perspectiva de justicia evaluativa, en la medida en que reconfiguran las formas de participación, los criterios de juicio y el sentido pedagógico de la evaluación (McArthur, 2023).

El cambio en las creencias del estudiantado podría estar condicionado por el impacto del liderazgo pedagógico. Cuando el profesorado asume el rol de líder de aprendizaje, trasciende la mera aplicación de instrumentos para gestionar integralmente el escenario de aprendizaje (Ravela et al., 2017; Sliwka et al., 2024). La decisión de ceder progresivamente el control del juicio evaluativo, guiar la reflexión y validar el error como parte constitutiva del aprendizaje permitió al estudiantado resignificar la evaluación, transitando desde una vivencia de evaluación como amenaza hacia una donde la evaluación es una herramienta de regulación interna (López-Pastor, 2017; Molina et al., 2023). El liderazgo docente resulta, así, decisivo para habilitar prácticas evaluativas percibidas como más justas y formativas (Sliwka et al., 2024).

El incremento observado en la dimensión Actores aporta evidencia a favor del desarrollo de la agencia del estudiante en el proceso evaluativo y es consistente con lo reportado por Gallardo-Fuentes et al. (2025), quienes identifican que la participación activa del alumnado en los procesos de evaluación constituye uno de los componentes más complejos, pero también más decisivos, para transformar la cultura evaluativa en Educación Física. Estos resultados permiten interpretar el fortalecimiento de la dimensión Actores como un primer indicador de redistribución del poder evaluativo, aun cuando se reconoce que su consolidación requiere que estas prácticas se sostengan y amplíen a nivel institucional.

Esta redistribución del poder se vio reforzada por la coherencia del proceso evaluativo en su conjunto, expresada en la explicitación de criterios, la claridad de los momentos evaluativos y la apertura de espacios de diálogo, condiciones que favorecieron evaluaciones percibidas como más comprensibles y legítimas, sin reducirlas a un momento final de calificación. Esta idea es consistente con enfoques que conciben la justicia evaluativa como un fenómeno relacional que se construye a lo largo de la experiencia evaluativa (Banasadi et al., 2023; Rasooli y DeLuca, 2024).

La literatura advierte sobre resistencias iniciales a la coevaluación vinculadas a dinámicas de poder, desconfianza o temor a la exposición (Harris et al., 2018). Los resultados obtenidos sugieren que estas barreras pueden ser mitigadas cuando el profesorado crea un clima de aula seguro, actúa como garante ético del proceso, entrega retroalimentación constructiva y desactiva el juicio punitivo; condiciones que propician el tránsito desde una participación pasiva hacia una agencia evaluativa activa y, con ello, el desarrollo de prácticas evaluativas percibidas como más justas, en línea con lo planteado por Treschman et al. (2025) sobre autorregulación y pensamiento crítico de orden superior.

En el contexto chileno, estos resultados aportan evidencia empírica a la implementación en aula de la política evaluativa chilena. Cuando el docente asume la responsabilidad de liderar prácticas evaluativas

orientadas a la justicia, se tensiona la inercia de la calificación tradicional y se restituye el sentido pedagógico de la evaluación acorde a lo orientado en los documentos normativos. Es en la calidad de las interacciones evaluativas cotidianas donde se define de forma colectiva qué aprendizajes se consideran valiosos y qué prácticas se perciben como justas (Wilkinson y Penney, 2021).

El estudio presenta limitaciones que deben ser consideradas. El diseño cuasiexperimental de grupo único impide atribuir causalidad exclusiva a la intervención, al no contar con un grupo de control aleatorizado. Aunque las acciones descritas en las Consideraciones sobre la Validez Interna reducen la plausibilidad de explicaciones alternativas, no es posible descartar por completo efectos de historia, maduración o de la propia prueba. Debe considerarse, además, la influencia del docente-investigador que implementó la intervención, ya que su presencia pudo inducir respuestas socialmente deseables en el estudiantado. Si bien el resguardo de identidad, la aplicación digital idéntica del cuestionario y la triangulación con las bitácoras contribuyen a mitigar este riesgo, no lo eliminan, por lo que se recomienda que futuros estudios incorporen aplicadores externos a la intervención.

