

## **Criterios de evaluación referidos, percibidos y utilizados por los maestros en matemáticas**

### **Assessment criteria cited, perceived, and used by mathematics teachers**

### **CrITÉrios de avaliaço referidos, percebidos y utilizados por os professores de Matemática**

Maria-Joo Marques  
Direco dos Servios de Educao e Juventude  
Macao, China  
[mariajoap@sapo.pt](mailto:mariajoap@sapo.pt)  
<https://orcid.org/0000-0002-8482-9662>

Manuel Montanero  
Universidad de Extremadura  
Badajoz, Espaa  
[mmontane@unex.es](mailto:mmontane@unex.es)  
<https://orcid.org/0000-0002-2153-1180>

**Resumen: Objetivo.** El principal objetivo de este trabajo es comparar los criterios de evaluaci3n que el profesorado de Educaci3n Primaria de matemáticas considera más relevantes, con la percepci3n que tienen sus propios estudiantes, así como con la observaci3n de los criterios que realmente utilizan en la prctica. **Metodologa.** Para ello, entrevistamos a 30 maestros y 120 estudiantes. Adems, los profesores participaron en una actividad de *evaluaci3n ciega* de un examen realizado por un alumno. **Resultados.** Los resultados sugieren que muchos maestros no tienen una conciencia suficientemente precisa de los criterios de evaluaci3n que realmente aplican. **Discusi3n.** Se discuten algunas implicaciones en la evaluaci3n de este rea curricular.

**Palabras clave:** Evaluaci3n educativa; Pensamiento del profesorado; Criterios de evaluaci3n; Evaluaci3n ciega; Matemáticas; Educaci3n Primaria.

**Abstract: Objective.** The main objective of this work was to compare the

assessment criteria considered most relevant by Mathematics teachers in Elementary Education, compared with those they actually used, and the perception their students have. **Method.** To this effect, we interviewed 30 teachers and 120 students. Additionally, the teachers participated in a *blinded assessment* of one student's exam. **Results.** The results suggest that many teachers are not sufficiently aware of the assessment criteria they are in fact using. **Discussion.** We discuss the influence of assessment in the subject of mathematics.

**Keywords:** Educational assessment; Teachers' thinking; Assessment criteria; Blinded assessment; Mathematics; Elementary Education.

**Resumo: Objetivo.** O principal objetivo deste trabalho é comparar os critérios de avaliação que os professores de Matemática do Ensino Básico mais valorizam, e os que efetivamente utilizam na prática, com a percepção que têm os seus próprios alunos. **Metodologia.** Para isso, entrevistamos 30 professores e 120 alunos. Os professores participaram, ainda, numa atividade de *avaliação cega* de uma prova realizada por um aluno. **Resultados.** Os resultados sugerem que muitos professores não têm uma consciência suficientemente precisa dos critérios de avaliação que efetivamente aplicam. **Discussão.** Finalmente, são discutidas algumas implicações na avaliação desta área curricular.

**Palavras chave:** Avaliação educacional; Pensamento do professor; Critérios de avaliação; Avaliação cega; Matemática; Ensino Básico.

## Introducción

La evaluación educativa es el proceso de recogida de información que conduce a un juicio de valor sobre algún aspecto de la enseñanza o el aprendizaje de los estudiantes, permite hacer públicos los resultados de esa valoración y tomar decisiones a partir de la misma (Coll, Barberà & Onrubia, 2000; Melo & Veiga, 2013). Se basa explícita o implícitamente en determinados criterios que los profesores tienen en cuenta como referencia para interpretar aquella información. Entendemos por *criterios de evaluación*, por tanto, los indicadores y las normas de referencia,

más o menos explícitas, que consideramos para valorar la consecución de los objetivos de aprendizaje. Los criterios de evaluación y calificación no tienen sólo una dimensión *normativa*, que tome como referencia el currículo oficial para cada nivel educativo. Es muy importante la dimensión *personalizada*, que toma como referencia al propio sujeto que se evalúa, comparando su progreso con respecto a momentos anteriores de evaluación (De Ory Azcárate & Ruiz Suárez, 2011).

