

UNIVERSITAT DE BARCELONA

TRABAJO FINAL DE MÁSTER

La evaluación en Educación Física en contextos predominantemente calificadores:
percepción del alumnado y bases para una propuesta progresiva de implementación de
evaluación formativa

Pedro José Maciá López

Tutor: Ramón Hidalgo Raya

Máster Formación del Profesorado de Secundaria: Especialidad Educación Física

Curso 2025/2026

Índice

Resumen y Palabras Clave.....	4
1. Introducción.....	7
1.1. Tema de Estudio.....	7
1.2. Justificación y Motivaciones Personales.....	7
1.3. Interés Científico, Social y Educativo.....	8
1.4. Estructura del Trabajo de Fin de Máster.....	8
2. Marco Teórico.....	9
2.1. Evaluación en el Contexto Educativo Actual.....	9
2.2. Evaluación Formativa: Fundamentos Teóricos y Conceptuales.....	13
2.2.1. Conceptualización.....	13
2.2.2. Impacto en el Aprendizaje.....	15
2.3. Evaluación Formativa en Educación Física.....	16
2.3.1. Evaluación Formativa y Modelos Pedagógicos en Educación Física.....	16
2.3.2. Estrategias para una Implementación Efectiva.....	20
2.3.3. Barreras y Condicionantes para la Implementación.....	23
3. Diseño Teórico del Estudio.....	24
3.1. Análisis del Contexto Escolar.....	24
3.2. Pregunta de Investigación.....	25
3.3. Hipótesis.....	25
3.4. Paradigma desde el que se Pretende Abordar.....	25
3.5. Objetivos.....	27
3.5.1. Objetivo General.....	27
3.5.2. Objetivos Específicos.....	27
4. Metodología de la Investigación.....	27
4.1. Cuadro del Modelo de Análisis.....	28
4.2. Fases de la Investigación.....	30
4.3. Instrumento.....	33
4.4. Muestra.....	38
5. Aplicación.....	39
6. Análisis de los Datos.....	40
7. Resultados.....	42
7.1. Dimensión 1. Concepciones sobre la Evaluación.....	42
7.2. Dimensión 2. Agentes Implicados en la Evaluación.....	44
7.3. Dimensión 3. Instrumentos y Momentos de Evaluación.....	47
7.4. Dimensión 4. Contenidos.....	49
8. Discusión.....	52
9. Propuesta de Implementación.....	58

10. Conclusiones.....	61
11. Limitaciones y Futuras Líneas de Investigación.....	62
12. Referencias.....	64
13. Anexos.....	74

Resumen y Palabras Clave

Resumen. Este Trabajo Final de Máster analiza la percepción del alumnado de secundaria y bachillerato sobre la evaluación en Educación Física en un contexto predominantemente calificador, comparándola con una investigación previa realizada en un centro con evaluación formativa habitual. Parte de la brecha entre el marco normativo competencial vigente, que apuesta por una evaluación continua, formativa e integradora, y las prácticas reales, reducidas a la certificación del rendimiento. Mediante un diseño descriptivo, comparativo e interpretativo no experimental, se administró a 165 estudiantes de un centro de Barcelona el cuestionario validado de Palacio et al. (2021), estructurado en cuatro dimensiones: concepciones sobre la evaluación, agentes implicados, instrumentos y momentos, y contenidos evaluados. Los datos se trataron estadísticamente y se triangularon con el estudio precedente ($n=137$). Los resultados muestran convergencia entre contextos en la centralidad de la calificación y la escasa participación del alumnado como evaluador, especialmente del profesorado. En cambio, hay diferencias marcadas en los momentos e instrumentos, con peso desproporcionado de la evaluación sumativa final, y en los contenidos, con mayor relevancia de aspectos actitudinales no académicos como la indumentaria. A partir de ello se diseña una propuesta de implementación progresiva, organizada en fases secuenciales que parten de explicitar criterios y avanzan hacia la autoevaluación, coevaluación y un sistema evaluativo compartido y corresponsable, adaptado al contexto.

Palabras clave: Evaluación formativa, Educación Física, modelos pedagógicos, secundaria, percepción, alumnado.

Resum. Aquest Treball Final de Màster analitza la percepció de l'alumnat de secundària i batxillerat sobre l'avaluació en Educació Física en un context predominantment qualificador, comparant-la amb una recerca prèvia realitzada en un centre amb avaluació formativa habitual. Part de la bretxa entre el marc normatiu competencial vigent, que aposta per una avaluació contínua, formativa i integradora, i les pràctiques reals, reduïdes a la certificació del rendiment. Mitjançant un disseny descriptiu, comparatiu i interpretatiu no experimental, es va administrar a 165 estudiants d'un centre de Barcelona el qüestionari validat de Palacio et al. (2021), estructurat en quatre dimensions: concepcions sobre l'avaluació, agents implicats, instruments i moments, i continguts avaluats. Les dades es van tractar estadísticament i es triangularon amb l'estudi precedent ($n=137$). Els resultats mostren convergència entre contextos en la centralitat de la qualificació i l'escassa participació de l'alumnat com a avaluador, especialment del professorat. En canvi, hi ha diferències marcades en els moments i instruments, amb pes desproporcionat de l'avaluació sumativa final, i en els continguts, amb major rellevància d'aspectes actitudinals no acadèmics com la indumentària. A partir d'això es dissenya una proposta d'implementació progressiva, organitzada en fases seqüencials que parteixen d'explicitar criteris i avancen cap a l'autoavaluació, coavaluació i un sistema avaluatiu compartit i corresponsable, adaptat al context.

Paraules clau: Evaluació formativa, Educació Física, models pedagògics, secundària, percepció, alumnat.

Abstract. This Master's Final Project analyzes the perception of secondary and upper secondary school students about the evaluation in Physical Education in a predominantly qualifying context, comparing it with a previous research carried out in a center with usual formative evaluation. Part of the gap between the current regulatory competence framework, which is committed to continuous, formative and integrative evaluation, and real practices, reduced to performance certification. Through a non-experimental descriptive, comparative and interpretative design, the validated de Palacio et al. (2021) questionnaire was administered to 165 students from a center in Barcelona, structured in four dimensions: conceptions about evaluation, agents involved, instruments and moments, and evaluated contents. The data were treated statistically and triangularly with the previous study (n=137). The results show convergence between contexts in the centrality of the qualification and the low participation of students as evaluator, especially teachers. On the other hand, there are marked differences in the moments and instruments, with disproportionate weight of the final summative evaluation, and in the contents, with greater relevance of non-academic attitude aspects such as clothing. From this, a proposal for progressive implementation is designed, organized in sequential phases that start from explaining criteria and move towards self-evaluation, co-evaluation and a shared and co-responsible evaluation system, adapted to the context.

Keywords. Formative evaluation, Physical Education, pedagogical models, secondary, perception, students.

1. Introducción

1.1. Tema de Estudio

La evaluación constituye uno de los elementos más relevantes y, al mismo tiempo, más complejos del proceso de enseñanza-aprendizaje. Su concepción, diseño e implementación determinan en gran medida la calidad de la experiencia educativa del alumnado y condicionan el tipo de aprendizaje que este desarrolla a lo largo de su escolarización (Black & Wiliam, 2018). En el ámbito de la Educación Física (EF), esta cuestión adquiere una dimensión específica, ya que la tensión entre evaluar para aprender y evaluar para calificar se manifiesta con especial intensidad en esta área, donde persiste una brecha significativa entre lo que el marco normativo vigente prescribe y lo que realmente ocurre en muchos centros educativos.

El presente Trabajo final de Máster (TFM) se enmarca en esta problemática y la continuidad de una investigación previa realizada en el Máster de Actividad Física y Educación de la Universidad de Barcelona, titulada “Evaluación formativa en Educación Física: un estudio sobre la percepción del alumnado y su influencia en el aprendizaje”. En este trabajo se analizó, mediante un cuestionario validado (Palacio et al., 2021) la percepción del alumnado en un centro donde la evaluación formativa se aplica de manera habitual. En las conclusiones de dicho trabajo se identificó como una futura línea de investigación la realización de estudios comparativos con contextos educativos de características evaluativas diferentes. De modo que, este trabajo recoge precisamente ese reto. Su objeto de estudio es la percepción del alumnado de Secundaria y Bachillerato sobre la evaluación en EF en un contexto predominantemente calificador y su contraste con los resultados obtenidos en el estudio anterior. A partir de este análisis comparativo, el trabajo culmina con el diseño de una propuesta progresiva de implementación de la evaluación formativa adaptada al contexto calificador estudiado, fundamentada en los hallazgos empíricos y en la literatura científica especializada.

1.2. Justificación y Motivaciones Personales

El interés por esta temática tiene un origen experiencial concreto. Durante las prácticas del Grado en Ciencias de la Actividad Física y el Deporte (CAFD) tuve mi primer contacto directo con la realidad evaluativa de la EF en un contexto predominantemente calificador, donde la evaluación se reducía fundamentalmente a la medición del rendimiento físico y a la asignación de notas. Esa experiencia me generó una inquietud sobre las posibilidades de transformar las

prácticas evaluativas en EF y sobre el efecto que dicha transformación podría tener sobre la experiencia del alumnado.

Durante el Máster de Actividad Física y Educación, realicé las prácticas en un centro con una cultura evaluativa radicalmente distinta, en el que la evaluación formativa se aplicaba de manera habitual y sistemática. Poder estudiar empíricamente la percepción del alumnado en ese contexto me permitió confirmar el valor pedagógico de este enfoque, pero también identificar sus límites y resistencias incluso en entornos favorables.

Para completar y dotar de mayor alcance esta investigación, consideré importante analizar también la percepción del alumnado en un contexto evaluativo radicalmente diferente. Solo a través de esa comparación sería posible determinar en qué medida las percepciones detectadas en el contexto formativo son específicas de ese entorno, o por el contrario, qué dimensiones de la relación del alumnado con la evaluación responden a patrones más generales.

1.3. Interés Científico, Social y Educativo

Desde el punto de vista científico, este trabajo ofrece una aportación metodológica relevante: el empleo del mismo instrumento validado en dos contextos evaluativos contrastados permite una comparación rigurosa y sistemática de la percepción del alumnado, dotando al estudio de una solidez que va más allá de la descripción de un caso único.

Desde el punto de vista social y educativo, la relevancia del estudio reside en la necesidad de disponer de propuestas concretas, graduadas y fundamentadas para la transformación de contextos evaluativos calificadores. La propuesta de implementación progresiva que se elabora a partir de los hallazgos empíricos comparados aspira a ser una herramienta útil para el profesorado de EF que trabaja en entornos similares al estudiado.

1.4. Estructura del Trabajo de Fin de Máster

El presente trabajo se organiza siguiendo la estructura propia de un proyecto de investigación educativa empírica. Se abre con un marco teórico que proporciona el sustento conceptual y científico necesario para comprender el problema de investigación en toda su complejidad. A partir de ahí, se realiza el diseño teórico del estudio donde se concreta los elementos que orientan la investigación (contexto analizado, participantes, pregunta de investigación, la hipótesis, el paradigma desde el que se pretende abordar y los objetivos). Estos

elementos configuran la base sobre la que se construye la metodología de la investigación, donde se describen y justifican las decisiones tomadas en cuanto al diseño del estudio, las fases del proceso investigador, los instrumentos y técnicas de recogida de datos y la muestra participante.

Una vez presentado el marco metodológico, el apartado de aplicación describe cómo se llevó a cabo la recogida de datos, dando paso al análisis de los datos, donde se explica el tratamiento aplicado a la información obtenida, con especial atención al proceso de triangulación entre los resultados del nuevo cuestionario y los del estudio anterior. Es precisamente en esta comparación en la que se articula el apartado de resultados, donde los hallazgos se presentan de forma descriptiva e interpretativa. Los resultados obtenidos nutren la discusión, donde se contrastan con la literatura científica para interpretar su significado y alcance. A partir de estas evidencias, la propuesta de implementación progresiva traduce los hallazgos en orientaciones concretas y graduadas para la transformación de las prácticas evaluativas en el contexto estudiado. Cerrando el trabajo con las conclusiones, que dan respuesta directa a los objetivos específicos planteados, y con el apartado de limitaciones y futuras líneas de investigación que reflexionan críticamente sobre los condicionantes del estudio y abre perspectivas para trabajos posteriores.

2. Marco Teórico

2.1. Evaluación en el Contexto Educativo Actual

La evaluación educativa conforma uno de los elementos más complejos y debatidos dentro del proceso de enseñanza-aprendizaje. Su conceptualización ha experimentado una transformación profunda a lo largo de las últimas décadas, pasando de entenderse como un instrumento de medición del rendimiento académico a concebirse como un proceso integrado al servicio del aprendizaje y la mejora de la práctica docente (Black & Wiliam, 2018; Carless & Boud, 2018). Esta evolución no responde únicamente a cambios en los enfoques pedagógicos, sino también a transformaciones en la manera de entender qué es aprender, para qué se evalúa y quién debe participar en dicho proceso.

Históricamente, la evaluación educativa se ha asociado a una función predominantemente sumativa orientada a certificar el nivel de logro alcanzado por el alumnado, situando al docente como único agente evaluador y al alumnado como receptor pasivo de juicios externos sobre su desempeño (Carless, 2015). Black & Wiliam (2018) señalan que este modelo proporciona una

información muy limitada sobre cómo aprende el alumnado, al operar fundamentalmente como un mecanismo de rendición de cuentas y no como una herramienta pedagógica. Esta revisión crítica del modelo tradicional, que se desarrollará con mayor profundidad en el siguiente apartado, ha ido acompañada de una creciente atención hacia la ampliación de los agentes evaluadores, incorporando al propio alumnado en el proceso, emergiendo la autoevaluación y la coevaluación como estrategias con un gran potencial para favorecer la autorregulación del aprendizaje. Panadero et al. (2017), en su metaanálisis, concluyen que su implementación sistemática produce mejoras tanto en el rendimiento académico como en la autoeficacia percibida, siempre que vaya acompañada de criterios claros y de orientación docente adecuada.

En el contexto español, la evolución del marco legislativo refleja también, aunque de forma desigual, este debate conceptual. La Ley Orgánica 2/2006, de 3 de mayo, de Educación (LOE) ya establecía el carácter continuo e integrador de la evaluación dentro del proceso de enseñanza-aprendizaje. La Ley Orgánica 8/2013, de 9 de diciembre (LOMCE), sin embargo, reforzó los mecanismos de evaluación externa y las pruebas estandarizadas, situando el énfasis nuevamente en la certificación del rendimiento. Es la Ley Orgánica 3/2020, de 29 de diciembre (LOMLOE), actualmente vigente, la que supone un giro hacia un enfoque competencial y continuo de la evaluación apostando explícitamente por una evaluación integrada en el proceso de aprendizaje, orientada al desarrollo de competencias y alejada de la simple acumulación de calificaciones.

A pesar de estos avances normativos y conceptuales, la literatura científica advierte de la persistencia de una brecha significativa entre el discurso evaluativo y las prácticas reales en los centros educativos, dificultada por la presión por la calificación, la tradición docente y las exigencias del sistema educativo (Carless, 2015; Black & Wiliam, 2018). Esta tensión entre evaluar para aprender y evaluar para calificar no es ajena al ámbito de la Educación Física (EF), donde se manifiesta con particularidades propias derivadas de su naturaleza práctica y de su histórica vinculación a la medición de capacidades físicas.

La trayectoria evaluativa de la EF refleja esta misma tensión entre rendimiento y aprendizaje, entre medición y comprensión (Martín et al., 2020). Sus orígenes en España estuvieron marcados por una concepción higienista e incluso militarista, en la que la evaluación tenía un carácter observacional orientado a comprobar la correcta ejecución gimnástica (Flix & Rius, 2020; Fernández-Truan & Muñoz, 2021). A partir de la segunda mitad del siglo XX, el

criterio evaluativo pasó a ser la medición objetiva de la condición física mediante baterías de tests estandarizados, en una lógica de legitimación científica de la asignatura que dejó una huella profunda y persistente en las prácticas del profesorado (Zubillaga-Olague & Cañadas, 2021; Ávila et al., 2022). Reformas posteriores como la LOGSE (1990) intentaron superar este modelo introduciendo el carácter continuo y formativo de la evaluación, pero la literatura constata que dicho cambio normativo tuvo un impacto limitado en la transformación real de las prácticas de aula, en parte por la evaluación cíclica de contenidos y orientaciones de la asignatura (Fernández-Truan & Muñoz, 2021) y en parte por la denominada deportivización de la EF, que orientó la evaluación hacia la medición del rendimiento deportivo y perpetuó dinámicas de exclusión del alumnado con menor condición física (Ávila et al., 2022).

La LOMLOE y su desarrollo curricular mediante el Real Decreto 217/2022, concretado en Cataluña a través del Decret 175/2022, introducen un nuevo marco de referencia que organiza los aprendizajes en torno a competencias específicas vinculadas a la salud, la motricidad y la gestión activa del ocio, desplazando el referente evaluativo desde los contenidos de cada unidad hacia el desempeño competencial del alumnado. Este enfoque exige que la evaluación sea continua, formativa e integradora, y reclama instrumentos coherentes con dicho propósito, como las rúbricas de desempeño, los registros de observación sistemática, los portafolios y la evaluación a través de situaciones de aprendizaje auténticas, frente a las pruebas físicas estandarizadas propias de modelos tradicionales (Pérez-Pueyo et al., 2022). La coherencia de este enfoque competencial reside también en la articulación de tres momentos evaluativos complementarios: la evaluación diagnóstica inicial, la continua y la sumativa. Siendo la evaluación formativa la que da sentido pedagógico al conjunto del proceso orientado hacia el aprendizaje y no únicamente hacia la calificación (Cañadas & Santos-Pastor, 2021), sin embargo, Otero-Saborido et al. (2020) constatan que los currículos de EF en secundaria reconocen formalmente estos tres momentos sin proporcionar siempre orientaciones metodológicas suficientes sobre cómo articularlos de forma coherente en la práctica docente.

Un elemento central en la concepción evaluativa del currículo competencial es la diversificación de los agentes implicados en la evaluación. Frente al modelo tradicional en el que el docente actúa como único agente evaluador, el marco normativo distingue tres modalidades complementarias: la heteroevaluación, en la que el docente evalúa el desempeño del alumnado; la autoevaluación, en la que el propio estudiante analiza y valora su proceso de aprendizaje en

relación con los criterios establecidos; y la coevaluación, en la que docente y alumnado, o los propios alumnos entre sí, valoran conjuntamente compartiendo la responsabilidad sobre los criterios y el proceso evaluativo (Mendoza et al., 2021). Tanto la LOMLOE (2020) como el Decret 175/2022 recogen explícitamente la necesidad de fomentar la autoevaluación y la coevaluación del alumnado como estrategias coherentes con el enfoque competencial y formativo de la evaluación. No obstante, las investigaciones recientes muestran que la distribución real de estas tres modalidades en los centros educativos está lejos de lo que el marco competencial prescribe: Zubillaga-Olague & Cañadas (2022) constatan que la heteroevaluación y la heterocalificación predominan de forma abrumadora sobre los procesos de participación del alumnado, especialmente en secundaria y entre el profesorado con menor formación en evaluación formativa, mientras que Grisales et al. (2023) identifican la heteroevaluación y la autoevaluación como las principales tendencias presentes en las aulas, con una presencia marginal de la coevaluación. Barba-Martín & Hortigüela-Alcalá (2022) sostienen que la implicación del alumnado en su propia evaluación no es solo pedagógicamente deseable, sino una condición necesaria para que esta cumpla una función formativa, ya que su participación activa favorece la comprensión de los criterios de logro y el desarrollo de la autorregulación. Estrechamente vinculada a esta diversificación se encuentra la percepción del alumnado sobre la equidad y la justicia de la evaluación docente, que Grisales et al. (2023) relacionan directamente con la claridad de los criterios compartidos, de modo que la transparencia evaluativa constituye no sólo un principio ético sino una condición pedagógica para que la evaluación cumpla su función formativa.

Pese a este fundamento normativo y pedagógico, la traducción de los principios competenciales en prácticas evaluativas reales compone, como se desarrollará en los apartados siguientes, uno de los principales desafíos pendientes de la disciplina, en la medida en que la adopción de un nuevo marco curricular no garantiza por sí misma una transformación equivalente de las concepciones y las rutinas evaluativas del profesorado. Comprender en profundidad qué implica esta transformación exige, por tanto, detenerse en los fundamentos teóricos y conceptuales de la evaluación formativa, que se desarrollan a continuación.

2.2. Evaluación Formativa: Fundamentos Teóricos y Conceptuales

2.2.1. Conceptualización

El concepto de evaluación formativa tiene sus raíces en los debates sobre diseño curricular e instrucción de la segunda mitad del siglo XX, cuando comenzó a establecerse la distinción fundamental entre evaluación sumativa y evaluación formativa. Black & Wiliam (2018) sitúan el origen de esta distinción en los trabajos de Scriven y Bloom durante la década de 1960 y 1970. Fue, sin embargo, la revisión realizada por Black & Wiliam (1998), ampliada y actualizada posteriormente por estos mismo autores en 2018, la que situó la evaluación formativa en el centro del debate pedagógico contemporáneo, al demostrar empíricamente que las prácticas evaluativas orientadas a proporcionar información continua sobre el aprendizaje producen mejoras significativas en el rendimiento del alumnado. Hallazgo que estudios posteriores como el de Sortwell et al. (2024) han confirmado de forma consciente en todos los niveles educativos.