De igual modo, los coeficientes de consistencia interna modestos de las dimensiones Concepciones ($\alpha = 0.61$) y Actores ($\alpha = 0.55$), en parte atribuibles al reducido número de ítems por dimensión (Dunn et al., 2014), obligan a interpretar los cambios en estas dimensiones con cautela. Asimismo, la duración acotada de la propuesta, aunque suficiente para observar cambios en las percepciones, no asegura la sostenibilidad de la agencia estudiantil a largo plazo. Futuras investigaciones deben reexaminar la estructura factorial del instrumento en población escolar chilena e incorporar diseños longitudinales y comparativos que permitan examinar la persistencia y transferencia de estos efectos. Si bien se reconoce la existencia de instrumentos orientados específicamente a medir percepciones de justicia evaluativa, en este estudio se optó por analizar la justicia evaluativa de manera indirecta, a partir de dimensiones pedagógicas situadas, coherentes con el enfoque de la Evaluación Formativa y Compartida, asumiéndola como un marco interpretativo emergente para el análisis de las prácticas evaluativas, por lo que las conclusiones referidas a la justicia evaluativa deben leerse como hipótesis interpretativas, a contrastar en diseños que incluyan una medición directa del constructo.

Conclusiones

Al examinar los efectos de una intervención pedagógica basada en la Evaluación Formativa y Compartida en las percepciones del estudiantado respecto de las dimensiones de Concepciones, Actores, Momentos-Instrumentos y Contenidos de la evaluación, y dado que en todas ellas se registraron cambios significativos, se concluye que la intervención produjo transformaciones relevantes en la manera en que el estudiantado comprende y experimenta la evaluación. En conjunto con la información cualitativa disponible, los hallazgos evidenciaron un desplazamiento desde una cultura evaluativa centrada en la calificación hacia una orientada al aprendizaje, la reflexión y la participación.

En el plano pedagógico, los resultados permitieron observar el potencial de la Evaluación Formativa y Compartida para incidir en la forma en que los estudiantes valoran y participan en los procesos evaluativos, más allá de sus efectos sobre el rendimiento. Los hallazgos sugieren que la intervención contribuyó a la construcción de percepciones favorables de justicia evaluativa, principalmente a través de la redistribución del poder evaluativo, la apertura de espacios de diálogo y el reconocimiento de trayectorias diversas de aprendizaje.

Afiliación de los autores

Mg. Walter Fernando Chavarría Gatica. Facultad de Educación. Universidad Católica de Temuco.

Lic. Michelle Acuña Guerra. Facultad de Educación. Universidad Católica de Temuco.

Dr. Ariane Alvarez Alvarez. Escuela de Diseño. Universidad Mayor.

Dr. Juan Francisco Cabrera Ramos. Facultad de Educación. Universidad Católica de Temuco.