Los criterios de evaluación pueden tener un sentido *formativo*, cuando los docentes centran su evaluación en regular el proceso de enseñanza y adecuar su ayuda a las necesidades de los alumnos (Black & Wiliam, 1998). En un sentido *sumativo y social*, se orienta también a acreditar las competencias y aprendizajes finalmente adquiridos, que son necesarios para participar con aprovechamiento en futuros procesos de formación (Onrubia & Lago, 2008).

En el caso particular de las matemáticas, algunos estudios han mostrado la dificultad de establecer un sistema de criterios, suficientemente simple y completo, que permita evaluar la competencia matemática (Teledahl, 2015). A pesar de ello, se han publicado diversas propuestas, entre las que cabe destacar los *principios y criterios para la educación matemática* (NCTM, 2000) y el *marco de alfabetización matemática* (OCDE, 2013). De acuerdo con el trabajo del grupo de expertos de PISA en esta área curricular los criterios de evaluación deberían fundamentarse concretamente en 6 elementos competenciales: (1) la comprensión y comunicación de enunciados con información matemática; (2) la comprensión y utilización de símbolos, formalismos propios del lenguaje matemático; (3) la matematización o traslación a modelos matemáticos de situaciones problemáticas; (4) la decodificación y utilización de representaciones matemáticas; (5) la selección y utilización de estrategias para resolver problemas matemáticos; (6) el razonamiento y la argumentación matemática (Turner et al., 2015).

El progresivo consenso acerca de este tipo de propuestas no supone, sin embargo, que el profesorado las asuma realmente en la práctica del aula. La toma

de conciencia y concreción de los criterios de evaluación es un requisito esencial de la calidad de la evaluación (Black & Wilian, 2009; Pinto & Santos, 2006) que tiene una influencia directa en los procesos y resultados de aprendizaje (Barbosa & Alaíz, 1994; Pacheco, 2002; Struyven et al., 2004). No cabe duda de que la precisión con que los estudiantes interpretan las demandas del profesor condiciona sus producciones en el aula (Hadji, 2001). Sin embargo, algunos estudios han encontrado evidencias de una fuerte discrepancia entre los criterios de evaluación que los profesores declaran tener en cuenta y los que realmente utilizan, así como los que sus alumnos perciben (Nunes, 1990).

En este sentido, se echan en falta trabajos que estudien el pensamiento y la práctica real del profesorado de matemáticas acerca de la evaluación, así como la percepción que tienen los propios estudiantes sobre lo que sus profesores valoran realmente en las actividades de evaluación. Se trata de una cuestión muy relevante, pero de difícil investigación, dada la dificultad de acceder al estudio de los criterios de evaluación que los docentes utilizan en su práctica educativa.

El procedimiento más extendido para obtener información sobre las prácticas evaluativas del profesorado de matemáticas se ha basado en entrevistas o cuestionarios (Riggan & Oláh, 2011; Suurtamm et al., 2010; Veldhuis et al., 2013). Los métodos de autoinforme están, sin embargo, fuertemente amenazados por el riesgo de que los profesores no tengan una conciencia suficientemente precisa de los criterios implícitos que guían realmente algunas de sus decisiones en la práctica del aula.

Una alternativa metodológica mucho menos frecuente consiste en la observación de las decisiones que los profesores toman en situaciones reales de evaluación. Onrubia & Lago (2008), por ejemplo, pidieron a profesores de matemáticas que verbalizaran las valoraciones y decisiones que tomaban durante la evaluación de una tarea. Las técnicas de pensamiento en voz alta requieren, no obstante, un considerable entrenamiento para conseguir que los informantes

automaticen la producción de verbalizaciones y que estas traduzcan con cierta fiabilidad las impresiones que les vienen a la mente durante la actividad de evaluación.

Otra técnica observacional, que no ha sido utilizada en matemáticas, consiste en la *evaluación ciega* de una prueba con una o más tareas curriculares, previamente resueltas por un alumno (Noizet & Caverni, 1985). A diferencia del pensamiento en voz alta, al profesor no se le permite ver las respuestas del estudiante, pero sí formular al entrevistador todas las cuestiones que considerasen necesarias y se les pide que justifiquen sus calificaciones. De este modo, es posible forzar la explicitación del pensamiento del profesorado acerca de lo que considera más importante en la tarea, de modo que el investigador pueda posteriormente inferir los criterios que tiene en cuenta para evaluarla y calificarla.