A partir de esta base, la evaluación formativa puede definirse como el conjunto de procesos mediante los cuales se recogen, interpretan y utilizan evidencias sobre el aprendizaje del alumnado, por parte del profesorado y del propio alumnado, con el fin de tomar decisiones que mejoren la enseñanza y el aprendizaje en curso (Black & Wiliam, 2018). López-Pastor & Pérez-Pueyo (2017) han desarrollado el concepto de evaluación formativa y compartida (EFyC), que añade a esta dimensión formativa una dimensión dialógica: la evaluación no solo debe mejorar el aprendizaje, sino hacerlo a través del diálogo y la participación activa del alumnado, compartiendo con él la responsabilidad sobre los criterios, los instrumentos y los juicios evaluativos. En esta misma línea, resulta pertinente la distinción entre evaluación para el aprendizaje, que designa las prácticas evaluativas integradas en el proceso de enseñanza con finalidad formativa, y evaluación del aprendizaje, que designa la evaluación con finalidad certificadora o sumativa (Carless & Boud, 2018); aunque ambas funciones son necesarias en los sistemas educativos, Black & Wiliam (2018) defienden que la primera debe predominar en los contextos de enseñanza obligatoria.

De la literatura científica revisada se desprende un conjunto de principios que caracterizan la evaluación formativa: la continuidad, entendida como su integración permanente en el proceso de enseñanza-aprendizaje para detectar dificultades y ajustar la enseñanza en tiempo real; la intencionalidad pedagógica, según la cual toda práctica educativa debe orientarse conscientemente a la mejora del aprendizaje; la transparencia de los criterios, que el alumnado

debe conocer y comprender; y la implicación activa del alumnado como agente evaluativo a través de la autoevaluación, la coevaluación y la reflexión sobre el propio aprendizaje (Black & Wiliam, 2018; Carless & Boud, 2018; López-Pastor & Pérez-Pueyo, 2017). En este marco, la retroalimentación ocupa un lugar central: Hattie & Timperley (2007), en su modelo conceptual del feedback educativo, sostienen que este solo tiene un impacto formativo significativo cuando responde a tres preguntas —¿hacia dónde voy?, ¿cómo voy? y ¿qué pasos debo dar a continuación?, y se orienta a la tarea, al proceso y a la autorregulación. Carless & Boud (2018) complementan este modelo con el concepto de feedback literacy, que hace referencia a la capacidad del alumnado para comprender, procesar y utilizar la retroalimentación de forma autónoma, subrayando que solo cuando el alumnado puede actuar sobre ella la retroalimentación cumple su función formativa; Morris et al. (2021), en su revisión sistemática sobre evaluación formativa y feedback en educación superior, corroboran que esta capacidad requiere ser enseñada explícitamente y no puede darse por supuesta. Black & Wiliam (2018) traducen estos principios en cinco estrategias clave: clarificar y compartir los criterios de éxito, diseñar actividades que evidencien el aprendizaje, proporcionar feedback orientado a la mejora, activar al alumnado como recurso evaluativo entre sí, y responsabilizarlo de su propio aprendizaje.

La aplicación coherente de estos principios no está, sin embargo, garantizada por el uso aislado de instrumentos formativos. Black & Wiliam (2018) distinguen entre una evaluación formativa superficial, que adopta dichos instrumentos sin transformar la concepción evaluativa de fondo, y una evaluación formativa profunda, que implica una reorientación genuina del rol del docente y del alumnado en el proceso evaluativo. Hortigüela et al. (2019) documentan esta distinción en el contexto específico de la Educación Física, identificando casos en los que el profesorado declara aplicar evaluación formativa pero se limita en la práctica a utilizar rúbricas o fichas de autoevaluación sin que exista un proceso real de retroalimentación y diálogo orientado a la mejora. Esta brecha entre la adopción formal de herramientas y la transformación efectiva de la práctica evaluativa resulta determinante para comprender por qué, como se expone a continuación, los efectos positivos de la evaluación formativa documentados en la literatura no se producen de manera automática, sino que dependen de las condiciones concretas en que esta se implementa.

2.2.2. Impacto en el Aprendizaje

Más allá de las mejoras en el rendimiento académico ya señaladas, la evaluación formativa incide de forma significativa sobre otras dimensiones del aprendizaje, siempre que el alumnado adopte en ella un rol activo y autorregulado, condición que la literatura considera clave para su efectividad (Black & Wiliam, 2018). En primer lugar, favorece el desarrollo de la autorregulación al proporcionar información continua sobre el progreso, lo que permite al alumnado identificar sus dificultades y ajustar sus estrategias de aprendizaje (Panadero et al., 2017). En segundo lugar, tiene efectos positivos sobre la motivación intrínseca, en la medida en que fomenta la autonomía del alumnado y proporciona retroalimentación orientada a la competencia, contribuyendo así a satisfacer las necesidades psicológicas básicas que, según la teoría de la autodeterminación, sustentan el compromiso académico sostenido (Bureau et al., 2022). En tercer lugar, estrategias específicas como la autoevaluación y la coevaluación producen efectos positivos tanto sobre el rendimiento como sobre la autoeficacia del alumnado, siempre que se implementen mediante criterios claros y un proceso formativo que desarrolle su capacidad para evaluarse con rigor (Panadero et al., 2017; Yan et al., 2022).

La evaluación formativa produce, además, efectos relevantes sobre la equidad educativa y sobre la práctica docente. Frente a los modelos sumativos, que tienden a reproducir y amplificar las desigualdades, tiene el potencial de reducir las barreras de rendimiento y favorecer el éxito de todo el alumnado con independencia de su punto de partida (Yan et al., 2022; Sortwell et al., 2024), un principio especialmente relevante en Educación Física, donde los modelos evaluativos basados en el rendimiento físico generan dinámicas de exclusión (Ávila et al., 2022). Desde la perspectiva del profesorado, integra asimismo un instrumento de desarrollo profesional, ya que la información generada a lo largo del proceso permite ajustar las estrategias de enseñanza y reflexionar de forma sistemática sobre la calidad de la propia práctica (Wisniewski et al., 2020).

No obstante, estos efectos positivos no son automáticos ni universales: su magnitud es, de hecho, significativa pero variable, y depende en gran medida de las condiciones en las que la evaluación formativa se implementa (Wisniewski et al., 2020). La claridad de los criterios compartidos con el alumnado, su nivel de formación en el uso de la autoevaluación y la coevaluación, la calidad del feedback proporcionado y la coherencia entre el sistema evaluativo y el contexto institucional son, en este sentido, factores que modulan directamente su efectividad

(Yan et al., 2022; Kingston et al., 2024; Wisniewski et al., 2020). Estas condiciones adquieren una relevancia particular en el ámbito de la Educación Física, donde la naturaleza práctica de la disciplina y su tradicional vinculación a la medición del rendimiento físico plantean exigencias específicas para la implementación de la evaluación formativa, tal como se examina en el apartado siguiente.

2.3. Evaluación Formativa en Educación Física

2.3.1. Evaluación Formativa y Modelos Pedagógicos en Educación Física

La Educación física ha experimentado durante las últimas décadas una profunda transformación pedagógica, impulsada por la crítica sistemática a los enfoques tradicionales de enseñanza, centrados en la instrucción directa, la repetición técnica descontextualizada y la evaluación del rendimiento motor, y por la aparición de propuestas alternativas orientadas a situar el aprendizaje del alumnado en el centro del proceso educativo (Fernández-Río & Iglesias, 2024). En este contexto, el enfoque denominado *Models-Based Practice* (MbP) o práctica basada en modelos, ha ganado protagonismo en la investigación y la práctica pedagógica en EF. Fernández-Río & Iglesias (2024), en su revisión sobre los modelos pedagógicos en Educación Física, definen la práctica basada en modelos como un enfoque pedagógico que utiliza modelos específicos como centro organizador de la programación y la enseñanza, con el objetivo de superar las limitaciones del enfoque multiactividad tradicional, el cual fragmenta el currículo en unidades breves y desconectadas sin posibilidad de un aprendizaje profundo y sostenido. Estos autores identifican como modelos básicos y consolidados el aprendizaje Cooperativo, el Modelo de Educación Deportiva, el Modelo Comprensivo (TGfU) y el Modelo de Responsabilidad Personal y Social, siendo precisamente estos los que presentan una relación más estrecha con los principios de la evaluación formativa (Fernández-Río & Iglesias, 2024; Zhang et al., 2024).

Un elemento de especial relevancia en la literatura científica sobre modelos pedagógicos en Educación Física es la constatación de que la evaluación conforma uno de los obstáculos más significativos para su implementación efectiva. Pérez-Pueyo et al. (2020) señalan que, aunque la adopción de modelos pedagógicos representa una transformación metodológica de primer orden, la coherencia de dicha transformación depende de que el sistema evaluativo utilizado sea acorde a los principios pedagógicos del modelo, de lo contrario, la evaluación se convierte en una barrera que limita los efectos formativos del modelo. Hernando-Garijo et al. (2021) subrayan que

la evaluación formativa y compartida es una herramienta clave para optimizar e implementar eficazmente la enseñanza basada en modelos, en la medida en que permite al alumnado tomar conciencia de su propio proceso de aprendizaje y al docente ajustar la enseñanza en función de las necesidades detectadas. En esta línea, Niubó Solé (2025), en su tesis doctoral sobre la influencia de la evaluación formadora en el área de EF sobre las funciones ejecutivas del alumnado, aporta evidencias sobre cómo la implementación de procesos evaluativos de carácter formativo en EF producen efectos positivos que van más allá del aprendizaje motor, incidiendo sobre capacidades cognitivas de orden superior estrechamente vinculadas al aprendizaje profundo y autorregulado. Por tanto, la evaluación formativa no es un complemento añadido a los modelos pedagógicos, sino una condición estructural para que estos alcancen sus objetivos de aprendizaje en toda su amplitud.

El modelo comprensivo de iniciación deportiva, conocido internacionalmente como *Teaching Games for Understanding* (TGfU), pretende transformar la enseñanza de los deportes colectivos desplazando el foco de la técnica descontextualizada hacia la comprensión táctica del juego (Ortiz et al., 2023). Su premisa fundamental es que el aprendizaje deportivo debe partir de la práctica del juego, modificado y adaptado al nivel del alumnado, para que sean los propios problemas tácticos emergentes los que justifiquen y orienten el desarrollo técnico (Ortiz et al., 2023). El modelo se articula en torno a cuatro principios pedagógicos: la transferencia, que permite identificar estructuras tácticas comunes entre deportes; la representación, que adapta la forma del juego a las características del alumnado; la exageración, que modifica las reglas para hacer más visibles los problemas tácticos; y la progresión en la complejidad táctica, que gradúa la dificultad del aprendizaje (Ortiz et al., 2023). Desde la perspectiva de la evaluación formativa, el TGfU presenta una afinidad estructural especialmente significativa al situar la comprensión táctica y la toma de decisiones como ejes del aprendizaje, lo que demanda formas de evaluación auténtica y de retroalimentación orientadas al proceso, que informen al alumnado sobre la calidad de sus decisiones y les ayuden a mejorar su comprensión del juego (Pérez-Pueyo et al., 2020). Estos mismos autores señalan que la evaluación auténtica en el TGfU significa evaluar al alumnado durante la práctica real del deporte que se está desarrollando y no en tareas aisladas y descontextualizadas, lo que conecta directamente con los principios de la evaluación formativa.

El Modelo de Educación Deportiva tiene como objetivo proporcionar al alumnado una experiencia deportiva auténtica, completa y educativamente significativa. Sus tres grandes

finalidades son el desarrollo de la competencia deportiva, la alfabetización deportiva (conocimiento de las reglas, las tradiciones y los valores del deporte) y el entusiasmo deportivo del alumnado (Guijarro et al., 2020). El modelo se estructura en torno a seis características que lo definen: la organización del trabajo en temporadas prolongadas, la afiliación de equipos estables, la asignación de roles más allá del de jugador, la competición formal, el registro de resultados y el evento culminante, los cuales en su conjunto configuran una experiencia deportiva escolar de mayor autenticidad y significado que la proporcionada por el enfoque multiactividad (Guijarro et al., 2020). Esta estructura hace del modelo de educación deportiva uno de los contextos más ricos para el desarrollo de la evaluación formativa, ya que los roles no jugadores requieren que el alumnado observe, analice y valore el rendimiento propio y ajeno de forma sistemática, lo que convierte la coevaluación y la autoevaluación en elementos naturales e integrados en la dinámica del modelo (Pérez-Pueyo et al., 2020). Manninen & Campbell (2022), en su revisión sistemática sobre los efectos del modelo de educación deportiva, concluyen que el modelo produce efectos positivos y significativos en la satisfacción de necesidades psicológicas básicas, la motivación intrínseca y las actitudes prosociales, efectos que se amplifican cuando el modelo se implementa con procesos evaluativos formativos que refuerzan el sentido de competencia y autonomía del alumnado (Manninen & Campbell, 2022).

El Aprendizaje Cooperativo (AC) es un modelo pedagógico que organiza el proceso de enseñanza-aprendizaje en torno a la interdependencia positiva entre los miembros del grupo, haciendo que el aprendizaje individual solo sea posible si el grupo aprende, lo que genera una dinámica de responsabilidad mutua y apoyo entre iguales (Bores-García et al., 2021). Casey & Goodyear (2015) identificaron en su revisión que el AC es capaz de promover aprendizajes en los cuatro dominios del aprendizaje en Educación Física (físico, cognitivo, social y afectivo), lo que lo convierte en uno de los modelos con mayor potencial para el desarrollo integral del alumnado. Bores-García et al. (2021) en su revisión sistemática sobre la literatura científica del AC en los últimos 5 años, concluyen que este modelo contribuye al desarrollo de todas las dimensiones de la personalidad del alumnado, con un impacto especialmente relevante en el dominio afectivo y social. Desde la perspectiva de la evaluación formativa, el AC presenta una articulación especialmente coherente donde los procesos de retroalimentación entre iguales son intrínsecos a su dinámica, en la medida en que el alumnado necesita información continua sobre el progreso del grupo para poder ajustar sus estrategias de trabajo cooperativo (Pérez-Pueyo et

al., 2020). Zach et al. (2023) señalan que la responsabilidad personal de cada alumnado dentro del grupo está directamente vinculada a la implementación de prácticas evaluativas formativas que hagan visible el aprendizaje individual dentro del proceso grupal.

Junto a los modelos anteriores, la literatura científica española ha prestado especial atención a otros dos modelos pedagógicos que tienen una estrecha relación con la evaluación formativa. El estilo actitudinal desarrollado por Pérez-Pueyo se basa en la idea de que el aprendizaje motor sólo adquiere pleno valor educativo cuando se vincula al desarrollo de actitudes positivas hacia la práctica y hacia los demás, para ello propone una secuenciación metodológica donde la evaluación formativa actúa como motor del aprendizaje actitudinal y del alumnado (Pérez-Pueyo et al., 2020). Por su parte, el Modelo de Responsabilidad Personal y Social (MRPS) tiene como finalidad central la transferencia progresiva de responsabilidad del docente al alumnado, tanto en el ámbito personal como en el social, articulándose en torno a cinco niveles jerárquicos que van desde el respeto por los derechos y sentimientos de los demás hasta la transferencia de los valores aprendidos fuera del contexto de la EF (Carreres-Ponsoda et al., 2021). Alcalá et al. (2020) evidencian que la evaluación formativa en el MRPS es especialmente pertinente, ya que la capacidad del alumnado para autoevaluarse y reflexionar sobre su nivel de responsabilidad integra el núcleo del modelo y no pueden desarrollarse sin procesos evaluativos que hagan visible dicho progreso. Ambos modelos muestran que la evaluación formativa no solo es compatible con los modelos pedagógicos en Educación Física, sino que es la condición que les permite alcanzar su máximo potencial transformador (Pérez-Pueyo et al., 2020).

Un aspecto emergente y de creciente relevancia en la investigación sobre los modelos pedagógicos en EF es la hibridación, término que hace referencia a la combinación deliberada de dos o más modelos para potenciar sus efectos educativos (Fernández-Río & Iglesias, 2024). Zhang et al. (2024), en su revisión sobre los efectos de los modelos híbridos en EF, identificaron que la combinación del Modelo de Educación Deportiva con el modelo comprensivo de iniciación deportiva es la hibridación más estudiada, generando efectos positivos sobre los resultados de aprendizaje en los dominios motor, cognitivo, emocional y social. Pérez-Pueyo et al. (2017) ya habían documentado hibridación entre Aprendizaje Cooperativo y otros modelos, señalando que la evaluación formativa actúa como elemento articulador entre los modelos en los procesos de hibridación, ya que proporciona al alumnado y al profesorado la información

necesaria para gestionar la mayor complejidad metodológica que implica trabajar simultáneamente con varios modelos. Fernández-Río & Iglesias (2024) subrayan que la práctica basada en modelos no supone la aplicación sucesiva de modelos aislados, sino la construcción de un programa coherente en el que múltiples modelos se complementan y se refuerzan mutuamente. En este sentido, la evaluación formativa emerge como el mecanismo que garantiza la coherencia del proceso, al proporcionar información continua sobre el progreso del alumnado. Por lo tanto, como señalan Pérez-Pueyo et al. (2020) y Fernández-Río & Iglesias (2024), la evaluación formativa y la enseñanza basada en modelos pedagógicos en EF son dos dimensiones de una misma apuesta por una Educación Física orientada al aprendizaje profundo y al desarrollo integral del alumnado.

2.3.2. Estrategias para una Implementación Efectiva

La implementación efectiva de la evaluación formativa en el contexto de la Educación Física no se produce de forma espontánea ni como consecuencia directa del conocimiento teórico del profesorado sobre sus fundamentos. La literatura científica muestra una brecha entre el discurso evaluativo y las prácticas reales del profesorado de EF en el aula, poniendo de manifiesto que la transformación evaluativa requiere estrategias específicas, sostenidas y contextualmente sensibles (Cañadas & Santos-Pastor, 2021; Zubillaga-Olague & Cañadas, 2022). Esta brecha no es atribuible únicamente a la falta de voluntad del profesorado, sino a una combinación de factores estructurales, formativos y culturales que condicionan las posibilidades reales de cambio (Hernando-Garijo et al., 2021). En consecuencia, la literatura científica ha identificado un conjunto de estrategias que, aplicadas de forma coherente y articulada, favorecen una implementación efectiva, progresiva y sostenible de la evaluación formativa en EF.

La primera y más fundamentada es la formación del profesorado, tanto en su dimensión inicial como continua. Cañadas & Santos-Pastor (2021) muestra cómo el profesorado de EF presenta concepciones evaluativas frecuentemente ancladas a modelos sumativos y calificadores, lo que supone una barrera para la implementación de la evaluación formativa. Cañadas (2023), en su estudio sobre la contribución de la evaluación formativa al desarrollo de competencias docentes, concluye que la vivencia directa de procesos de evaluación formativa durante la formación inicial se trata del mecanismo más efectivo para que el futuro profesorado comprenda sus principios, interiorice sus procedimientos y desarrolle la disposición necesaria para

implementarlos. Black & Wiliam (2018) subrayan que la formación docente en evaluación formativa sólo resulta efectiva cuando combinan la fundamentación teórica con la práctica reflexiva en contextos reales de enseñanza. Esta idea es coherente con los hallazgos de Slingerland et al. (2024), quienes evidencian que el desarrollo de la competencia evaluativa formativa requiere un proceso de formación continua situado en el propio contexto de enseñanza, que combine la reflexión sobre la práctica real con el acceso a conocimiento especializado y al apoyo entre iguales.

Una segunda estrategia de especial relevancia para la implementación efectiva de la evaluación formativa en EF es la promoción sistemática de la autoevaluación y la coevaluación como prácticas habituales e integradas en el proceso de enseñanza-aprendizaje. Martín (2022), en su revisión sistemática sobre los procesos de autoevaluación y coevaluación en EF, concluye que ambos procedimientos son efectivos para el desarrollo de la autonomía y la responsabilidad del alumnado en contextos de EF, siempre que se implementen con criterios claros, progresividad adecuada y acompañamiento docente suficiente. La autoevaluación permite al alumnado reflexionar sobre su propio proceso de aprendizaje, identificar fortalezas y áreas de mejora, y desarrollar la capacidad metacognitiva necesaria para autorregular su aprendizaje (Panadero et al., 2017). La coevaluación, por su parte, promueve un ambiente de aprendizaje colaborativo en el que el alumnado desarrolla habilidades críticas y de observación, aprende a formular retroalimentación constructiva y se corresponsabiliza de los procesos de aprendizaje del grupo (Yan et al., 2022). En ambos casos, la investigación advierte que la efectividad de estas prácticas depende en gran medida de la formación previa del alumnado en su uso, ya que sin una introducción progresiva y explícita a los criterios de evaluación y a los procedimientos de auto y coevaluación, el alumnado tiende a valorar de forma poco fiable o a reproducir dinámicas de evaluación superficial (Martín, 2022; Yan et al., 2022).

Una tercera estrategia determinante para la implementación efectiva de la evaluación formativa en EF es el diseño y uso de instrumentos de evaluación auténtica que evalúen al alumnado en situaciones reales o próximas a la práctica real, integrando la evaluación en la propia actividad motriz y no en tareas artificiales o descontextualizadas (Pérez-Pueyo et al., 2020). Entre los instrumentos más coherentes con los principios de la evaluación formativa en EF destacan las rúbricas de desempeño (describen los niveles de logro esperados de forma clara y comprensible para el alumnado), los registros de observación sistemática (permiten al docente

recoger evidencias de aprendizaje durante la práctica), los portafolios (documentan el proceso de aprendizaje a lo largo del tiempo) y las hojas de autoevaluación y coevaluación vinculadas a los criterios de evaluación establecidos (López-Pastor & Pérez-Pueyo, 2017). Tapia et al. (2022), evidencian que el uso de rúbricas compartidas con el alumnado mejora significativamente la calidad del aprendizaje y el nivel de implicación del alumnado en el proceso evaluativo. La transparencia de los criterios de evaluación, materializada a través de estos instrumentos, no solo orienta el aprendizaje del alumnado, sino que contribuye a reducir la percepción de arbitrariedad evaluativa que con frecuencia caracteriza la EF en contextos calificadores (Zubillaga-Olague & Cañadas, 2022).

