

Artículo de Investigación

Análisis de la evaluación y la calificación. ¿Es posible una coordinación pedagógica entre ambos enfoques?

Analysis of evaluation and grading. Is pedagogical coordination between both approaches possible?

Luis Miguel Romero-González¹: Universidad de Cádiz, España.

luismiguel.romerogonzalez@alum.uca.es

Fecha de Recepción: 11/06/2024

Fecha de Aceptación: 28/06/2024

Fecha de Publicación: 09/07/2024

Cómo citar el artículo (APA 7^a):

Romero-González, L. M. (2024). Análisis de la evaluación y la calificación. ¿Es posible una coordinación pedagógica entre ambos enfoques? [Analysis of evaluation and qualification. Is pedagogical coordination between both approaches possible?]. *European Public & Social Innovation Review*, 10, 1-21. <https://doi.org/10.31637/epsir-2024-327>

Resumen:

Introducción: Actualmente, existe una confusión terminológica entre evaluación y calificación que implica, además de equivocaciones conceptuales, errores en la forma de entender la evaluación educativa. La evaluación se considera formativa y al servicio del aprendizaje, mientras que la calificación tiene una función acreditativa para medir el aprendizaje.

Metodología: Para la elaboración de este estudio se llevó a cabo una revisión no sistemática de la bibliografía existente en diferentes bases de datos y obras impresas con un doble objetivo: analizar si la literatura concibe como posible la sinergia entre evaluación y calificación y localizar estrategias que las coordinen. **Resultados:** Se encontró que ambos procesos, evaluación y calificación, tienen en la literatura científica su lugar diferenciado, aunque en la práctica suelen confundirse. Esta confusión se debe principalmente a la tradición calificadora de la evaluación y a la asignación de múltiples funciones a la evaluación, algunas incluso contradictorias, lo que confunde a los docentes. Muchos docentes han tenido la experiencia de ser calificados, no evaluados, durante su etapa como estudiantes. **Discusión:** La confusión entre evaluación y calificación tiene raíces profundas en las tradiciones educativas y en la multiplicidad de funciones asignadas a la evaluación. Esto provoca que, en la práctica, ambos términos se utilicen indistintamente, afectando la claridad y efectividad de los procesos de enseñanza y aprendizaje. **Conclusiones:** El profesorado puede combinar la evaluación y

¹ Autor Correspondiente: Luis Miguel Romero-González. Universidad de Cádiz (España).

calificación mediante innovadoras y creativas estrategias a favor del proceso de aprendizaje, superando así la tradicional confusión y potenciando su impacto positivo en el aprendizaje de los estudiantes.

Palabras clave: calificación; evaluación sumativa; medición; evaluación; evaluación formativa; retroalimentación; evaluación para el aprendizaje; evaluación en el aula.

Abstract:

Introduction: Currently, there is a terminological confusion between evaluation and qualification that implies, in addition to conceptual mistakes, errors in the way of understanding educational evaluation. Assessment is considered formative and serves learning, while grading has an accrediting function to measure learning. **Methodology:** For this study, a non-systematic review of existing literature in various databases and printed works was conducted with a dual objective: to analyze if the literature conceives the possibility of synergy between assessment and grading and to locate strategies that coordinate them. **Results:** It was found that both processes, assessment and grading, have their differentiated place in scientific literature, although in practice they are often confused. This confusion is mainly due to the grading tradition of assessment and the imposition of multiple functions on assessment, some even contradictory, which confuses teachers. Many teachers have experienced being graded, not assessed, during their student years. **Discussion:** The confusion between assessment and grading has deep roots in educational traditions and the multiplicity of functions assigned to assessment. This causes both terms to be used interchangeably in practice, affecting the clarity and effectiveness of teaching and learning processes. **Conclusions:** Teachers can combine assessment and grading through innovative and creative strategies in favor of the learning process, thus overcoming the traditional confusion and enhancing their positive impact on student learning.

Keywords: grading; summative assessment; measurement; assessment; formative assessment; feedback; assessment for learning; classroom assessment.

1. Introducción

La investigación sobre evaluación, especialmente en el ámbito educativo, es uno de los campos más controvertidos en esta ciencia (Arribas, 2017). Dos de los aspectos más investigados son, por un lado, la evolución que ha tenido la evaluación desde sus orígenes y, por otro, cómo esta evolución ha ido generando diferentes tipos de evaluación, según la función que desempeña, llegando al momento en que se produce la separación entre evaluación y calificación, cada una con sus características, instrumentos y finalidades.

En este primer apartado de introducción del presente trabajo se aborda el estado de la cuestión, realizando una síntesis de la evolución de la evaluación durante las últimas décadas hasta llegar al momento actual, en el que evaluación y calificación se entienden dos formas diferentes de comprender el proceso evaluativo, procediendo a definir las y a exponer aspectos relevantes de la evaluación y, especialmente, de la calificación. A partir de ahí, dos son los objetivos fundamentales de este estudio que se abordarán en los resultados y discusión. El primer objetivo consistirá en analizar mediante una revisión bibliográfica referencias sobre la posibilidad de que la evaluación y la calificación puedan, o no, trabajar de manera conjunta en el proceso educativo, especialmente en la educación obligatoria. El segundo objetivo, en caso de ser posible esta sinergia, se localizarán y valorarán algunas propuestas concretas.

1.1. Un repaso a las últimas décadas: la evaluación como calificación y su evolución

No se pretende hacer un exhaustivo estudio histórico sobre la evolución de la evaluación y la calificación en el ámbito educativo, aunque sí consideramos importante conocer su desarrollo durante el último siglo para comprender el origen de la confusión entre ambos términos y por qué, a pesar de que la evaluación educativa es uno de los temas más investigados y con mayor proliferación de trabajos en las últimas décadas (Arribas, 2017), esta confusión persiste en la comunidad educativa.

La calificación tiene su origen en los años treinta, momento en que surgieron corrientes de pensamiento que trataban de medir y cuantificar características de los objetos o sujetos para poder establecer comparaciones, siendo ésta la única finalidad de la evaluación en esos momentos (Alcaraz, 2015). En estos años, personas expertas en mediciones pensaron que podrían usarse pruebas nuevas, objetivas, para mejorar los resultados de la educación, incrementando la precisión en la toma de decisiones de los docentes ante la demostrada falta de confiabilidad de los exámenes que por aquel entonces se realizaban. Así, la evaluación de esta época solo tenía el objetivo de valorar los conocimientos adquiridos tras una enseñanza (Rosales, 2014), comprobar el desempeño del estudiante al final de un proceso educativo, con la intención de certificar o acreditar la consecución de uno o varios objetivos, a través de la calificación. Por ello, el examen se alzó como el instrumento clave. De este modo, empezaron a enseñar a docentes en activo cómo hacer pruebas siguiendo principios científicos de medición, abordando conceptos como la validez y fiabilidad, el análisis de reactivos, análisis estadísticos, etc. aunque se hicieron críticas acerca de que las pruebas objetivas solo medían fragmentos de los hechos, lo cierto es que venció la tendencia de las pruebas estandarizadas (Shepard, 2006).

En los años 70 y en adelante, destacando las aportaciones sobre “evaluación educativa” de Ralph Tyler, quien acuñó el término, se produce una separación entre evaluación y medición, debido al nacimiento de tendencias opuestas a las vigentes que defendían la evaluación como algo distinto de la calificación, evolucionando hasta lo que hoy conocemos como evaluación formativa (Alcaraz, 2015a). Entre los años 70 y 80 se produce una proliferación de modelos de evaluación fruto de un aumento de la preocupación y sensibilización sobre la evaluación educativa, produciéndose una gran pluralidad de conceptos, métodos, finalidades... así, aparece la “evaluación democrática” de MacDonald, la “evaluación iluminativa” de Parlett y Hamilton, la “evaluación como crítica artística” de Eisner, etc., hasta llegar al nacimiento de un nuevo paradigma naturalista que supuso un cambio en la forma de entender la realidad, no solo la educativa, que apuesta por técnicas hermenéuticas, la dialéctica, las técnicas cualitativas para entender la realidad, la cual es múltiple y no única ni exacta, dando importancia al contexto. Además, por estos años hay un aumento de la responsabilización, esto es, de hacer responsables a las escuelas, dirección y profesorado del progreso académico de los estudiantes, que junto a las conclusiones sobre los efectos negativos de las pruebas cerradas para la calidad del proceso de enseñanza-aprendizaje, allanaron el camino para el auge de la evaluación frente a la calificación, apoyado por los avances de la investigación en psicología cognitiva y motivacional.

