



Towards a formative and affective feedback in Physical Education: a Delphi analysis of its key attributes

Hacia un feedback formativo y afectivo en Educación Física: un análisis Delphi de sus atributos clave

Autores

Ana Casares Polo ¹
Alicia M^a Alonso Martínez ²
M^a Rosario Romero Martín ³
Nerea Estrada Marcén ⁴
Sonia Asún Dieste ⁵

^{1,2} Universidad Pública de Navarra
(España)

^{3,4,5} Universidad de
Zaragoza (España)

Autora de correspondencia:
Ana Casares Polo
anacasarespolo71@gmail.com

Received: 05-02-26

Accepted: 30-05-26

How to cite in APA

Casares Polo, A., Asún-Dieste, S., Alonso Martínez, A. M., Romero Martín, M. R., & Estrada Marcén, N. (2026). Towards a Formative-Affective Feedback in Physical Education: a Delphi Analysis of its Key Attributes. *Retos*, 82, 701-714.
<https://doi.org/10.47197/retos.v82.118735>

Abstract

Introduction: feedback in educational context is a resource used to convey information about learning in order to regulate and motivate it throughout the process; and it is considered a central element of formative assessment. In Physical Education (PE), the motor nature of tasks and social interaction with peers foster an affective-emotional climate that supports the development of feedback. Although there is a substantial body of research on feedback, its characteristics are presented in diverse ways and are sometimes insufficiently systematized. **Objective:** the aim of this study is to identify the attributes that define a formative and affective feedback in PE.

Methodology: to this end, the Delphi method was applied through two rounds of consultation with 14 experts.

Results: data revealed high levels of representativeness and consensus regarding six attributes related to: (1) adaptation to students; (2) information about learning, successes, and errors; (3) promotion of dialogue, active listening, and respect; (4) integration into peer-assessment strategies; (5) inclusion in self-assessment actions; (6) orientation toward personal improvement, self-confidence, and commitment.

Discussion: the Delphi study validates the structure of formative and affective feedback (FAF), integrating instructional, interactional, and self-regulatory dimensions within secondary PE.

Conclusions: these attributes constitute the conceptual framework of FAF and represent an essential preliminary phase for the subsequent development of research instruments, that are useful both for generating scientific evidence, and for improving initial and continuous teacher education in PE.

Keywords

Affective dimension; feedback; formative assessment; physical education.

Resumen

Introducción: el feedback en el ámbito educativo constituye un recurso para transmitir información sobre el aprendizaje, con el fin de regularlo y motivar al alumnado durante el proceso. Se considera un elemento central de la evaluación formativa. En Educación Física (EF), la naturaleza motriz de las tareas y la interacción social con los compañeros/as propician un clima afectivo-emocional para desarrollar este feedback. Las numerosas investigaciones sobre feedback describen características heterogéneas y a veces poco sistematizadas.

Objetivo: identificar los atributos que definan un feedback formativo y afectivo en EF.

Metodología: se aplicó el método Delphi a través de dos rondas de consulta con 14 expertos/as.

Resultados: los datos mostraron altos niveles de representatividad y consenso en los seis atributos siguientes: (1) adaptación al alumnado; (2) información sobre el aprendizaje, aciertos y errores; (3) fomento del diálogo, escucha activa y respeto; (4) integración en estrategias de coevaluación; (5) inclusión en acciones de autoevaluación; y (6) orientación hacia la mejora personal, la autoconfianza y el compromiso.

Discusión: el Delphi valida la estructura de un feedback formativo afectivo (FFA), que integra lo instructivo, interaccional y autorregulatorio, en el contexto de la EF en secundaria.

Conclusiones: estos atributos configuran un patrón conceptual de un FFA, y constituyen una fase previa esencial para el posterior desarrollo de instrumentos de investigación útiles, tanto para la evidencia científica como para la mejora en la formación inicial y continua del profesorado de EF.

Palabras clave

Dimensión afectiva; educación física; evaluación formativa; feedback.

Introducción

Hoy en día y dentro del ámbito educativo, la evaluación ha experimentado cambios importantes, al desplazarse desde un enfoque centrado en el papel del docente para calificar, hacia otro que propone una evaluación formativa con el alumnado como protagonista activo de su aprendizaje. Esta evaluación formativa se entiende como un proceso continuo que permite la recogida de información relevante para regular el aprendizaje y orientar la acción pedagógica (Sanmartí, 2010). Frente a la evaluación sumativa, centrada en la medición del rendimiento y la asignación de calificaciones, la evaluación formativa pone el énfasis en el acompañamiento del aprendizaje y en la construcción de significado por parte del alumnado (Adarkwah, 2021). Dentro de esta, el feedback o la retroalimentación se convierte en un recurso esencial que aporta información relevante sobre el aprendizaje con la intención de regularlo.

En este contexto, cabría preguntarse qué se entiende por feedback. Son numerosos los estudios y las definiciones. Algunos de los modelos iniciales y que han tenido gran impacto son los de Hattie y Timperley (2007) y Brookhart (2017). Ambos priorizan la eficacia del mensaje hacia el logro de unos objetivos de aprendizaje. Para Hattie y Timperley (2007), el feedback es la información que da un agente (profesor, compañero, libro, padre) sobre aspectos del desempeño del estudiante para reducir la brecha entre el estado actual y la meta (enfoque cognitivo). Establecen cuatro niveles en los que focalizar el feedback: la tarea, el proceso, la autorregulación y la persona. En una línea similar, Brookhart (2017) propone un concepto de feedback que ayuda al alumnado a saber qué hizo bien, qué necesita mejorar del aprendizaje y cómo puede lograr los objetivos. En relación con las características del *feedback*, esta autora propone la valencia (positivo/negativo), la temporalidad (inmediato/diferido), el foco (en la tarea o en la persona) y el modo (verbal/no verbal). Estas aportaciones se centran en el papel del docente como emisor y en la vertiente más instructiva o cognitiva del feedback, no tanto en la parte afectiva.

En contraste con estos autores, Carless (2016) plantea el feedback como un diálogo donde las dos partes (emisor y receptor del feedback), aportan información y dan significado al aprendizaje. El docente no solo corrige errores, sino que interactúa con el alumnado de forma bidireccional. Desde este marco, destacan dos aportaciones recientes que proponen un feedback más enfocado a la percepción del alumnado y los aspectos afectivos del mensaje. La primera es de Mandouit & Hattie (2023). Estos autores hablan del potencial del feedback desde la perspectiva del estudiante, afirmando que, independientemente de cómo el docente diseñe el feedback, lo importante es lo que pasa por la cabeza del alumnado cuando lo recibe.

La segunda aportación es de Carless (2025). Este autor plantea que el alumnado puede percibir el feedback de tres formas: como una evaluación de su rendimiento a partir de unos estándares de aprendizaje, como un reconocimiento y aprecio por lo realizado, y como consejos y sugerencias para crecer y mejorar. Por ejemplo, cuando una persona espera el reconocimiento, pero recibe consejos, puede surgir un choque en la comunicación.

Otra aportación relevante es la revisión sistemática de Lipnevich y Panadero (2022) y su propuesta del modelo MISCA formado por cinco elementos que interactúan desde ambas vertientes (cognitiva y afectiva). La información técnica, por muy precisa que sea, puede resultar ineficaz si no se considera la disposición afectiva del estudiante como receptor del feedback (Lipnevich & Panadero, 2022). Estos autores mencionan la falta de modelos que unifiquen ambas dimensiones.