Referencias

- Aquario, D. (2021). Through the lens of justice. A systematic review on equity and fairness in learning assessment. *Education Sciences & Society*, 12(2). <https://doi.org/10.3280/ess2-2021oa12405>
- Asun-Dieste, S., & Guío Carrera, M. (2023). Use of formative assessment in Physical Education teacher education in Secondary School: A case study. *Cultura, Ciencia Y Deporte*, 18(55), 171-190. <https://doi.org/10.12800/ccd.v18i55.1918>
- Baniasadi, A., Salehi, K., Khodaie, E., Bagheri Noaparast, K., & Izanloo, B. (2023). Fairness in classroom assessment: A systematic review. *The Asia-Pacific Education Researcher*, 32(1), 91-109. <https://doi.org/10.1007/s40299-021-00636-z>
- Barrientos Hernán, E. J., López-Pastor, V. M., Lorente-Catalán, E., & Kirk, D. (2023). 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>
- Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology*, 3(2), 77-101. <https://doi.org/10.1191/1478088706qp0630a>
- Canabal, C., & Margalef, L. (2017). La retroalimentación: La clave para una evaluación orientada al aprendizaje. *Profesorado. Revista de Currículum y Formación de Profesorado*, 21(2), 149-170. <https://doi.org/10.30827/profesorado.v21i2.10329>
- Dunn, T. J., Baguley, T., & Brunnsden, V. (2014). From alpha to omega: A practical solution to the pervasive problem of internal consistency estimation. *British Journal of Psychology*, 105(3), 399-412. <https://doi.org/10.1111/bjop.12046>
- Fraile-Aranda, A., Aparicio-Herguedas, J. L., Asún-Dieste, S., & Romero-Martín, R. (2018). La evaluación formativa de las competencias genéricas en la formación del profesorado de Educación Física. *Estudios Pedagógicos*, 44(2), 39-53. <https://dx.doi.org/10.4067/S0718-07052018000200039>
- Gallardo-Fuentes, F., Carter-Thuillier, B., Martínez-Angulo, C., Gallardo-Fuentes, J., & Peña-Troncoso, S. (2025). Percepciones del alumnado en dos cohortes de la formación inicial en Educación Física: consistencia, compromiso y retos en la evaluación. *Retos*, 70, 39-50. <https://doi.org/10.47197/retos.v70.113463>
- González-Palacio, E. V., Chaverra-Fernández, B. E., Ramírez-Martínez, L. F., Gaviria-Cortes, D. F., Arboleda-Serna, V. H., Arango-Paternina, C. M., Bustamante-Castaño, S. A., & Muriel-Echavarría, J. M. (2022). Percepción del estudiantado hacia la enseñanza del profesor de Educación Física: Diseño y validación de un cuestionario. *Educación Física y Ciencia*, 24(1), Artículo e210. <https://doi.org/10.24215/23142561e210>
- Harris, L. R., Brown, G. T. L., & Dargusch, J. (2018). Not playing the game: Student assessment resistance as a form of agency. *The Australian Educational Researcher*, 45(1), 125-140. <https://doi.org/10.1007/s13384-018-0264-0>
- Hernández-Sampieri, R., Fernández-Collado, C., & Baptista-Lucio, P. (2014). *Metodología de la investigación*. México: McGraw-Hill
- Hernández-Sampieri, R., y Mendoza, C. P. (2018). *Metodología de la investigación: Las rutas cuantitativa, cualitativa y mixta*. McGraw-Hill.
- Herrero-González, D., López-Pastor, V. M., Manrique-Arribas, J. C., & Moura, A. (2023). Formative and shared assessment: Literature review on the main contributions in physical education and physical education teacher education. *European Physical Education Review*, 30(3), 493-510. <https://doi.org/10.1177/1356336X231220995>
- Hortigüela-Alcalá, D., Palacios-Picos, A., & López-Pastor, V. (2019). The impact of formative and shared or co-assessment on the acquisition of transversal competences in higher education. *Assessment & Evaluation in Higher Education*, 44(6), 933-945. <https://doi.org/10.1080/02602938.2018.1530341>
- Junta Nacional de Auxilio Escolar y Becas. (2024). *Índices de vulnerabilidad escolar y vulnerabilidad multidimensional*. <https://www.junaeb.cl/imprescindibles-para-realizar-un-analisis-de-las-comunidades-educativas-del-pais-junaeb-publica-indices-de-vulnerabilidad-escolar-2024/>
- Kakara, A. H. L., Govaerts, M., Abdulla, L., Balmer, D. F., Busari, J. O., & West, D. C. (2025). Clarifying and expanding equity in assessment by considering three orientations: Fairness, inclusion and justice. *Medical Education*, 59(5), 494-502. <https://doi.org/10.1111/medu.15534>