Es por esta época cuando Scriven (1967) acuña los términos de “evaluación formativa” y “evaluación sumativa”, en un intento de separar definitivamente la evaluación de la calificación. Sin embargo, todavía hoy se confunden los términos y se usan de manera sinónima (Alcaraz, 2015). De hecho, Martínez (2021) nos explica que con demasiada frecuencia cuando un docente habla de evaluación, realmente está hablando de calificación. De hecho, esta función calificadora, que muchos profesionales de la educación asocian a una evaluación descontextualizada dentro del marco pedagógico actual sigue estando vigente en nuestros días con mayor presencia que la evaluación formativa (Zubillaga-Olague y Cañadas, 2021).

Alcaraz (2015a) plantea que el profesorado actual pertenece a una “generación perdida o ecléctica” (p. 22), haciendo referencia a que nos encontramos en un momento en el que hay una extensa bibliografía científica que diferencia la evaluación formativa y calificación, se conocen sus funciones, sus métodos, sus finalidades, hasta cómo hacer que trabajen juntas en favor del aprendizaje, etc., pero la realidad es que se mantienen las prácticas de evaluación tradicionales, como mucho, enmascaradas bajo nuevos términos, pero basadas principalmente en la calificación. Cambia el nombre, la esencia se mantiene.

1.2. Evaluación y calificación, procesos distintos

A lo largo de los años, muchos autores y autoras han definido con sus propias palabras (Álvarez-Méndez, 2001; Llamoca *et al.*, 2022; Rosales, 2014) o han recopilado palabras de colegas de profesión para perfilar conceptos relacionados con la evaluación y la calificación (Cruzado, 2022). Gracias a estos trabajos, se pueden encontrar múltiples definiciones o acotaciones tanto de evaluación como de calificación.

A continuación, se exponen definiciones de evaluación y calificación, pero también de evaluación formativa y evaluación sumativa, pues se entiende en este artículo que se asocian respectivamente la primera con la segunda.

Según Rosales (2014), quien incluye en su trabajo un pequeño glosario de conceptos clave, recoge que la evaluación formativa debe facilitar la toma de decisiones, basándose sobre todo en la observación y análisis de tareas, no solo de tareas o actividades formales de evaluación, sino en el quehacer diario del aula, recogiendo información no solo de resultados, sino del proceso, conociendo mejor al estudiante, destacando la función reguladora. Sería aquella que se usa para adaptar el proceso didáctico a los progresos y necesidades del estudiante para tomar decisiones en base a las dificultades y potencialidades detectadas para lograr el éxito. Por otro lado, entiende la evaluación sumativa como aquella que se realiza al término de un proceso para verificar los resultados, determinar si se lograron los objetivos estipulados y en qué medida, para calificar al estudiante o programa educativo, poniendo el acento en la recogida de información mediante instrumentos que permitan obtener valores fiables, comparables, y hacer mediciones. Puede coincidir con la evaluación formativa en que también puede pretender mejorar a partir de los resultados.

En varios de sus trabajos, Alcaraz (2015a; 2015b) y Alcaraz y Fernández (2024) hablan de ambos conceptos, evaluación y calificación. Concluyen, tras sus estudios, que el concepto de evaluación ha ido evolucionando con el paso de los años, aunque a día de hoy sigue confundándose con la calificación, una confusión que no solo se queda a nivel terminológico, sino que afecta a la forma de entender todo el proceso evaluativo en educación. Citando a autores, acudiendo incluso a la RAE, determinan que la calificación se caracteriza por la obsesión de medir el aprendizaje, algo que se sabe imposible, pero que a pesar de ello se intenta desesperadamente en un intento de buscar la precisión. Así, la calificación gasta su energía en comprobar continuamente si se alcanzan los objetivos, mientras que la evaluación pretende

reorientar las situaciones para que el aprendizaje se produzca. La calificación es, en palabras de la autora, el arma más poderosa para clasificar a las personas entre las que valen y las que no. Los docentes comentan que evalúan para comprobar los aprendizajes, pero la evaluación no persigue comprobar un aprendizaje, sino investigar las condiciones para que este se produzca, en favor y al servicio del mismo, y no al revés. Esta sería, según la autora y el autor, la principal diferencia entre evaluar y calificar: la evaluación está al servicio del aprendizaje, pero el aprendizaje suele estar, en la práctica, al servicio de la calificación.

Álvarez-Méndez (2001) comienza haciendo una clara distinción entre evaluar, con intención formativa, y calificar, con intención de medir. Evaluar puede implicar actividades como corregir, examinar, pasar pruebas, medir, etc., pero no debe confundirse con ellas, y la manera de evitar confusiones es atender a los recursos y fines a los que sirven. De estas actividades (corregir, examinar, medir, etc.) no se aprende, y es precisamente al punto donde no llegan estas actividades donde empieza la evaluación. El docente evalúa para comprender y mejorar su práctica y el aprendizaje del estudiante, el cual mejora gracias a la información que recibe sobre su desempeño desde un punto de vista crítico y argumentado sin intención penalizadora. La calificación, por su parte, llega tarde para el aprendizaje, solo pudiendo clasificar, seleccionar y excluir. El autor explica que la confusión entre evaluar y calificar se debe a la mezcla de funciones que cada vez más se le asignan a la evaluación educativa: funciones relacionadas con el aprendizaje, pero también con la selección y certificación, funciones administrativas, motivacionales, etc.

En el artículo de Shepard (2006), la principal preocupación de la evaluación formativa es la mejora del proceso de aprendizaje del alumnado, mientras que la evaluación sumativa se preocupa por alcanzar un nivel adecuado de validez y fiabilidad. Tanto la validez como la fiabilidad son importantes para garantizar una adecuada evaluación. Estos factores de validez (que los exámenes midan lo que se pretende medir) y fiabilidad (los resultados son reflejo de lo recogido en el contexto abordado) son los llamados “problemas evaluativos”, sobre los cuáles profundizamos en un apartado posterior.

Zubillaga-Olague y Cañadas (2021) nos hablan de los tres momentos clave de un proceso educativo y los tres tipos de evaluación que se aplicarían en cada momento: la evaluación al inicio, llamada evaluación diagnóstica; durante el proceso, con una finalidad formativa; y al final, hablando de una evaluación sumativa. Sin embargo, como apunta López-Pastor (2005), coincidiendo con Morales-Lobo y Fernández (2022), no toda evaluación continua, aplicada durante el proceso, es *per se* formativa, ya que puede darse el caso en el que un docente aplique de manera continuada una sucesión de evaluaciones sumativas.

Por su parte, Hortiguëla *et al.* (2019) exponen que la evaluación (formativa) debe aplicarse de manera obligatoria en educación, señalando cinco aspectos clave que la caracterizan: permite ser más consciente de lo que se aprende, trabaja la capacidad de autorregulación, permite amoldar el proceso de enseñanza-aprendizaje al estudiante, tiene la posibilidad de dar *feedback* por varios canales y mejora la práctica docente.

1.3. La calificación, un elemento importante

Dedicamos un subapartado exclusivo a exponer aspectos importantes sobre la calificación, en primer lugar, porque el objeto de este estudio es ubicarla dentro de nuestro sistema educativo como un proceso necesario, indagando en la literatura de evaluación si puede compaginarse y de qué manera con la evaluación; y en segundo lugar, porque el auge de la evaluación formativa en los últimos años ha provocado que la cantidad de obras científicas dedicados a la calificación sea mucho menor de los centrados en la evaluación (Mellado *et al.*, 2021).

Con respecto a la calificación, se pueden encontrar estudios que usan el término “calificación” exclusivamente y otros que usan “evaluación sumativa”. Ruiz (2009), en su trabajo sobre la evaluación sumativa, manifiesta que suele usarse más para la evaluación de productos terminados para determinar su valía en función del uso que desea hacer con dicho producto.

A la hora de calificar a sus estudiantes, los docentes buscan hacerlo de la manera más justa posible (López-Pastor, 2005; Salcedo *et al.*, 2013). Sin embargo, muchos autores (Cruzado, 2022; Rodríguez *et al.*, 2018) revelan los problemas que puede encontrar la evaluación sumativa a la hora de decidir las calificaciones de los estudiantes. Se pudiera pensar que al tener la evaluación sumativa un enfoque más positivista, cuantitativo, objetivo, etc., será más exacta, justa, y gozará de mayor validez y fiabilidad que la evaluación formativa, de carácter subjetiva. Por ejemplo, en pruebas o exámenes, la calificación se establece en función de las respuestas de los estudiantes en el momento concreto de ese examen. Además, dicho examen será corregido por un docente desde su subjetivo punto de vista (Cruzado, 2022). En etapas preuniversitarias, donde las rúbricas tienen cada vez más presencia en un intento de buscar alternativas al tradicional examen, es difícil que una misma práctica o actividad evaluable sea corregida por varios evaluadores, lo cual invalida la opción de comprobar su fiabilidad inter-evaluadores (Rodríguez *et al.*, 2018). Otro aspecto a recalcar es que muchos docentes deciden calificar al alumnado atendiendo no solo a la demostración de su desempeño en relación a un objetivo, por ejemplo, sumando o restando a la calificación valor en función del esfuerzo o la constancia (Shepard, 2006), aspectos difícilmente medibles que afectan a la validez de la calificación (dejaría de medirse lo que en principio pretendía medir).