Desde esta perspectiva, se plantea en el actual estudio el concepto de un feedback formativo y afectivo (FFA). A diferencia del feedback socioemocional, más centrado en el refuerzo positivo genérico o el ánimo personal (Lipnevich & Panadero, 2022), el FFA integra la información sobre el aprendizaje con el soporte afectivo y la autorregulación. Se define como una acción comunicativa, intencional y dialógica, cuyo mensaje combina la información técnica de la tarea motriz con la gestión consciente de la energía emocional. Su objetivo es doble: por un lado, permitir el progreso en el aprendizaje mediante información precisa, y por otro, fortalecer la confianza del estudiante, su autorregulación y el compromiso con la práctica.

En relación con su aplicación práctica, y con el soporte de la literatura anteriormente mencionada, este FFA podría estructurarse en tres dimensiones: a) la dimensión instruccional, centrada en la precisión del mensaje; b) la dimensión interaccional, orientada a la vertiente más afectiva de la comunicación, y vinculada al cuidado y la seguridad del alumnado; y c) la dimensión autorregulatoria, que promueve la autorreflexión y el autocontrol del alumnado en torno a sus aprendizajes.

La dimensión instruccional se relaciona con los niveles de tarea y proceso cognitivo de Hattie y Timperley (2007). Por su parte, las dimensiones interaccional y autorregulatoria encuentran sustento en los componentes del modelo MISCA (Lipnevich & Panadero, 2022). La dimensión interaccional se vincula con los elementos de *Agente*, *Contexto* y *Estudiante*, para generar un clima emocionalmente favorable, de confianza y seguridad, y en la naturaleza dialógica del feedback descrita por Carless (2016). Este autor propone el feedback como un intercambio interactivo que construye significados a través de un diálogo fluido. La dimensión autorregulatoria responde al componente *Estudiante* del MISCA; el alumnado procesa la información desde factores cognitivos, metacognitivos y motivacionales. Además, el procesamiento interno de autorregulación se vincula con el *feedback literacy* (Carless & Boud, 2018), entendido como la competencia del alumnado para comprender la información, gestionar el impacto emocional del mensaje y transformarlo en acciones concretas para la mejora de sus desempeños motrices.

De alguna manera, la vertiente cognitiva y la afectiva del feedback se interconectan para abordar las tres dimensiones de forma holística, en la línea de las corrientes actuales que vinculan los estilos de enseñanza con la mejora del clima social-emocional del aula y la satisfacción del alumnado (León-Díaz et al., 2023).

El presente estudio traslada este enfoque al contexto de la EF, donde la parte instructiva y afectiva convergen en un entorno de práctica motriz y exposición pública constante del cuerpo, factores generadores de vulnerabilidad emocional en la infancia y adolescencia. En este sentido, Almonacid-Fierro et al. (2024), destacan la importancia de la interacción profesor-alumno, el apoyo emocional, la comunicación efectiva y la regulación emocional en situaciones pospandemia. Estos autores reconocen la importancia de la afectividad y la motivación para promover el aprendizaje en un clima de clase saludable. En este contexto, el FFA puede ser un recurso que ayude a crear una atmósfera emocionalmente favorable para el aprendizaje.

Por otra parte, además de los enfoques cognitivos ya mencionados (Hattie & Timperley, 2007; Brokhart, 2017), existen dos estudios que refuerzan la importancia de estudiar este FFA en el contexto motriz de las clases de EF. El primero es la revisión sistemática de Zhou et al. (2021), donde se evidencia mayor efectividad en el aprendizaje de habilidades motrices en el alumnado que recibió feedback frente al que no lo recibió. El segundo estudio es el meta-análisis de Han et al. (2022) donde se confirma que, aunque el feedback en EF mejora de forma significativa las habilidades motrices, los formatos más tradicionales verbales, informativos o aislados pierden impacto si no se optimizan metodológicamente. En este sentido, será importante planificar cómo, cuándo y a quién dar este feedback. A partir de estos dos estudios emergen nuevas cuestiones en relación a los diferentes formatos del FFA y su aplicación en las clases de EF.

Por otra parte, si bien la pedagogía moderna reconoce la importancia de las emociones, las guías prácticas para docentes de EF siguen priorizando la corrección del error motriz sobre la gestión del impacto emocional. A partir de las aportaciones de Pérez-Pueyo et al. (2024), se puede concluir que el profesorado carece de recursos para sistematizar la práctica del feedback, lo que deriva en una aplicación intuitiva, y a veces escasa o genérica, teniendo en cuenta la realidad de las clases, con grupos numerosos y poco tiempo de compromiso motor.

Teniendo en cuenta estas consideraciones, el presente estudio tiene como objetivo identificar y consensuar, mediante el método Delphi, los atributos esenciales de un FFA en EF. La delimitación de este patrón conceptual, formado por los atributos del FFA, puede servir de base teórica y operativa para que, en etapas posteriores, se pueda abordar el diseño y la validación de instrumentos útiles para la práctica docente diaria.

Método

Diseño

Para la identificación, depuración y consenso de los atributos de un FFA se empleó un diseño no experimental basado en el método Delphi. Este método permite integrar el juicio experto mediante un proceso estructurado de consulta anónima y sucesivas rondas de retroalimentación, siendo especialmente adecuado cuando el objetivo es consensuar y operacionalizar constructos complejos procedentes de marcos teóricos diversos (Keeney et al., 2025).

El estudio partió de la literatura actual sobre el feedback para sistematizar y priorizar atributos vinculados a las tres dimensiones mencionadas (instruccional, interaccional y autorregulatoria). A partir de una revisión bibliográfica inicial, el Delphi incorporó valoraciones expertas sobre la relevancia, la claridad y la pertinencia de dichos atributos en contextos reales de EF.

El proceso se estructuró en tres fases —preliminar, exploratoria y final— siguiendo propuestas metodológicas consolidadas (García & Suárez, 2013).

Muestra

La selección de las personas expertas se realizó mediante un muestreo intencional, atendiendo a criterios de idoneidad, experiencia y reconocimiento en evaluación formativa en EF. Se establecieron los siguientes criterios de inclusión: (a) experiencia docente y/o investigadora contrastada en evaluación en EF; (b) participación en proyectos, publicaciones o redes académicas relacionadas con evaluación formativa; y (c) conocimiento específico del contexto educativo español.

Para ello se contó con 14 personas expertas, docentes y ex docentes de EF. Este tamaño es adecuado según la literatura, que prioriza la competencia sobre el número de participantes (Powell, 2003; Skulmoski et al., 2007). Asimismo, se garantizó la estabilidad del panel durante las rondas, condición clave para la validez del proceso Delphi. Los expertos pertenecían a la Red Nacional de Evaluación Formativa y Compartida ya que se priorizó la necesidad de homogeneidad epistémica y especificidad profesional. Al acotar el panel a la principal red científica de la temática en el contexto nacional, se aseguró un nivel óptimo de competencia experta y una alta sintonía conceptual, requisitos avalados por la literatura cuando el Delphi se orienta a la validación y operativización de constructos complejos (Hsu y Sandford, 2007; Landeta, 2006). El objetivo no era confrontar paradigmas divergentes, sino contar con un juicio cualificado en torno a los atributos de un *feedback* formativo y afectivo, evitando una dispersión semántica que comprometiera el constructo (Powell, 2003). Para neutralizar el sesgo de conformidad por la filiación académica compartida, se aplicó un estricto anonimato entre los participantes en todas las rondas, impidiendo presiones sociales o liderazgos internos (Diamond et al., 2014). Además, la pluralidad institucional y geográfica del panel garantizó la diversidad de contextos y formas de aplicación práctica del *feedback*. Los expertos procedían de diferentes universidades e institutos de cuatro comunidades autónomas (Aragón, 29%; Cataluña, 29%; Castilla-La Mancha, 21%; y Murcia, 21%). Se aseguró además un equilibrio de género (43% mujeres, 57% hombres) y una amplia representatividad intergeneracional (Tabla 1).