Nuestra investigación se fundamentó en esta última estrategia para contrastar el pensamiento y la práctica del profesorado en la evaluación de tareas matemática en la Educación Primaria. En concreto, nos planteamos dos objetivos específicos de investigación:

1.- Documentar la diversidad de criterios que el profesorado de los primeros cursos de Educación Primaria refiere explícitamente como más relevantes para evaluar la solución de problemas aritméticos.

2.- Analizar la posible discrepancia con la percepción que manifiestan los estudiantes de esos mismos maestros, así como con los criterios que estos realmente aplican en una actividad de *evaluación ciega*.

## **Método**

### **Participantes**

El estudio fue realizado en 14 escuelas de Enseñanza Básica del distrito de Lisboa (Portugal). Se seleccionaron por conveniencia 30 maestros, con edades

comprendidas entre los 27 y los 50 años y una experiencia docente relativamente amplia (entre los 5 y los 22 años). La mitad eran profesores de segundo curso (13 mujeres y 2 hombres) y la otra mitad de cuarto (también 13 mujeres y 2 hombres).

De cada uno de los 30 grupos-clase en los que impartía docencia la muestra de profesorado se seleccionaron también 4 estudiantes: 2 con rendimiento académico normal y 2 con rendimiento bajo (“no satisfactorio”) en matemáticas, de acuerdo con la primera evaluación del curso. La muestra de estudiantes estuvo, por tanto, compuesta por un total de 120 estudiantes (60 niños y 60 niñas). Ninguno de ellos había repetido curso escolar. La media de edad era de 7,2 en 2º curso y 10,3 años en 4º. Provenían de un entorno social y familiares de nivel socioeconómico similar (medio-bajo).

## **Instrumentos y materiales**

### *Entrevistas semiestructuradas*

Para el proceso de recogida de datos fueron construidas dos entrevistas semiestructuradas, una direccionada a los maestros y otra a los estudiantes. La entrevista a los docentes tenía como objetivo identificar los criterios que consideran más relevantes para la evaluación de la competencia matemática, a partir de las siguientes cuestiones: ¿a qué le da más importancia al evaluar una prueba de matemáticas?; ¿qué criterios utiliza en la evaluación de dichas pruebas?; ¿qué criterios de evaluación considera más importantes en matemáticas?

La entrevista a los alumnos pretendía recoger información acerca de los criterios de evaluación que perciben y los que ellos mismos consideran relevantes. Fueron formuladas las siguientes cuestiones abiertas: ¿qué piensas que es preciso hacer para conseguir una buena calificación en un examen de matemáticas?; ¿qué consejos darías a un compañero para tener una buena calificación en una prueba de matemáticas?

### *Prueba de evaluación*

A partir de la revisión de un conjunto de actividades de evaluación propuestas en libros de texto de 2º y 4º curso, se elaboraron sendas pruebas de evaluación, una para 2º y otra para 4º curso. Las pruebas incluían ejercicios de lectura y escritura de números, cálculo aritmético y solución de problemas. Fueron pre-testadas en dos grupos de 2º y 4º año (con alumnos que no participaron en el estudio); fruto de lo cual se efectuaron ligeras alteraciones para equilibrar su dificultad.

### **Procedimiento de recogida de datos**

#### *Entrevista a los maestros (criterios referidos)*

La entrevista individual con los profesores se centró en conocer los criterios de evaluación referidos como más importantes en la evaluación, de acuerdo con las preguntas que enunciábamos anteriormente. Las respuestas fueron grabadas y transcritas. Veamos un ejemplo de las respuestas de uno de los maestros a dichas preguntas:

Entrevistador: ¿Qué valora más en la evaluación de una prueba de matemáticas?

Profesor 1: A lo que más valor le doy es a la solución de problemas.

Entrevistador: ¿Qué criterios utiliza en la evaluación de una prueba de matemáticas?

Profesor 1: Pues... considero el razonamiento, la resolución de operaciones, si están bien escritos los números (las centenas, decenas, unidades...)

Entrevistador: Pero de todos ellos, cuál es el criterio de evaluación que considera más importante

Profesor 1: El razonamiento.