La calificación como tal, no acompañada de un proceso de reflexión, no tiene sentido en el sistema educativo. Es decir, “la nota ofrecida a un alumno o alumna sin más no tiene valor didáctico ni educativo” (Fernández Sierra, 1996, p. 3). Por el contrario, discutir con ellos las calificaciones, su interpretación, la tarea desempeñada... es enseñarles a aprender. En su artículo, Fernández (1996) concluye que la prevalencia de la calificación se debe a intereses sociopolíticos y económicos, obligando al docente a prestar más atención a calificar que a evaluar, a certificar que a mejorar el aprendizaje; pero también a intereses gremiales, poco explicitados, pero existentes, pues muchos docentes no desean renunciar al poder que les otorga el poder calificar a sus estudiantes y hacer diferentes usos de la calificación como mecanismo de supervivencia en el aula, funciones coercitivas, punitivas y sancionadoras.

Algunas de las críticas más señaladas por diversos autores a la calificación serían:

- Se estudia para aprobar el examen: lo cual implica no solo estudiar lo que entre en el examen porque es lo único importante, sino otros aprendizajes, como descubrir lo que el profesor de turno quiere oír, la memorización en detrimento de la comprensión y la réplica, etc.
- La educación al servicio del examen, en vez de al contrario: la educación deja de ser un servicio o sistema para aprender para ser un recurso para prepararse de manera continua para cada examen. No es que el examen deba prohibirse, pues bien utilizado es un recurso aceptable, pero no debe condicionar el currículum ni la enseñanza (Álvarez-Méndez, 2001).
- Lo que no se califica, no es importante. “Para el alumnado queda claro que aquello que no se pueda canjear por unas buenas puntuaciones no merece la pena” (Alcaraz, 2015b, p. 215). Incluso se llega a pensar que el docente que realiza pocas evaluaciones sumativas que conlleven calificaciones, es peor docente o, al menos, es motivo de desconfianza (López-Pastor, 2005). En vez de usar el examen como un recurso para mejorar la educación (lo cual es posible), se usa la educación para mejorar los

exámenes. En palabras de Álvarez-Méndez (2001), “en las prácticas habituales lo más valioso suele identificarse con aquello que más puntúa” (p. 11).

- La forma de concebir el éxito y el error: el error es una fuente de aprendizaje muy valiosa (Morales-Lobo y Fernández, 2022). Sin embargo, en el contexto de la evaluación sumativa, el error penaliza, resta nota en la calificación total, con lo cual, hay que evitar equivocarse a toda costa, incluso recurriendo a las trampas. Además, el fracaso escolar se suele asociar a la valía del estudiante, de tal manera que bajas calificaciones implica que no se vale para estudiar, lo cual tiene un peso enorme para el autoconcepto y autoestima de los estudiantes (López-Pastor, 2005), cuando en realidad este fracaso escolar está asociado, en muchas ocasiones, con circunstancias socioeconómicas y culturales que para nada son elegibles por los estudiantes, además de la imposibilidad de la escuela para dar una respuesta diversificada (Gallego-Noche, 2019) en contraposición al pensamiento de que el fracaso escolar se produce por una mala decisión personal en pro de otros proyectos personales (Fernández, 1986, como se citó en López-Pastor, 2005).
- La desprofesionalización docente: la cultura de la evaluación externa para el control y la era de rendición de cuentas en la que nos encontramos (Shepard, 2006) conlleva a que el docente se convierta en un mero ejecutor técnico de estas “evaluaciones”. Para que una evaluación externa, ejemplo máximo de evaluación sumativa, pudiera ser útil en un contexto formativo, debería estar alineada con los objetivos de aprendizaje de cada programación didáctica en la que se inserta, pudiendo usar las calificaciones obtenidas para enriquecer la retroalimentación al alumnado.

1.4. Validez y fiabilidad de la evaluación y la calificación

Una de las diferencias analizadas en la literatura entre la evaluación y la calificación es la relativa a la validez y fiabilidad en sus procedimientos y en sus resultados. De hecho, la búsqueda de validez y fiabilidad es de las principales preocupaciones de la evaluación sumativa (Arribas, 2017) y una de las críticas más importantes que se le acusan a la evaluación formativa (Álvarez-Méndez, 2001). No obstante, esta crítica no tiene fundamento pues, en actividades evaluativas formativas se puede perseguir la validez y fiabilidad por vías distintas a las pruebas calificadoras objetivas de mentalidad positivista, tal y como expone Álvarez-Méndez (2001):

En este nuevo contexto de entendimiento, podemos llegar por otros caminos a los valores que representan la validez y la fiabilidad matemática para la mentalidad positivista, no necesariamente estadísticos: debate, diálogo, discusión, negociación, consenso, deliberación, argumentación, credibilidad, confiabilidad, responsabilidad, razonamiento (p. 45).

Aunque hay que tener en cuenta que, si bien la evaluación formativa, más subjetiva y cualitativa, puede ser menos fiable que la evaluación sumativa, más objetiva y cuantitativa, es cierto que la formativa, al desarrollarse de manera continua en el día a día de clase, tiene la oportunidad de corregir apreciaciones erróneas en el desempeño de un estudiante mediante la información recogida día a día, llegando a tener un conocimiento más individualizado de cada estudiante que la propia evaluación sumativa (Shepard, 2006). Asimismo, la evaluación sumativa, desarrollada principalmente a través de pruebas o exámenes, puede no ser tan objetiva en la práctica, puesto que las calificaciones que los docentes determinasen hacen en base a las evidencias que han podido recoger y valorar, con la subjetividad que conlleva (Cruzado, 2022).

Salcedo *et al.* (2013), tras su estudio, concluyen que para que un examen sea aplicado como una herramienta eficaz en favor de una adecuada evaluación sumativa debe explicitar las “claves de calificación”, entendidas estas como el elemento que garantiza la objetividad y la calificación justa de los exámenes. Estas claves de calificación deben incluir:

- Los objetivos que se pretende evaluar en cada pregunta.
- Los indicadores de evaluación.
- Los criterios de calificación, es decir, el valor asignado a cada respuesta posible en cada pregunta. Por supuesto, en caso de preguntas abiertas, los autores aclaran que no hay que incluir todas las respuestas posibles, bastando señalar lo que se exigirá para otorgar una puntuación u otra dentro de una escala.
- Indicaciones para la calificación global del examen. Aquí se debe explicitar cómo se obtendrá la calificación final, que puede incluir aspectos más allá del sumatorio de las calificaciones de las preguntas, por ejemplo, atender a las faltas de ortografía.

Esta forma tan exhaustiva de preparar exámenes sin duda conlleva una inversión de tiempo y esfuerzo para el docente, aunque no es ese el factor más criticable. Aunque se pueda pensar que este método aumenta la objetividad, validez y fiabilidad de los exámenes, el último apartado de “indicaciones para la calificación global” puede chocar frontalmente con autores como Polo (2011), que aclaran que el hecho de atender a aspectos no incluidos en los referentes de evaluación puede alterar la evaluación (medición), en cuanto que no se mide exactamente aquello que se quiere medir.

1.5. El porqué de la confusión que aún persiste

Uno de los problemas en el campo de la evaluación es la confusión terminológica a la que está arraigada (Alcaraz, 2015b; López-Pastor, 2005). Una de las confusiones más destacadas es la equiparación de “evaluación” con “calificación”, llegando a usarse en muchas situaciones como términos sinónimos, copando la calificación un espacio que bien podría ser aprovechado por la evaluación (formativa, educativa, pedagógica) para provocar mejoras en el proceso de enseñanza-aprendizaje. Esta superposición de la calificación sobre la evaluación provoca, todavía más, que el alumnado (y también el profesorado) esté más preocupado (o únicamente preocupado) por las notas, las calificaciones, que por su propio aprendizaje. Lo que importa es el resultado.