Con el fin de asegurar la calidad del panel, se calculó el índice de competencia experta (K), considerando el conocimiento declarado y la fundamentación de dicho conocimiento. Los valores obtenidos oscilaron entre 0.625 y 0.975. Siguiendo la clasificación de Almenara y Osuna (2013), el 71% de las personas expertas presentó un nivel alto de competencia ($K > .80$) y el 29% un nivel medio ($.50 \leq K \leq .80$).

Tabla 1. Características de las personas expertas.

Sexo		Edad					Competencia experta	
H	M	<30	31-40	41-50	51-60	>60	media(.50-.80)	alta(>.80)
43%	57%	0%	36%	57%	0%	7%	29%	71%

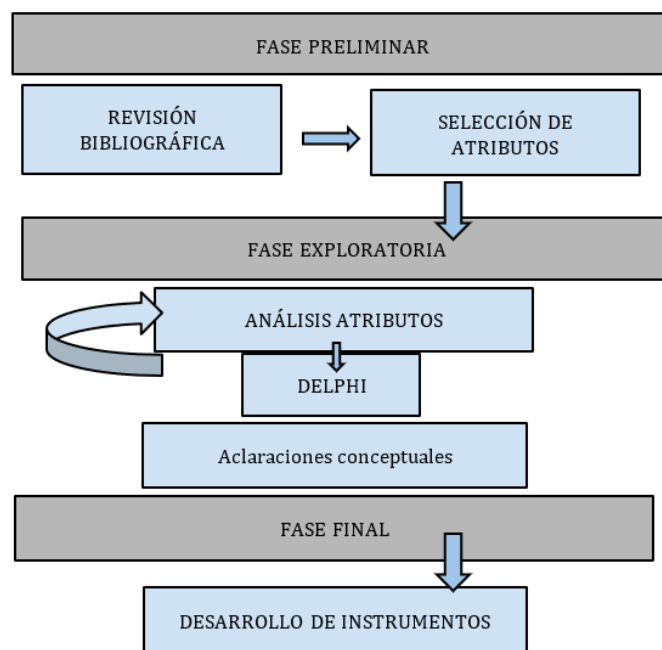
Nota: El porcentaje está calculado respecto al total de las personas participantes

Procedimiento

La investigación se estructuró de forma secuencial en tres fases diferenciadas: (1) fase preliminar, orientada a la acotación teórica y la selección y codificación de los atributos iniciales del feedback; (2) fase exploratoria, constituida por el desarrollo de dos rondas Delphi con el panel de expertos para la valoración cuantitativa y cualitativa de los indicadores; y (3) fase final, dedicada a la determinación del consenso y la redacción del informe definitivo.

Para el análisis estadístico de los datos cuantitativos se utilizó el programa IBM SPSS Statistics 30.0.

Figura 1. Fases de la investigación



Fase preliminar: selección de atributos

El objetivo de esta fase fue identificar y ordenar qué atributos podrían ser propios de un FFA en el contexto de la EF. Para ello, se llevó a cabo una revisión documental estructurada a partir de Google Scholar, Scopus y Dialnet, utilizando los términos de búsqueda: "evaluación formativa", "retroalimentación", "feedback formativo", "afectividad" y "educación física".

Para la selección de la literatura, se aplicaron de forma sistemática tres criterios de inclusión: (1) ser monografías o artículos de revisión de impacto internacional que definieran modelos de feedback vigentes como Brookhart (2008) y Hattie y Timperley (2007); (2) ser investigaciones empíricas o teóricas que situaran la retroalimentación en el plano de la interacción social, la evaluación compartida y el *feedback literacy* como Carless (2012, 2013, 2025) y Carless & Boud (2018); y (3) ser publicaciones de alto impacto de los últimos seis años orientadas a la integración de componentes socioemocionales y autorregulatorios de la evaluación como Berlanga y Juárez-Hernández, (2020), y Panadero y Lipnevich (2022). Estas obras se categorizaron a partir de las tres dimensiones de un FFA (instruccional, interaccional y autorregulatoria), establecidas tras el análisis, deliberación y consenso de cinco expertos en Didáctica de dos universidades españolas. Se realizó un vaciado de aportaciones y sugerencias textuales de la bibliografía, hasta sintetizar un listado de 25 atributos, para iniciar la primera ronda de consulta con el panel de expertos.

Tabla 2. Referencias bibliográficas, dimensiones y selección inicial de atributos del feedback.

Dimensión del FFA/ Referencias	Atributos iniciales (AT)
Dimensión instruccional: calidad técnica y pragmática del mensaje	
Brookhart, S. (2017)	1, 2, 7, 8, 9, 13, 14, 19, 21
Hattie, J. y Timperley, H. (2007)	3, 4, 5, 6, 19, 22
Wisniewski et al. (2020)	3,4,6
Dimensión interaccional: diálogo, clima de aula y vertiente afectiva	
Carless, D. (2012)	1,6,15
Berlanga, M., & Juárez-Hernández, L. G. (2020)	1, 3, 7, 14, 20
Tobón, S. (2013)	1,3,14,20
Dimensión autorregulatoria: autonomía, metacognición y <i>feedback literacy</i>	
Carless, D. (2013)	16,17-
Carless, D. y Boud, D. (2018)	1,15,16,17
Carless, D. (2025)	1,15,21,22
Panadero, E. y Lipnevich, A. A. (2022)	10, 15, 16, 17, 19, 20, 21, 22

Nota. Los números corresponden a los 25 atributos iniciales (AT) para la primera ronda de consulta del Delphi

Fase exploratoria

Concluida la delimitación del listado inicial, se procedió a evaluar el grado de idoneidad de los 25 atributos mediante el juicio de expertos. Para ello, se aplicó un método Delphi de dos rondas consecutivas al panel de expertos, quienes trabajaron de manera independiente y anónima. En ambas rondas, se les solicitó valorar mediante una escala tipo Likert de 1 a 5 puntos la claridad, pertinencia y relevancia de cada atributo (donde 1 = valor mínimo y 5 = valor máximo), disponiendo asimismo de un espacio cualitativo abierto para observaciones.

Para garantizar la consistencia interna en el proceso de toma de decisiones y la robustez del constructo, se aplicaron criterios estadísticos a partir de las medidas de tendencia central con puntos de corte en la media y mediana ($\bar{X} \geq 4.2$ y $Me \geq 4.0$), (Landeta, 2006). Así, solo aquellos atributos considerados como "muy importantes" por el panel superaban la criba (Varela-Ruiz et al., 2012).

Como criterio de estabilidad y de dispersión se empleó el Coeficiente de Variación de Pearson (CV), determinándose un umbral de $CV < 20\%$ para reflejar un consenso alto y dispersión interna baja que permitieran aceptar el atributo (Chang et al., 2010).

Los ítems con un CV situado en una franja moderada (entre el 21% y el 30%) fueron derivados a revisión cualitativa, mientras que aquellos con un $CV > 30\%$ fueron descartados por demasiada dispersión en las respuestas.