### *Entrevista a los alumnos (criterios percibidos)*

En la entrevista individual a cada uno de los estudiantes de la muestra el entrevistador, después de un breve período de familiarización, planteaba preguntas similares sobre la percepción de los criterios más importantes en la evaluación en matemáticas, como en el siguiente fragmento con un estudiante de 4º con rendimiento medio-alto:

Entrevistador: ¿Qué piensas que es necesario hacer para tener una buena nota en un examen de matemáticas?

Alumno 2: Tienes que mirar con mucha atención los ejercicios, las frases que están escrita encima para saber qué es lo que tenemos que hacer en ese ejercicio; si no, lo tienes luego todo mal.

Entrevistador: ¿Qué consejos darías a otro niño para conseguir una buena nota en una prueba de matemáticas?

Alumno 2: Prestar mucha atención a las cuentas, intentar no tener fallos, hacer bien los números, hacer los ejercicios con cuidado, mirar bien lo que hay escrito...

Las respuestas fueron también grabadas e íntegramente transcritas para su posterior análisis.

### *Evaluación ciega de una prueba de matemáticas (criterios utilizados)*

Para poder observar en la práctica los criterios de evaluación de tareas matemáticas que realmente utilizaban los maestros diseñamos una actividad de *evaluación ciega*, basada en el *método de descubrimiento de criterios*, propuesto originalmente por [Noizet y Caverni \(1985\)](#), y que ha sido utilizada en actividades de evaluación de la lectoescritura (véase [Alves Martins, Matta, Mata & Nunes, 1991](#); [Nunes, 1990](#)).

El entrevistador se desplazó de nuevo al centro escolar para solicitar a cada

uno de los profesores que había entrevistado que participase en un “juego” que consistía en calificar una prueba con varias tareas matemáticas, que había sido resuelta por un alumno, sin poder verla. Los maestros podían formular al entrevistador todas las cuestiones que considerasen necesarias (tanto sobre los enunciados de las tareas como sobre las respuestas registradas por el estudiante), sin límite de tiempo, pero sin poder acceder a la visualización directa de la prueba hasta el final. El entrevistador respondía de forma concreta y escueta cuando la pregunta era clara. En cambio, solicitaba una reformulación de la pregunta cuando esta era ambigua o imprecisa. Finalizadas las preguntas, el profesor calificaba la ficha y, sólo después, se le permitía verla. Finalmente se le preguntaba: (a) si mantenía la calificación otorgada antes de verla y (b) qué criterios había tenido en cuenta para dicha calificación final. Toda la interacción verbal fue grabada y transcrita, como en el siguiente ejemplo, uno de los más breves, correspondiente al profesor 29 de 4º curso:

[...]

Profesor: ¿Hay problemas en los que haya que escribir números decimales o no?

Entrevistador: Sí hay.

Profesor: ¿Las operaciones están incluidas en las situaciones problemáticas?

Entrevistador: Sí, pero también fuera de ellas

Profesor: ¿Cómo las ha resuelto? ¿Están todas correctas?

Entrevistador: Hay dos partes: una está correcta y la otra no.

Profesor: ¿Cuántas cuestiones tiene la primera parte?

Entrevistador: 6 cuestiones

Profesor: ¿Y problemas?

Entrevistador: 3.

Profesor: ¿Las operaciones son con número decimales o no?

Entrevistador: Hay una multiplicación con números decimales y una división con números enteros.

Profesor: ¿Las ha hecho bien?

Entrevistador: Una está bien y la otra no.

Profesor: ¿Y las 5 cuestiones cómo están?

Entrevistador: La escritura de números está bien.

Profesor: ¿Y los problemas están todos correctos?

Entrevistador: Sí, bueno, dos de ellos. En el tercero el resultado no está

correcto.

Profesor: No necesito saber más. Le pondría un “Muy bien”.

Entrevistador: Entonces ahora vas a ver la prueba. Después me dices si mantienes la calificación o la cambias y por qué. [El profesor dedica un par de minutos a leer la hoja de la prueba]

Profesor: Le pondría un “Bien”, pero no “Muy bien”, por que debía haber hecho mejor las cuentas, hacer una lectura de dos números correcta y saber que este número decimal era mayor.