¿De dónde viene esta confusión entre evaluación y calificación? López-Pastor (2005) nos explica que la evaluación del estudiante tiene, en principio, una doble función:

- Ofrecer información interna, en el contexto de aula, para mejorar los procesos de enseñanza-aprendizaje a través del seguimiento, monitorio continuo, la retroalimentación y la toma de decisiones durante el proceso, para aprender de los hallazgos encontrados.
- Ofrecer información externa, de cara a la sociedad, con intención acreditativa y certificadora, que sirva como boleto para promocionar.

Claramente, la primera de las funciones descritas está relacionada con la evaluación formativa, más relacionada con la enseñanza y el aprendizaje, mientras que la segunda hace referencia a la evaluación sumativa, más directamente conectada con la calificación, la notificación y certificación social. En este contexto, en una sociedad meritocrática, en la que lo que prevalece es la acumulación de acreditaciones, el alumnado, especialmente en la Educación Secundaria y postobligatoria, debe alcanzar unas calificaciones determinadas para competir con sus compañeras y compañeros por el acceso a determinados estudios. Esta idea atormenta no solo

a los estudiantes, sino a las familias, que desean que sus hijas e hijos tengan éxito académico, asociando este éxito a altas calificaciones; y a los propios docentes, que intentan aplicar una evaluación lo más justa que pueden, enfrentándose a un proceso de evaluación con tantas funciones que muchas incluso se contradicen (Alcaraz, 2015a; López-Pastor, 2005). La filosofía actual de la evaluación, recogida en muchas normativas estatales y autonómicas, no solo en España, sino en muchos países (Gallardo-Fuentes *et al.*, 2023), de manera explícita recogen como primordial, pedagógicamente y normativamente hablando, la aplicación de una evaluación formativa, para el aprendizaje, que promueva la participación del alumnado y su implicación en su propio aprendizaje mediante la autorregulación favorecida por la retroalimentación constante del docente, que evalúa de manera continua durante el proceso de aprendizaje usando diferentes técnicas e instrumentos; pero, a su vez, explícitamente solicitan que no se abandone la evaluación sumativa, calificadora, certificadora, con función burocrática-administrativa, , que permita acreditar el aprobado, la promoción y titulación del alumnado, desde los primeros cursos de la educación obligatoria. Además, de manera implícita, se le da mucha más importancia a esta última, a la evaluación sumativa y la calificación, pues es la que puede ser objetivo de revisiones y reclamaciones, la que es utilizada para la acreditación, como llave para el acceso a cursos superiores o etapas educativas posteriores, para seleccionar al alumnado en régimen de competitividad, etc., y no solo en etapas postobligatorias, que hasta cierto punto tiene su justificación y, como dice Santos-Guerra (2024), es debatible. Esta competición y obsesión por la calificación por parte de la administración ocurre ya desde el primer curso de la Educación Primaria.

Álvarez-Méndez (2001) denuncia igualmente que a una misma actividad de evaluación se le puede asignar funciones muy diferentes, incluso contrarias:

Parte de la confusión que se produce en la evaluación del rendimiento de los alumnos se debe a la mezcla de funciones que se asignan a la evaluación educativa. Es evidente que desempeña muchas y muy variadas, más de las que formalmente se le reconocen: formación, selección, certificación, ejercicio de autoridad, mejora de la práctica docente; funciones relacionadas con la motivación y la orientación; funciones administrativas, académicas de promoción o de recuperación; de información y de retroalimentación, de control. Aumentan en cantidad en la misma proporción en que pierden en transparencia. [...] En la confusión, un mismo acto, un mismo concepto, una misma actividad de evaluación desempeñan funciones contrarias para las que no están pensados (p. 24).

Toda esta gran contradicción presiona al profesorado, quien puede ser conocedor de las bondades de la evaluación formativa y la no utilización de exámenes para favorecer el aprendizaje, pero que debe atender a esta demanda administrativa, burocrática y social, a través de exámenes (o variantes), lo que, por otro lado, es comprensible, pues como nos recuerda Arribas (2017), la evaluación en el ámbito académico tiene la función, además de la mejora de la propia educación, de rendir cuentas, además de servir como instrumento de acreditación social. Fernández Sierra, ya en su artículo de 1996, explicaba que el hecho de requerir calificaciones para acreditar la obtención de títulos y el acceso a otros estudios es algo lógico, no discutible, pero no debería ser aplicable dentro de la Educación Secundaria Obligatoria. En este contexto, el autor hace una reflexión utópica, aunque prefiere no renunciar a ella: la posibilidad de que todos los estudiantes reciban el certificado al terminar la etapa, en un intento de evitar la discriminación de aquellos que terminan con título de los que no y, de paso, liberando a los docentes de las presiones de las calificaciones y usando la evaluación para reorientar la enseñanza y reconducir el aprendizaje. Sin embargo, es consciente de que hoy en día es algo imposible. De hecho, el sistema productivo en el que vivimos, cada vez más especializado y competitivo (Gallego-Noche, 2019), obliga a la escuela a hacer una tarea poco

educativa: seleccionar y clasificar a los estudiantes, controlando la calidad de los sujetos que pasan por la educación. Así, los estudiantes se irán diferenciando y clasificando según su éxito escolar y las decisiones que tomen. De este modo, la calificación actúa como un mecanismo de clasificación social, adquiriendo el docente un papel muy importante, de alta responsabilidad, no solo profesional, sino social y ética, obligando a los docentes a desempeñar el rol de educador y, a la vez, juez.

Un aspecto esencial en el debate científico sobre el trabajo conjunto entre la evaluación formativa y la sumativa es en la retroalimentación que se ofrece al alumnado, especialmente la naturaleza e intención de la misma (Ruiz-Cuéllar, 2021). López y Osorio-Sánchez (2016) describen la retroalimentación como el punto neurálgico de la evaluación para que los estudiantes tengan conocimiento sobre qué metas deben lograr, si lo han hecho y qué deben hacer para alcanzarlas.

En este sentido, Llamoca *et al.* (2022) conciben la evaluación como un proceso en continuo crecimiento cuyo objetivo principal es el desarrollo de aprendizajes para orientar, guiar y hacer de ese aprendizaje una realidad, dando a conocer el nivel de aprendizaje al estudiante mediante la retroalimentación, buscando involucrar al estudiante para que regule su propio aprendizaje. Este aprendizaje mejora, según los autores, por la autorreflexión a la que llegan los estudiantes, estimulando su motivación por aprender en lugar de incitar procesos evaluativos que provocan ansiedad o miedo en el estudiante.

Morales-Lobo y Fernández (2022) y Sanmartí (2020) explican en sus respectivos libros que al acompañar la retroalimentación que se da al alumnado con calificaciones, dentro de un enfoque formativo, la esencia y capacidad de esa retroalimentación para mejorar el aprendizaje pierde mucho valor, pues el estudiante centrará su atención en la calificación. Con respecto a esta situación, Arribas (2017) se pregunta si la calificación puede actuar como un reforzador extrínseco o de verdad pervierte el aprendizaje. Sin negar el importante valor certificador que posee la calificación, en un contexto de evaluación sumativa o certificadora, es cierto que el aprendizaje es algo que no puede observarse ni medirse de manera directa, con lo cual, el intento de medición del citado aprendizaje es algo muy criticado por muchos autores (Alcaraz, 2015b). Arribas (2017) aclara, ante esta afirmación, que aunque ciertamente el aprendizaje no sea algo cuantificable ni calificable de manera directa, los docentes miden los criterios e indicadores de evaluación de los mismos:

Tales constructos, al igual que las competencias, no pueden ser medidos directamente, se miden los indicadores que ponen de manifiesto esas competencias; luego es imprescindible, en primer lugar, disponer de unos indicadores de los que se infiera, realmente, la existencia de aquellos (Arribas, 2017, p. 394).

2. Metodología

El presente estudio fue realizado aplicando una metodología cualitativa de revisión no sistemática o revisión narrativa (Ortiz, 2005) de literatura científica, consultando artículos y libros relacionados con el objeto de estudio. A pesar de seguir un protocolo estructurado de pasos para la puesta en marcha de esta revisión bibliográfica, no se corresponde con todas las premisas de una revisión sistemática (Moreno *et al.*, 2018), aunque sí se han tenido en cuenta muchas de sus características, expuestas en Anguera (2023). De esta forma, en primer, lugar se diseñaron preguntas de investigación para decidir, posteriormente, las palabras claves y operadores booleanos *AND* y *OR* que se usaron para la búsqueda de artículos. Así, se hicieron búsquedas como las siguientes: “evaluación sumativa” *OR* “summative assessment”, “evaluación formativa” *OR* “formative assessment”, “evaluación formativa *AND* evaluación

sumativa”, “evaluación AND calificación”, para encontrar artículos centrados en la evaluación formativa o sumativa, o artículos que estudiaran la relación entre ambos enfoques. Se aplicaron como criterios de exclusión la fecha de publicación, para centrar la búsqueda en los últimos 20 años, aunque adicionalmente se amplió la búsqueda para incluir artículos reconocidos en la temática para tener una perspectiva histórica más amplia. Asimismo, no se limitó la búsqueda a artículos referidos al sistema educativo español, para tener una perspectiva comparativa de la evaluación con otros países de habla hispana.