Respecto a la dinámica adoptada entre las rondas y el consecuente análisis de concordancia, el proceso se inició (Ronda 1) evaluando los 25 atributos originales (cuestionario de 25 ítems), con las pautas estadísticas mencionadas. Aquellos ítems que cumplieron estrictamente el criterio de estabilidad y tendencia central fueron aceptados como atributos clave. Aquellos que se situaron en la franja de consenso moderado o que recibieron comentarios cualitativos críticos fueron sometidos a un proceso de ajuste (unificación de ítems redundantes o modificaciones de la redacción). Este proceso de depuración dio lugar a una propuesta refinada de 10 atributos para ser evaluados de nuevo.

En la Ronda 2 se envió a los expertos un nuevo cuestionario de 10 ítems acompañado del feedback informativo de la primera ronda, que consistió en exponer las estadísticas descriptivas obtenidas junto a todos los comentarios realizados por los expertos.

Tras la recogida de datos de esta segunda ronda, se volvieron a aplicar los mismos criterios de corte ($\bar{X} \geq 4.2$, $Me \geq 4.0$ y $CV < 20\%$). Con el fin de aportar solidez a la estadística en el acuerdo entre expertos, se aplicó la prueba W de Kendall para evaluar la convergencia y la fuerza del acuerdo global del panel una vez que este ha sido sometido a la retroalimentación informativa de una ronda previa (Hsu & Sandford, 2007; Schmidt, 1997). Se consideró un valor de significación estructural ($p < .05$) junto con un nivel de concordancia moderado-alto ($W \geq .60$) como criterios cuantitativos de control. En consonancia con las directrices metodológicas para procesos de consenso (Diamond et al., 2014), y las aportaciones de Hasson et al. (2025) que defienden el Delphi como un método de investigación reflexiva y de naturaleza mixta, estos umbrales guiaron y enriquecieron el posterior análisis cualitativo de las aportaciones abiertas de los expertos. Las respuestas se organizaron en torno a categorías semánticas alineadas con las tres dimensiones que sustentan el patrón conceptual del estudio (instruccional, interaccional y autorregulatoria), lo que permitió reajustar, unificar o matizar algunos atributos.

Tabla 3. Resumen fase exploratoria.

Ronda	Atributos	Método	Análisis	Resultado
1	25	1-5 + abierta	$\bar{x}$, Me , DT, CV	10 atributos
2	10	1-5 informada	$\bar{x}$, Me , DT, CV, W Kendall	Consenso final seis atributos

Fase final: redacción del informe

Tras el procedimiento mixto de análisis de la segunda ronda del Delphi se obtuvo un consenso definitivo de seis atributos que definen un FFA en EF, y se procedió a redactar un informe final con intención de crear un patrón conceptual.

Resultados

Primera Ronda del Delphi

La Tabla 4 presenta los estadísticos descriptivos de los 25 atributos iniciales (AT) tras la primera ronda de consulta, junto a la decisión adoptada y el atributo resultante (At) para pasar a la segunda ronda.

Tabla 4. Resultados tras la primera ronda.

Atributos iniciales (AT). El feedback...	$\bar{x}$	Me	DT	CV	Decisión	Atributo para 2ª ronda (At)
AT1. Se adapta a las características personales del alumnado	4.47	4.5	0.11	2.46%	Unificar con AT5 y AT12	At1
AT2. Se realiza justo después de la tarea y en un momento de atención plena de los estudiantes	4.22	4.0	0.10	2.37%	Unificar con AT3, AT4 y AT7	At2
AT3. Aporta información sobre hechos observables, sin realizar juicios personales	4.25	4.0	0.22	5.18%	Unificar con AT2 AT3 y AT7	At2
AT4. Incluye información referida al objetivo a lograr	4.14	4.0	0.10	2.42%	Unificar con AT2 AT3 y AT7	At2
AT5. Aporta información sobre cómo aprende y progresa el estudiante	4.81	5.0	0.19	3.95%	Unificar con AT1 y AT12	At1
AT6. Aporta información sobre qué es lo siguiente a aprender	3.92	4.0	0.08	2.04%	Descartar	
AT7. Se realiza cuando el alumnado tiene tiempo y margen para actuar/rectificar	4.17	4.0	0.22	5.28%	Unificar con AT2 AT3 y AT4	At2
AT8. Centra la información en dos o tres puntos clave/diana del aprendizaje	4.06	4.0	0.13	3.20%	Descartar	
AT9. Si es escrito debe estar alineado con el oral y enfocarse en aspectos concretos	3.67	4.0	0.08	2.18%	Descartar	
AT10. Aporta información sólida, argumentada y basada en instrumentos de evaluación descriptivos	4.11	4.0	0.27	6.57%	Descartar	
AT11. Se realiza al final de la sesión para no interrumpir la tarea y en un momento de atención	3.08	3.0	0.52	16.8%	Descartar	
AT12. Se inicia indicando los aciertos/éxitos	4.28	4.0	0.10	2.34%	Unificar con AT1 y AT5	At1
AT13. Se expresa a través de un lenguaje positivo y crítico constructivo	4.58	5.0	0.10	2.18%	Mantener	At3
AT14. Menciona los errores como oportunidades de mejora	4.58	5.0	0.25	5.46%	Mantener	At4
AT15. Estimula el diálogo en un ambiente respetuoso	4.69	5.0	0.26	5.54%	Mantener	At5
AT16. Se incluye en una estrategia de coevaluación a través del diálogo ordenado	4.42	4.0	0.39	8.82%	Mantener	At6
AT17. Se incluye en una estrategia de autoevaluación con propuestas de mejora	4.53	5.0	0.45	9.93%	Mantener	At7
AT18. Aporta información desde hechos observables con un juicio personal	4.14	4.0	0.15	3.62%	Descartar	
AT19. Se formula con preguntas abiertas sobre futuras estrategias	3.39	3.0	0.08	2.36%	Descartar	
AT20. Se comunica de forma cordial y con escucha activa	4.78	5.0	0.03	0.63%	Mantener	At8
AT21. Se comunica desde el respeto a los sentimientos de los estudiantes	4.50	5.0	0.35	7.78%	Mantener	At9
AT22. Estimula la autoconfianza y el compromiso con el aprendizaje	4.69	5.0	0.14	2.99%	Unificar con AT23	At10
AT23. Utiliza el elogio centrado en el aprendizaje y el esfuerzo	4.25	4.0	0.06	1.41%	Unificar con AT22	At10
AT24. Menciona errores en primer lugar como oportunidades de mejora	2.83	3.0	0.05	1.77%	Descartar	
AT25. Combina el mensaje oral diario con la auto/coevaluación posterior	3.44	3.0	0.14	4.07%	Descartar	

Se dio mucha importancia a los comentarios personales que sugerían la unificación. Según Nasa et al. (2021), puntuaciones cuantitativas sub-umbral pueden no provocar la eliminación automática de un ítem si el análisis cualitativo detecta problemas de redundancia o solapamiento conceptual.

A partir de estas consideraciones, los ítems AT1, AT5 y AT12 se unificaron en el At1.

Los atributos AT4 ($\bar{x}=4.14$) y AT7 ($\bar{x}=4.17$), quedaron en un principio descartados por situarse ligeramente por debajo del punto de corte cuantitativo, pero los expertos señalaron la necesidad de sintetizar el mensaje. *El feedback depende mucho del tipo de tarea y es muy importante que la retroalimentación sea muy concreta* (Experto/a 3). El panel coincidió en la relevancia de vincular la información con el objetivo y con el tiempo de actuación del estudiante, y sugirió unificar los cuatro ítems. Por tanto, y de acuerdo a los principios de complementariedad y optimización metodológica (Nasa et al., 2021), se procedió a unificar el AT4 y AT7 junto al AT2 y AT3 para dar forma al At-2, relacionado con la vertiente más instruccional del mensaje.