## **Procedimiento de análisis de datos**

### *Análisis del contenido de las transcripciones*

A partir de las transcripciones de las respuestas registradas durante las entrevistas, así como durante la actividad de evaluación ciega, se realizó un análisis de contenido, para categorizar los criterios referidos (entrevista a los maestros), percibidos (entrevista a los alumnos) y utilizados (evaluación ciega).

En primer lugar, las transcripciones de las respuestas fueron segmentadas semánticamente ideas. En segundo lugar, se seleccionaron las ideas que hacían directamente referencia a criterios de evaluación, es decir, a indicadores o estándares implícitos que se tomaron como referencia para evaluar o calificar la solución de problemas matemáticos. En tercer lugar, se categorizaron dichas ideas. El sistema de categorías emergente se revisó en un proceso inductivo-deductivo de *comparación permanente* en el que, tras un primer ciclo de categorización, se examinaron las relaciones de semejanza entre algunas categorías, con objeto de modificar su definición, reagruparlas, diversificarlas y jerarquizarlas, hasta que se alcanzó una clasificación exhaustiva de categorías mutuamente excluyentes. Finalmente se estimó la fiabilidad de la categorización, calculando el grado de concordancia entre dos codificadores (división del número de acuerdos entre codificadores por el total de categorizaciones efectuadas). Se obtuvo un 81% de acuerdos, resolviéndose las discrepancias por consenso.

## Medidas

Para cuantificar la importancia concedida a cada uno de los criterios de evaluación (referidos por los maestros o percibidos por los alumnos) se calculó la proporción de profesorado y alumnado que los mencionaron al menos una vez, es decir, el porcentaje de entrevistas en el que encontramos al menos un registro verbal de una idea clasificada en una categoría criterial determinada.

Igualmente, para cuantificar la utilización de cada uno de los criterios utilizados durante la evaluación ciega se calculó la proporción de profesores que verbalizó al menos un comentario o pregunta que fuera clasificada en una de las categorías.

Para medir la discrepancia entre las categorías criterioales que encontramos en las entrevistas (criterios referidos por los maestros o percibidos por los alumnos) y los realmente utilizados en la actividad de evaluación ciega empleamos el *Índice de Jaccard* (Jaccard, 1901). La adaptación de esta medida al ámbito de la evaluación educativa (Alves Martins *et al.*, 1991; Nunes, 1990) consiste en una estimación cuantitativa que mide la proximidad entre las distribuciones de los criterios referidos y utilizados por cada uno de los profesores, mediante el siguiente cálculo:  $a/(a+b+c)$ ; donde “a” es igual al número de criterios referidos y utilizados por cada profesor, “b” es el número de criterios no referidos y utilizados, y “c” es el número de criterios referidos y no utilizados.

En el caso de los alumnos la fórmula es la siguiente:  $a/(a+b)$ . En este caso, “a” representa el número de criterios utilizados por los profesores y percibidos por el alumno; “b” es el número de criterios utilizados por los profesores y no percibidos por los alumnos.

En ambos, casos, cuanto mayor es el valor resultante, cabe inferir una mayor convergencia entre los criterios referidos y los utilizados por los maestros o

percibidos por los alumnos.

## Resultados

El análisis de las transcripciones de las entrevistas, así como de las verbalizaciones grabadas durante la tarea de evaluación ciega, generó un total de 21 categorías de criterios de evaluación, que posteriormente se agruparon en 6 dimensiones. La Tabla 1 recoge el sistema de categorías que emergió al final del procedimiento inductivo de análisis de contenido, junto con algunos ejemplos de registros verbales.

**Tabla 1:** Sistema de categorías construido a partir de los registros verbales grabados en las entrevistas a los maestros (criterios referidos), a los alumnos (criterios percibidos), así como durante la actividad de evaluación ciega (criterios utilizados).