Las búsquedas se realizaron en varias plataformas: Scielo, Scopus, Web Of Sites, Proquest y Google Scholar, descartando aquellas revistas que no fueran del ámbito de la educación.

Una vez seleccionados los artículos suficientes, se comenzó el análisis de todos ellos, categorizándolos según si el eje central de cada artículo se centraba en uno u otro enfoque de evaluación, si abordaba la relación entre ambos, si eran de investigación o de revisión y si hacían referencia a etapas educativas preuniversitarias o universitarias, dado que este artículo pretendía centrarse en la etapa preuniversitaria, concretamente, en la educación obligatoria.

Analizados los artículos, se procedió a la síntesis de todos ellos, construyendo un corpus unificado con los aspectos que compartían sobre ambos enfoques de evaluación, formativa y sumativa. De este modo, de acuerdo con Llamoca *et al.* (2022), se trató de recolectar, recopilar y seleccionar información relevante de múltiples fuentes para tener una visión panorámica.

3. Resultados

Tras la búsqueda de publicaciones se seleccionaron un total de 57 artículos de revistas de Educación y 6 manuales. Del total de artículos, se consideraron 29 artículos y los 6 manuales que hacían alusión a la evaluación formativa, a la evaluación sumativa, a la calificación o a la relación entre varios de estos procesos. Todos los trabajos se clasificaron según su año de publicación de la siguiente manera (nótese que no se han computado en este recuento dos de los artículos más recientes cuya consulta se realizó para fundamentar la metodología empleada, no siendo artículos relacionados con el objeto de estudio):

Tabla 1.

Trabajos seleccionados para el análisis

Antigüedad	N.º de trabajos
Últimos 10 años	24
Entre 20 y 10 años de antigüedad	8
Más de 20 años de antigüedad	3
Total	35

Fuente: Elaboración propia (2024).

De este análisis se extraen los siguientes resultados en base a los dos objetivos planteados inicialmente: por un lado, descubrir en la literatura si existe la posibilidad de que la evaluación y la calificación trabajen juntas en el proceso educativo; y por otro lado, en caso de encontrar referencias a favor, exponer estrategias concretas para lograrlo.

3.1. Evaluación y calificación, ¿irreconciliables?

Una vez conocidas las diferencias entre evaluación y calificación, se aborda la cuestión sobre si ambos procesos pueden trabajar juntos, o si sus diferencias son irreconciliables.

Según Martínez (2021), las diferencias que contraponen a ambos procesos son tales que el término “sumativa” (evaluación asociada a las calificaciones) ha adquirido connotaciones peyorativas, pues se supone lo contrario de lo pedagógicamente correcto: valora solo el final, no el proceso; pretende calificar, en lugar de reorientar; es injusta al ser estandarizada, en lugar de adaptarse a cada individuo; no se preocupa por la mejora, mientras que la formativa, sí:

En el colectivo docente, el término de evaluación sumativa, tiene connotaciones negativas ya que se contrapone a la lógica de la evaluación formativa debido a que se percibe como una imposición administrativa del sistema. Es por esta razón se prefiere hablar de evaluación para la certificación, y no evaluación sumativa (Martínez, 2021, p. 12).

Uno de los trabajos que analiza esta relación en términos de *amor-odio* es el de Shepard (2006), donde se explica que la evaluación sumativa y la calificación conforman “una seria amenaza para los objetivos de aprendizaje declarados por la evaluación formativa” (p. 30), dado que las calificaciones pueden tener efectos muy negativos en la motivación intrínseca de los estudiantes para aprender, haciendo que los estudiantes estudien solo aquello que se califica, aspecto con el que concuerda Alcaraz (2015b), quien opina que estudiantes curiosos y creativos, que podrían estar motivados de manera intrínseca hacia el aprendizaje, se pueden volver aburridos y entrenados para buscar el éxito académico en forma de altas calificaciones, preocupados solo por aquello que es calificable y que suma para “la nota”. Las calificaciones pueden también afectar negativamente a aquellos estudiantes que estiman imposible alcanzar un aprobado en un objetivo concreto, abandonando el estudio del mismo. Otro efecto pernicioso sería que las calificaciones, a través de las cuales suele crearse un entorno competitivo por las notas, pueden minar la buena voluntad entre compañeras y compañeros para ayudarse o aprender de otros. Sin embargo, esta misma autora recopila en el mismo trabajo otros aspectos positivos de la calificación, como el hecho de que hay cierta evidencia extraída de la investigación motivacional que avala el uso de las calificaciones. Por ejemplo, parece que los estudiantes estudian y se interesan más cuando esperan que se les realice una prueba con nota. Además, obliga a los estudiantes a estudiar y repasar, así como redirige la atención del estudiante hacia lo que al docente le interesa. Por otro lado, se les da la oportunidad a los estudiantes de poner a prueba sus conocimientos.

Martínez-Rizo (2013) expone que el impacto (en el aprendizaje) de las evaluaciones del enfoque formativo es más positivo que las del enfoque sumativo, aunque advierte que no hay que tener una visión simplista de esta afirmación, pues se puede cometer el error de pensar que la mera aplicación (o el intento de aplicación) de la evaluación formativa conllevará una mejora automática del proceso de aprendizaje de un grupo de estudiantes.

3.2. Trabajando juntas

Podemos encontrar alusiones en la literatura revisada al trabajo conjunto de la evaluación y calificación que mejoran el proceso de aprendizaje. En algunas ocasiones, se trata de dotar a la evaluación formativa de características, finalidades o recursos de la evaluación sumativa y, a veces, hacer lo contrario: hacer que la evaluación sumativa, en principio solo calificadora, asuma aspectos clave de la evaluación formativa, creando un híbrido en pro del aprendizaje.

Rodríguez-Negro y Zulaika-Isasti (2016) investigaron, en un estudio cuantitativo, las formas de evaluar del profesorado de Educación Primaria y Secundaria en la materia de Educación Física para analizar las posibles diferencias entre la teoría y la práctica. Sus conclusiones fueron que el profesorado está cambiando su forma tradicional de evaluar, abrazando poco a poco formas alternativas de evaluación, aunque sin descartar la primera. Estas conclusiones van en sintonía con las que obtuvieron Zubillaga-Olague y Cañadas (2021) en su estudio, del cual extraen que el profesorado de Educación Física en Educación Primaria y Secundaria aplica diferentes modelos de evaluación (inicial, continua y final), aunque la evaluación con fines de calificación sigue teniendo, desde el punto de vista de los profesores y profesoras participantes en el estudio, mucho peso.

Una estrategia descrita en Martínez-Rizo (2013) y en Zubillaga-Olague y Cañadas (2021) consiste en acompañar las calificaciones de las evaluaciones sumativas con retroalimentación, lo cual otorga un matiz formativo que será más o menos rico en función de la calidad, adecuación y dedicación de dicha retroalimentación. Esta retroalimentación, para que tenga de verdad carácter formativo, no puede limitarse a comentarios vacíos como *excelente trabajo* o *necesita más citas*, pues se deben ofrecer comentarios relacionados con grados de desempeño previamente definidos y estrategias para mejorar, yendo más allá de la mera transmisión de información (Nicol y Macfarlane-Dick, 2006).

Del trabajo de Bartolomé (1990, como se citó en Rosales, 2014), hace más de 30 años, se desprende que el autor ya asumía como imprescindible la compaginación de la evaluación formativa y sumativa para una adecuada evaluación educativa. En esta línea, Shepard (2006) coincide en que la mejor estrategia es aquella en la que se consigue alinear la evaluación formativa y la sumativa con los objetivos educativos:

De hecho, desde una perspectiva cognitiva, el mejor sistema sería aquel en el que las evaluaciones sumativa y formativa estuvieran mutuamente alineadas con objetivos de aprendizaje orientados conceptualmente, y en el que las evaluaciones sumativas se utilizaran como momentos importantes de logro (quizás reconocidos por la familia y los amigos) después de felices periodos de aprendizaje reforzados por la evaluación formativa (p. 35).