Aquellos ítems cuya aportación cualitativa era escasa, poco relevante, o que no alcanzaron los umbrales estadísticos fueron descartados (AT6, AT8, AT9, AT10 y AT18, AT19, AT25), al igual que los ítems control utilizados como filtro de validez (AT11 y AT24).

Con respecto al ítem del feedback como estrategia de coevaluación (AT16, $\bar{x}=4.42$), el panel de expertos introdujo una advertencia a tener en cuenta en la etapa de educación secundaria: *"la coevaluación requiere de formación y madurez en el alumnado"* (Experto/a 6). En consecuencia, el ítem se mantuvo pero reformuló la semántica para la segunda ronda (At-6), haciendo referencia al feedback entre iguales.

En cuanto a la autoevaluación (AT17), los jueces ratificaron su solidez cuantitativa expresando que: *"es clave que el alumnado identifique sus áreas de mejora y proponga soluciones concretas"* (Experto/a 9), y dio lugar al At7 para la segunda ronda.

Asimismo, se destacaron de forma consistente la relevancia de la escucha activa, la estimulación de la autoconfianza y el compromiso con el aprendizaje (AT20 y AT21), dando lugar al At8 y At9 para la segunda ronda.

Por último, el ítem AT23 (*Utiliza el elogio...*), que había superado los puntos de corte cuantitativos ($\bar{X}=4.25$, $CV=1.41\%$), se integró con el AT22 (At-10), al entender el panel que el elogio se puede asociar con estimular la autoconfianza, siempre que se vincule con el aprendizaje, no con la persona.

Tras este primer proceso de depuración tanto cuantitativa como cualitativa, los 25 ítems iniciales dieron lugar a 10 atributos (At) para una nueva validación en la segunda ronda Delphi (Tabla 5).

Segunda Ronda del Delphi

En la segunda ronda se evaluó la estabilidad de los 10 atributos resultantes de la primera ronda. Los resultados cuantitativos, detallados en la Tabla 5, registraron medias muy elevadas ($\bar{X}\geq 4.26$) y coeficientes de variación muy bajos ($CV<11\%$), lo que constata la alta representatividad de las puntuaciones de las personas expertas.

Tabla 5. Resultados tras la segunda ronda.

Atributo evaluado (At). El feedback...	$\bar{X}$	DT	CV	Me	W	p_valor	Consenso Decisión	Atributo final (ATF)
At1. Se adapta a la actitud personal y capacidad de aprendizaje del alumnado.	4.55	0.16	3.51%	5	.239	.035	Aceptable. Mantener, Afinar.	ATF-1
At2. Aporta información sobre los aprendizajes conseguidos y el progreso del alumnado.	4.64	0.04	0.86%	5	.214	.050	Aceptable. Unificar con At3 y At4.	ATF-2
At3. Se expresa verbal o no verbalmente a través de mensajes claros y honestos, de forma positiva y asertiva.	4.33	0.37	8.54%	5	.765	<.001	Alto. Unificar con At2 y At4	ATF-2
At4. Analiza los errores de forma constructiva, valorándolos como oportunidades de mejora.	4.52	0.18	3.98%	5	.334	.009	Moderado. Unificar con At2 y At3.	ATF-2
At5. Estimula el diálogo, opiniones y sugerencias en un ambiente relajado, respetuoso y de mutuo apoyo.	4.67	0.38	8.13%	5	.481	.001	Moderado. Unificar con At8 y At9.	ATF-3
At6. Se incluye en estrategias de coevaluación a través del diálogo ordenado y respetuoso entre compañeros/as.	4.29	0.20	4.66%	5	.571	<.001	Alto. Mantener, Afinar.	ATF-4
At7. Se incluye en situaciones de autoevaluación en las que el alumnado identifica áreas de mejora y propone soluciones.	4.26	0.11	2.58%	5	.251	.030	Aceptable. Mantener, Afinar.	ATF-5
At8. Su mensaje se comunica de forma cordial y con actitud de escucha activa.	4.79	0.34	7.09%	5	.286	.018	Aceptable. Unificar con At9 y At5.	ATF-3
At9. Su mensaje respeta en todo momento los sentimientos y pensamientos del alumnado al comunicar el mensaje.	4.43	0.11	2.48%	5	.125	.174	*Bajo, Rescate. Unificar con At4, At8 y At5.	ATF-3
At10. Estimula la mejora personal, la autoconfianza y el compromiso con el aprendizaje	4.33	0.44	10.16%	5	.643	<.001	Alto. Mantener. Afinar.	ATF-6

Nota. * Para la unificación del At9 se tuvieron en cuenta los datos cuantitativos y la relevancia de las discusiones iterativas y comentarios cualitativos del panel de expertos, de acuerdo a la naturaleza mixta del Delphi (Diamond et al., 2014; Nasa et al., 2021)

Para determinar la concordancia interjueces, el coeficiente W de Kendall se interpretó bajo los criterios de García y Suárez (2013) adaptados a estudios exploratorios. Los datos revelaron un consenso alto o moderado y estadísticamente significativo en los atributos At-3, At-4, At-5, At-6 y At-10. Por su parte, los atributos At-1, At-2 y At-7 exhibieron valores de W más discretos pero válidos, respaldados por una dispersión interna mínima ($CV < 4\%$) y significación estadística ($p \leq .05$).

El atributo At-9 ("el mensaje respeta los sentimientos y pensamientos del alumnado"), obtuvo un coeficiente $W = .125$ no significativo ($p = .174$). Al contrastar estos datos estadísticos con las observaciones cualitativas, se constató que los expertos consideraban el contenido del At9 esencial, y relacionado con los atributos At-5 y At-8. Los expertos señalaron que la escucha activa implicaba el respeto a sentimientos y pensamientos, en un clima favorable de clase, y sugerían vincular estos ítems. Siguiendo los criterios de Diamond et al. (2014), en procesos de validación donde el constructo presenta implicación afectiva o social, la valoración cualitativa del panel de expertos puede considerarse prioritaria para asegurar una robustez teórica final. En la misma línea, Nasa et al. (2021). afirman que la coexistencia de un bajo grado de concordancia experta (W) con datos cuantitativos favorables, puede ser un indicador de redundancia o solapamiento conceptual entre ítems. Así, el atributo At-9 se unificó con el At-5 y At-8, todos ellos relacionados con la vertiente más afectiva del feedback. La estabilidad de las métricas cuantitativas y la firmeza de los comentarios, hicieron innecesaria una tercera ronda. El proceso dio como resultado una estructura final compuesto por seis atributos propios de un FFA.

Informe final

Tras la segunda ronda Delphi y el análisis integrado cuantitativo-cualitativo, se elaboró el informe final para definir seis atributos clave, fruto de mantener o unificar los 10 atributos previos.

El FFA en EF:

1. se adapta a los conocimientos, habilidades y actitudes del alumnado (At1)
2. aporta información al alumnado sobre los aprendizajes adquiridos, los aciertos y los aspectos a mejorar con un lenguaje respetuoso (At2, At3 y At4);
3. fomenta el diálogo, la escucha activa, el apoyo mutuo, la cordialidad y el respeto (At5, At8 y At9);
4. se integra en estrategias de coevaluación entre iguales (At6);
5. se incluye en acciones de autoevaluación (At7); y
6. se orienta hacia la mejora personal, promoviendo la autoconfianza y el compromiso con el aprendizaje (At10).