Dimensiones	Categorías	Definición	Ejemplos de registros verbales
A. Aspectos procedimentales	A1 Operaciones aritméticas	Realización de algoritmos aritméticos de suma, resta, multiplicación y división	Referido por el prof. 16: "Es fundamental hacer bien las operaciones: el cálculo es muy importante".
	A2 Tablas	Memorización de hechos numéricos y tablas (como las tablas de multiplicar)	Percibido por el alum. 65: "Lo más importante es saberse bien las tablas".
	A3 Cálculo mental	Resolución "mental" de operaciones aritméticas (sin papel ni ningún otro apoyo)	Referido por el prof. 2: «El cálculo mental es también muy importante y también lo valoro mucho".
	A4 Escritura de números	Escritura correcta de la representación numérica de cantidades	Utilizado por el prof. 3: "Ha escrito bien los números en los huecos?"
	A5 Reducciones	Transformación de unidades de medida en múltiplos (más grandes) o submúltiplos (más pequeñas)	Percibido por el alum. 83: "Tienes que tener mucho cuidado con pasar bien de metros a centímetros".
	A6 Geometría	Designación de propiedades sobre la posición o la forma de objetos en el espacio	Utilizado por el prof 3: "¿Ha dibujado figuras geométricas?"
	A7 Automatización de técnicas	Apropiación y dominio de algoritmos y técnicas "mecánicas"	Percibido por el alum.120: "Es muy importante saber colocar los números, saber dividir y saber qué números hay que poner en cuenta".
B. Aspectos cognitivos	B1 Razonamiento matemático	Habilidad para relacionar, argumentar y extraer conclusiones	Utilizado por el prof. 14: "Me interesa saber si acertó en el razonamiento aunque errara las cuentas".
	B2	Comprensión del enunciado del	Referido por el prof. 7: "Es

	Comprensión del problema	ejercicio (qué datos se dan y qué se pide)	importante la comprensión del problema”.
	B3 Creatividad	Razonamiento sobre alternativas o estrategias innovadoras para resolver un problema	Referido por el prof. 16: “Yo valoro que se utilicen formas alternativas de resolver un problema”.
	B4 Generalización	Transferencia de una estrategia a otro caso o tarea similar	Referido por el prof. 19: “Hacemos siempre cosas semejantes para ver si consiguen resolver de otra manera otros problemas”.
	B5 Selección de operación	Planificación o selección de los pasos u operaciones a realizar para resolver un problema	Referido por el prof. 21: “Es importante si consiguen saber cuándo tienen que hacer una resta, por ejemplo”.
	B6 Capacidad de análisis	Descomposición de la tarea o problema en partes	Referido por el prof. 7: “Es importante que les ayudemos a pensar más cosas, a analizar el problema”.
	B7 Capacidad de atención	Focalización o distribución de la atención sobre elementos relevantes de la tarea	Percibido por el alum. 62: “Tenemos que estar atentos, con mucha atención”.
C. Aspectos formales	C1 Presentación	Limpieza y presentación formal del ejercicio, conforme a las normas establecidas	Referido por el prof. 2 : “La presentación también tiene importancia. No me gusta que un niño entregue el examen todo sucio, todo emborronado...”
	C2 Notación-Indicación	Otras cuestiones relacionadas con la escritura o las normas convencionales de representación simbólica de algoritmos	Percibido por el alum. 1: “Tenemos que hacer bien los números”.
D. Resultados	D1 Resultados finales	Corrección del resultado obtenido	Utilizado por el prof. 1: En las otras preguntas, ¿cuántos resultados tiene mal?”
E. Auto-regulación	E1 Tiempo de resolución	Ajuste al tiempo disponible	Percibido por el alum. 4: “Hay que hacerlo todo rápido”.
	E2 Autonomía	Resolución independiente, sin ayuda	Referido por el prof. 13: “La autonomía de ellos: cuando consiguen hacer la ficha solos, tiene más valor”.
	E3 Esfuerzo-atención	Persistencia y mantenimiento de la atención	Referido por el prof. 18: “Tengo que ver [...] la atención del alumno”.
F. Otros	F 1 Otros	Otros criterios ambiguos o inapropiados	Referido por el prof. 16: “Un alumno puede llegar a una conclusión y yo entender lo que va a hacer, pero no ser una situación...”