La retroalimentación conforma el mecanismo central a través del cual la evaluación formativa incide sobre el aprendizaje del alumnado, y su gestión efectiva representa una de las competencias docentes más exigentes para el profesorado de EF. En esta área concreta, el feedback adopta formas y temporalidades particulares. Esto se debe a que al producirse durante la práctica motriz, debe ser inmediato, preciso y orientado a la mejora de la ejecución en curso, lo que requiere una alta capacidad de observación y análisis por parte del docente (Hattie & Timperley, 2007). Barrientos et al. (2023), en su estudio sobre los desafíos del uso de la evaluación auténtica y formativa en EF desde la perspectiva del profesorado experimentado, identifican la gestión del feedback como uno de los principales retos de la implementación. Muestran cómo el profesorado señala dificultades para proporcionar retroalimentación individualizada y de calidad a todos los alumnos simultáneamente durante la práctica, lo que pone de manifiesto la necesidad de incorporar el feedback entre iguales y el feedback a través de instrumentos escritos como estrategias complementarias. Carless y Boud (2018) subrayan que la calidad del feedback no depende únicamente de su contenido, sino de la capacidad del alumnado para comprenderlo, procesarlo y utilizarlo para mejorar su práctica, lo que sitúa el desarrollo de la *feedback literacy* del alumnado como una dimensión clave de la implementación efectiva de la evaluación formativa en EF.

Finalmente, los estudios realizados sobre evaluación formativa señalan que ninguna de las estrategias anteriores puede producir sus efectos si no se implementa de forma progresiva, coherente y sensible con el contexto específico al que se aplica. Hernando-Garijo et al. (2021) evidencian que los intentos de implementación brusca o total de la evaluación formativa tienden al fracaso y generan resistencias tanto en el profesorado como en el alumnado. En la misma

línea, Sortwell et al. (2024) señalan que la coherencia entre el sistema de evaluación formativa y las expectativas del alumnado, la cultura evaluativa del centro y las demandas del sistema de calificación es una condición determinante de su efectividad. Esta consideración adquiere especial relevancia en los contextos predominantemente calificadores, donde el alumnado ha interiorizado una cultura evaluativa centrada en la nota y puede percibir las prácticas de evaluación formativa como una amenaza a su calificación o como una tarea adicional sin valor intrínseco (Zubillaga-Olague & Cañadas, 2022). La implementación progresiva supone, en consecuencia, una estrategia gradual que comienza por hacer explícitos los criterios de evaluación, incorpora paulatinamente la autoevaluación y la coevaluación, y culmina en un sistema evaluativo compartido y corresponsable, adaptando en cada fase el grado de demanda a las posibilidades reales del alumnado y del contexto institucional (López-Pastor & Pérez-Pueyo, 2017; Hernando-Garijo et al., 2021).

2.3.3. Barreras y Condicionantes para la Implementación

A pesar del respaldo científico acumulado sobre los beneficios de la evaluación formativa en EF y de las estrategias disponibles para su implementación, la investigación desarrollada en el contexto español evidencia que las prácticas evaluativas reales del profesorado de EF siguen ancladas mayoritariamente en un modelo tradicional centrado en la calificación. Hortigüela-Alcalá et al. (2021), en su estudio sobre la percepción del profesorado de EF español en diferentes etapas educativas, constatan que el profesorado de secundaria tiende a confundir evaluación con calificación, concentra sus prácticas en la valoración de la condición física y los tests deportivos, y concibe la evaluación como una tarea de implementación exclusivamente docente sin implicación del alumnado. Esta confusión conceptual no es una cuestión individual, sino el reflejo de una cultura evaluativa institucional profundamente arraigada que otorga a la calificación un papel de legitimación social que ninguna otra práctica evaluativa ha logrado disputar en la EF de secundaria (Pastor et al., 2006; Herrero-González et al., 2024). La consecuencia directa es la persistencia de una brecha entre el discurso evaluativo, que favorece la evaluación formativa, y las prácticas reales, que reproduce el modelo tradicional incluso entre docentes que declaran conocer y valorar los principios de la evaluación formativa.

Una segunda barrera de primer orden identificada es la insuficiente formación del profesorado de EF en evaluación formativa, tanto en su dimensión inicial como permanente.

Herrero-González et al. (2021), en su estudio sobre la incidencia de la formación inicial y permanente en la aplicación de la evaluación formativa y compartida en EF, concluyen que la competencia evaluativa del profesorado está directamente condicionada por el tipo de formación recibida. En el mismo sentido, Hernán et al. (2019) documentan, a través de entrevistas en profundidad a profesorado de EF en activo en España, que la decisión de implementar evaluación formativa y compartida está fuertemente motivada por la formación recibida y por la participación en grupos de trabajo y redes de formación permanente, lo que subraya la necesidad de articular una oferta formativa específica, continuada y situada en la práctica docente real como condición necesaria para superar esta barrera.

Junto a los condicionantes formativos, también se identifican una serie de barreras de naturaleza contextual que operan en el nivel del centro educativo y del aula. La ratio elevada del alumnado en EF, la presión institucional por calificar con nota numérica en un tiempo limitado, y la percepción del alumnado de secundaria sobre la evaluación como sinónimo de nota condicionan el margen de actuación real del profesorado que intenta implementar procesos evaluativos formativos (Herrero-González et al., 2024; Cañadas & Santos-Pastor, 2021). A estas barreras se suma la resistencia cultural del propio alumnado que ha crecido en un modelo evaluativo calificador a lo largo de toda su escolarización, el alumnado de Educación Secundaria Obligatoria (ESO) tiende a percibir las prácticas de evaluación formativa como tareas adicionales de escaso valor instrumental para su calificación final (Zubillaga-Olague & Cañadas, 2022). Esta resistencia no es atribuible a una actitud negativa del alumnado, sino al resultado lógico de una cultura evaluativa que ha condicionado sistemáticamente su relación con la evaluación. Conocer con precisión cómo el alumnado percibe y experimenta la evaluación en sus clases de EF constituye, por tanto, una condición previa y necesaria para el diseño de cualquier propuesta de cambio evaluativo que aspire a ser efectiva y sostenible en contextos predominantemente calificadores.

3. Diseño Teórico del Estudio

3.1. Análisis del Contexto Escolar

El estudio se desarrolla en un centro educativo privado concertado de la ciudad de Barcelona que imparte todas las etapas de la ESO y Bachillerato. El centro se inscribe en el marco del enfoque competencial establecido por la normativa educativa vigente en Cataluña

(Decret 175/2022, de 27 de setembre), que apuesta explícitamente por una evaluación continua, formativa e integradora orientada al desarrollo de competencias y al aprendizaje del alumnado.

No obstante, el diagnóstico previo realizado del centro mediante un instrumento de observación estructurado evidenció que las prácticas evaluativas reales desarrolladas en el área de EF se alejan significativamente de este marco normativo, respondiendo a un modelo predominantemente calificador en el que la evaluación opera fundamentalmente como mecanismo de certificación del rendimiento, con una presencia escasa de estrategias orientadas a la autorregulación del aprendizaje o a la participación activa del alumnado en su propia evaluación. Esta brecha entre el discurso normativo y la práctica evaluativa real, ampliamente documentada con la literatura científica especializada supone el punto de partida empírico del presente estudio y justifica la pertinencia de la investigación. El desarrollo detallado del instrumento de observación y de los resultados obtenidos se presentan en la fase de exploración del apartado de metodología de la investigación.

3.2. Pregunta de Investigación

¿Cómo percibe el alumnado de Secundaria y Bachillerato la evaluación en Educación Física en un contexto predominantemente calificador y que implicaciones tiene esta percepción para el diseño de una propuesta progresiva de implementación de evaluación formativa, a partir de la comparación con contextos contrastados?

3.3. Hipótesis

Se parte de la hipótesis de que el alumnado de secundaria y bachillerato perteneciente a un contexto predominantemente calificador presenta una percepción de la evaluación en Educación Física menos orientada al aprendizaje, a la autorregulación y a la mejora continua que el alumnado que se encuentra en contextos donde se aplica la evaluación formativa.

3.4. Paradigma desde el que se Pretende Abordar

Toda investigación educativa se inscribe en un marco epistemológico que determina la concepción del conocimiento que la sustenta, la naturaleza de la realidad que pretende estudiar y los procedimientos a través de los cuales se genera y valida dicho conocimiento (Bisquerra, 2022). La literatura sobre metodología de la investigación educativa distingue habitualmente tres

grandes paradigmas: el positivista, el interpretativo y el sociocrítico (Bisquerra, 2022; Latorre et al., 2021).

El paradigma positivista parte de una concepción objetivista de la realidad, asumiendo que esta puede ser conocida, medida y explicada a través de procedimientos cuantitativos y leyes generalizables, con la finalidad de predecir y controlar los fenómenos educativos mediante la verificación empírica de hipótesis (Bisquerra, 2022). El paradigma sociocrítico, por su parte, incorpora una dimensión transformadora a la investigación, ya mediante este no solo se pretende comprender la realidad educativa, sino cuestionar y contribuir activamente a su transformación a través de procesos participativos que implican a los propios sujetos (Latorre et al., 2021). Frente a ambos, el paradigma interpretativo se caracteriza, según Bisquerra (2022), por su interés en comprender los significados, las percepciones y las experiencias que los propios actores educativos construyen sobre la realidad en la que participan, asumiendo que dicho conocimiento es de naturaleza subjetiva, situada y contextualizada, y que no puede ser reducido a mediciones objetivas ni a generalizaciones estadísticas (Latorre et al., 2021).

El presente estudio se sitúa dentro del paradigma interpretativo como marco epistemológico dominante, si bien su diseño presenta una naturaleza metodológica mixta que conviene precisar. El propósito central de la investigación es comprender cómo percibe el alumnado de secundaria y bachillerato la evaluación en EF en un contexto predominantemente calificador, lo que responde plenamente a los fundamentos del paradigma interpretativo (Bisquerra, 2022). No obstante, para alcanzar este propósito el estudio incorpora, en primer lugar, una fase previa de recogida y análisis de datos cuantitativos mediante un cuestionario con escala Likert, cuyos resultados se trataron estadísticamente con una finalidad descriptiva y no explicativa ni predictiva. En segundo lugar, el diseño incluye una fase de triangulación y comparación de datos entre los resultados obtenidos en el presente estudio y los del estudio realizado anteriormente en un contexto evaluativo contrastado, lo que permite interpretar las diferencias y convergencias detectadas en la percepción del alumnado de ambos contextos. Finalmente, a partir de los hallazgos obtenidos, el estudio culmina con una fase propositiva orientada al diseño de una propuesta progresiva de implementación de la evaluación formativa en el centro estudiado. En este sentido, aunque los instrumentos cuantitativos están al servicio de una finalidad interpretativa y comparativa, la complejidad del diseño exige reconocer su carácter mixto dentro de un marco epistemológico predominantemente interpretativo (Latorre et al.,

2021). La investigación adopta, en consecuencia, un diseño descriptivo, comparativo e interpretativo de corte no experimental, coherente con los fundamentos epistemológicos del paradigma interpretativo (Bisquerra, 2022; Latorre et al., 2021).

3.5. Objetivos

3.5.1. Objetivo General

El objetivo general de esta investigación es analizar la percepción del alumnado de secundaria y bachillerato sobre la evaluación en Educación Física en un contexto predominantemente calificador.

3.5.2. Objetivos Específicos

Los objetivos específicos planteados para este estudio son los siguientes:

- Comparar la percepción del alumnado del centro de prácticas con la del alumnado de un contexto educativo con evaluación formativa.
- Analizar las principales diferencias detectadas en la percepción del alumnado en relación con los principios de la evaluación formativa.
- Explorar la relación entre la percepción del alumnado sobre la evaluación y su influencia percibida en el aprendizaje en Educación Física en un contexto predominantemente calificador..
- Diseñar una propuesta progresiva de implementación de la evaluación formativa en Educación Física, contextualizada en el centro de estudio.

4. Metodología de la Investigación

La presente investigación adopta un diseño descriptivo, comparativo e interpretativo de corte no experimental, coherente con el paradigma interpretativo desde el que se aborda el estudio (Bisquerra, 2022; Latorre et al., 2021). El propósito no es establecer relaciones causales ni generalizar resultados a poblaciones más amplias, sino comprender y comparar la percepción del alumnado sobre la evaluación en EF en dos contextos evaluativos contrastados, con el fin de extraer interpretaciones significativas que fundamentan una propuesta de transformación evaluativa contextualizada (Creswell & Creswell, 2017).

Para ello, la metodología se articula en torno a tres técnicas de recogida y análisis de información complementarias: la observación estructurada mediante una lista de control, el cuestionario validado administrado al alumnado y el análisis documental del estudio precedente. La triangulación de estas tres fuentes de información constituye el procedimiento central a través del cual se obtienen los resultados del estudio (Cohen et al., 2002).

4.1. Cuadro del Modelo de Análisis

El cuadro de modelo de análisis conforma el elemento articular entre el marco teórico y el diseño metodológico de la investigación, ya que permite operacionalizar los conceptos clave desarrollados en la literatura científica traduciéndose en dimensiones, variables e indicadores observables que orientan de forma sistemática tanto la recogida como el análisis de los datos (Bisquerra, 2022). El presente estudio, cuyo objetivo es la percepción del alumnado sobre la evaluación en EF en un contexto predominantemente calificador y su contraste con los resultados obtenidos en el TFM precedente (Maciá López, 2025), el cuadro de modelo de análisis se estructura en torno a las cuatro dimensiones del cuestionario validado de Palacio et al. (2021): concepciones sobre la evaluación, agentes implicados, instrumentos y momentos, y contenidos evaluados. La adopción de estas mismas dimensiones responde a una exigencia de coherencia metodológica: dado que el instrumento de recogida de datos es idéntico en ambos estudios, la equivalencia del modelo de análisis supone una condición necesaria para garantizar la validez de la comparación entre contextos. Cada dimensión se encuentra sustentada por los fundamentos teóricos desarrollados en el marco teórico, de modo que las variables e indicadores que la integran no responden únicamente a la estructura del cuestionario sino también a los conceptos y evidencias científicas que lo sustentan.

Tabla 1

Cuadro del Modelo de Análisis

Dimensiones	Variables	Indicadores
Concepciones sobre la evaluación	Valor percibido de la evaluación	Reconocimiento de la evaluación como herramienta útil para el aprendizaje y la mejora personal (ítem 1)
	Función de la evaluación	Identificación del propósito que cumple la evaluación en el proceso educativo (ítem 3)

	Relevancia de la calificación	Importancia atribuida a la nota o resultado cuantitativo (ítem 2)
	Evaluación compartida	Valoración del papel del alumnado en la evaluación propia y de sus compañeros (ítems 4 y 5)
Agentes implicados en la evaluación	Participación del alumnado	Grado de implicación del alumnado en su propia evaluación y la de sus compañeros: autoevaluación (ítem 7), coevaluación (ítem 10) y evaluación del profesorado (ítem 9)
	Percepción del docente como evaluador	Juicio sobre la imparcialidad, transparencia y equidad del profesorado en los procesos evaluativos (ítems 6 y 8)
	Percepción de justicia evaluativa	Grado en que el alumnado percibe que el profesorado evalúa de forma coherente, transparente e imparcial (ítem 26)
	Evaluación diagnóstica	Presencia de mecanismos para conocer el nivel previo del alumnado antes de iniciar un nuevo contenido (ítem 11)
Instrumentos y momentos de la evaluación	Evaluación continua	Regularidad y seguimiento sistemático del aprendizaje a lo largo del proceso: revisión de contenidos previos (ítem 16) y calificación de lo trabajado en la sesión (ítem 17)
	Evaluación final	Valoración global del aprendizaje una vez finaliza el proceso: examen final (ítem 13) y test físico final (ítem 14)
	Técnicas e instrumentos de evaluación	Diversidad y adecuación de los métodos utilizados para recopilar información sobre el aprendizaje: evaluación escrita (ítem 19) y explicación de criterios al inicio del periodo (ítem 12)
	Contenidos conceptuales	Evaluación de conocimientos teóricos vinculados a la educación física: criterios de evaluación por sesión (ítem 15)
Contenidos evaluados	Contenidos procedimentales	Evaluación del desarrollo práctico y motriz del alumnado: práctica (ítem 18), demostraciones (ítem 23) y correcciones (ítem 24)
	Contenidos actitudinales	Evaluación de actitudes, comportamientos y valores expresados durante la práctica: participación (ítem 20), comportamiento (ítem 22), uso del uniforme (ítem 21) y reconocimiento del esfuerzo (ítem 25)

Nota: Entre paréntesis se indica el ítem o ítems correspondiente del cuestionario correspondiente a cada indicador.

4.2. Fases de la Investigación

El diseño de toda investigación educativa exige una estructuración sistemática y justificada del proceso a seguir, de modo que las decisiones metodológicas adoptadas respondan a una lógica interna coherente orientada a dar respuesta a la pregunta de investigación y a los objetivos planteados (Bisquerra, 2022). La literatura científica sobre metodología de investigación educativa propone diferentes modelos de fases o etapas que permiten organizar el proceso investigador de forma rigurosa. Latorre et al. (2021) identifican en todo proceso de investigación educativa una secuencia que incluye, como mínimo, una fase de planteamiento del problema, una fase de recogida de información, una fase de análisis e interpretación de los datos y una fase de elaboración de conclusiones. En una línea similar, Cohen et al. (2002) proponen que cualquier diseño de investigación descriptiva debe articularse en torno a una fase de diagnóstico inicial, una fase de recogida sistemática de datos, una fase de análisis comparativo y una fase de síntesis e interpretación de resultados. Creswell & Creswell (2017), por su parte, subrayan que en los diseños de investigación de métodos mixtos, como el que se adopta en el presente estudio, la secuenciación de fases debe explicitar con claridad el orden en que se recogen y analizan los datos de naturaleza cuantitativa e interpretativa, así como el momento en que se produce la integración de ambos tipos de información.

A partir de estos referentes metodológicos y atendiendo a las características específicas del presente estudio, (que combina un diagnóstico observacional previo, la administración de un cuestionario en un nuevo contexto, la triangulación con los datos del TFM anterior y el diseño de una propuesta de implementación), se propone una estructura de siete fases que articula de forma coherente el proceso investigador. Esta propuesta adapta los modelos de referencia citados a la lógica específica del estudio, respondiendo a la pregunta central de la investigación. Cada fase se describe en términos de su objetivo específico, la técnica e instrumento empleados y la intención que la justifica dentro del conjunto del proceso, tal y como se refleja en la Tabla 2.

Tabla 2

Fases de la Investigación

Fase	Objetivo	Técnica e instrumento	Intención
Fase 1: Exploración y diagnóstico	Identificar y caracterizar el contexto evaluativo del centro como predominantemente calificador e	Lista de control (instrumento de observación estructurado)	Justificar empíricamente la pertinencia del estudio y establecer el punto de partida comparativo con el TFM

inicial	identificar los ítems más sensibles para detectar la ausencia de evaluación formativa		precedente
Fase 2: Planificación y diseño	Seleccionar el instrumento de recogida de datos y garantizar su equivalencia con el estudio precedente	Análisis documental del TFM anterior	Asegurar la coherencia metodológica entre ambos estudios como condición de validez comparativa
Fase 3: Recogida de datos	Administrar el cuestionario al alumnado del centro de estudio	Cuestionario validado de Palacio et al. (2021) con escala Likert de 28 ítems	Obtener datos comparables con los del TFM anterior para posibilitar la triangulación
Fase 4: Análisis y comparación de resultados	Analizar los datos obtenidos y contrastarlos con los del TFM anterior mediante triangulación	Análisis estadístico descriptivo y análisis documental comparativo del TFM anterior	Identificar convergencias y divergencias en la percepción del alumnado entre ambos contextos
Fase 5: Discusión e interpretación	Contrastar los resultados comparativos con la literatura científica del marco teórico para interpretar su alcance y significado	Síntesis interpretativa de los resultados comparativos en diálogo con la literatura científica	Fundamentar la propuesta de implementación sobre una base empírica e interpretativa sólida
Fase 6: Propuesta de implementación	Diseñar una propuesta de implementación progresiva de evaluación formativa contextualizada en el centro estudiado	Síntesis de evidencias empíricas obtenidas y literatura científica especializada	Traducir los hallazgos comparativos en orientaciones prácticas, graduadas y fundamentadas para el centro
Fase 7: Conclusiones y reflexión	Dar respuesta directa a los objetivos específicos y reflexionar críticamente sobre las limitaciones del estudio	Síntesis interpretativa del conjunto del proceso investigador	Cerrar el proceso investigador, evaluar el alcance de los resultados y abrir futuras líneas de interpretación

La Fase 1, de exploración y diagnóstico inicial, conforma el punto de partida empírico de la investigación. Durante un primer periodo se aplicó un instrumento de observación estructurado que permitió sistematizar las evidencias recogidas sobre las prácticas evaluativas del área de EF en el centro estudiado. El instrumento de observación estructurado empleado en esta fase se recoge íntegramente en el Anexo 1, donde puede consultarse la lista de control utilizada, los indicadores observados y las evidencias registradas que fundamentan la caracterización del contexto como predominantemente calificador. Los resultados de dicho instrumento evidenciaron que la evaluación es concebida predominantemente como un mecanismo calificador, con la presencia marginal de estrategias orientadas a la autorregulación del aprendizaje o a la participación activa del alumnado en su propia evaluación. A partir de este diagnóstico inicial, y en contraste con los datos aportados por el TFM precedente (Maciá López, 2025), se procedió a identificar qué ítems del cuestionario de Palacio et al. (2021) resultan más sensibles para detectar la ausencia de prácticas evaluativas formativas en el nuevo contexto.