- Koo, T. K., & Li, M. Y. (2016). A guideline of selecting and reporting intraclass correlation coefficients for reliability research. *Journal of Chiropractic Medicine*, 15(2), 155-163. <https://doi.org/10.1016/j.jcm.2016.02.012>
- López-Pastor, V. M. (2017). Evaluación formativa y compartida: evaluar para aprender y la implicación del alumnado en los procesos de evaluación y aprendizaje. En V. M. López-Pastor & Á. Pérez-Pueyo (Eds.), *Buenas prácticas docentes. Evaluación formativa y compartida en educación: experiencias de éxito en todas las etapas educativas* (pp. 34-68). Universidad de León.
- McArthur, J. (2023). Rethinking authentic assessment: work, well-being, and society. *Higher Education*, 85(1), 85-101. <https://doi.org/10.1007/s10734-022-00822-y>
- Ministerio de Educación de Chile. (2018). Decreto 67. *Biblioteca del Congreso Nacional de Chile*. <https://www.bcn.cl/leychile/navegar?idNorma=1127255>
- Molina, S. 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>
- Moura, A., MacPhail, A., Graça, A., & Batista, P. (2024). Encouraging students to co-construct and co- and self-regulate their learning within a cooperative learning environment in physical education. *Journal of Teaching in Physical Education*, 43(1), 11-20. <https://doi.org/10.1123/jtpe.2022-0104>
- Nieminen, J. H. (2022). Inclusive assessment, exclusive academy. En R. Ajjawi, J. Tai, D. Boud & T. Jorre de St Jorre (Eds.), *Assessment for inclusion in higher education: Promoting equity and social justice in assessment* (pp. 63-73). Routledge. <https://doi.org/10.4324/9781003293101-8>
- 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>
- Leão Pereira, A. F., Maldonado, D. T., & Lorente-Catalán, E. (2025). Quality physical education and critical pedagogy in the post-pandemic era: Reflections on European and Brazilian production. *Movimento*, 31, Article e31019. <https://doi.org/10.22456/1982-8918.144329>
- Pérez-Norambuena, S., Mora-Donoso, M., Peña-Troncoso, S., & Gallardo-Fuentes, F. (2025). Perspectiva emancipadora en las orientaciones evaluativas en el currículum de Educación Física en Chile. *Alteridad*, 20(2), 272-286. <https://doi.org/10.17163/alt.v20n2.2025.09>
- Randall, J., Poe, M., Slomp, D., & Oliveri, M. E. (2024). Our validity looks like justice. Does yours? *Language Testing*, 41(1), 203-219. <https://doi.org/10.1177/02655322231202947>
- Rasooli, A., & DeLuca, C. (2024). A critical review of fairness from multiple perspectives: Implications for classroom assessment theory. *Applied Measurement in Education*, 37(2), 148-164. <https://doi.org/10.1080/08957347.2024.2345594>
- Ravela, P., Picaroni, B., & Loureiro, G. (2017). *¿Cómo mejorar la evaluación en el aula? Reflexiones y propuestas de trabajo para docentes*. Grupo Magro Editores.
- Rubiana, I., Purwanto, S., & Anwar, M. H. (2025). Educación física inclusiva en la enseñanza secundaria: Un análisis bibliométrico de publicaciones de 2000 a 2024. *Retos*, 64, 233-241. <https://doi.org/10.47197/retos.v64.111747>
- Ruiz Lara, E., Romero Martín, M. R., & Ureña Ortín, N. (2023). Formative and shared assessment in Physical Education. Assessment to involve, dialogue, share and learn. *Cultura, Ciencia y Deporte*, 18(55), i-iii. <https://doi.org/10.12800/ccd.v18i55.2024>
- Ruz López, D., Pérez Arredondo, C., & Bañales Faz, G. (2024). Evaluación, calificación y promoción escolar en el Decreto 67 en Chile: Análisis discursivo de las perspectivas conceptuales y de los roles de los actores educacionales. *Discurso & Sociedad*, 18(2), 337-367. <https://doi.org/10.14198/dissoc.18.2.7>
- Sanmartí, N. (2020). *Evaluar y aprender: Un único proceso*. Octaedro.
- Santos Guerra, M. A. (2019). Errores a troche y moche sobre la evaluación educativa. En A. de la Herrán Gascón, J. M. Valle & J. L. Villena (Coords), *¿Qué estamos haciendo mal en educación?* (pp.131-156). Octaedro
- Santos Guerra, M. A. (2024). ¿Qué hacemos con el cuchillo? Las finalidades de la evaluación. *Revista Iberoamericana de Evaluación Educativa*, 17(1), 11-24. <https://doi.org/10.15366/riee2024.17.1.001>



- Slingerland, M., Weeldenburg, G., & Borghouts, L. (2024). Formative assessment in physical education: Teachers' experiences when designing and implementing formative assessment activities. *European Physical Education Review*, 30(4), 620-637. <https://doi.org/10.1177/1356336X241237398>
- Sliwka, A., Klopsch, B., Beigel, J., & Tung, L. (2024). Transformational leadership for deeper learning: Shaping innovative school practices for enhanced learning. *Journal of Educational Administration*, 62(1), 103-121. <https://doi.org/10.1108/JEA-03-2023-0049>
- Streiner, D. L., & Norman, G. R. (2008). *Health measurement scales: A practical guide to their development and use* (4.^a ed.). Oxford University Press.
- Tolgfors, B. (2018). Different versions of assessment for learning in the subject of physical education. *Physical Education and Sport Pedagogy*, 23(3), 311-327. <https://doi.org/10.1080/17408989.2018.1429589>
- 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>
- Velásquez Díaz, W. S. (2025). Evaluación formativa en secundaria: avances y desafíos 2025. *Revista Tribunal*, 5(12), 581-606. <https://doi.org/10.59659/revistatribunal.v5i12.225>
- Wilkinson, S. D., & Penney, D. (2021). Setting policy and student agency in physical education: Students as policy actors. *Sport, Education and Society*, 26(3), 267-280. <https://doi.org/10.1080/13573322.2020.1722625>

Datos de los autores:

Walter Fernando Chavarría Gatica
Michelle Acuña Guerra
Ariane Alvarez Alvarez
Juan Francisco Cabrera Ramos

walter.chavarria.gatica@gmail.com
michelleee.acg@gmail.com
ariane.alvarez@umayor.cl
jcabrera@uct.cl

Autor
Autora
Autora
Autor