En conjunto, estos seis atributos delimitan un patrón conceptual de este FFA, que puede organizarse en tres dimensiones instruccional, interaccional y autorregulatoria (Tabla 6). Su definición consensuada mediante juicio experto, proporciona una base para diseñar y validar instrumentos de utilidad para los docentes de EF, y siempre en contextos de evaluación formativa.

Tabla 6. Atributos finales (ATF) y dimensiones de un FFA en EF.

Atributos clave del FFA	El feedback...	Dimensión predominante
ATF-1. Se adapta a los conocimientos, habilidades y actitudes del alumnado.		instruccional
ATF-2. Aporta información al alumnado sobre los aprendizajes adquiridos, los aciertos y los aspectos a mejorar.		instruccional
ATF-3. Fomenta el diálogo, la escucha activa, el apoyo mutuo, la cordialidad y el respeto.		interaccional
ATF-4 Se integra en estrategias de coevaluación entre iguales.		autorregulatoria
ATF-5 Se incluye en acciones de autoevaluación.		autorregulatoria
ATF-6 Se orienta hacia la mejora personal, promoviendo la autoconfianza y el compromiso con el aprendizaje.		interaccional

Discusión

El principal hallazgo de este estudio es la identificación mediante el método Delphi de una estructura tridimensional del FFA en EF compuesta por las dimensiones instruccional, interaccional y autorregulatoria. Se aporta así relevancia y estructura a la vertiente afectiva del feedback. Aunque estudios previos destacan la importancia de la calidad del mensaje y de los procesos de autorregulación asociados al feedback (Hattie y Timperley, 2007; Carless y Boud, 2018), los resultados obtenidos sugieren que la dimensión afectiva no constituye un elemento accesorio, sino un componente transversal que condiciona la eficacia de la retroalimentación en contextos de EF.

La dimensión instruccional queda configurada por el ATF-1 (adaptación al alumnado) y el ATF-2 (información sobre aciertos, errores y aspectos a mejorar). Las valoraciones de los expertos indicaron que un mensaje genérico o descontextualizado anula el sentido del feedback, lo que se alinea con el meta-análisis de Hattie y Timperley (2007) y el de Wisniewski et al. (2020). Estos autores destacan que el impacto de la retroalimentación depende de cómo el alumnado procesa la información recibida, la cual debe ser útil, precisa y adaptada.

Sin embargo, la realidad de las clases de EF en secundaria, caracterizada por altas ratios y falta de tiempo, da lugar a que el feedback personalizado sea un reto difícil para el profesorado. Los grupos rondan los 25 estudiantes y la duración de las sesiones es de 55 minutos, tiempo de vestuario y traslados incluido. Esto obliga a los docentes a dar comentarios grupales y generales.

En este sentido Pérez-Pueyo et al. (2021), apuntan que el feedback dentro de la evaluación formativa, debería plantearse como algo ágil y operativo, que facilitase al docente su aplicación, dentro del contexto a veces complicado de sus clases.

La dimensión interaccional se relaciona con el ATF-3 (diálogo, escucha activa y respeto) y el ATF-6 (orientación hacia la mejora personal y autoconfianza). El clima de respeto y diálogo horizontal que promueve el ATF-3 resulta clave para fortalecer los vínculos sociales en el aula, mientras que el foco en

la superación personal del ATF-6 afianza la percepción de competencia y la seguridad del estudiante. A partir de este feedback dialogado se favorece que el alumnado actúe como protagonista y agente de su propio aprendizaje. Al atender de manera integrada estas necesidades afectivas, la dimensión interaccional actúa como un motor indispensable para potenciar la motivación intrínseca y el compromiso con la práctica. En esta línea, Hinojosa-Torres et al. (2025) apuntan que la retroalimentación formativa es crucial para promover el aprendizaje autodirigido.

Esta dimensión interaccional cobra especial relevancia en la realidad de las clases de EF, diferente del resto de asignaturas teóricas. Se imparte en espacios abiertos, las habilidades motrices son visibles para todos, y el cuerpo, como ya se ha mencionado anteriormente, está expuesto de forma constante, lo que puede generar vulnerabilidad emocional e inseguridad en el alumnado adolescente. Por ello el ATF-3 y el ATF-6 pueden funcionar como "colchón emocional" que minimice los riesgos de dicha exposición, transformando la evaluación en un acto de cuidado y responsabilidad afectiva hacia una práctica motriz verdaderamente significativa (Beni et al., 2017).

Por su parte, la dimensión autorregulatoria queda englobada en el ATF-4 (coevaluación entre iguales) y el ATF-5 (autoevaluación), otorgando un papel activo al estudiante. Esto se alinea con el enfoque del *feedback literacy* (alfabetización del feedback) de Carless (2025), y con las recientes aportaciones de Yan et al. (2026) sobre los roles del alumnado como agente y receptor del feedback en los procesos de evaluación entre iguales.

Sin embargo, en la práctica de la EF en secundaria se percibe la dificultad del alumnado para desarrollar estos procesos, evidenciándose una falta de *feedback literacy* (Carless & Boud, 2018). Estos autores resaltan que los estudiantes necesitan desarrollar la capacidad de comprender la información y gestionar sus emociones para tomar decisiones de mejora. En ocasiones, el alumnado adolescente muestra resistencia o falta de sentido crítico para evaluarse. Yan et al. (2026) enfatizan que las prácticas de coevaluación (ATF-4) requieren un entrenamiento previo y explícito enfocado en la seguridad y el clima de aula para evitar conflictos interpersonales y gestionar las barreras emocionales del alumnado.

En relación con los atributos que presentaron consenso bajo o moderado tanto en la primera ronda (AT4 y AT7) como en la segunda (At9) se tuvo en cuenta la naturaleza mixta e iterativa del método Delphi donde el análisis cualitativo permite refinar y dar cohesión al constructo (Landeta, 2006; Diamond et al., 2014). Así, ítems de la primera ronda como el antiguo AT4 (información referida al objetivo a lograr) y el AT7 (margen temporal para rectificar), fueron unificados de acuerdo a la opinión de los expertos e integrados por su complementariedad en el actual ATF-2, dentro de la dimensión instruccional. Del mismo modo, el At9 de la segunda ronda (respeto a los sentimientos y pensamientos del alumnado al comunicar el mensaje) se incluyó para dar lugar al ATF-3 dentro de la dimensión interaccional.

Por otra parte, el patrón de tres dimensiones del FFA se puede alinear con los cinco componentes del modelo MISCA (Panadero y Lipnevich, 2022): Mensaje, Implementación, Estudiante, Contexto y Agente. Los seis atributos conceptualizan la vertiente afectiva como transversal a esos cinco componentes. En primer lugar, impactan en el *Mensaje*, que ha de adaptarse y aportar información con un lenguaje respetuoso (ATF-1 y ATF-2). Además, hacen referencia a los roles del *Agente* (docente) y del *Estudiante* otorgándoles un papel activo y dialogante a través de la coevaluación y autoevaluación (ATF-4 y ATF-5). Por último, orientan la *Implementación* hacia la mejora personal (ATF-6); y actúan directamente sobre el *Contexto* de la EF, transformando el espacio motriz abierto en un entorno seguro a través del diálogo y la escucha activa (ATF-3).