La Fase 2, de planificación y diseño, tiene como propósito central garantizar la coherencia metodológica entre el presente estudio y el estudio anterior (Maciá López, 2025).

Para ello se realiza un análisis documental del estudio anterior, entendido como técnica de investigación que permite extraer información sistemática y rigurosa a partir de documentos ya existentes (Bisquerra, 2022). Este análisis documental cumple una doble función: por un lado, permite verificar que el instrumento de recogida de datos es el mismo en ambos estudios, condición necesaria para la validez comparativa; y por otro lado, permite identificar los resultados más relevantes del estudio precedente que actuarán como referente en la fase de comparación.

La Fase 3, de recogida de datos, consiste en la administración del cuestionario validado de Palacio et al. (2021) al alumnado del centro predominantemente calificador. El procedimiento de aplicación, las condiciones de recogida y las consideraciones éticas relativas al tratamiento de datos de menores de edad se detallan en el apartado de aplicación. La Fase 4, de análisis y comparación de resultados, supone el núcleo metodológico del estudio. Los datos obtenidos mediante el cuestionario se someten a un análisis estadístico de carácter descriptivo y, a continuación, se contrastan con los resultados del estudio anterior a través de un proceso de triangulación de datos (Creswell & Creswell, 2017). La triangulación, entendida como el uso de múltiples fuentes de información para contrastar y enriquecer la interpretación de los hallazgos, permite en este caso identificar las convergencias y divergencias más significativas en la percepción del alumnado entre el contexto formativo y el calificador, dotando al estudio de una solidez interpretativa que trasciende el análisis de un caso único (Latorre et al., 2021).

La Fase 5, de discusión e interpretación, es el puente necesario entre los resultados comparativos obtenidos y la propuesta de implementación que los sucede. En esta fase, los hallazgos derivados de la triangulación se contrastan sistemáticamente con la literatura científica revisada en el marco teórico, con el objetivo de interpretar su alcance y significado en relación con los principios de la evaluación formativa y con las evidencias disponibles sobre los contextos evaluativos calificadores. Este proceso de discusión es una condición imprescindible para que la propuesta de implementación posterior no se limite a una descripción de diferencias entre contextos sino que se fundamente sobre una interpretación rigurosa de su significado pedagógico.

La Fase 6, de propuesta de implementación progresiva, traduce los hallazgos empíricos y las interpretaciones derivadas de la fase anterior en orientaciones concretas, graduadas y contextualizadas para la transformación progresiva de las prácticas evaluativas en el centro

estudiado. Esta fase se nutre tanto de los resultados de la comparación entre contextos como de la literatura científica revisada en el marco teórico, articulando así la dimensión empírica y la dimensión teórica del estudio en una propuesta de carácter aplicado. Finalmente, la Fase 7, de conclusiones y reflexión, cierra el proceso investigador dando respuesta directa a los objetivos específicos planteados, evaluando el alcance y las limitaciones del estudio y abriendo perspectivas para futuras líneas de investigación. Esta fase responde al principio metodológico que afirma que las conclusiones de toda investigación educativa deben guardar una correspondencia explícita y verificable con los objetivos que la orientan (Bisquerra, 2022; Latorre et al., 2021).

4.3. Instrumento

El presente estudio emplea tres instrumentos de recogida y análisis de información que se articulan de forma complementaria a lo largo de las diferentes fases de la investigación, tal y como se recoge en la Tabla 2: una lista de control aplicada durante un periodo inicial de observación, el análisis documental del trabajo anterior y un cuestionario administrado al alumnado del centro estudiado. La utilización de múltiples instrumentos responde a la lógica del diseño metodológico mixto adoptado y permite abordar el objeto de estudio desde perspectivas complementarias, dotando al proceso de triangulación de una mayor solidez interpretativa (Cohen et al., 2002; Creswell & Creswell, 2017).

La lista de control constituye el primer instrumento empleado en la investigación y corresponde a la Fase 1 de exploración y diagnóstico inicial. Se trata de un instrumento de observación estructurada aplicada durante un periodo de observación previo, diseñado para recoger de forma sistemática evidencias sobre las prácticas innovadoras del centro, y específicamente, sobre el uso de la evaluación como herramienta de aprendizaje en el área de Educación Física. La observación estructurada, entendida como la recogida sistemática de información sobre la realidad educativa a través de categorías e indicadores predefinidos, conforma una técnica ampliamente reconocida en la investigación educativa de carácter interpretativo por su capacidad para generar evidencias contextualizadas y directamente vinculadas a la práctica real del aula (Bisquerra, 2022; Cohen et al., 2002).

El indicador directamente relevante para el presente estudio es el número 5: *“Utiliza la evaluación como herramienta de autorregulación del aprendizaje”*, que obtuvo una valoración

negativa, con la evidencia registrada de que la evaluación es concebida predominantemente como una herramienta calificadora y no como un instrumento al servicio del aprendizaje. Este resultado compone la evidencia empírica que fundamenta la caracterización del contexto como predominantemente calificador y justifica la pertinencia del presente estudio. El instrumento completo se recoge en el Anexo 1.

El análisis documental del estudio previo integra el segundo instrumento de la investigación. Esta es una técnica de investigación que consiste en la localización, selección, lectura y análisis de documentos con el fin de extraer información relevante para los objetivos del estudio (Bisquerra, 2022; Latorre et al., 2021). En el presente estudio, este instrumento opera en tres fases distintas del proceso investigador: en la Fase 2, para verificar la equivalencia del instrumento de recogida de datos entre ambos estudios; en la Fase 4, como fuente de los datos del contexto formativo que se contrastan con los del nuevo contexto mediante triangulación; y en la Fase 5, como referente interpretativo para la discusión de las convergencias y divergencias detectadas entre ambos contextos.

El cuestionario supone el tercer instrumento de la investigación y el principal dispositivo de recogida de datos en la Fase 3. Su utilización se justifica por razones tanto metodológicas como contextuales. Desde el punto de vista metodológico, el cuestionario estandarizado constituye el único instrumento que garantiza la comparabilidad sistemática de las respuestas entre dos grupos pertenecientes a contextos diferentes, condición que resulta indispensable para la validez del diseño comparativo adoptado (Bisquerra, 2022). Desde el punto de vista contextual, el anonimato que ofrece el cuestionario resulta especialmente pertinente en contextos predominantemente calificadores, donde el alumnado puede tender a ajustar sus respuestas a las expectativas percibidas del docente, reduciendo así el sesgo de deseabilidad social que podría comprometer la veracidad de los datos (Palacio et al., 2021).

El instrumento empleado es el cuestionario ya validado de Palacio et al. (2021), diseñado específicamente para recoger las concepciones y percepciones del alumnado sobre la evaluación en Educación Física. Su elección responde a una exigencia de coherencia metodológica con el estudio precedente (Maciá López, 2025). Dado que el objeto del presente estudio es la comparación entre dos contextos evaluativos contrastados, la utilización del mismo cuestionario en ambos estudios es una condición indispensable para garantizar la validez comparativa de los resultados. El cuestionario consta de 28 ítems en total: 26 ítems originales del instrumento

validado que conforman la escala de medición de la percepción evaluativa, y 2 ítems adicionales de caracterización sociodemográfica (referentes al curso al que pertenece el alumnado y si ha cursado todos los cursos de la ESO en el centro), cuya función es exclusivamente descriptiva y que no forma parte de la escala de medición. Los 26 ítems de la escala se responden con opciones de respuesta tipo Likert de cinco niveles: los ítems del 1 al 5 con la escala (1) Muy en desacuerdo, (2) En desacuerdo, (3) Indiferente, (4) De Acuerdo y (5) Muy de acuerdo; y los ítems 6 a 26 con la escala (1) Nunca, (2) Casi nunca, (3) A veces, (4) Casi siempre y (5) Siempre. La distribución de estos 26 ítems de la escala por dimensión se muestran en la Tabla 3.

Tabla 3

Dimensiones del Cuestionario y sus Respectivos Ítems

Nº	Ítem	Dimensión
1	La evaluación que realiza el profesor aporta a tu aprendizaje	Concepciones sobre la evaluación
2	La calificación es importante para ti	
3	La evaluación sirve para medir tu nivel de aprendizaje	
4	Te parece importante la evaluación que realiza el profesor	
5	Tus compañeros(as) de clase aprenderían más, si tu haces parte de su evaluación	
6	La calificación que te pone el profesor refleja tu nivel de aprendizaje	Agentes implicados en la evaluación
7	El profesor permite que te evalúes tu mismo	
8	El profesor les pregunta cómo les pareció la clase	
9	Evalúas a tu profesor	
10	Tus compañeros(as) de clase hacen parte de tu evaluación	
26	El profesor es justo cuando te evalúa	Instrumentos y momentos de evaluación
11	Antes de empezar un nuevo tema, el profesor les pregunta qué saben del tema	
13	Al final del periodo el profesor realiza un examen para ver qué han aprendido	
14	Al final del periodo el profesor realiza un test físico para ver que han mejorado	
16	En cada clase el profesor pregunta lo que vieron en la clase anterior	
17	Al finalizar la clase el profesor califica lo trabajado en la sesión	
19	El profesor te evalúa de forma escrita (cuaderno, talleres)	Instrumentos y momentos de evaluación
12	Al inicio del periodo el profesor les explica qué va a evaluar	

15	Al inicio de la clase el profesor les dice los criterios de evaluación que van a tener en clase	Contenidos
18	El profesor te evalúa lo que haces en la práctica (ejercicios, juegos, deportes, etc.)	
20	El profesor evalúa tu participación en clase	
21	La calificación de la clase depende de llevar correctamente el uniforme	
22	La calificación de la clase depende de tu comportamiento durante las actividades	
23	El profesor te solicita demostrar los ejercicios durante la clase	
24	El profesor te hace correcciones durante la clase	
25	El profesor te felicita cuando haces algo bien hecho	

<https://forms.gle/YkfHvyDy2riLpRDG8>

Con el fin de garantizar la consistencia interna del cuestionario, se calculó el coeficiente Alfa de Cronbach a partir de los 26 ítems originales del instrumento, obteniéndose un valor de (.838), que indica un nivel de fiabilidad alto (Cervantes, 2005; Tavakol & Dennick, 2011). Valor el cual es casi idéntico al obtenido en anterior estudio, lo que confirma que el instrumento mantiene su consistencia interna en un contexto evaluativo diferente y refuerza la solidez metodológica del diseño comparativo. Además, para profundizar en el análisis de la consistencia interna del instrumento, la Tabla 4 recoge las estadísticas de fiabilidad para cada uno de los 26 ítems de la escala, mostrando el valor de alfa de Cronbach que se obtendría si cada elemento fuera eliminado, así como la correlación elemento-total corregida de cada ítem. Los resultados confirman la solidez del instrumento en el nuevo contexto: los valores de alfa de Cronbach si se elimina el elemento oscilan entre (.829) y (.843), un rango muy estrecho y próximo al valor global de (.838), lo que indica que ningún ítem compromete ni mejora sustancialmente la consistencia interna de la escala y que todos contribuyen de forma homogénea a la medición del constructo. En cuanto a la correlación elemento-total corregida, la mayoría de los ítems presentan valores moderados o altos, siendo los ítems 5 (coevaluación entre compañeros/as) y 19 (evaluación escrita) los que muestran una correlación más baja (.155) y (.193) respectivamente, resultado que concuerda con los obtenidos en el estudio precedente y que refleja la menor homogeneidad de estos ítems respecto al resto de la escala, sin comprometer en ningún caso la fiabilidad global del instrumento. En conjunto, estos resultados confirman que el cuestionario de

Palacio et al. (2021) mantiene unas propiedades psicométricas equivalentes en el nuevo contexto, lo que valida su uso como instrumento comparativo entre ambos estudios.

Tabla 4

Estadísticas de Valoración Total si se Elimina un Elemento

Ítem	Alfa de Cronbach si el elemento se ha suprimido	Correlación elemento-total corregida
1	.834	.436
2	.839	.228
3	.834	.398
4	.835	.365
5	.842	.155
6	.833	.428
7	.834	.391
8	.832	.460
9	.840	.229
10	.834	.400
11	.833	.431
12	.835	.385
13	.838	.275
14	.840	.279
15	.834	.408
16	.830	.504
17	.829	.539
18	.834	.409
19	.843	.193
20	.830	.524
21	.840	.246
22	.834	.414
23	.837	.333

24	.833	.422
25	.829	.518
26	.830	.533

En cumplimiento del marco legal y sociolingüístico vigente en Cataluña, el cuestionario fue administrado en catalán, de acuerdo con lo establecido por la Llei 1/1998, de 7 de gener, de política lingüística, y la Llei 12/2009, del 10 de juliol, d'educació de Catalunya, que establecen el catalán como lengua vehicular habitual en los centros educativos. El cuestionario fue administrado de forma anónima y con el consentimiento informado de las familias (Anexo 2), de conformidad con el artículo 7 de la Ley orgánica 3/2018, de protección de Datos Personales y garantía de derechos digitales.

4.4. Muestra

La muestra del presente estudio fue seleccionada mediante un muestreo intencional, el cual es un procedimiento de muestreo no probabilístico en el que la selección de los participantes se realiza de forma deliberada en función de criterio específicos directamente vinculados al objeto de estudio (Palinkas et al., 2015). Este tipo de muestreo es el más adecuado para investigaciones de carácter interpretativo y comparativo como esta, ya que permite seleccionar a los participantes que mejor pueden dar respuesta a la pregunta de investigación (Bisquerra, 2022). En este estudio, el criterio de selección fue la pertenencia de la muestra a un centro educativo con un modelo evaluativo predominantemente calificador en el área de EF, caracterización que se estableció en la Fase 1 de exploración mediante el instrumento de observación.

En cuanto al tamaño de la muestra, siguiendo las recomendaciones de Carretero-Dios & Pérez (2005), quienes establecen que para garantizar la fiabilidad de un cuestionario se requieren entre cinco y diez participantes por ítem, el umbral mínimo para un instrumento de 28 ítems se situaría entre 140 y 280 participantes. La muestra final del presente estudio estuvo compuesta por 165 estudiantes, cifra que se encuentra dentro del rango recomendado y que supera la obtenida en el anterior estudio donde participaron 137 estudiantes (Maciá López, 2025). Esta mayor amplitud muestral refuerza la representatividad de los datos obtenidos y la solidez del diseño comparativo.

Los 165 participantes cursan estudios en las etapas de Educación Secundaria Obligatoria (ESO) y Bachillerato, abarcando desde 1º de la ESO hasta 1º de Bachillerato, ya que la asignatura de Educación física no se imparte en 2º de Bachillerato. Esta distribución permite obtener una visión amplia y diversa de la percepción evaluativa del alumnado a lo largo de toda la trayectoria formativa en el centro, abarcando diferentes momentos de desarrollo académico y distintos niveles de experiencia con el modelo evaluativo predominante. Con el fin de contextualizar la muestra, se incorporó al cuestionario una pregunta de caracterización referida a la continuidad de la escolarización en el centro: el 81,2% del alumnado participante (n=134) ha cursado toda la etapa en el mismo centro, mientras que el 18,8% restante (n=31) procede de otros centros educativos. Este dato es metodológicamente relevante, ya que la mayoría del alumnado ha estado expuesto de forma continuada al modelo de evaluación del centro a lo largo de toda su escolarización, lo que refuerza la validez de los datos obtenidos como reflejo fiel de la cultura evaluativa predominante en ese contexto.

Dado que la totalidad de los participantes son menores de edad, la participación en el estudio estuvo supeditada a la obtención del consentimiento informado de las familias, en cumplimiento de lo dispuesto en el artículo 7 de la Ley Orgánica 3/2018, de Protección de Datos Personales y garantía de los derechos digitales. La participación fue completamente voluntaria y anónima, y los datos fueron tratados en todo momento garantizando la confidencialidad de los participantes.

5. Aplicación

El proceso de recogida de datos se articuló en torno a los dos instrumentos de recogida de datos primarios del estudio. La lista de control fue aplicada mediante observación directa y sistemática en un primer periodo de observación, registrando de forma manual las evidencias sobre las prácticas evaluativas del área de EF conforme a los indicadores predefinidos del instrumento. La administración del cuestionario al alumnado se realizó en formato digital mediante la herramienta de Google Forms, plataforma que permite centralizar las respuestas de forma inmediata, garantizar el anonimato de los participantes y facilitar la posterior exportación de los datos para su tratamiento estadístico (Cohen et al., 2002). El enlace de acceso fue facilitado al alumnado mediante correo electrónico, siendo cumplimentado de forma individual y voluntaria.

Una vez finalizado el periodo de recogida, las respuestas fueron exportadas desde Google Forms a Microsoft Excel, donde se procedió a la codificación numérica de las opciones de respuesta de la escala Likert. Los datos codificados fueron posteriormente importados al *software* JASP 0.97.1 para el análisis de la consistencia interna del instrumento y el cálculo de las frecuencias, medias y desviaciones típicas que constituyen la base de análisis comparativo presentado en el apartado de resultados.

6. Análisis de los Datos

El análisis de los datos del presente estudio se estructura en tres niveles complementarios que se corresponden con los instrumentos empleados y con las fases del proceso investigador: el análisis estadístico descriptivo de los datos del cuestionario, el análisis documental del estudio precedente y la triangulación de ambas fuentes de información. Este diseño de análisis responde a los principios del paradigma interpretativo adoptado, según el cual la comprensión de la realidad educativa requiere la integración y el contraste de múltiples fuentes de información (Bisquerra, 2022; Creswell & Creswell, 2017).

En primer lugar, los datos obtenidos mediante el cuestionario fueron sometidos a un análisis estadístico de carácter descriptivo mediante el *software* JASP0.97.1, la cual es una herramienta específica de análisis estadístico que permite obtener con rigor los estadísticos descriptivos necesarios para el estudio. Para cada uno de los 26 ítems se calcularon la media, la desviación típica y la distribución de frecuencias de las cinco opciones de respuesta, expresada en términos de porcentajes. Este tipo de análisis descriptivo permite caracterizar de forma precisa y sistemática la percepción del alumnado en cada uno de los ítems y en cada una de las cuatro dimensiones del instrumento, sin pretender establecer relaciones causales ni generalizar los resultados más allá del contexto estudiado (Cohen et al., 2002). Los resultados se presentan mediante tablas de frecuencias y medias, complementadas con representaciones gráficas que facilitan la interpretación visual de los datos y la identificación de patrones de respuesta de cada dimensión. Esta doble presentación sigue la misma estructura que la investigación previa realizada con el fin de facilitar la comparación sistemática entre ambos estudios.

En segundo lugar, se llevó a cabo el análisis documental de la investigación precedente, Maciá López (2025), como fuente de datos del contexto formativo de referencia. El análisis documental es una técnica de investigación sistemática que permite extraer, organizar e

interpretar información relevante a partir de documentos existentes, siendo especialmente adecuada en investigaciones comparativas donde uno de los conjuntos de datos proviene de un estudio previo (Bowen, 2009; Bisquerra, 2022). Latorre et al. (2021) subrayan que esta técnica, correctamente aplicada, proporciona datos tan válidos y rigurosos como los obtenidos mediante recogida directa, siempre que el documento analizado cumpla los criterios de fiabilidad y pertinencia requeridos por el objeto de estudio. En el presente caso, ambos criterios quedan garantizados por el hecho de que el estudio precedente empleó el mismo instrumento validado, con una muestra equivalente y en un contexto claramente definido. De dicho estudio se extrajeron las medidas, desviaciones típicas y distribuciones de frecuencia de los 26 ítems del cuestionario obtenidas con una muestra de 137 estudiantes, que construyeron el referente comparativo con el que se contrastan los datos del presente estudio.

En tercer lugar, el núcleo del análisis lo constituye el proceso de triangulación de datos entre ambas fuentes de información. La triangulación, definida como el uso de múltiples fuentes de datos con el fin de contrastar y enriquecer la interpretación de los hallazgos obtenidos, permite identificar las convergencias y divergencias más significativas en la percepción del alumnado entre el contexto predominantemente calificador y el formativo (Latorre et al., 2021; Creswell & Creswell, 2017). El proceso de triangulación se desarrolla de forma sistemática dimensión a dimensión e ítem a ítem, seleccionando para la comparación aquellos ítems que resultan más sensibles para detectar las diferencias entre los modelos evaluativos. Los resultados comparativos se muestran mediante tablas y gráficos que permiten visualizar de forma clara las diferencias entre ambos contextos, facilitando una lectura interpretativa de los patrones de convergencia y divergencia detectados. Estos hallazgos son posteriormente interpretados con el sustento de la literatura científica en el apartado de discusión, donde se contrasta su significado pedagógico con las evidencias disponibles sobre evaluación formativa y contextos calificadores. Finalmente, las conclusiones derivadas de este proceso analítico e interpretativo constituyen la base sobre la que se construye la propuesta de implementación progresiva, que traduce los hallazgos empíricos comparativos en orientaciones concretas, graduadas y contextualizadas para la transformación de las prácticas evaluativas en el centro estudiado.