El autor está igualmente de acuerdo con la posibilidad de utilizar recursos propios de la evaluación sumativa con carácter formativo:

Aun si uno toma tiempo para un examen formal, los resultados pueden utilizarse para un diagnóstico de la enseñanza con el fin de decidir qué conceptos necesitan más análisis y trabajo. Y los estudiantes pueden llegar a entender que tales evaluaciones tienen una finalidad de aprendizaje (p. 19).

Sin embargo, sigue exponiendo el autor, hay que tratar de no desdibujar las diferencias que hay entre ambas. La evaluación formativa se aplica para aprender y la sumativa para demostrar los logros alcanzados. Un error común que hay que evitar es que las evaluaciones sumativas no deben ser una mera repetición de tareas que previamente se han hecho iguales pero con enfoque formativo, sino que deben permitir al estudiante aplicar su aprendizaje. Es decir, una buena estrategia sería desarrollar una evaluación formativa durante una unidad didáctica, insertando durante su desarrollo y, al final de la misma, evaluaciones sumativas con calificaciones que permitan al estudiante probar y conocer su capacidad para aplicar lo aprendido. No es recomendable introducir calificaciones durante el proceso de una evaluación formativa o, de hacerlo, presentarlas como “calificaciones hipotéticas” que apoyen la retroalimentación al estudiante para reorientar su aprendizaje.

En su artículo de los años noventa, Alonso *et al.* (1996) comparaban la evaluación con la calificación, como dos enfoques distintos, pero complementarios. Evaluar no es calificar, pero ambas pueden trabajar de la mano. Los autores no niegan que se puede aprender, y mucho, solo con la presencia de la evaluación formativa y dejando las calificaciones al margen. Sin embargo, reconocen que las calificaciones constituyen una manera rápida de dar información al estudiante sobre su progreso, por ejemplo, en la Educación Secundaria, donde el alumnado debe conocer su progreso de cara a la calificación final de curso. El docente debe aplicar la evaluación (formativa) para dar retroalimentación al estudiante durante el curso para que vaya mejorando su propio aprendizaje, a lo que puede acompañar sucesivas calificaciones cualitativas significativas acompañadas de reflexiones que favorezcan la autorregulación.

Aunque no analiza en profundidad estrategias concretas para abordar de manera compaginada la evaluación formativa y sumativa, Achega (2013) sí extrae los beneficios que supondría esta compenetración, destacando las siguientes: una retroalimentación más completa mediante el feedback de la evaluación formativa de manera constante, acompañada de instantáneas sobre el rendimiento general mediante evaluaciones sumativas; mayor participación de los estudiantes, gracias a la interacción constante docente-estudiante que facilita el enfoque formativo, unido a la sensación de logros y cierre de las evaluaciones sumativas; mejora del aprendizaje personalizado, al usar tanto el seguimiento continuo formativo como las calificaciones para mejorar la experiencia del estudiante; evaluación integral desde los dos enfoques, más cualitativo y cuantitativo; enseñar y adaptar el proceso educativo basándose en datos del rendimiento.

Reflexionando más en la práctica, Nicol (2010) determina que, en el mundo real, los docentes rara vez tienen tiempo para dar sesiones de tutoría individuales para dar una retroalimentación de calidad, proponiendo como solución la creación de rutinas de aula que garanticen interacciones mínimas con todo el alumnado, haciendo uso de comentarios por parte del docente y también entre compañeras y compañeros de clase. La autora propone también el uso de mapas de progreso para mejorar la experiencia de aprendizaje del alumnado. A diferencia de los informes de evaluación, normalmente se constituye una lista de verificación sobre la consecución o no de diferentes objetivos, un mapa de progreso muestra de manera más pormenorizada qué aspectos ha desarrollado más y cuáles menos, similares a una rúbrica analítica.

Santos (2016) considera que la articulación entre la evaluación formativa y sumativa es más bien una tarea compleja, pero no imposible. Propone, para lograrlo, la creación de un Plan de Trabajo Individual (PTI), en el que se oriente y apoye a cada estudiante de manera personalizada, usando ambas evaluaciones para monitorear el progreso individual de cada estudiante, ofrecer comentarios oportunos a cada uno, combinando la retroalimentación continua con las evaluaciones finales para mejorar tanto el desarrollo como el rendimiento. Las claves de los PTI se basan en esa orientación y apoyo personalizados, así como que ambos tipos de evaluación pueden nutrirse entre ellas: la evaluación formativa puede manifestar los progresos de los estudiantes, los cuales demuestran dicho progreso en las sucesivas evaluaciones sumativas; por otro lado, las calificaciones de las evaluaciones sumativas pueden verse modificadas, siguiendo unas reglas predefinidas, atendiendo a los progresos registrados en la evaluación formativa.

Otra estrategia efectiva y cada vez con mayor presencia es el modelo de evaluación sostenible, una manera de aplicar la evaluación educativa en cuyos principios y bases se encuentran los principios de la evaluación formativa y sumativa (Boud, 2000; Boud y Soler, 2016). La evaluación sostenible es aquella que, además de cumplir los requisitos actuales de la evaluación para el aprendizaje y del aprendizaje, prepara a los estudiantes para aprendizajes

futuros. Busca fomentar la reflexión, la autoevaluación y la toma de decisiones sobre el propio aprendizaje, destrezas que permiten al estudiante adaptarse a desafíos futuros y seguir aprendiendo durante toda la vida. La evaluación sumativa se considera una importante fuente de información que puede ayudar en la evaluación sostenible al ofrecer información valiosa al estudiante sobre su rendimiento y ayudarlo a ser crítico consigo mismo, desarrollando la autoevaluación y autorregulación. Por su parte, el enfoque formativo está totalmente presente en la evaluación sostenible, en cuanto que contempla la retroalimentación constante al alumnado para mejorar su aprendizaje, así como el desarrollo, como se ha comentado, de la autorregulación y autoevaluación. Por tanto, la evaluación sostenible persigue integrar la evaluación formativa y sumativa de manera que se refuercen entre sí, no como procesos separados.

Independientemente del modelo o estrategia, es importante evitar lo que Hortiguéla *et al.* (2019) consideran errores clave a la hora de aplicar la evaluación formativa, de los cuales rescatamos tres, por estar conectadas con la calificación:

- No permitir entregas previas: es cierto que la calificación no debe realizarse en base a una actividad que se haya hecho previamente de manera idéntica varias veces, a modo de entrenamiento, aunque sí se deben permitir varias entregas previas a la definitiva que será objeto de calificación para poder aplicar la retroalimentación obligada en la evaluación formativa.
- No garantizar el salto de la evaluación a la calificación de manera pedagógica: para evitarlo, se debe llegar a acuerdos previos sobre cómo se realizará la calificación, y las retroalimentaciones realizadas durante el proceso deben tener impacto en la calificación.
- No asociar procedimientos de evaluación con sus respectivos instrumentos: a la hora de evaluar, se debe planificar qué procedimiento se usará para evaluar y qué instrumento para calificar.

Como síntesis de los datos extraídos de toda la documentación analizada, la calificación usada de manera aislada resulta un planteamiento reduccionista sin efectos educativos para el alumnado, pudiendo ser compatible, a través del diseño de un plan sistemático de trabajo (como el Plan de Trabajo Individual o la evaluación sostenible), con otros elementos relacionados con la evaluación formativa, como el seguimiento individual y monitoreo de la actividad del alumnado, la retroalimentación continuada, la autoevaluación y la autorregulación. Esta integración evaluación-calificación, ya defendida por autoras y autores hace más de 30 años, podría suponer una tarea laboriosa para el profesorado, quien debe guardar un equilibrio entre la aplicación de la evaluación para el aprendizaje y la calificación para la acreditación, ambos procesos obligatorios en la educación.

4. Discusión

La variedad de investigaciones y revisiones consultadas evidencia que, efectivamente, la evaluación y calificación se encuentran sumidas en una importante confusión terminológica, usando ambos conceptos de manera indiferenciada. Mientras que la evaluación persigue recabar datos de manera continua, durante el proceso de enseñanza-aprendizaje, con una clara intención de servir al aprendizaje, reorientarlo y mejorarlo a través de la retroalimentación continua, la calificación, bajo el enfoque sumativo de la evaluación, pretende únicamente medir los resultados alcanzados en base a unos objetivos definidos para poder llevar a cabo, a partir de la información obtenida, acciones de categorización, clasificación, selección (y exclusión), siendo el instrumento preferencial el examen, pruebas u otros instrumentos similares.