A partir de estos hallazgos, y teniendo en cuenta los límites metodológicos del estudio, se sugiere que el FFA en sus tres dimensiones aproxima las vertientes instructiva y afectiva del feedback como recurso clave de la evaluación formativa. Un estudio Delphi identifica, depura y delimita los atributos de un constructo a partir de un alto grado de consenso interjueces, pero no constituye una validación empírica de su eficacia pedagógica, ni de su impacto directo sobre el desempeño motriz o el bienestar psicológico del alumnado.

La constatación del acuerdo experto justifica la validez de contenido y la pertinencia y relevancia del patrón de seis atributos, pero su efectividad en las clases de EF y la confirmación de las tres dimensiones del FFA, quedan supeditadas a futuras investigaciones en contextos reales.

Que los expertos debatieran de forma activa sobre la parte emocional demuestra que el componente

afectivo es un aspecto clave y pendiente de sistematizar. Así, la integración de los comentarios cualitativos dota a los seis atributos finales de una valiosa transferencia práctica y validez ecológica, acordes a la realidad del profesorado de EF en secundaria.

A partir del análisis de las justificaciones y aportaciones cualitativas de este estudio, se presentan tres necesidades a tener en cuenta las clases de EF:

En primer lugar, los expertos discutieron la viabilidad del ATF-1 y ATF-2, en relación con la dificultad de personalizar el feedback ante las altas ratios y tiempos reducidos de las clases de EF. Por ello se presenta la necesidad de diseñar recursos concretos, operativos y ágiles para individualizar el feedback sin perder la vertiente humana y afectiva, (tal vez plantillas y/o rúbricas), y ofrecer información eficaz sin restar tiempo de práctica motriz.

En segundo lugar, los expertos constataron la importancia de perfilar la dimensión autorregulatoria con el ATF-4 y ATF-5, pero destacaron la falta de conocimientos y hábito en el alumnado adolescente a la hora de participar en acciones de auto y coevaluación. Se presenta así la necesidad de alfabetizar y entrenar al alumnado en torno al feedback y las acciones de auto y coevaluación a partir de programas dirigidos a valorar el propio aprendizaje y el de sus iguales desde el respeto, el cuidado y la seguridad emocional.

En tercer lugar, el fuerte consenso alcanzado por los jueces en torno al diálogo la escucha activa, el respeto y el fomento de la confianza como aspectos esenciales del FFA, argumentaron el ATF-3 y ATF-6. Así, se sugiere la necesidad de formación del profesorado en competencias comunicativas y gestión emocional de cara a garantizar que el feedback se transmita desde una vertiente afectiva, empática y respetuosa.

Como principal limitación se menciona el sesgo de homogeneidad del panel de expertos, dado que los 14 jueces pertenecen a la Red Nacional de Evaluación Formativa y Compartida. Si bien esta adscripción fue una decisión estratégica que garantizó un marco conceptual afín, optimizando la validez de contenido sobre un constructo tan específico, introduce un cierto sesgo que pudo atenuar la aparición de visiones alternativas de otros docentes en activo con enfoques metodológicos diferentes. Por ello, y como prospectiva, las líneas futuras de investigación en torno al diseño de instrumentos a partir de los seis atributos deberán tener en cuenta muestras de profesorado más amplias y diversas.

A partir de aquí, el verdadero reto puede ser no tanto comprender el feedback a nivel teórico, sino conocer cómo lo percibe el alumnado adolescente, para poder sistematizarlo como recurso didáctico. La delimitación de estos seis atributos en tres dimensiones como patrón conceptual ofrece conocimientos y recursos para el profesorado. Se genera así un marco de actuación que permite continuar la investigación hacia el diseño y validación de instrumentos de utilidad para los docentes a la hora de poner en práctica un FFA.

Conclusiones

A través del método Delphi se ha podido identificar y consensuar un patrón conceptual de seis atributos clave de un FFA en EF, estructurado en tres dimensiones: instruccional, interaccional y autorregulatoria. Este patrón conceptual resuelve en cierta medida la separación entre la vertiente instructiva y la afectiva del mensaje, quedando unificados en una única acción comunicativa que es el FFA.

En EF, en secundaria (12-16 años), y en un entorno de exposición pública del cuerpo y desarrollo de las habilidades motrices, este FFA puede favorecer la creación de un clima emocional seguro y saludable, que favorezca un aprendizaje con sentido.

Los atributos consensuados proporcionan una base teórica estable y operativa de cara a que en el futuro, y dando continuidad a este estudio, se puedan diseñar y validar instrumentos para mejorar la práctica docente.

Asimismo, se propone explorar la transferencia del instrumento hacia otras materias y etapas educativas, así como hacia el ámbito de la iniciación deportiva, de cara a lograr una validez ecológica y perspectivas diferentes que completen este patrón conceptual.

Agradecimientos

Las autoras agradecen la colaboración de las 14 personas expertas que participaron en el estudio Delphi, así como a los cinco docentes universitarios expertos en Didáctica, cuya disposición y aportaciones fueron fundamentales para realizar la investigación.

Referencias

- Adarkwah, M. A. (2021). The power of assessment feedback in teaching and learning: a narrative review and synthesis of the literature. *SN Social Sciences*, 1(3), 75.
<https://doi.org/10.1007/s43545-021-00086-w>
- Almenara, J. C., & Osuna, J. M. B. (2013). La utilización del juicio de experto para la evaluación de TIC: el coeficiente de competencia experta. *Bordón: Revista de pedagogía*, 65(2), 25-38.
<https://doi.org/10.13042/brp.2013.65202>
- Almonacid-Fierro, A., Aguilar-Valdés, M., de Carvalho, R. S., & Merellano-Navarro, E. (2024). Emocionalidad en el aula de primaria de Educación Física postpandemia (Emotionality in the post-pandemic primary physical education classroom). *Retos*, 53, 608-617.
<https://doi.org/10.47197/retos.v53.100235>
- Beni, S., Fletcher, T., & Chróinín, D. N. (2017). Meaningful experiences in physical education: A systematic review of literature. *Quest*, 69(3), 291-312.
<https://doi.org/10.1080/00336297.2016.1224192>
- Berlanga, M., & Juárez-Hernández, L. G. (2020). Evaluation paradigms: From the traditional to the socio-formative. *Diálogos sobre educación. Temas actuales en investigación educativa*, 11(21). <https://doi.org/10.32870/dse.v0i21.646>
- Brookhart, S. M. (2017). *How to give effective feedback to your students* (2nd ed.). ASCD.
- Carless, D. (2012). Trust and its role in facilitating dialogic feedback. In D. Boud & E. Molloy (Eds.), *Feedback in higher and professional education: Understanding it and doing it well* (pp. 90-103). Routledge. <https://doi.org/10.4324/9780203074336>
- Carless, D. (2013). Sustainable feedback and the development of student self-evaluative capacities. In S. Merry, M. Price, D. Carless, & M. Taras (Eds.), *Reconceptualising feedback in higher education: Developing dialogue with students* (pp. 113-122). Routledge.
<https://doi.org/10.4324/9780203081341>
- Carless, D. (2016). Feedback as dialogue. *Encyclopedia of Educational Philosophy and Theory*, 1-6.
https://doi.org/10.1007/978-981-287-532-7_389-1
- Carless, D., & Boud, D. (2018). The development of student feedback literacy: Enabling uptake of feedback. *Assessment & Evaluation in Higher Education*, 43(8), 1315-1325.
<https://doi.org/10.1080/02602938.2018.1463354>
- Carless, D. (2025). Feedback literacy concepts and practices: Toward academic feedback literacy. *Change: The Magazine of Higher Learning*, 57(5), 5-11.
<https://doi.org/10.1080/00091383.2025.2539038>
- Chang, A. M., Gardner, G. E., Duffield, C., & Ramis, M. A. (2010). A Delphi study to validate an advanced practice nursing role evaluation tool. *Journal of Advanced Nursing*, 66(10), 2320-2330.
<https://doi.org/10.1111/j.1365-2648.2010.05367.x>
- Diamond, I. R., Grant, R. C., Feldman, B. M., Pencharz, P. B., Ling, S. C., Moore, A. M., & Wales, P. W. (2014). Defining consensus: A systematic review recommends methodologic criteria for reporting of Delphi studies. *Journal of Clinical Epidemiology*, 67(4), 401-409.
<https://doi.org/10.1016/j.jclinepi.2013.12.002>
- García, M., & Suárez, M. (2013). El método Delphi para la consulta a expertos en la investigación científica. *Revista Cubana de Salud Pública*, 39(2), 253-267.
http://scielo.sld.cu/scielo.php?script=sci_arttext&pid=S0864-34662013000200007
- Hattie, J., & Timperley, H. (2007). The power of feedback. *Review of Educational Research*, 77(1), 81-112. <https://doi.org/10.3102/003465430298487>