7. Resultados

Este apartado expone los resultados obtenidos mediante el cuestionario administrado al alumnado del centro, organizado en torno a cuatro dimensiones del modelo de análisis. Para cada dimensión se presenta una tabla comparativa con las medias (M) y desviaciones estándar (DE) de los 26 ítems en ambos contextos (contexto predominantemente calificador, presente estudio con una muestra de 165 y el contexto formativo del estudio anterior realizado por Maciá López (2025), con una muestra de 137), seguida de una descripción sistemática de los resultados de cada ítem con los datos de frecuencia más relevantes y una representación gráfica comparativa. La tabla completa con las medias, desviaciones estándar y distribución de frecuencias de todos los ítems del cuestionario aplicado en el contexto actual de estudio se recogen en el Anexo 3.

7.1. Dimensión 1. Concepciones sobre la Evaluación

Esta dimensión recoge la percepción del alumnado sobre el valor, la función y la relevancia de la evaluación en su proceso de aprendizaje, así como su disposición hacia la evaluación compartida (Tabla 5).

Tabla 5

Comparación Resultados Medias y Desviación Estándar Dimensión 1

Ítem	Descripción	Contexto calificador		Contexto formativo	
		M	DE	M	DE
1	La evaluación que realiza el profesor aporta a tu aprendizaje	3.97	0.69	4.02	0.88
2	La calificación es importante para ti	4.30	0.73	4.36	0.74
3	La evaluación sirve para medir tu nivel de aprendizaje	3.85	0.88	3.85	0.95
4	Te parece importante la evaluación que realiza el profesor	4.10	0.83	4.23	0.83
5	Tus compañeros(as) de clase aprenderían más, si tu haces parte de su evaluación	3.07	0.79	2.96	0.85

Nota. **M**=media; **DE**= desviación estándar; **1:** Muy en desacuerdo; **2:** En desacuerdo; **3:** Indiferente; **4:** De acuerdo; **5:** Muy de acuerdo.

El rasgo más destacable de esta dimensión es la notable convergencia entre ambos contextos, ya que ninguno de los cinco ítems presenta una diferencia de medias superior a 0.13 puntos y las desviaciones estándar son reducidas en la mayoría de los casos, lo que indica respuestas homogéneas en los dos grupos. Esta convergencia es especialmente llamativa entre los ítems que miden la valoración general de la evaluación docente y la importancia de la calificación, que obtienen las medias más elevadas de la dimensión en ambos contextos.

La calificación como referente central del proceso evaluativo se confirma en el ítem 2, que registra la media más alta de la dimensión tanto en el contexto predominantemente calificador ($M=4.30$; $DE=0.73$) como en el contexto formativo ($M=4.36$; $DE=0.74$), con el 87,2% del alumnado del contexto calificador posicionándose en los niveles “de acuerdo” y “muy de acuerdo” y ninguna respuesta en el nivel más bajo de la escala. En sintonía con este resultado, los ítems 1 y 4, referidos respectivamente a si la evaluación del profesor aporta al aprendizaje y a si esta resulta importante para el alumnado, obtienen medias igualmente elevadas en el contexto calificador ($M=3.97$; $DE=0.69$ y $M=4.10$; $DE=0.83$ respectivamente), con el 80% y el 84,2% de las respuestas concentradas en los niveles 4 y 5. En ambos casos las diferencias con el contexto formativo son mínimas, 0.05 y 0.13 puntos respectivamente, siendo el ítem 4 el que registra la mayor diferencia de toda la dimensión. Resulta destacable que la desviación estándar del ítem 1 en el contexto formativo ($DE=0.88$) es notablemente superior a la del calificador ($DE=0.69$), lo que indica una mayor dispersión de respuestas en el contexto formativo a pesar de obtener una media ligeramente más alta.

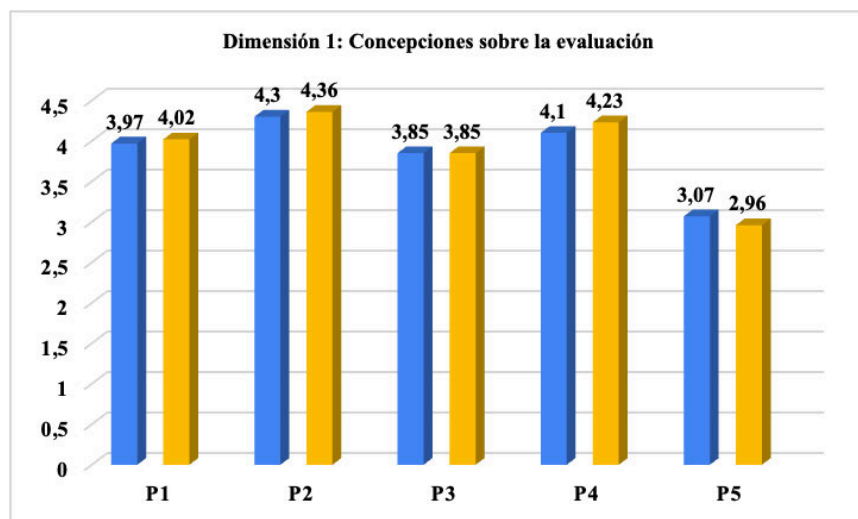
El ítem 3, referido a si la evaluación sirve para medir el nivel de aprendizaje, es el único que obtienen exactamente la misma media en ambos contextos, 3.85, con desviaciones estándar muy similares ($DE=0.88$ en el calificador y $DE=0.95$ en el formativo) y con el 73,3% de las respuestas del contexto calificador situadas en los niveles 4 y 5, lo que muestra un patrón de respuesta prácticamente idéntico al del resto de ítems de valoración positiva de esta dimensión.

El único ítem que rompe con la tónica de medias elevadas y convergencia entre contextos es el ítem 5, que hace referencia a si el alumnado cree que sus compañeros de clase aprenderían más si estos forman parte de su evaluación. Con una media de 3.07 ($DE=0.79$) en el contexto calificador y de 2.96 ($DE=0.85$) en el formativo. Este ítem registra las medias más bajas de la dimensión en los dos grupos y una distribución de respuestas marcadamente diferenciada al resto: el 63% del alumnado del contexto calificador se sitúa en el nivel “indiferente”, con tal solo

el 21,2% en los niveles “de acuerdo” y “muy de acuerdo”. Cabe señalar, además, que este es el único ítem de la dimensión en que el contexto calificador obtiene una media ligeramente superior al formativo, con una diferencia de 0.11 puntos.

Figura 1

Comparación de Medias de la Dimensión 1 entre Ambos Contextos



7.2. Dimensión 2. Agentes Implicados en la Evaluación

Esta dimensión recoge la percepción del alumnado sobre quién participa en los procesos evaluativos: el papel del docente como agente evaluador, el grado de implicación del propio alumnado mediante autoevaluación y coevaluación, la posibilidad de evaluar al profesorado y la percepción de justicia en la evaluación docente. Los resultados comparativos de esta dimensión se muestran en la Tabla 6.

Tabla 6

Comparación Resultados Medias y Desviación Estándar Dimensión 2

Ítem	Descripción	Contexto calificador		Contexto formativo	
		M	DE	M	DE
6	La calificación que te pone el profesor refleja tu nivel de aprendizaje	3.58	0.84	3.56	0.88
7	El profesor permite que te evalúes tu mismo	3.39	1.03	3.00	1.08

8	El profesor les pregunta cómo les pareció la clase	2.75	1.03	2.91	1.16
9	Evalúas a tu profesor	1.70	1.00	1.64	1.01
10	Tus compañeros(as) de clase hacen parte de tu evaluación	2.65	1.00	2.36	0.99
26	El profesor es justo cuando te evalúa	4.04	0.87	4.11	0.87

Nota. **M**=media; **DE**= desviación estándar; **1**: Nunca; **2**: Casi nunca; **3**: A veces; **4**: Casi siempre; **5**: Siempre.

A diferencia de la dimensión 1, los resultados de esta dimensión muestran un perfil de respuestas más heterogéneo, con medias que oscilan entre 1.70 y 4.04 en el contexto calificador, y con patrones de distribución que varían de forma notable entre los ítems. El rasgo más destacado es la coexistencia de una percepción muy positiva del docente como agente evaluador justo (ítem 26) con valores muy bajos en los ítems que miden la participación del alumnado en el proceso evaluativo, especialmente en lo que respecta a la evaluación del profesorado (ítem 9) y la coevaluación entre iguales (ítem 10).

La percepción de justicia evaluativa (ítem 26) registra la media más elevada en ambos contextos, 4.04 (DE=0.87) en el calificador y 4.11 (DE=0.87) en el formativo, con una diferencia de tan solo 0.07 puntos y desviaciones estándar idénticas, lo que indica una homogeneidad equivalente en los dos contextos. En sintonía con este resultado, el ítem 6 sobre si la calificación del profesorado refleja el nivel de aprendizaje del alumnado obtiene una media de 3.58 (DE=0.84) en el calificador y de 3.56 (DE=0.88) en el formativo, con una diferencia de 0.02 que la convierte en la diferencia más pequeña de toda la dimensión. Ambos ítems describen conjuntamente una percepción generalizada de confianza en la actuación del docente como agente evaluador, que se mantiene de forma prácticamente idéntica en ambos contextos.

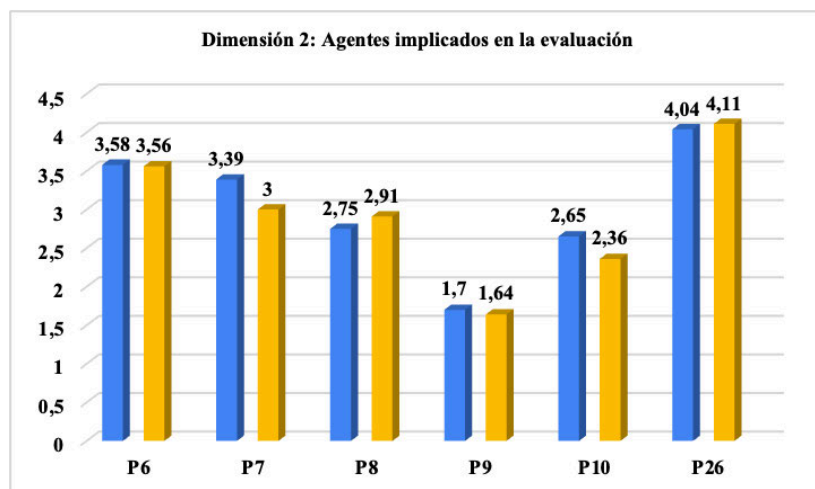
En contraste con esta valoración positiva del docente, los ítems que miden la participación activa del alumnado en los procesos evaluativos presentan medias notablemente más bajas y diferencias más pronunciadas entre contextos. El ítem 7, referido a si el profesor permite al alumnado que se evalúe a sí mismo, obtiene una media de 3.39 (DE=1.03) en el contexto calificador, 0.39 puntos superior a la del contexto formativo (M=3.00;DE=1.08), siendo la mayor diferencia de la dimensión y una de las más significativas del conjunto del cuestionario. La elevada desviación estándar en ambos contextos (superior a 1.00) refleja una notable

dispersión de respuestas, con el alumnado posicionándose de forma muy heterogénea ante esta práctica evaluativa. En la misma línea, el ítem 8 sobre si el profesor pregunta al alumnado cómo les pareció la clase obtiene una media de 2.75 (DE=1.03) en el calificador, ligeramente inferior a la del formativo (M=2.91; DE=1.16), con una diferencia de 0.16. En ambos contextos la media se sitúa por debajo del valor central de la escala, con una disposición elevada que indica respuestas muy variables entre el alumnado.

Los resultados más bajos de la dimensión corresponden a los ítems que miden las formas de evaluación más participativas. El ítem 10, referido a si los compañeros forman parte de la evaluación del alumnado, obtiene una media de 2.65 (DE=1.00) en el calificador frente a 2.36 (DE=0.99) en el formativo, con una diferencia de 0.29 puntos favorable al contexto calificador, con el 53,9% de las respuestas en “a veces” y un 35,8% en los niveles “nunca” y “casi nunca”. El ítem 9, que hace referencia a si el alumnado evalúa al profesor, registra la media más baja de toda la dimensión y del conjunto del cuestionario en ambos contextos: 1.70 (DE=1.00) en el calificador y 1.64 (DE=1.01) en el formativo, con una diferencia de 0.06 puntos. En el contexto calificador, el 59,4% del alumnado responde “nunca” y el 18,8% “casi nunca”, acumulando el 78,2% de las respuestas en los dos niveles más bajos de la escala. Las medias y distribuciones prácticamente idénticas en ambos contextos confirman que la evaluación del profesorado por parte del alumnado es una práctica ausente de forma generalizada con independencia del modelo evaluativo predominante.

Figura 2

Comparación de Medias de la Dimensión 2 entre Ambos Contextos



7.3. Dimensión 3. Instrumentos y Momentos de Evaluación

Esta dimensión recoge la percepción del alumnado sobre los momentos en que se lleva a cabo la evaluación y los instrumentos y técnicas que el profesorado emplea para recoger información sobre el aprendizaje (Tabla 7).

Tabla 7

Comparación Resultados Medias y Desviación Estándar Dimensión 3

Ítem	Descripción	Contexto calificador		Contexto formativo	
		M	DE	M	DE
11	Antes de empezar un nuevo tema, el profesor les pregunta qué saben del tema	3.84	1.06	3.99	0.97
13	Al final del periodo el profesor realiza un examen para ver qué han aprendido	4.66	0.79	2.99	1.39
14	Al final del periodo el profesor realiza un test físico para ver que han mejorado	3.51	1.26	2.80	1.41
16	En cada clase el profesor pregunta lo que vieron en la clase anterior	3.21	1.06	2.96	1.00
17	Al finalizar la clase el profesor califica lo trabajado en la sesión	3.55	0.97	3.22	1.17
19	El profesor te evalúa de forma escrita (cuaderno, talleres)	3.36	1.12	2.67	1.21

Nota. **M**=media; **DE**= desviación estándar; **1**: Nunca; **2**: Casi nunca; **3**: A veces; **4**: Casi siempre; **5**: Siempre.

Esta dimensión presenta el perfil comparativo más diferenciado de todo el cuestionario. A diferencia de las dimensiones anteriores, las medias del contexto calificador superan de forma sistemática a las del formativo en cinco de los seis ítems, con diferencias que oscilan entre 0.25 y 1.67 puntos. El único ítem en que el contexto formativo obtiene una media superior es el 11, referido a la evaluación diagnóstica inicial. El patrón resultante muestra un contexto en el que la evaluación se concreta de forma muy marcada en los momentos finales del periodo y en instrumentos de carácter sumativo, con una presencia más limitada de prácticas de seguimiento continuo del aprendizaje.

El dato más destacado de la dimensión es el registrado por el ítem 13, referente a la realización de un examen al final del periodo para comprobar lo aprendido. Con una media de

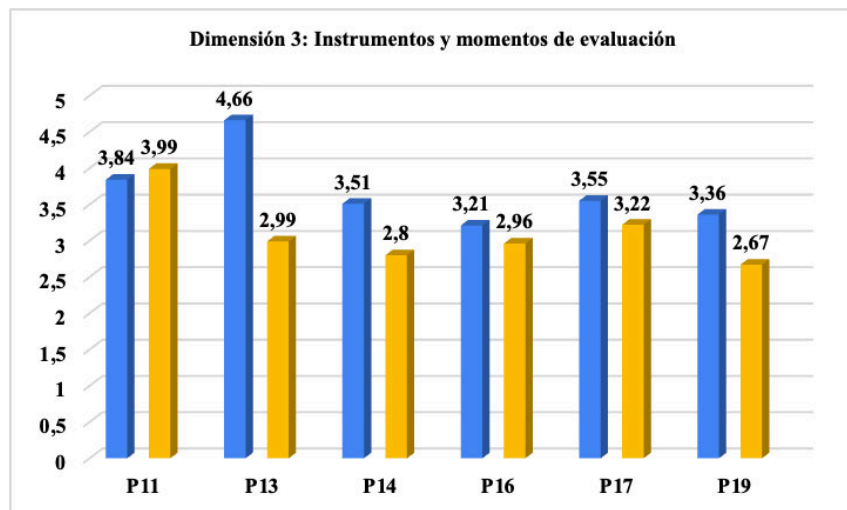
4.66 (DE=0.79) en el contexto calificador, este ítem presenta la diferencia más elevada de todo el estudio respecto al contexto formativo ($M=2.99$; $DE=1.39$), con una diferencia de 1.67 puntos. En el contexto calificador, el 80,6% del alumnado responde “siempre” y el 8,5% “casi siempre”, acumulando el 89,1% de las respuestas en los dos niveles más altos de la escala. Por su parte, en el contexto formativo las respuestas se distribuyen de forma mucho más dispersa a lo largo de toda la escala, lo que se refleja en una desviación estándar de 1.39, notablemente superior a la del calificador (DE=0.79). En la misma línea, el ítem 14, que trata la realización de tests físicos al final del periodo, obtiene una media de 3.51 (DE=1.26) en el calificador frente a 2.80 (DE=1.41) en el formativo, con una diferencia de 0.71 puntos. Juntos, estos dos ítems configuran el rasgo más distintivo del contexto calificador: una evaluación final de carácter sumativo sistemáticamente presente en la percepción del alumnado.

En el extremo opuesto de la dimensión se sitúa el único ítem en el que el contexto formativo supera al calificador: el ítem 11, referido a si el profesorado pregunta al alumnado qué saben sobre el tema antes de empezar. Con una media de 3.84 (DE=1.06) en el calificador y de 3.99 (DE=0.97) en el formativo, siendo la diferencia de tan solo 0.15 puntos y ambas medidas se sitúan próximas al nivel de “casi siempre”. La ligeramente mayor homogeneidad del contexto formativo (DE=0.97 frente a 1.06) sugiere una práctica de evaluación diagnóstica más consolidada y regular en ese contexto.

Los ítems 16, 17 y 19 completan el perfil de la dimensión con diferencias moderadas entre contextos, todas ellas favorables al calificador. El ítem 16, que hace referencia a la pregunta del profesor al inicio de la clase por lo visto en la sesión anterior, obtiene una media de 3.21 (DE=1.06) en el calificador frente a 2.96 (DE=1.00) en el formativo, con una diferencia de 0.25 puntos. El ítem 17, sobre si el profesor califica al finalizar cada sesión lo trabajado, registra una media de 3.55 (DE=0.97) en el calificador y de 3.22 (DE=1.17) en el formativo, con una diferencia de 0.33 puntos y una mayor dispersión de respuestas en el contexto formativo. Finalmente, el ítem 19, referido a si el profesorado evalúa de forma escrita, obtiene una media de 3.36 (DE=1.12) en el calificador frente a 2.67 (DE=1.21) en el formativo, con una diferencia de 0.69 puntos que lo convierte en el tercer ítem con mayor diferencia entre contextos de esta dimensión.

Figura 3

Comparación de Medias de la Dimensión 3 entre Ambos Contextos



7.4. Dimensión 4. Contenidos

Esta dimensión recoge la percepción del alumnado sobre los aspectos del aprendizaje que son objeto de evaluación por parte del profesorado, distinguiendo entre contenidos conceptuales, procedimentales y actitudinales (Tabla 8).

Tabla 8

Comparación Resultados Medias y Desviación Estándar Dimensión 4

Ítem	Descripción	Contexto calificador		Contexto formativo	
		M	DE	M	DE
12	Al inicio del periodo el profesor les explica qué va a evaluar	4.50	0.76	3.72	1.08
15	Al inicio de la clase el profesor les dice los criterios de evaluación que van a tener en clase	3.72	1.24	3.18	1.24
18	El profesor te evalúa lo que haces en la práctica (ejercicios, juegos, deportes,etc.)	4.26	0.73	3.72	0.89
20	El profesor evalúa tu participación en clase	4.22	0.82	4.12	0.97
21	La calificación de la clase depende de llevar correctamente el uniforme	4.09	1.05	2.86	1.23

22	La calificación de la clase depende de tu comportamiento durante las actividades	4.18	0.90	4.09	0.97
23	El profesor te solicita demostrar los ejercicios durante la clase	3.29	0.99	3.03	1.02
24	El profesor te hace correcciones durante la clase	3.41	0.95	3.46	1.02
25	El profesor te felicita cuando haces algo bien hecho	3.35	1.14	3.86	1.02

Nota. **M**=media; **DE**= desviación estándar; **1**: Nunca; **2**: Casi nunca; **3**: A veces; **4**: Casi siempre; **5**: Siempre

Esta dimensión presenta, junto a la dimensión 3, uno de los perfiles comparativos más diferenciados del cuestionario, si bien con un sentido inverso: en siete de los nueve ítems en contexto calificador registra medias superiores al formativo, con diferencias que oscilan entre 0.09 y 1.23 puntos, mientras que únicamente dos ítems (24 y 25) presentan medias ligeramente superiores en el contexto formativo. El patrón resultante describe un contexto en el que el alumnado percibe de forma muy consistente que se le explican los criterios de evaluación, que se evalúa su participación, su comportamiento y su indumentaria, mientras que las prácticas más vinculadas al refuerzo positivo individualizado presentan una posición algo más moderada en comparación con el contexto formativo.

El dato más destacado de esta dimensión corresponde al ítem 21, referido a si la calificación de la clase depende de llevar correctamente el uniforme. Con una media de 4.09 (DE=1.05) en el contexto calificador, este ítem presenta la segunda diferencia más elevada de todo el estudio respecto al contexto formativo (M=2.86; DE=1.23), con una diferencia de 1.23 puntos. En el contexto calificador, el 72,8% del alumnado se sitúa en los niveles “casi siempre” y “siempre”, frente a tan solo el 7,2% que responde “nunca” o “casi nunca”. La desviación estándar, más elevada en el contexto formativo, indica además una mayor dispersión de respuestas en dicho contexto en torno a este ítem.