**Nota:** Elaboración propia

## Criterios de evaluación referidos por los maestros en la entrevista

El *razonamiento* matemático fue el único criterio de evaluación referido en las entrevistas por la casi totalidad del profesorado. Por el ejemplo, el prof. 17 declara lo siguiente: “Pueden equivocarse en una cuenta, que eso para mí no tiene mucha importancia, si muestran un razonamiento correcto”. Algunos profesores destacan, además, el valor esencialmente *formativo*, más que sumativo, de este criterio, como se aprecia en el siguiente fragmento de la entrevista a este mismo profesor:

Prof. 17: Es importante entender el razonamiento del niño, por que muchas veces ellos van por un camino que no es el correcto, pero es el que ellos pensaban. Por eso hay siempre una cierta lógica en el pensamiento de los niños, que muchas veces es diferente del nuestro o que no siempre es el más correcto. Pero es fundamental entender cuál fue esa lógica que siguieron, para poder explicarles y cambiarles algunas cosas.

La mayoría de los maestros aludieron también a la *comprensión* de los problemas y al *cálculo mental* como criterios relevantes. Así, el profesor 7 expresa: “Es importante la comprensión. El objetivo principal de las matemáticas es siempre la resolución de situaciones problemáticas. Si el niño comprende, consigue llegar a un resultados favorable”. Por su parte, el profesor 6 afirma: “Para la calificación de las cuestiones tengo en cuenta el cálculo mental: pienso que el cálculo mental es la base”.

Por último, la mitad de los profesores reconocen prestar mucha importancia también al *resultado final* de la tarea:

Prof. 5: Hago el sumatorio de las preguntas que fallan y al final pongo un Excelente, Muy bien, Bien o Insuficiente. Califico cada pregunta. Pongo Bien a un alumno que tiene todos los resultados correctos, pero no le pongo una nota superior si tuve que corregir algunos.

Algunos profesores matizan, sin embargo, que dicho criterio se pondera en función del razonamiento mostrado por el estudiante, como se aprecia en el siguiente fragmento:

Prof. 21: Califico cada pregunta dependiendo del alumno. Si fuera Juan, por ejemplo, si el razonamiento es acertado, a pesar de que la cuenta esté mal, lo valoro, por que lo entendió. Pero a Bruno ya no le doy la misma facilidad si el resultado estuviera completamente equivocado o no sé de dónde viene. El valor que le daría sería muy inferior al de Juan.

En torno a una cuarta parte del profesorado menciona también otros criterios procedimentales, relacionados con la corrección y automatización de las operaciones aritméticas o la escritura de números. El resto de criterios de evaluación son referidos sólo por una pequeña minoría de maestros.

### **Criterios de evaluación percibidos por los alumnos en la entrevista**

Los criterios de evaluación a los que aluden la mayoría de los estudiantes en las entrevistas son la corrección de las *operaciones* aritméticas y de los *resultados finales*, como se aprecia en el siguiente fragmento:

Alum. 17: Tienes que hacer las cuentas todas bien y no tener ningún error para sacar todo el examen. Tenemos que tener mucho cuidado con las cuentas por que pueden ser un poquito difíciles. Algunas veces escribo las cuentas en el papel y otras veces hago las cuentas con los dedos.[...]  
“La gente tiene que sabe las cuentas y sacar los resultados siempre..., para tenerlo todo bien”.

No obstante, un 41% y un 35% del alumnado señaló más o menos explícitamente cuestiones relacionadas con el *razonamiento matemático* o la *comprensión* de los problemas, respectivamente. El alumno 67, por ejemplo, alude a los riesgos de resolver mecánicamente los problemas, sin analizar antes lo que realmente se pide: “Es necesario pensar, leer bien los problemas, porque a veces

no son como parecen”.

Bastantes estudiantes aluden también a la importancia de la concentración en la tarea (44%), la escritura correcta de números enteros y decimales (34%) y la memorización de la tabla de multiplicar (24%). El resto de los criterios son mencionados por menos de una quinta parte de los estudiantes.

### **Criterios de evaluación utilizados en la actividad de evaluación ciega**

El *razonamiento* es también el criterio más utilizado por la práctica totalidad de los profesores en la situación de evaluación ciega, como se aprecia en el siguiente comentario que uno de los docentes verbaliza al final de la actividad de evaluación ciega: “Yo le pondría más nota, por que ella ha razonado bien la operación de cálculo. Quizá se equivocó en la cuenta, por no saberse bien la tabla [de multiplicar]” (Prof. 16).