- Han, Y., Syed Ali, S. K. B., & Ji, L. (2022). Feedback for Promoting Motor Skill Learning in Physical Education: A Trial Sequential Meta-Analysis. *International Journal of Environmental Research and Public Health*, 19(22), 15361. <https://doi.org/10.3390/ijerph192215361>
- Hasson, F., Keeney, S., & McKenna, H. (2025). Revisiting the Delphi technique-Research thinking and practice: A discussion paper. *International Journal of Nursing Studies*, 168, 105119. <https://doi.org/10.1016/j.ijnurstu.2025.105119>
- Hinojosa-Torres, C., Zavala-Crichton, J. P., Serey-Galindo, F., Hurtado-Guerrero, M., Gajardo-Vergara, X., Álvarez-Valenzuela, S., Cantallops-Jerez, F., Espoz-Lazo, S., & Yáñez-Sepúlveda, R. (2025). Retroalimentación formativa: percepción docente sobre su implementación y resultados en el proceso de enseñanza y aprendizaje en educación física. *Retos*, 62, 872-882. <https://doi.org/10.47197/retos.v62.110706>
- Hsu, C.-C., & Sandford, B. A. (2007). The Delphi technique: Making sense of consensus. *Practical Assessment, Research, and Evaluation*, 12(1), 10. <https://doi.org/10.7275/pdz9-th90>
- Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics*, 33(1), 159–174. <https://doi.org/10.2307/2529310>
- Landeta, J. (2006). Current validity of the Delphi method in social sciences. *Technological Forecasting and Social Change*, 73(5), 467–482. <https://doi.org/10.1016/j.techfore.2005.09.002>
- León-Díaz, Ó., Martínez-Muñoz, L. F., & Santos-Pastor, M. L. (2023). Metodologías activas en la Educación Física. Una mirada desde la realidad práctica. *Retos*, 48, 647–656. <https://doi.org/10.47197/retos.v48.96661>
- Lipnevich, A. A., & Panadero, E. (2021). A review of feedback models and typologies: Towards an integrative model of feedback elements. *Educational Research Review*, 35, 100416. <https://doi.org/10.1016/j.edurev.2021.100416>
- Mandouit, L., & Hattie, J. (2023). Revisiting “The Power of Feedback” from the perspective of the learner. *Learning and Instruction*, 84, 101718. <https://doi.org/10.1016/j.learninstruc.2022.101718>
- Marín-Almenara, J., & Osuna-Rodríguez, M. (2013). La competencia experta en los procesos Delphi: cálculo y significación del coeficiente K. *Revista de Investigación Educativa*, 31(2), 345–361. <https://revistas.um.es/rie/article/view/155701>
- Nasa, P., Azoulay, E., Khanna, A. K., Jain, R., Gupta, S., Javeri, Y., ... & Myatra, S. N. (2021). Expert consensus statements for the management of COVID-19-related acute respiratory failure using a Delphi method. *Critical Care*, 25(1), 106. <https://doi.org/10.1186/s13054-021-03491-y>
- Panadero, E., Broadbent, J., Boud, D., & Lodge, J. M. (2019). Using formative assessment to influence self- and co-regulated learning: The role of evaluative judgement. *European Journal of Psychology of Education*, 34(3), 535–557. <https://doi.org/10.1007/s10212-018-0407-8>
- Panadero, E., & Lipnevich, A. A. (2022). A review of feedback models and typologies: Towards an integrative model of feedback elements. *Educational Research Review*, 35, Article 100416. <https://doi.org/10.1016/j.edurev.2021.100416>
- Pérez-Pueyo, Á., Hortigüela-Alcalá, D., Gutiérrez-García, C., & Barba-Martín, R. A. (2024). *Evaluación formativa y compartida en Educación Física: fundamentos y experiencias prácticas en todas las etapas educativas*. Servicio de Publicaciones de la Universidad de León. ISBN: 978-84-19682-79-6.
- Powell, C. (2003). The Delphi technique: Myths and realities. *Journal of Advanced Nursing*, 41(4), 376–382. <https://doi.org/10.1046/j.1365-2648.2003.02537.x>
- Sanmartí, N. (2010). *Evaluar para aprender: 10 ideas clave*. Graó.
- Schmidt, R. C. (1997). Managing Delphi surveys using nonparametric statistical techniques. *Decision Sciences*, 28(3), 763–774. <https://doi.org/10.1111/j.1540-5915.1997.tb01330.x>
- Skulmoski, G. J., Hartman, F. T., & Krahn, J. (2007). The Delphi method for graduate research. *Journal of Information Technology Education: Research*, 6(1), 1–21. <https://doi.org/10.28945/199>
- Tobón, S. (2013). *Evaluación de las competencias en la educación básica*. Santillana.
- Varela-Ruiz, M., Díaz-Bravo, L., & García-Durán, R. (2012). Descripción y usos de la técnica Delphi en investigaciones del área de la salud. *Investigación en Educación Médica*, 1(2), 90–95. http://www.scielo.org.mx/scielo.php?script=sci_arttext&pid=S2007-50572012000200007
- Wisniewski, B., Zierer, K., & Hattie, J. (2020). The power of feedback revisited: A meta-analysis of educational feedback research. *Frontiers in Psychology*, 10, Article 3087. <https://doi.org/10.3389/fpsyg.2019.03087>

- Yan, Z., Lao, H., Panadero, E., & Fernández-Castilla, B. (2026). A meta-analytical review of the effects of self- assessment and peer-assessment interventions on self-regulated learning strategies and affective outcomes. *Assessment in Education: Principles, Policy & Practice*, 1-39. <https://doi.org/10.1080/0969594X.2026.2681630>
- Zhou, Y., Shao, W. D., & Wang, L. (2021). Effects of Feedback on Students' Motor Skill Learning in Physical Education: A Systematic Review. *International Journal of Environmental Research and Public Health*, 18(12), 6281. <https://doi.org/10.3390/ijerph18126281>

Datos de las autoras y traductora:

Ana Casares Polo	anacasarespolo71@gmail.com	Autora
Alicia M ^a Alonso Martínez	aliciamaria.alonso@unavarra.es	Autora
M ^a Rosario Romero Martín	rromero@unizar.es	Autora
Nerea Cristina Estrada Marcén	nereaes@unizar.es	Autora
Sonia Asún Dieste	sonasun@unizar.es	Autora
Loreto Alonso Martínez	loretoalonso@gmail.com	Traductora