Una de las confusiones más expandidas, destacadas por varios autores en esta revisión (Hortiguéla *et al.*, 2019; López-Pastor, 2007; Martínez-Rizo, 2013; Morales-Lobo y Fernández, 2022), es la de equiparar la evaluación formativa con la evaluación continua, lo cual es un error, porque se está mezclando dos tipos de evaluación, la formativa (según el para qué) y la continua (según el momento de aplicación). La equivocación se basa en pensar que la evaluación formativa lo es por ser continua, igualándola a la calificación continua. Para no confundir la evaluación formativa con la evaluación sumativa (calificación) continua, podemos hacer la siguiente distinción: la evaluación será formativa cuando se lleva a cabo una evaluación para el aprendizaje, y será sumativa cuando se realiza una evaluación del aprendizaje.

Una de ellas, al ser el objetivo tradicional de la evaluación desde sus inicios la calificación de los estudiantes para determinar el grado de asimilación de conocimientos tras una lección, la tradición y la costumbre ha provocado que actualmente se siga asociando la calificación como lo más importante en la escuela. Otra de las causas, relacionada con la anterior, es la experiencia que como estudiantes han vivido los docentes actuales, que han sido evaluados (calificados) durante toda su vida y les cuesta romper con esa herencia. Se puede añadir como causa de esta confusión la variedad de funciones que poco a poco se le han ido otorgando a la evaluación en la educación, siendo originalmente calificar, pero luego la de orientar al estudiante, detectar dificultades, reconducir el proceso de enseñanza, etc., recogiendo, incluso en la normativa, la imposición de aplicar funciones que se antojan contradictorias, como evaluar para aprender y evaluar el aprendizaje. Estas contradicciones confunden al profesorado, que no encuentra en esa normativa criterios claros para convertir la evaluación educativa en calificaciones, como ya denunciaba Fernández (1996) y como sigue ocurriendo en la actualidad.

Se puede afirmar, además, en base a los trabajos revisados, que esta confusión entre los términos “evaluación” y “calificación” no es lo más grave, pues esta confusión es solo la punta del iceberg, ya que debajo encontramos toda una confusión en la forma de entender la evaluación en el sistema educativo.

La corriente pedagógica moderna, analizada en este trabajo en artículos desde hace 30 años hasta la actualidad, apuesta por la evaluación formativa como aquella que posibilita el aprendizaje, a la vez que expone los aspectos negativos potenciales que conlleva realizar calificaciones al alumnado. No obstante, muchos otros autores, incluso algunos que exponen estas consecuencias de la calificación, también reconocen que, bien usados los exámenes y otros métodos de calificación, pueden obtenerse beneficios de cara al aprendizaje del alumnado, aunque todas estas estrategias se basan en que la evaluación sumativa o calificaciones asuman aspectos de la evaluación formativa. Podemos deducir, de esta forma, que para que la calificación tenga algún sentido o valor pedagógico, tiene que ir de la mano de la evaluación formativa, ya que por sí sola solo puede asignar una nota a un desempeño, en un intento (inútil, como se ha expuesto) de medir el aprendizaje.

Aunque cueste admitirlo, la calificación se presume necesaria, incluso en etapas obligatorias de la enseñanza, pues es el mecanismo que tradicionalmente se ha usado y ha perdurado hasta nuestros días para la categorización y selección del alumnado. Sin embargo, debería usarse con prudencia y de manera acorde a la edad del estudiante: en etapas como Educación Infantil o Educación Primaria, la calificación no debería condicionar el futuro de las niñas y los niños. El objetivo debería ser aprender y desarrollarse según sus capacidades, intereses, motivaciones y necesidades. No obstante, en Educación Secundaria Obligatoria, aunque el interés pedagógico será, igualmente, aprender y crecer como persona, la calificación que obtenga el estudiante en cada una de las materias y a lo largo de la etapa determinarán una nota media

que sí les condicionará para futuros itinerarios académicos, tristemente. Es por ello que, aunque en la educación obligatoria, lo importante debería ser aprender, no las notas, es comprensible que los estudiantes (y sus familias y docentes) se preocupen por las calificaciones. Para que la Educación Secundaria Obligatoria pueda tener un fin verdaderamente pedagógico, este uso seleccionador de las calificaciones debería desaparecer. Los demás usos que se hagan de las calificaciones (seguimiento del progreso, apoyo a la retroalimentación, información concisa y entendible a las familias frente a extensos y técnicos informes) podrían, bien aplicados, apoyar a la evaluación formativa para favorecer el aprendizaje.

De esta relación entre evaluación y calificación pueden surgir interesantes propuestas prácticas para aplicar en el aula. La evaluación sostenible descrita por Boud y Soler (2016) o el Plan de Trabajo Individual (PTI) expuesto por Santos (2016) son claros ejemplos. Lo cierto es que se debe contar con un plan estructurado para que la evaluación y la calificación trabajen en línea con los objetivos, pues no es una tarea sencilla. De hecho, varias investigaciones concluyen que la evaluación (formativa) o no se está aplicando, a pesar de las reformas pedagógicas y normativas, o se hace muy lentamente y con muchas dificultades (Rodríguez-Negro y Zulaika-Isasti, 2016; Zubillaga-Olague y Cañadas, 2021).

5. Conclusiones

La literatura sitúa de manera indiscutible a la evaluación (formativa) como imprescindible en el proceso educativo (Hortiguëla *et al.*, 2019), para que se produzca el aprendizaje, a través de todos sus elementos, especialmente, la retroalimentación, que favorece el autoconocimiento, autorregulación y metacognición del alumnado en su propio proceso de aprendizaje. Sin embargo, a la hora de hablar de calificación, surge el debate: ¿puede favorecer el aprendizaje? ¿Lo entorpece? Simplemente... ¿no tiene efecto en el aprendizaje? Igualmente, hay mucha literatura sobre cómo evaluar al alumnado, pero no tanta sobre cómo calificarlo (Mellado *et al.*, 2021). Es decir, ¿qué se debe tener en cuenta para calificar al alumnado? ¿Solo la consecución de los objetivos, para ser lo más objetivos posible? ¿Es eso lo más justo? ¿Es justo que a todos los estudiantes, especialmente en la educación obligatoria, se los compare con los mismos objetivos? En esta línea, el profesor Santos-Guerra comparte desde hace años sus metáforas a partir de ilustraciones para visibilizar la injusticia de la evaluación, como aquella en la que muchos animales (un pez en su pecera, un pájaro, un elefante, un gato, etc.) se disponen a hacer una misma carrera para ver quien trepa primero a un árbol (Santos-Guerra, 2024). ¿No está claro que primero llegará el pájaro, y que el pez ni siquiera entenderá las instrucciones?

Es cierto que la calificación en educación no puede realizar con total validez y fiabilidad la medición de lo que un estudiante ha aprendido, aunque presuma de ello, aunque no por ello debemos descartarla del sistema educativo, al menos, hoy por hoy. Mientras el sistema educativo obedezca a otros intereses no educativos, como intereses sociopolíticos y económicos, la calificación seguirá usándose para determinar la valía de un estudiante frente a otro, quién merece más el acceso a una determinada enseñanza y quién tiene que repetir curso. Por ello, dado que actualmente, el profesorado debe convivir con estas funciones contrarias ya descritas de la evaluación, lo más sensato será recurrir a estrategias creativas e innovadoras para aplicar de manera conjunta la evaluación y calificación.