El ítem 12, relativo a si el profesorado explica al inicio del periodo qué va a evaluar, registra la media más alta de toda la dimensión en el contexto calificador (M=4.50; DE=0.76), con una diferencia de 0.78 puntos respecto al contexto formativo (M=3.72; DE=1.08). El 87,2% de las respuestas del contexto calificador se concentran en los niveles 4 y 5, y la desviación estándar es la más baja de toda la dimensión, lo que indica una percepción muy homogénea del

alumnado en torno a la comunicación inicial de los criterios de evaluación. En cambio, el ítem 15, que traslada esta misma cuestión al ámbito de cada sesión concreta, obtiene una media notablemente inferior ($M=3.72$; $DE=1.24$), aunque sigue siendo superior a la del contexto formativo ($M=3.18$; $DE=1.24$) en 0.54 puntos. En el contexto calificador, el 57% de las respuestas se sitúan en los niveles 4 y 5, frente a un 15,8% en los niveles 1 y 2, y ambos contextos comparten idéntica desviación estándar ($DE=1.24$), lo que sugiere un patrón de dispersión equivalente en la percepción de esta práctica a nivel de sesión.

Los ítems 18, 20 y 22 muestran un patrón convergente de medias elevadas en ambos contextos, con diferencias moderadas favorables al calificador. El ítem 18, referido a si el profesorado evalúa lo que el alumnado hace en la práctica, obtiene una media de 4.26 ($DE=0.73$) en el calificador frente a 3.72 ($DE=0.89$) en el formativo, con una diferencia de 0.54 puntos y el 85,4% de las respuestas del contexto calificador situadas en los niveles 4 y 5. El ítem 20, sobre si el profesorado evalúa la participación del alumnado en clase, presenta la diferencia más reducida de toda la dimensión entre ambos contextos: 4.22 ($DE=0.82$) en el calificador frente a 4.12 ($DE=0.97$) en el formativo, una diferencia de tan solo 0.10 puntos, con el 80% de las respuestas en los niveles 4 y 5. De forma similar, el ítem 22, referido a si la calificación depende del comportamiento del alumnado durante las actividades, registra una media de 4.18 ($DE=0.90$) en el calificador y de 4.09 ($DE=0.97$) en el formativo, con una diferencia de 0.09 puntos, la segunda más pequeña de la dimensión, y el 79,4% de las respuestas concentradas en los niveles más altos de la escala.

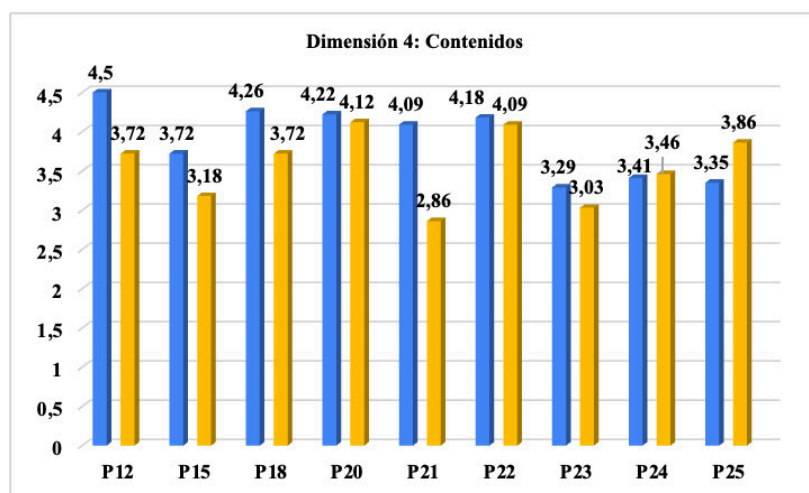
El ítem 23, que hace referencia a si el profesorado solicita al alumnado demostrar los ejercicios durante la clase, obtiene una media de 3.29 ($DE=0.99$) en el calificador frente a 3.03 ($DE=1.02$) en el formativo, con una diferencia de 0.26 puntos. En el contexto calificador, el 41,2% de las respuestas se concentran en el nivel "a veces", mientras que el 40,6% se sitúa en los niveles 4 y 5, lo que indica una distribución relativamente equilibrada entre los niveles centrales y altos de la escala.

Los dos únicos ítems en los que el contexto formativo supera al calificador cierran la dimensión. El ítem 24, sobre si el profesorado realiza correcciones durante la clase, presenta una media de 3.41 ($DE=0.95$) en el calificador frente a 3.46 ($DE=1.02$) en el formativo, con una diferencia de tan solo 0.05 puntos a favor del contexto formativo, la más pequeña de toda la dimensión. En el contexto calificador, el 44,2% de las respuestas se sitúan en el nivel "a veces" y

el 41,9% en los niveles 4 y 5. Finalmente, el ítem 25, referido a si el profesorado felicita al alumnado cuando hace algo bien hecho, registra la diferencia más elevada de la dimensión a favor del contexto formativo: 3.35 (DE=1.14) en el calificador frente a 3.86 (DE=1.02) en el formativo, con una diferencia de 0.51 puntos. En el contexto calificador, el 44,3% de las respuestas se concentran en los niveles 4 y 5, frente a un 18,8% en los niveles 1 y 2, siendo además el ítem con la desviación estándar más elevada de toda la dimensión (DE=1.14).

Figura 4

Comparación de Medias de la Dimensión 4 entre Ambos Contextos



8. Discusión

Los resultados obtenidos mediante la triangulación de los datos del presente estudio con los del estudio anterior de Maciá López (2025) permiten interpretar en qué medida la percepción del alumnado sobre la evaluación en Educación Física varía, o no, en función del contexto evaluativo predominante. La hipótesis de partida planteaba que el alumnado de un contexto predominantemente calificador presentaría una percepción de la evaluación menos orientada al aprendizaje, a la autorregulación y a la mejora continua que el alumnado de un contexto donde se aplica la evaluación formativa. Como muestran los resultados obtenidos, esta hipótesis se confirma sólo parcialmente, ya que las diferencias entre ambos contextos no se distribuyen de manera uniforme entre las cuatro dimensiones del cuestionario, sino que se concentran de forma muy marcada en aspectos concretos relacionados con el momento y la función de la evaluación,

mientras que en otras dimensiones la percepción del alumnado resulta sorprendentemente convergente o incluso favorable al contexto calificador.

El hallazgo más destacado de la dimensión de concepciones sobre la evaluación es precisamente esta convergencia entre contextos. El alumnado de ambos grupos otorga una importancia muy elevada a la calificación ($M=4.30$ en el calificador y $M=4.36$ en el formativo) y valora de forma prácticamente idéntica que la evaluación sirve para medir su nivel de aprendizaje ($M=3.85$ en ambos casos), con diferencias mínimas también en la percepción de que la evaluación del profesorado aporta a su aprendizaje. Esta coincidencia resulta especialmente relevante porque sugiere que la centralidad de la calificación en la experiencia evaluativa del alumnado no es un rasgo exclusivo de los contextos predominantemente calificadores, sino un fenómeno más transversal y arraigado en la cultura evaluativa de la Educación Secundaria. Esta interpretación es coherente con lo señalado por Carless (2015) y por Black & Wiliam (2018), quienes advierten que la asociación entre evaluar y calificar conforman uno de los esquemas más persistentes del pensamiento evaluativo, y con Zubillaga-Olague & Cañadas (2022), que atribuyen a la calificación un papel de legitimación social que ninguna otra práctica evaluativa logra disputar en la Educación Física de secundaria. El hecho de que esta percepción se mantenga prácticamente inalterada incluso en un contexto donde la evaluación formativa se aplica de manera habitual indica que la transformación de las prácticas evaluativas del profesorado no se traduce automáticamente en una transformación equivalente de las concepciones del alumnado sobre el sentido último de la evaluación, lo que conecta con la advertencia de Hortigüela et al. (2019) sobre el riesgo de que la evaluación formativa se implemente de forma superficial, sin que ello modifique sustancialmente la vivencia subjetiva del alumnado respecto a la centralidad de la nota. En esta misma dimensión, el ítem referido a si los compañeros aprenderían más si el alumnado formara parte de su evaluación obtiene las medias más bajas en ambos contextos ($M=3.07$ en el calificador y $M=2.96$ en el formativo), con el 63% de las respuestas del contexto calificador situadas en el nivel “indiferente”. Esta baja valoración inicial de la coevaluación, presente incluso en el contexto formativo, anticipa ya un patrón que se confirma con mayor claridad en la dimensión siguiente.

En relación con el tercer objetivo específico del estudio, que plantea explorar la relación entre la percepción del alumnado sobre la evaluación y su influencia percibida en el aprendizaje, los resultados de esta dimensión ofrecen una primera aproximación relevante. El hecho de que el

alumnado del contexto calificador perciba de forma muy mayoritaria que la evaluación del profesorado aporta a su aprendizaje ($M=3.97$) y que esta sirve para medir su nivel de aprendizaje ($M=3.85$), en un contexto donde, como se discutirá a continuación, predomina una evaluación de carácter marcadamente sumativo y final, sugiere que dicha percepción positiva no presupone necesariamente la existencia de los mecanismos de retroalimentación continua y de uso activo de la información evaluativa que, según Boud & Molloy (2013), resultan imprescindibles para que el ciclo de la evaluación formativa se complete. Es decir, el alumnado puede percibir que la evaluación contribuye a su aprendizaje en un sentido genérico o retrospectivo, sin que ello implique que perciba la evaluación como un proceso que orienta de forma activa y continua su aprendizaje futuro, distinción que resulta central para Carless & Boud (2018) y que será retomada en el análisis conjunto de las dimensiones siguientes.

En relación con los agentes implicados en la evaluación, los resultados muestran un patrón que matiza de forma importante la hipótesis inicial. Por un lado, la percepción de justicia evaluativa del profesorado es elevada y prácticamente equivalente en ambos contextos ($M=4.04$ en el calificador y $M=4.11$ en el formativo), al igual que la percepción de que la calificación refleja el nivel de aprendizaje del alumnado ($M=3.58$ frente a $M=3.56$). Por otro lado, y de forma menos esperable, los ítems referidos a la autoevaluación y a la coevaluación entre compañeros obtienen medias iguales o ligeramente superiores en el contexto calificador ($M=3.39$ frente a $M=3.00$ en autoevaluación, y $M=2.65$ frente a $M=2.36$ en coevaluación), lo que invierte parcialmente el sentido de la hipótesis para estos indicadores concretos. Sin embargo, el dato más relevante de esta dimensión es la práctica ausencia, en ambos contextos, de la evaluación del profesorado por parte del alumnado ($M=1.70$ en el calificador y $M=1.64$ en el formativo), con más del 78% de las respuestas del contexto calificador concentradas en los niveles “nunca” y “casi nunca”. Esta coincidencia resulta especialmente significativa porque sugiere que la escasa implicación del alumnado en determinadas modalidades evaluativas no compone un rasgo diferencial del contexto calificador, sino una barrera de carácter más general que afecta también a contextos donde la evaluación formativa está consolidada. Esta interpretación es coherente con los hallazgos de Zubillaga-Olague & Cañadas (2022), quienes constatan que la heteroevaluación y la heterocalificación predominan de forma abrumadora sobre los procesos de participación del alumnado en Educación Física de secundaria con independencia del modelo evaluativo declarado, y con Grisales et al. (2023), quienes identifican la heteroevaluación y la

autoevaluación como las tendencias predominantes en las aulas, frente a una presencia marginal de la coevaluación, exactamente el patrón observado en ambos contextos del presente estudio. La explicación de este fenómeno puede encontrarse en los planteamientos de Slingerland et al. (2024), quienes evidencian que el profesorado implementa estas estrategias con dificultad debido a la falta de formación específica y de tiempo, una limitación que, de acuerdo con Herrero-González et al. (2021), está directamente condicionada por la formación inicial y permanente recibida y que, por tanto, puede operar de forma transversal e independiente del modelo evaluativo predominante en cada centro. En conjunto, estos resultados sugieren que, si bien el alumnado de ambos contextos confía en la justicia evaluativa del profesorado, su implicación activa como agente evaluador, y en particular su capacidad para evaluar al propio profesorado, sigue siendo una dimensión escasamente desarrollada que la literatura señala como una de las más resistentes al cambio (Barba-Martín & Hortigüela-Alcalá, 2022).

La dimensión de instrumentos y momentos de evaluación es la que ofrece la evidencia más clara a favor de la hipótesis de partida. El hallazgo más destacado de todo el estudio se concentra precisamente en esta dimensión: la realización de un examen al final del periodo para comprobar lo aprendido obtiene una media de 4.66 en el contexto calificador frente a 2.99 en el formativo, con una diferencia de 1.67 puntos, la mayor de todo el cuestionario, y con el 89,1% de las respuestas del contexto calificador concentradas en los niveles “casi siempre” y “siempre”. En la misma línea, la realización de un test físico final para comprobar la mejora ($M=3.51$ frente a $M=2.80$) y la evaluación de forma escrita mediante cuaderno o talleres ($M=3.36$ frente a $M=2.67$) presentan también diferencias sustanciales a favor del contexto calificador. Estos resultados son plenamente coherentes con lo descrito por Ávila et al. (2022), quienes constatan que el profesorado de Educación Física en activo continúa empleando de forma preferente la observación del rendimiento físico y la calificación mediante pruebas como principales instrumentos evaluativos, y con Hortigüela-Alcalá et al. (2021), quienes señalan que el profesorado de secundaria concentra sus prácticas evaluativas en la valoración de la condición física y los tests deportivos, confundiendo de este modo evaluación y calificación. El fuerte peso de la evaluación sumativa final detectado en el contexto calificador representa, por tanto, la materialización empírica de esta tendencia descrita en la literatura. No obstante, resulta especialmente relevante que el único ítem de esta dimensión en que el contexto formativo supera al calificador sea el referido a la evaluación diagnóstica inicial, es decir, si el profesorado

pregunta al alumnado qué sabe sobre un tema antes de empezarlo ($M=3.84$ frente a $M=3.99$), aunque con una diferencia reducida de 0.15 puntos y ambas medias situadas en niveles elevados. Esta circunstancia es coherente con el planteamiento de Cañadas & Santos-Pastor (2021), quienes subrayan que la evaluación diagnóstica, la continua y la sumativa deben operar de forma complementaria para que el conjunto del proceso evaluativo tenga sentido pedagógico, y con Otero-Saborido et al. (2020), quienes constatan que, si bien los currículos de Educación Física reconocen formalmente estos tres momentos, no siempre proporcionan orientaciones suficientes sobre cómo articularlos de forma coherente en la práctica docente. Los datos del contexto calificador estudiado parecen confirmar esta falta de articulación: la evaluación diagnóstica inicial está presente de forma notable, pero queda eclipsada por el peso desproporcionado de la evaluación final, lo que reproduce el desequilibrio entre momentos evaluativos que la literatura científica atribuye a los contextos predominantemente calificadores.

Finalmente, la dimensión de contenidos evaluados aporta una de las interpretaciones más matizadas del conjunto del estudio. Por un lado, el contexto calificador presenta una mayor explicitación de los criterios de evaluación, tanto al inicio del periodo ($M=4.50$ frente a $M=3.72$) como al inicio de cada sesión ($M=3.72$ frente a $M=3.18$), así como una mayor frecuencia percibida de evaluación de la práctica motriz ($M=4.26$ frente a $M=3.72$). Esta mayor transparencia de los criterios resulta, en principio, coherente con uno de los principios fundamentales de la evaluación formativa, según el cual el alumnado debe conocer y comprender los criterios que se aplican en su evaluación (Carless & Boud, 2018; López-Pastor & Pérez-Pueyo, 2017), y con los hallazgos de Tapia et al. (2022), quienes evidencian que el uso de rúbricas y criterios compartidos con el alumnado mejora la calidad del aprendizaje y su nivel de implicación en el proceso evaluativo. Sin embargo, esta mayor explicitación de criterios convive en el contexto calificador con dos rasgos que apuntan en sentido contrario. En primer lugar, la calificación de la clase en función de llevar correctamente el uniforme presenta la segunda diferencia más elevada de todo el estudio ($M=4.09$ frente a $M=2.86$, una diferencia de 1.23 puntos), lo que ilustra de forma muy concreta lo señalado por Zubillaga-Olague & Cañadas (2022) respecto a que, en los contextos evaluativos predominantemente calificadores, la evaluación tiende a concentrarse de forma desproporcionada en contenidos actitudinales superficiales, alejados de los aprendizajes competenciales que el marco curricular vigente pretende priorizar. En segundo lugar, el ítem referido a si el profesorado felicita al alumnado

cuando hace algo bien hecho es el único de toda la dimensión en que el contexto formativo supera claramente al calificador ($M=3.86$ frente a $M=3.35$, una diferencia de 0.51 puntos), lo que podría indicar una menor presencia de retroalimentación de carácter motivacional y orientada al refuerzo positivo en el contexto calificador, en contraste con el papel central que la literatura otorga al feedback como mecanismo de regulación del aprendizaje (Hattie & Timperley, 2007; Carless & Boud, 2018). La combinación de estos resultados, mayor transparencia formal de los criterios junto con un mayor peso de contenidos actitudinales no académicos y un feedback aparentemente menos orientado al reconocimiento, resulta coherente con la distinción planteada por Black & Wiliam (2018) y documentada empíricamente por Hortigüela et al. (2019) entre una evaluación formativa superficial, que adopta determinados instrumentos o discursos formativos sin transformar la lógica de fondo del proceso evaluativo, y una evaluación formativa profunda, que implica una reorientación genuina del rol del docente y del alumnado.

En conjunto, la interpretación de estos mediante la literatura científica revisada en el marco teórico permite matizar de forma sustancial la hipótesis de partida. El contexto predominantemente calificador estudiado no presenta un déficit generalizado y homogéneo respecto al contexto formativo estudiado previamente en todas las dimensiones de la evaluación, sino un perfil específico caracterizado por un desequilibrio claro entre los momentos del proceso evaluativo, con un peso desproporcionado de la evaluación sumativa final, y por una evaluación de los contenidos que, pese a una notable explicitación formal de los criterios, sigue otorgando un papel relevante a aspectos actitudinales no estrictamente académicos y ofrece un menor reconocimiento positivo del progreso del alumnado. Por el contrario, en las dimensiones referidas a las concepciones generales sobre la evaluación y a la implicación del alumnado como agente evaluador, las diferencias entre contextos son mínimas o incluso favorables al contexto calificador, lo que sugiere que determinadas barreras para la evaluación formativa, en particular la centralidad de la calificación en la cultura evaluativa del alumnado y la escasa participación del alumnado en la evaluación del profesorado, operan de forma transversal con independencia del modelo evaluativo predominante en el centro. Esta interpretación resulta coherente con la idea defendida por Hernando-Garijo et al. (2021) y por Sortwell et al. (2024), de que la implementación de la evaluación formativa no puede entenderse como una transformación total e inmediata, sino como un proceso progresivo en el que determinadas dimensiones del proceso evaluativo pueden avanzar de forma desigual en función de las condiciones específicas de cada

contexto. Estas conclusiones constituyen el fundamento sobre el que se articula la propuesta de implementación progresiva que se desarrolla en el siguiente apartado, la cual debe atender de manera prioritaria a las dimensiones donde se han detectado las divergencias más marcadas, sin descuidar aquellas barreras de carácter más general que el presente estudio ha permitido identificar como comunes a ambos contextos evaluativos.

9. Propuesta de Implementación

A partir de las evidencias obtenidas y de su interpretación en el apartado de discusión, se presenta una propuesta de implementación progresiva de la evaluación formativa en el área de Educación Física del centro estudiado. La propuesta se fundamenta en dos premisas. La primera es que la implementación de la evaluación formativa debe ser gradual y sensible al contexto, ya que las transformaciones radicales o totales tienden al fracaso y generan resistencia (Hernando-Garrido et al., 2021; Sortwell et al., 2024). La segunda premisa es que dicha gradualidad debe seguir una secuencia que parte de la explicitación de los criterios de evaluación, incorporar progresivamente la autoevaluación y la coevaluación y culmina en un sistema evaluativo compartido y corresponsable (López-Pastor & Pérez-Pueyo, 2017; Hernando-Garrido et al., 2021).

La aplicación de esta secuencia al centro estudiado parte de una base favorable como es la explicitación de criterios de evaluación, la cual tiene una presencia muy elevada en la percepción del alumnado ($M=4.50$ al inicio del periodo). Esta propuesta aprovecha dicha base para concentrar el esfuerzo de cambio en las dimensiones donde los resultados revelan mayores divergencias, como por ejemplo: el peso desproporcionado de la evaluación sumativa final, la práctica ausencia de coevaluación y de evaluación del profesorado, el papel relevante de contenidos actitudinales no académicos en la calificación, y un menor reconocimiento positivo del progreso del alumnado.

La propuesta se organiza en cuatro fases secuenciales y dos orientaciones transversales aplicables desde el inicio. La secuenciación es orientada y debe adaptarse al ritmo del centro, comenzando por intervenciones de bajo coste y alta probabilidad de éxito y avanzando hacia las que requieren una transformación más profunda de la cultura evaluativa.

La fase previa y transversal debe dirigirse a la formación del profesorado en evaluación formativa y compartida, condición estructural para evitar lo que Black & Wiliam (2018) y

Hortigüela et al. (2019) denominan evaluación formativa superficial, es decir, la incorporación de instrumentos formativos sin una reorientación real del rol del docente y del alumnado. Herrero-González et al. (2021) concluyen que la competencia evaluativa del profesorado está directamente condicionada por la formación recibida, y Slingerland et al. (2024) evidencian que su desarrollo requiere formación continua situada en la propia práctica, combinando reflexión, conocimiento especializado y apoyo entre iguales. En términos concretos, se propone constituir un espacio periódico de formación y reflexión del departamento de Educación Física en el que se revisen los fundamentos de la evaluación formativa, se analicen las prácticas actuales a la luz de los resultados de este estudio y se planifique la introducción de las fases siguientes.