6. Referencias

- Achega, I. (2013). *Os diferentes tipos de avaliação: avaliação formativa e avaliação sumativa* [Tesis de maestría, Universidade Nova de Lisboa]. Repositorio Institucional RUN. <http://hdl.handle.net/10362/13803>
- Alcaraz, N. (2015a). Aproximación Histórica a la Evaluación Educativa: De la Generación de la Medición a la Generación Ecléctica. *Revista Iberoamericana De Evaluación Educativa*, 8(1), 11-25. <https://revistas.uam.es/riee/article/view/2973>
- Alcaraz, N. (2015b). Evaluación versus calificación. *Aula De Encuentro*, 17(2), 209-236. <https://revistaselectronicas.ujaen.es/index.php/ADE/article/view/2662>
- Alcaraz, N. y Fernández, M. (2024). Presentación. Evaluación y calificación ¿una relación toxica? Funciones y finalidades diferenciadas y modelos de convivencia. *Revista Iberoamericana de Evaluación Educativa*, 17(1), 7-9. <https://revistas.uam.es/riee/article/view/19158/17098>
- Alonso, M., Gil, D. y Martínez, J. (1996). Evaluar no es calificar. La evaluación y la calificación en una enseñanza constructiva de las ciencias. *Revista Investigación en la Escuela*, 30, 15-26. <https://revistascientificas.us.es/index.php/IE/article/view/8060>
- Álvarez-Méndez, J. M. (2001). *Evaluar para conocer, examinar para excluir*. Morata.
- Anguera, M. T. (2023). Revisitando las revisiones sistemáticas desde la perspectiva metodológica. *RELIEVE*, 29(1), 1-25. <https://doi.org/10.30827/relieve.v29i1.27758>
- Arribas, J. M. (2017). La evaluación de los aprendizajes. Problemas y soluciones. *Profesorado, Revista De Currículum Y Formación Del Profesorado*, 21(4), 381-404. <https://doi.org/10.30827/profesorado.v21i4.10061>
- Boud, D. (2000). Sustainable assessment: rethinking assessment for the learning society. *Studies in Continuing Education*, 22(2), 151-167. <https://doi.org/10.1080/713695728>
- Boud, D. y Soler, R. (2015). Sustainable assessment revisited. *Assessment & Evaluation in Higher Education*, 41(3), 400-413. <https://doi.org/10.1080/02602938.2015.1018133>
- Cruzado, J. J. (2022). La evaluación formativa en la educación. *Comuni@cción: Revista De Investigación En Comunicación Y Desarrollo*, 13(2), 149-160. <https://doi.org/10.33595/2226-1478.13.2.672>
- Gallardo-Fuentes, F., Carter-Thuillier, B., Peña-Troncoso, S., Martínez-Angulo, C. y López-Pastor, V. (2023). Analizando críticamente la incorporación de las actuales normativas sobre evaluación, calificación y promoción escolar en la formación inicial del profesorado en educación física en Chile. *Interciencia*, 48(10), 544-551. <https://bit.ly/4b4H7DE>
- Gallego-Noche, B. (2019). *El buen hacer en educación: narrativas contrahegemónicas y prácticas inclusivas*. Editorial UCA. Universidad de Cádiz.

- Heritage, M. (2021). *Formative assessment: Making it happen in the classroom*. Corwin Press.
- Hortiguëla, D., Pérez-Pueyo, Á. y 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), 13-27. <https://doi.org/10.15366/riee2019.12.1.001>
- Llamoca, M. C., Vera, L. V. y Yarin, E. T. (2022). Evaluación formativa en los estudiantes de secundaria. *TecnoHumanismo*, 2(3), 1-21. <https://doi.org/10.53673/th.v2i9.163>
- López-Pastor, V. M., (2005). La evaluación como sinónimo de calificación. Implicaciones y efectos en la Educación y en la Formación del Profesorado. *Revista Electrónica Interuniversitaria de Formación del Profesorado*, 8(4), 1-7. <https://www.redalyc.org/articulo.oa?id=217017876007>
- López, A. A. y Osorio-Sánchez, K. (2016). Percepciones de estudiantes sobre la retroalimentación formativa en el proceso de evaluación. *Actualidades Pedagógicas*, 68, 43-64. <http://dx.doi.org/10.19052/ap.2829>
- Martínez-Rizo, F. (2013). Dificultades para implementar la evaluación formativa. Revisión de literatura. *Perfiles Educativos*, XXXV (139), 128-150. <https://www.redalyc.org/articulo.oa?id=13225611007>
- Martínez, L. (2021). *La evaluación como práctica contradictoria*. ANEP. Consejo de Formación en Educación. <https://repositorio.cfe.edu.uy/handle/123456789/1725>
- Mellado-Moreno, P.C., Sánchez-Antolín, P. y Blanco-García, M. (2021). Tendencias de la evaluación formativa y sumativa del alumnado en Web of Sciences. *Alteridad*, 16(2), 170-183. <https://doi.org/10.17163/alt.v16n2.2021.01>
- Morales-Lobo, M. M. y Fernández, J. G. (2022). *La evaluación formativa: estrategias eficaces para regular el aprendizaje*. Ediciones SM España.
- Moreno, B., Muñoz, M., Cuellar, J., Domancic, S. y Villanueva, J. (2018). Revisiones Sistemáticas: definición y nociones básicas. *Revista clínica de periodoncia, implantología y rehabilitación oral*, 11(3), 184-186. <https://dx.doi.org/10.4067/S0719-01072018000300184>
- Nicol, D. J. (2010). From monologue to dialogue: improving written feedback processes in mass higher education. *Assessment & Evaluation in Higher Education*, 35(5), 501-517. <https://bit.ly/4clz2LU>
- Nicol, D. J., y Macfarlane-Dick, D. (2006). Formative assessment and self-regulated learning: A model and seven principles of good feedback practice. *Studies in higher education*, 31(2), 199-218. <https://bit.ly/3xDaoaG>
- Ortiz, Z. (2005). Revisiones sistemáticas. *Bioquímica y Patología Clínica*, 69(2), 34-36. <https://www.redalyc.org/pdf/651/65169204.pdf>
- Polo, I. (2011). Los criterios de evaluación como detonante de la programación didáctica. *Avances En Supervisión Educativa*, 14. <https://doi.org/10.23824/ase.v0i14.482>

- Rodríguez-Negro, J. y Zulaika-Isasti, D. L. M. (2016). Evaluación en Educación Física. Análisis comparativo entre la teoría oficial y la praxis cotidiana. *Sportis. Scientific Journal of School Sport, Physical Education and Psychomotricity*, 2(3), 421-438. <https://doi.org/10.17979/sportis.2016.2.3.1448>
- Rosales, M. (2014). Proceso evaluativo: evaluación sumativa, evaluación formativa y Assesment su impacto en la educación actual. En *Congreso Iberoamericano de Ciencia, Tecnología, Innovación y Educación* (Vol. 4, p. 1-13). <https://bit.ly/3VqqKeP>
- Ruiz-Cuéllar, G. (2021). Evaluación formativa del aprendizaje. Uno de los tantos desafíos que trajo consigo la pandemia. *Revista mexicana de investigación educativa*, 26(90), 655-661. <https://bit.ly/3Vu73CG>
- Salcedo, F. E., Ramos de las Heras, A., Esperance Matamoros, M. y González Rivero, O. (2013). La elaboración de los exámenes y claves de calificación. *Didáctica Y Educación*, 4(6), 171-184. <https://revistas.ult.edu.cu/index.php/didascalía/article/view/238>
- Sanmartí, N. (2020). *Evaluar y aprender: un único proceso*. Ediciones Octaedro.
- Santos-Guerra, M. A. (2015). *La evaluación como aprendizaje: cuando la flecha impacta en la diana*. Narcea.
- Santos-Guerra, M. A. (2024). ¿Qué hacemos con el cuchillo? Las finalidades de la evaluación). *Revista Iberoamericana de Evaluación Educativa*, 17(1), 11-24. https://revistas.uam.es/riee/article/view/riee17_1_001
- Santos, L. (2016). A articulação entre a avaliação somativa e a formativa, na prática pedagógica: uma impossibilidade ou um desafio? *Ensaio: avaliação e políticas públicas em Educação*, 24, 637-669. <https://doi.org/10.1590/S0104-40362016000300006>
- Scriven, M. (1967). The methodology of evaluation. En M. Scriven (Ed.), *Perspectives of Curriculum Evaluation* (pp. 39-83). Rand McNally and Company.
- Shepard, L. (2006). La evaluación en el aula. En R.L. Brennan (Ed.), *Educational Measurement* (4ª ed., pp. 623-646). ACE/ Praeger Westport. <https://bit.ly/3VRheCS>
- Zubillaga-Olague, M. y Cañadas, L. (2021). Finalidades de los procesos de evaluación y calificación en Educación Física. *Espiral. Cuadernos del profesorado*, 14 (29), 124-135.

CONTRIBUCIONES DE AUTORES/AS, FINANCIACIÓN Y AGRADECIMIENTOS

AUTOR/ES:

Luis Miguel Romero González
Universidad de Cádiz

Doctorando en la Universidad de Cádiz, cuenta con una amplia trayectoria en el ámbito educativo. Diplomado en Magisterio de Educación Física y licenciado en Psicopedagogía, ha trabajado como maestro de Educación Primaria y orientador educativo en la ESO, desempeñándose además como Jefe del Departamento de Orientación. Con una pasión por la innovación pedagógica y las nuevas tecnologías, ha sido tutor y ponente en diversos cursos y talleres sobre competencia digital, evaluación formativa y Diseño Universal para el Aprendizaje. Ha coordinado numerosos proyectos y programas educativos, recibiendo premios y menciones por su labor innovadora. Su investigación se centra en el fracaso escolar, el abandono prematuro y la competencia digital docente.

luismiguel.romerogonzalez@alum.uca.es

Google Scholar: <https://scholar.google.es/citations?user=GaeAud4AAAAJ&hl=es>