La primera fase de intervención directa debe consolidar la explicitación de criterios como herramienta de autorregulación e iniciar el rebalanceo entre los momentos evaluativos. Aunque el profesorado explica qué se va a evaluar al inicio del periodo de forma muy frecuente ($M=4.50$), esta comunicación es percibida con menor frecuencia a nivel de cada sesión ($M=3.72$), lo que sugiere que los criterios no siempre se traducen en referentes operativos para cada tarea. Black & Wiliam (2018) señalan que clarificar y compartir los criterios de éxito es una estrategia clave de la evaluación formativa, y Tapia et al. (2022) constatan que el uso de rúbricas compartidas mejora la calidad del aprendizaje y la implicación del alumnado. Se propone, por tanto, elaborar rúbricas de desempeño sencillas para las situaciones de aprendizaje más relevantes de cada unidad, compartidas al inicio de cada sesión. En paralelo, debe iniciarse un rebalanceo de los momentos evaluativos: el examen final es, con diferencia, la práctica más presente en la percepción del alumnado ($M=4.66$, la diferencia más amplia de todo el estudio respecto al contexto formativo), lo que, según Cañadas & Santos-Pastor (2021), indica un desequilibrio entre los tres momentos evaluativos que deberían ser complementarios. Sin plantear su eliminación, se propone incrementar progresivamente el peso de la evaluación continua, apoyada en registros de observación y en las rúbricas elaboradas, de modo que la evaluación final represente una parte proporcionalmente menor del conjunto. Esta fase se considera prioritaria por su bajo coste organizativo y por apoyarse en una práctica ya consolidada.

La segunda fase se dirige a sistematizar la autoevaluación, que ya presenta una presencia moderada en el centro ($M=3.39$), superior incluso a la del contexto formativo de referencia. Panadero et al. (2017) concluyen que la autoevaluación produce mejoras en el rendimiento y en

la autoeficacia cuando se acompaña de criterios claros y de orientación docente, y Martín (2022) advierte de que, sin una introducción progresiva, el alumnado tiende a valorarse de forma poco fiable. Se propone introducir hojas de autoevaluación vinculadas a las rúbricas elaboradas en la fase anterior, a completar otras tareas significativas, acompañadas de un breve momento de reflexión guiada.

La tercera fase aborda la introducción progresiva de la coevaluación, una de las dimensiones con menor presencia en el centro ($M=2.65$). Yan et al. (2022) constatan los efectos positivos de la coevaluación sobre el rendimiento cuando se implementa sistemáticamente, y Barba-Martín & Hortigüela-Alcalá (2022) la consideran una condición necesaria para que la evaluación cumpla una función formativa. Para introducirla de forma andamiada, se propone apoyarse en los modelos pedagógicos descritos en el marco teórico: en el Modelo de Educación Deportiva, la asignación de roles no jugadores (árbitro, anotador, observador) convierte la coevaluación en un elemento natural de la dinámica (Pérez-Pueyo et al., 2020; Guijarro et al., 2020), y en el Aprendizaje Cooperativo, la retroalimentación entre iguales es intrínseca al trabajo en grupo (Bores-García et al., 2021). En ambos casos resulta imprescindible trabajar de forma explícita la *feedback literacy* del alumnado, entendida como su capacidad para comprender, procesar y proporcionar retroalimentación constructiva (Carless & Boud, 2018).

La cuarta fase tiene como horizonte un sistema evaluativo compartido y corresponsable, e incluye la introducción de mecanismos para que el alumnado evalúe la práctica docente. Este es el ítem de menor presencia de todo el cuestionario, tanto en el centro estudiado ($M=1.70$) como en el contexto formativo de referencia ($M=1.64$), con más del 78% de respuestas en "nunca" o "casi nunca" en ambos casos. Esta coincidencia, discutida como una barrera transversal especialmente resistente al cambio, justifica que esta dimensión se aborde en último lugar, una vez que las fases previas hayan familiarizado al alumnado con su papel como agente evaluativo. En coherencia con la LOMLOE (2020) y el Decret 175/2022, se propone introducir cuestionarios breves y anónimos de valoración del proceso al finalizar cada unidad didáctica, acompañados de una comunicación explícita sobre el uso que se hará de dicha información.

De forma transversal, la propuesta incorpora dos orientaciones adicionales. La primera es el refuerzo del feedback positivo: aunque las correcciones durante la práctica son frecuentes ($M=3.41$), las felicitaciones por un desempeño correcto lo son sensiblemente menos ($M=3.35$, frente a $M=3.86$ en el contexto formativo). Hattie & Timperley (2007) y Carless & Boud (2018)

subrayan que el feedback sólo cumple su función formativa cuando el alumnado puede comprenderlo y utilizarlo para mejorar, lo que requiere un equilibrio entre corrección y reconocimiento del progreso. La segunda orientación es la revisión del peso de los contenidos actitudinales no académicos en la calificación: llevar correctamente el uniforme presenta una de las diferencias más elevadas de todo el estudio ($M=4.09$, frente a $M=2.86$ en el contexto formativo), lo que, según Zubillaga-Olague & Cañadas (2022), ilustra la tendencia de los contextos calificadores a concentrar la evaluación en contenidos actitudinales superficiales, alejados de las competencias que el Real Decreto 217/2022 y el Decret 175/2022 priorizan. Se recomienda que el departamento revise el peso relativo de estos aspectos en favor de los criterios vinculados a las competencias específicas del área.

Toda esta propuesta progresiva de implementación desarrollada en este apartado aparece recogida de forma sintetizada en el Anexo 4, donde se recoge para cada fase el hallazgo que la motiva, su objetivo, las acciones propuestas y su fundamentación.

10. Conclusiones

Este trabajo de fin de máster ha permitido analizar la percepción del alumnado de secundaria y bachillerato sobre la evaluación en Educación Física en un contexto predominantemente calificador. Por lo que respecta a la comparación de esta percepción con la del alumnado de un contexto donde la evaluación formativa se aplica de manera habitual, el análisis confirma que el contexto calificador estudiado no se configura como un escenario opuesto en todos sus extremos a los principios de la evaluación formativa, sino como un contexto con un perfil propio, en el que determinadas concepciones y prácticas evaluativas resultan equiparables a las del contexto formativo, mientras que otras presentan divergencias muy marcadas.

En cuanto a las principales diferencias detectadas en relación con los principios de la evaluación formativa, estas se concentran en el peso y la función de los distintos momentos e instrumentos de evaluación. El alumnado percibe que el profesorado explicita con claridad los criterios de evaluación y confía en la justicia con la que este evalúa, lo que supone una base coherente con dichos principios. Sin embargo, la evaluación sigue centrándose de forma desproporcionada en la función sumativa y final, y otorga un peso relevante a aspectos actitudinales no estrictamente vinculados al desarrollo competencial, como la indumentaria del

alumnado. Esta combinación refleja una transformación evaluativa todavía incompleta, en la que conviven elementos propios de un enfoque formativo con la persistencia de una lógica fundamentalmente calificadora. A ello se le suma que la participación del alumnado como agente evaluador, tanto de sí mismo como de sus iguales y del propio profesorado, sigue siendo limitada.

Respecto a la relación entre la percepción del alumnado sobre la evaluación y su influencia percibida en el aprendizaje, los resultados muestran que el alumnado, en ambos contextos, reconoce que la evaluación contribuye a su aprendizaje y le permite conocer su nivel de logro. No obstante, esta valoración convive con un modelo en el que predomina la certificación del rendimiento, lo que sugiere que dicha percepción responde a una concepción de la evaluación más vinculada a la constatación de resultados que a su uso activo y continuo como herramienta de mejora.

En lo que cierra al diseño de una propuesta progresiva de implementación de la evaluación formativa contextualizada en el centro de estudio, los hallazgos anteriores evidencian que dicha transformación no debe partir de cero, sino apoyarse en las fortalezas ya presentes en las prácticas evaluativas del centro. Avanzar hacia un modelo más equilibrado, continuo y participativo, en el que el alumnado asuma progresivamente un papel más activo en su propia evaluación y en la de sus iguales, constituye el horizonte en el que debe orientarse dicha transformación, y es precisamente este horizonte el que fundamenta la propuesta de implementación progresiva desarrollada en el presente trabajo.

En definitiva, este trabajo integra una nueva aportación a la línea de investigación iniciada en el estudio anterior, al ofrecer una caracterización empírica y contrastada de cómo se vive la tensión entre evaluar para aprender y evaluar para calificar en un contexto evaluativo concreto. Esta caracterización, lejos de agotar dicha tensión, sienta las bases para que su transformación pueda abordarse de forma fundamentada, progresiva y adaptada a la realidad del contexto.

11. Limitaciones y Futuras Líneas de Investigación

La principal limitación de este estudio se deriva de su propio diseño comparativo. La triangulación se ha realizado entre datos recogidos en momentos temporales distintos y mediante administraciones no simultáneas del cuestionario, lo que impide controlar posibles cambios contextuales o normativos producidos entre ambas recogidas. Asimismo, la caracterización del

centro estudiado como predominantemente calificador se ha basado en un instrumento de observación aplicado durante un periodo limitado, por lo que las diferencias detectadas respecto al contexto formativo de referencia deben interpretarse como asociadas a dicho perfil, sin que sea posible establecer una relación causal directa entre el modelo evaluativo y la percepción del alumnado.

En segundo lugar, el hecho de que la comparación se establezca entre dos estudios de caso impide determinar si el patrón de convergencias y divergencias identificado se trata de un patrón generalizable entre contextos calificadores y formativos, o responde a las características particulares de los dos centros analizados.

Finalmente, la propuesta de implementación progresiva elaborada en este trabajo tiene un carácter exclusivamente teórico y propositivo, fundamentado en los hallazgos empíricos y en la literatura científica. Su diseño constituye el punto de llegada del presente trabajo, sin que se haya contemplado su aplicación ni su valoración por parte del profesorado del centro o de personas expertas en evaluación formativa en Educación Física. En consecuencia, se plantea como futura línea de investigación, ajena al alcance de este trabajo, la implementación de dicha propuesta y la evaluación de su impacto mediante un diseño cuasiexperimental. Este podría adoptar dos formas complementarias: por un lado, la comparación entre un grupo experimental, en el que se aplique la propuesta de implementación progresiva, y un grupo control que mantenga las prácticas evaluativas actuales, analizando posteriormente las diferencias en la percepción y las concepciones del alumnado sobre la evaluación entre ambos grupos; y por otro lado, un diseño de medidas repetidas con la misma muestra, mediante la aplicación del cuestionario antes y después de la implementación de la propuesta, que permita analizar la evolución de dicha percepción a lo largo del tiempo en un mismo grupo de alumnado. Ambos diseños permitirían aportar evidencia empírica sobre el impacto real de la propuesta más allá de su fundamentación teórica, y constituirían asimismo una vía para ampliar este diseño comparativo y determinar la replicabilidad del patrón detectado en el presente estudio.

12. Referencias

- Alcalá, D. H., Pérez-Pueyo, Á., & Fernández-Río, J. (2020). Evaluación formativa y modelos pedagógicos: modelo de responsabilidad personal y social y de autoconstrucción de materiales. *Revista Española de Educación Física y Deportes*, (430), ág-23. <https://doi.org/10.55166/reefd.vi430.919>
- Ávila, C. R. L., Arcila-Rodríguez, W. O., & Agudelo, J. E. B. (2022). Prácticas evaluativas en la clase de Educación Física: un asunto de desconocimiento (Evaluation practices in the Physical Education Class: An issue of lack of knowledge). *Retos*, 44, 77-86. <https://doi.org/10.47197/retos.v44i0.88630>
- Barba-Martín, R. A., & Hortigüela-Alcalá, D. (2022). Si la Evaluación es Aprendizaje, he de Formar parte de la misma. Razones que Justifican la Implicación del Alumnado. *Revista Iberoamericana de Evaluación Educativa*, 15(1).
- 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>
- Bisquerra Alzina, R. (2022). Metodología de la investigación educativa.
- Black, P., & Wiliam, D. (1998). Assessment and classroom learning. *Assessment in Education: principles, policy & practice*, 5(1), 7-74. <https://doi.org/10.1080/0969595980050102>
- Black, P., & Wiliam, D. (2018). Classroom assessment and pedagogy. *Assessment in education: Principles, policy & practice*, 25(6), 551-575. <https://doi.org/10.1080/0969594X.2018.1441807>
- Bores-García, D., Hortigüela-Alcalá, D., Fernandez-Rio, F. J., Gonzalez-Calvo, G., & Barba-Martin, R. (2021). Research on cooperative learning in physical education:

- Systematic review of the last five years. *Research quarterly for exercise and sport*, 92(1), 146-155. <https://doi.org/10.1080/02701367.2020.1719276>
- Boud, D., & Molloy, E. (2013). Rethinking models of feedback for learning: the challenge of design. *Assessment & Evaluation in higher education*, 38(6), 698-712. <https://doi.org/10.1080/02602938.2012.691462>
- Bowen, G. A. (2009). Document analysis as a qualitative research method. *Qualitative research journal*, 9(2), 27-40. <https://doi.org/10.3316/ORJ0902027>
- Bureau, J. S., Howard, J. L., Chong, J. X., & Guay, F. (2022). Pathways to student motivation: A meta-analysis of antecedents of autonomous and controlled motivations. *Review of educational research*, 92(1), 46-72. <https://doi.org/10.3102/00346543211042426>
- Cañadas, L., & Santos-Pastor, M. L. (2021). La evaluación formativa desde la perspectiva de docentes noveles en las clases de educación física en primaria y secundaria. *Revista Electrónica Educare*, 25(3), 452-471. <http://dx.doi.org/10.15359/ree.25-3.25>
- Cañadas, L. (2023). Contribution of formative assessment for developing teaching competences in teacher education. *European Journal of Teacher Education*, 46(3), 516-532. <https://doi.org/10.1080/02619768.2021.1950684>
- Carless, D. (2015). *Excellence in university assessment: Learning from award-winning practice*. Routledge.
- 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>
- Carreres-Ponsoda, F., Escartí, A., Jimenez-Olmedo, J. M., & Cortell-Tormo, J. M. (2021). Effects of a teaching personal and social responsibility model intervention in competitive youth sport. *Frontiers in psychology*, 12, 624018. <https://doi.org/10.3389/fpsyg.2021.624018>

- Carretero-Dios, H., & Pérez, C. (2005). Normas para el desarrollo y revisión de estudios instrumentales. *International Journal of clinical and health psychology*, 5(3), 521-551.
- Casey, A., & Goodyear, V. A. (2015). Can cooperative learning achieve the four learning outcomes of physical education? A review of literature. *Quest*, 67(1), 56-72.
<https://doi.org/10.1080/00336297.2014.984733>
- Cervantes, V. (2005). Interpretaciones del coeficiente alpha de Cronbach. *Avances en medición*, 3(1), 9-28.
- Cohen, L., Manion, L., & Morrison, K. (2002). *Research methods in education*. Routledge.
- Creswell, J. W., & Creswell, J. D. (2017). *Research design: Qualitative, quantitative, and mixed methods approaches*. Sage publications.
- Decret 175/2022, de 27 de setembre, d'ordenació dels ensenyaments de l'educació bàsica . *Diari Oficial de la Generalitat de Catalunya*.
<https://portaljuridic.gencat.cat/eli/es-ct/d/2022/09/27/175>
- Fernández-Truan, J. C., & Muñoz, R. F. (2021). Desarrollo Legislativo de los Contenidos de Educación Física en Educación Secundaria en España. *Kronos*, 20(01-02).
<https://doi.org/10.64197/Kronos.20.01-02.906>
- Flix, X. T., & Rius, J. B. (2020). Los preámbulos fundacionales de la educación física escolar en el entorno español. Una mirada de los textos en educación del siglo XVIII. *Materiales para la Historia del Deporte*, (20), 16-38. <https://doi.org/10.20868/mhd.2020.20.4387>
- Grisales, N. E. M., Palacio, E. V. G., Montoya, L. F. J., Vélez, C. P., & Vélez, H. H. (2023). La evaluación en la clase de Educación Física desde la percepción de los estudiantes de secundaria. *Revista Virtual Universidad Católica del Norte*, (68), 94-119.
<https://doi.org/10.35575/rvucn.n68a5>

- Guijarro, E., Rocamora, I., Evangelio, C., & Villora, S. G. (2020). El modelo de Educación Deportiva en España: una revisión sistemática (Sport Education Model in Spain: a systematic review). *Retos*, 38, 886-894. <https://doi.org/10.47197/retos.v38i38.77249>
- Hattie, J., & Timperley, H. (2007). The power of feedback. *Review of educational research*, 77(1), 81-112. <https://doi.org/10.3102/003465430298487>
- Hernán, E. J. B., Pastor, V. M. L., & Brunicardi, D. P. (2019). ¿ Por qué hago evaluación formativa y compartida y/o evaluación para el aprendizaje en EF?: La influencia de la formación inicial y permanente del profesorado. *Retos: nuevas tendencias en educación física, deporte y recreación*, (36), 37-43.
- Hernando-Garijo, A., Hortigüela-Alcalá, D., Sánchez-Miguel, P. A., & González-Villora, S. (2021). Fundamental pedagogical aspects for the implementation of models-based practice in physical education. *International Journal of Environmental Research and Public Health*, 18(13), 7152. <https://doi.org/10.3390/ijerph18137152>
- Herrero-González, D., Arribas, J. C. M., & López-Pastor, V. M. (2021). Incidencia de la Formación Inicial y Permanente del Profesorado en la aplicación de la Evaluación Formativa y Compartida en Educación Física (Incidence of Pre-service and In-service Teacher Education in the application of Formative and Shared Assessment. *Retos*, 41, 533-543. <https://doi.org/10.47197/retos.v0i41.86090>
- Herrero-González, D., López-Pastor, V. M., Manrique-Arribas, J. C., & Moura, A. (2024). 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, D., Pérez-Pueyo, Á., & González-Calvo, G. (2019). Pero...¿ A qué nos referimos realmente con la Evaluación Formativa y Compartida?: Confusiones Habituales y Reflexiones Prácticas. *Revista Iberoamericana de Evaluación Educativa*, 12(1).

Hortigüela-Alcalá, D., González-Villora, S., & Hernando-Garijo, A. (2021). ¿Evaluamos realmente el aprendizaje en educación física? Percepción de los profesores en diferentes etapas educativas (Do we really assess learning in physical education? Teachers' perceptions at different educational stages). *Retos*, 42, 643-654.
<https://doi.org/10.47197/retos.v42i0.88686>

Latorre, A., Arnal, J., & Del Rincón, D. (2021). Bases metodológicas de la investigación educativa.

Ley Orgánica 1/1990, de 3 de octubre, de Ordenación General del Sistema Educativo. *Boletín Oficial del Estado*, 238, de 4 de octubre de 1990.
<https://www.boe.es/eli/es/lo/1990/10/03/1>

Ley Orgánica 2/2006, de 3 de mayo, de Educación. *Boletín Oficial del Estado*, 106, de 4 de mayo de 2006. <https://www.boe.es/eli/es/lo/2006/05/03/2/con>

Ley Orgánica 8/2013, de 9 de diciembre, para la mejora de la calidad educativa. *Boletín Oficial del Estado*, 295, de 10 de diciembre de 2013.
<https://www.boe.es/eli/es/lo/2013/12/09/8/con>

Ley Orgánica 3/2018, de 5 de diciembre, de protección de datos personales y garantía de los derechos digitales. *Boletín Oficial del Estado*, núm. 294, 6 de diciembre de 2018.
<https://www.boe.es/eli/es/lo/2018/12/05/3>

Ley Orgánica 3/2020, de 29 de diciembre, por la que se modifica la Ley Orgánica 2/2006, de 3 de mayo, de Educación. *Boletín Oficial del Estado*, 340, de 30 de diciembre de 2020.
<https://www.boe.es/eli/es/lo/2020/12/29/3>

Llei 1/1998, de 7 de gener, de política lingüística. *Diari Oficial de la Generalitat de Catalunya*, núm. 2553, 9 de gener de 1998. <https://portaljuridic.gencat.cat/eli/es-ct/l/1998/01/07/1>

Llei 12/2009, del 10 de juliol, d'educació de Catalunya. *Diari Oficial de la Generalitat de Catalunya*, núm. 5422, 16 de juliol de 2009.

<https://portaljuridic.gencat.cat/eli/es-ct/1/2009/07/10/12>

López Pastor, V. M., & Pérez Pueyo, Á. (2017). Evaluación formativa y compartida en educación: experiencias de éxito en todas las etapas educativas.

Maciá López, P.J. (2025). Evaluación formativa en Educación Física: un estudio sobre la percepción del alumnado y su influencia en el aprendizaje. (Trabajo Final de Máster no publicado). Universidad de Barcelona.

Manninen, M., & Campbell, S. (2022). The effect of the Sport Education Model on basic needs, intrinsic motivation and prosocial attitudes: A systematic review and multilevel meta-analysis. *European Physical Education Review*, 28(1), 78-99.
<https://doi.org/10.1177/1356336X211017938>

Martín, R. B., Alcalá, D. H., & Pueyo, Á. P. (2020). Evaluar en Educación Física: análisis de la tensiones existentes y justificación del empleo de la evaluación formativa y compartida. *Educación Física y Deporte*, 39(1), 3.

Martín, L. C. (2022). Procesos de auto-evaluación y co-evaluación en Educación Física. Una revisión sistemática. *Revista Iberoamericana de Evaluación educativa*, 15(1), 161-176.

Mendoza, S. T. B., Cedeño, J. A. M., Espinales, A. N. V., & Gámez, M. R. (2021). Autoevaluación, Coevaluación y Heteroevaluación como enfoque innovador en la práctica pedagógica y su efecto en el proceso de enseñanza-aprendizaje. *Polo del Conocimiento: Revista científico-profesional*, 6(3), 828-845.

Morris, R., Perry, T., & Wardle, L. (2021). Formative assessment and feedback for learning in higher education: A systematic review. *Review of education*, 9(3), e3292.
<https://doi.org/10.1002/rev3.3292> Digital Object Identifier (DOI)

- Niubó Solé, J. (2025). Influència de l'avaluació formadora en l'àrea d'educació física sobre les funcions executives. <http://hdl.handle.net/10803/696329>
- Ortega, E. M., & Coll, C. (2023). Estructura y componentes de los currículos de la LOMLOE. In *Nuevo currículo, nuevos desafíos educativos: Tus proyectos, tus recursos* (pp. 23-45). Fundación SM.
- Ortiz, M., Merono, L., Morales-Belando, M. T., Vaquero-Cristobal, R., & Gonzalez-Galvez, N. (2023). Teaching games for understanding in game performance and psychosocial variables: systematic review and meta-analysis of randomized control trial. *Children*, 10(3), 573. <https://doi.org/10.3390/children10030573>
- Otero-Saborido, F. M., Vázquez-Ramos, F. J., & González-Jurado, J. A. (2020). Análisis de la evaluación en los currículos de Educación Física en Secundaria. *Educación XXI*, 23(2), 259-282. <https://doi.org/10.5944/educXXI.25662>
- Palacio, E. V. G., Fernández, B. C., Castaño, S. A. B., & Toro, C. A. (2021). Diseño y validación de un cuestionario sobre las concepciones y percepción de los estudiantes sobre la evaluación en Educación Física. *Retos: nuevas tendencias en educación física, deporte y recreación*, (40), 317-325.
- Palinkas, L. A., Horwitz, S. M., Green, C. A., Wisdom, J. P., Duan, N., & Hoagwood, K. (2015). Purposeful sampling for qualitative data collection and analysis in mixed method implementation research. *Administration and policy in mental health and mental health services research*, 42(5), 533-544.
- Panadero, E., Jonsson, A., & Botella, J. (2017). Effects of self-assessment on self-regulated learning and self-efficacy: Four meta-analyses. *Educational research review*, 22, 74-98. <https://doi.org/10.1016/j.edurev.2017.08.004>

- Pastor, V. M. L., Aguado, R. M., García, J. G., Pastor, E. L., Pinela, J. M., Badiola, J. G., ... & García, L. M. (2006). La Evaluación en Educación Física. Revisión de los modelos tradicionales y planteamiento de una alternativa: la evaluación formativa y compartida. *Retos*, 10, 31-40. <https://doi.org/10.47197/retos.v0i10.35061>
- Pérez-Pueyo, Á., Alcalá, D. H., Álvarez, I. H., Cobo, D. V., Bernardino, C. H., García, J. G., ... & Garijo, A. H. (2017). La hibridación de modelos pedagógicos en educación física y la evaluación formativa. *Revista Infancia, Educación y Aprendizaje*, 3(2), 411-418.
- Pérez-Pueyo, Á., Alcalá, D. H., & Fernandez-Río, J. (2020). Evaluación formativa y modelos pedagógicos: Estilo actitudinal, aprendizaje cooperativo, modelo comprensivo y educación deportiva. *Revista española de educación física y deportes*, (428), ág-47. <https://doi.org/10.55166/reefd.vi428.881>
- Pérez-Pueyo, Á., Alcalá, D. H., Berrocal, O. C., Bernardino, C. H., & Álvarez, I. H. (2022). Análisis y reflexión sobre el nuevo currículo de educación física. *Revista Española de Educación Física y Deportes*, (463 (3)). [https://doi.org/10.55166/reefd.vi463\(3\).1073](https://doi.org/10.55166/reefd.vi463(3).1073)
- Real Decreto 217/2022, de 29 de marzo, por el que se establece la ordenación y las enseñanzas mínimas de la Educación Secundaria Obligatoria. *Boletín Oficial del Estado*, 76, de 30 de marzo de 2022. <https://www.boe.es/eli/es/rd/2022/03/29/217/con>
- Salvador, C. C., & Ortega, E. M. (2021). La LOMLOE, una oportunidad para la modernización curricular. *Avances en supervisión educativa*, (35). <https://doi.org/10.23824/ase.v0i35.731>
- 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>

- Sortwell, A., Trimble, K., Ferraz, R., Geelan, D. R., Hine, G., Ramirez-Campillo, R., ... & Xuan, Q. (2024). A systematic review of meta-analyses on the impact of formative assessment on K-12 students' learning: Toward sustainable quality education. *Sustainability*, 16(17), 7826. <https://doi.org/10.3390/su16177826>
- Tapia, M. A. V., Bastias, M. F., Aguilar, L. I., & Valderrama, F. P. (2022). Implementación de la evaluación formativa por estudiantes de pedagogía en Educación Física. *Retos: nuevas tendencias en educación física, deporte y recreación*, (43), 916-922.
- Tavakol, M., & Dennick, R. (2011). Making sense of Cronbach's alpha. *International journal of medical education*, 2, 53. [10.5116/ijme.4dfb.8dfd](https://doi.org/10.5116/ijme.4dfb.8dfd)
- Wisniewski, B., Zierer, K., & Hattie, J. (2020). The power of feedback revisited: A meta-analysis of educational feedback research. *Frontiers in psychology*, 10, 487662. <https://doi.org/10.3389/fpsyg.2019.03087>
- Yan, Z., Lao, H., Panadero, E., Fernández-Castilla, B., Yang, L., & Yang, M. (2022). Effects of self-assessment and peer-assessment interventions on academic performance: A meta-analysis. *Educational Research Review*, 37, 100484. <https://doi.org/10.1016/j.edurev.2022.100484>
- Zach, S., Shoval, E., & Shulruf, B. (2023). Cooperative learning in physical education lessons-literature review. In *Frontiers in Education* (Vol. 8, p. 1273423). Frontiers Media SA. <https://doi.org/10.3389/feduc.2023.1273423>
- Zhang, J., Geok Soh, K., Bai, X., Mohd Anuar, M. A., & Xiao, W. (2024). Optimizing learning outcomes in physical education: A comprehensive systematic review of hybrid pedagogical models integrated with the Sport Education Model. *PloS one*, 19(12), e0311957. <https://doi.org/10.1371/journal.pone.0311957>

Zubillaga-Olague, M., & Cañadas, L. (2021). Diseño y validación del cuestionario “# EvalEF” para conocer el proceso de evaluación desarrollado por los docentes de educación física. *Retos*, 42, 47-55.

Zubillaga-Olague, M., & Cañadas, L. (2022). Agentes participantes en los procesos de evaluación y calificación en Educación Física. *Contextos Educativos. Revista de Educación*, (30), 83-98. <https://doi.org/10.18172/con.5371>

13. Anexos

Anexo 1

Instrumento de Observación Estructurado: Lista de Control

INSTRUMENTO DE OBSERVACIÓN			
Indicadores de Innovación a Observar	Educación Física	Centro	Evidencias
Ámbito 1: Metodologías y recursos para la mejora de la enseñanza y el aprendizaje			
1. Fomentan metodologías activas y experienciales: ABR, ABP, trabajo cooperativo, gamificación, APS...	No	Si	En Educación Física no observé apenas la utilización de este tipo de metodologías. Si que se utilizó de forma esporádica el trabajo cooperativo pero no se fomenta de forma continuada. En el caso del centro sí que se fomenta, ya que hay una asignatura que se llama Proyectos donde se utiliza el trabajo cooperativo y el Aprendizaje Basado en Proyectos.
2. Desarrollo de la educación emocional	No	Si	En Educación Física no vi ninguna situación donde se trabaje este aspecto, pero a nivel de centro sí que pude observar cómo se trabajaba, principalmente en la asignatura de tutoría, mediante dinámicas y reflexiones realizadas en clase.
3. Fomentan el espíritu crítico y reflexivo (con preguntas sobre temáticas de la vida cotidiana, mediante debates o conversaciones sobre noticias de actualidad de nuestro ámbito...)	No	Si	Esto es algo que he podido observar, sobre todo, en las horas de tutoría. Donde se trataban diferentes temas relevantes y de actualidad con el objetivo de fomentar la reflexión del alumnado.
4. Utiliza la tecnología para la mejora del aprendizaje.	Si	Si	En el centro los alumnos tienen cada uno un chromebook que utilizan durante las clases. A este no le dan un uso únicamente para tener el temario sino que también lo utilizan para preparar canvas y presentaciones individuales y en grupo sobre determinados temas. En Educación Física la utilización de esta herramienta es un poco más limitada, pero sí que se utiliza en ocasiones para realizar trabajos de investigación sobre ciertos temas y en bachillerato para la elaboración de videos relacionados con diferentes contenidos como por ejemplo el análisis individual de la técnica de carrera de cada alumno.
5. Utiliza la evaluación como herramienta de autorregulación del aprendizaje (no para calificar sino para ayudar y acompañar a aprender, momentos de feedback entre iguales, aterrizar objetivos, uso de instrumentos de manera formativa...)	No	No	Por lo que he podido observar durante este primer periodo, la evaluación es concebida más como una herramienta calificadora que como una herramienta a través de la cual se regule el aprendizaje.
6. La Educación Física colabora en algún proyecto transdisciplinar o globalizado junto	No		

con otras materias.

Ámbito 2: Organización y gestión

7. Se impulsa el trabajo colaborativo entre docentes.	Si	Si	En Educación Física los docentes trabajan de forma conjunta tanto en el diseño de las Situaciones de Aprendizaje como en la implementación de tareas durante las sesiones. En el centro, siguiendo el ejemplo de la asignatura de Proyectos, se trabaja de forma conjunta y coordinada con varios profesores de diferentes asignaturas.
8. Se combinan reuniones pedagógicas con reuniones organizativas (más burocráticas)	No	Si	En Educación Física estas reuniones organizativas suelen ser más improvisadas y realizadas de forma puntual en los momentos donde coinciden los profesores (durante el curso), evidentemente se realiza una reunión más formal y planificada al inicio del curso. En cuanto a nivel de centro si que se realiza de forma planificada (1 vez por semana) donde se reúnen los coordinadores de cada nivel (1º, 2º...) con los tutores de cada grupo de ese nivel.
9. Se planifican tiempos y espacios para compartir objetivos ya sean del centro, del departamento o del equipo docente.	No	Si	A nivel de centro esto se produce durante las reuniones semanales que se mencionan en el punto anterior.
10. Se evalúan y se hace seguimiento de estos objetivos de centro, del equipo docente o de departamento.	Si	Si	En el caso de EF se realiza un seguimiento diario, siguiendo la estructura improvisada y espontánea que se ha mencionado anteriormente, y se hacen reuniones al inicio para plantear objetivos y al final donde se habla de la consecución de estos objetivos. A nivel de centro esto se realiza de forma trimestral, donde al final de cada trimestre se realiza la reunión de evaluación donde acuden todos los profesores del grupo y el tutor.
11. Se utilizan diversos espacios del centro para favorecer la interacción docente sobre temas pedagógicos.	Si	Si	Esta interacción se realiza en diversos espacios como puede ser el despacho de coordinación de cada nivel, en aulas donde se realizan las reuniones, en la sala de profesores e incluso en los espacios de almuerzo y comidas.
12. Se impulsa el trabajo por proyectos en relación a las competencias transversales.	No	Si	Si que se realiza a nivel de centro, aunque es cierto que solo se realiza en la asignatura de proyectos. En el resto de casos no se realiza, por tanto se puede decir que es una práctica puntual al realizarse solamente en una asignatura.
13. Se promueve el aprendizaje del alumnado de forma diversa y flexible.	No	No	No diría que se promueve este tipo de aprendizaje. Si que se intenta adecuar los contenidos, en el caso de Educación física

al nivel del alumnado, pero utilizando el mismo tipo de contenidos para toda la clase. El único caso donde he podido ver que si ocurría esto es en un examen teórico de EF, donde a una alumna con dificultades de aprendizaje se le daba la oportunidad de, una vez entregado el examen corregido, llevárselo a su casa y volver a hacerlo de nuevo, siendo la nota de esta segunda oportunidad la que obtendría.

Ámbito 3: Relación con el entorno (escuela, familia y entorno)

14. Se colabora con las familias o con entidades cercanas al centro (sociales, culturales, empresariales) con el fin de conectar aprendizajes formales y no formales.

No

No

En este primer periodo no he observado que se produzca ninguna de las colaboraciones que se manifiestan en este punto.

15. Se identifican necesidades, situaciones y contextos de aprendizaje donde se puedan aplicar los recursos metodológicos que se disponen.

No

Si

A nivel de centro existe el aula de acogida donde el alumnado procedente de otros países acude un número de horas a la semana, en función de la necesidad, para trabajar en grupos reducidos contenidos muy básicos para que aprendan el idioma.

Anexo 2

Documento de Consentimiento Informado

Estimados/as alumnos/as, familias y tutores legales:

Me dirijo a ustedes para informarles sobre la realización de un cuestionario anónimo que forma parte de mi Trabajo de Fin de Máster (TFM) en el Máster de Formación del Profesorado de Secundaria Obligatoria y Bachillerato, Formación Profesional y Enseñanza de Idiomas de la Universidad de Barcelona. El objetivo del cuestionario es recopilar información relacionada con las percepciones del alumnado de Educación Secundaria Obligatoria (ESO) y Bachillerato en relación con la evaluación en Educación Física.

La participación en este cuestionario es completamente voluntaria y el alumnado puede abandonar el cuestionario en cualquier momento sin que esto suponga ningún perjuicio. El cuestionario no recopila datos personales identificables (como nombre, dirección, correo electrónico, etc.), por lo que todas las respuestas serán tratadas de forma totalmente anónima y confidencial. Los datos recogidos serán utilizados únicamente con fines académicos y de investigación en el marco del TFM, y serán analizados de forma agregada. Este estudio cumple con la normativa vigente sobre protección de datos personales.

Al aceptar participar, se entiende que han sido informados/as adecuadamente sobre el estudio y que dan su consentimiento para la participación del alumno/a en el mismo. Si el/la alumno/a es menor de 14 años, se requiere el consentimiento expreso del padre/madre o tutor/a legal.

Gracias por su colaboración.

Firma padre, madre o tutor legal

Pedro José Maciá López



Universidad de Barcelona

Anexo 3

Tabla Completa de Resultados del Contexto Actual Estudiado

D	Ítems	M	DE	1	2	3	4	5
D1	1. La evaluación que realiza el profesor aporta a tu aprendizaje	3.97	0.69	0.6	1.2	18.2	60.6	19.4
	2. La calificación es importante para ti	4.30	0.73	0.0	1.8	10.9	43.0	44.2
	3. La evaluación sirve para medir tu nivel de aprendizaje	3.85	0.88	1.2	7.3	18.2	52.1	21.2
	4. Te parece importante la evaluación que realiza el profesor	4.10	0.83	1.8	2.4	11.5	52.1	32.1
	5. Tus compañeros(as) de clase aprenderían más, si tu haces parte de su evaluación	3.07	0.79	3.6	12.1	63.0	16.4	4.8
D2	6. La calificación que te pone el profesor refleja tu nivel de aprendizaje	3.58	0.84	1.2	7.9	33.9	46.1	10.9
	7. El profesor permite que te evalúes tu mismo	3.39	1.03	4.2	10.3	46.1	21.2	18.2
	8. El profesor les pregunta cómo les pareció la clase	2.75	1.03	13.3	22.4	45.5	13.3	5.5
	9. Evalúas a tu profesor	1.70	1.00	59.4	18.8	17.0	2.4	2.4
	10. Tus compañeros(as) de clase hacen parte de tu evaluación	2.65	1.00	15.8	20.0	53.9	4.2	6.1
D3	26. El profesor es justo cuando te evalúa	4.04	0.87	1.8	1.8	19.4	44.8	32.1
	11. Antes de empezar un nuevo tema, el profesor les pregunta qué saben del tema	3.84	1.06	3.0	6.7	26.7	30.9	32.7
	13. Al final del periodo el profesor realiza un examen para ver qué han aprendido	4.66	0.79	0.6	3.0	7.3	8.5	80.6
	14. Al final del periodo el profesor realiza un test físico para ver que han mejorado	3.51	1.26	9.7	12.1	20.6	32.7	24.8
	16. En cada clase el profesor pregunta lo que vieron en la clase anterior	3.21	1.06	6.7	16.4	37.0	29.1	10.9
D4	17. Al finalizar la clase el profesor califica lo trabajado en la sesión	3.55	0.97	2.4	9.7	35.2	35.8	17.0
	19. El profesor te evalúa de forma escrita (cuaderno, talleres)	3.36	1.12	6.1	15.2	32.1	29.7	17.0
	12. Al inicio del periodo el profesor les explica qué va a evaluar	4.50	0.76	0.0	1.8	10.9	23.0	64.2
	15. Al inicio de la clase el profesor les dice los criterios de evaluación que van a tener en clase	3.72	1.24	6.7	9.1	27.3	19.4	37.6
	18. El profesor te evalúa lo que haces en la práctica (ejercicios, juegos, deportes, etc.)	4.26	0.73	0.0	1.2	13.3	43.6	41.8

20. El profesor evalúa tu participación en clase	4.22	0.82	0.0	2.4	17.6	35.2	44.8
21. La calificación de la clase depende de llevar correctamente el uniforme	4.09	1.05	3.0	4.2	20.0	26.1	46.7
22. La calificación de la clase depende de tu comportamiento durante las actividades	4.18	0.90	1.2	3.0	16.4	35.8	43.6
23. El profesor te solicita demostrar los ejercicios durante la clase	3.29	0.99	4.8	13.3	41.2	29.7	10.9
24. El profesor te hace correcciones durante la clase	3.41	0.95	1.8	12.1	44.2	26.7	15.2
25. El profesor te felicita cuando haces algo bien hecho	3.35	1.14	8.5	10.3	37.0	26.7	17.6

Nota. **D:** dimensión; **M:** media; **DE:** desviación estándar; **D1:** concepciones; **D2:** agentes implicados; **D3:** instrumentos y momentos; **D4:** contenidos; **1:** muy en desacuerdo (preguntas de 1 a 5) o nunca (preguntas de 6 a 26); **2:** en desacuerdo (preguntas de 1 a 5) o casi nunca (preguntas de 6 a 26); **3:** indiferente (preguntas de 1 a 5) o a veces (pregunta de 6 a 26); **4:** de acuerdo (preguntas de 1 a 5) o casi siempre (preguntas de 6 a 26); **5:** muy de acuerdo (preguntas de 1 a 5) o siempre (preguntas de 6 a 26).

Anexo 4

Propuesta de Implementación Progresiva de la Evaluación Formativa

Fase	Hallazgo que la motiva	Objetivo	Acciones concretas	Fundamentación
Fase 0. Formación del profesorado	Riesgo de implementación superficial sin formación específica	Dotar al profesorado de los fundamentos y procedimientos de la evaluación formativa y compartida	Espacio periódico de formación y reflexión del departamento; análisis de las prácticas actuales	Herrero-González et al. (2021); Slingerland et al. (2024); Black & Wiliam (2018); Hortigüela et al. (2019)
Fase 1. Criterios como herramienta de autorregulación y rebalanceo de momentos evaluativos	Criterios explícitos al inicio del periodo (M=4.50) pero menos a nivel de sesión (M=3.72); predominio del examen final (M=4.66)	Convertir la transparencia formal de criterios en una herramienta operativa; reducir el peso relativo de la evaluación sumativa final	Rúbricas de desempeño por situación de aprendizaje; incremento progresivo del peso de la evaluación continua	Black & Wiliam (2018); Tapia et al. (2022); Cañadas & Santos-Pastor (2021)
Fase 2. Sistematización de la autoevaluación	Presencia moderada de autoevaluación (M=3.39)	Consolidar y sistematizar la autoevaluación	Hojas de autoevaluación vinculadas a las rúbricas; reflexión guiada tras tareas significativas	Panadero et al. (2017); Martín (2022)
Fase 3. Introducción progresiva de la coevaluación	Coevaluación marginal (M=2.65)	Introducir la coevaluación de forma andamiada	Roles no jugadores (modelo de educación deportiva) y Aprendizaje cooperativo; desarrollo de <i>feedback literacy</i>	Pérez-Pueyo et al. (2020); Guijarro et al. (2020); Bores-García et al. (2021); Yan et al. (2022); Barba-Martín & Hortigüela-Alcalá (2022); Carless & Boud (2018)
Fase 4. Evaluación compartida y corresponsable	Evaluación del profesorado prácticamente inexistente (M=1.70)	Incorporar la valoración del alumnado sobre el proceso de enseñanza	Cuestionarios breves y anónimos al finalizar cada unidad, con retorno explícito de su uso	LOMLOE (2020); Decret 175/2022
Orientación A. Refuerzo del feedback positivo	Menor frecuencia de felicitaciones (M=3.35)	Equilibrar feedback correctivo y reconocimiento del progreso	Momentos deliberados de reconocimiento del progreso individual y grupal	Hattie & Timperley (2007); Carless & Boud (2018)
Orientación B. Revisión de contenidos actitudinales no académicos	Elevado peso del uniforme en la calificación (M=4.09)	Equilibrar contenidos conceptuales, procedimentales y actitudinales	Revisión del peso relativo de aspectos no académicos en la calificación	Zubillaga-Olague & Cañadas (2022); Real Decreto 217/2022; Decret 175/2022

Nota. **M:** media del contexto calificador