No obstante, casi todo el profesorado pregunta también por la corrección del resultado final. Además, este criterio que tiene mucha influencia en la calificación que se otorga a la prueba, como se aprecia en el comentario final del profesor 19, refiriéndose a errores en los resultados numéricos de algunas tareas: “Yo le pondría una nota baja; la prueba es muy sencilla; hay cosas que no se perdonan”. En este mismo sentido, un 70% de los profesores solicita también información acerca de las operaciones aritméticas (Prof. 16: “¿Y las operaciones las hizo correctamente?”); y un 80% pregunta explícitamente cómo están escritos los números (Prof. 1: “¿Hay ejercicios sobre escritura de números? ¿Los ha hecho bien o no?”).

Otros criterios que son utilizados con cierta frecuencia tienen que ver con el cálculo mental (26%) o el conocimiento de la tabla de multiplicar (36%). El resto de criterios tiene una presencia muy minoritaria.

## **Comparación global entre los criterios de evaluación referidos, percibidos y utilizados**

En la Tabla 2 se compara la frecuencia de ocurrencia de cada una de las anteriores categorías de registros verbales, junto con el porcentaje de sujetos que los verbalizó al menos una vez.

Tabla 2. Frecuencia y proporción de maestros y estudiantes que refirieron/percibieron cada criterio de evaluación en las entrevistas, así como de los maestros que los utilizaron en la actividad de evaluación ciega.

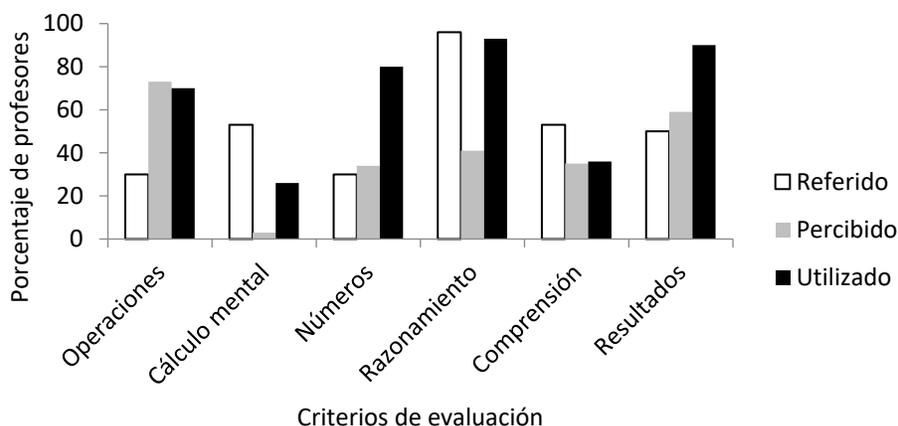
Dimensiones	Categorías (criterios)	Referidos			Percibidos			Utilizados		
		2º	4º	Total	2º	4º	Total	2º	4º	Total
A. Aspectos procedimentales	A1 Operaciones aritméticas	6 (0,4)	3 (0,2)	9 (0,3)	36 (0,6)	52 (0,86)	88 (0,73)	10 (0,66)	11 (0,73)	21 (0,7)
	A2 Tablas	2 (0,13)	0	2 (0,06)	7 (0,11)	22 (0,36)	29 (0,24)	5 (0,33)	6 (0,4)	11 (0,36)
	A3 Cálculo mental	7 (0,46)	9 (0,6)	16 (0,53)	1 (0,01)	3 (0,05)	4 (0,03)	4 (0,26)	4 (0,26)	8 (0,26)
	A4 Escritura de números	6 (0,4)	3 (0,2)	9 (0,3)	23 (0,38)	18 (0,3)	41 (0,34)	13 (0,86)	11 (0,73)	24 (0,8)
	A5 Reducciones	0	1 (0,06)	1 (0,03)	0	10 (0,16)	10 (0,08)	1 (0,06)	8 (0,53)	9 (0,3)
	A6 Geometría	1 (0,06)	0	1 (0,03)	0	2 (0,03)	2 (0,01)	3 (0,2)	4 (0,26)	7 (0,23)
	A7 Automatización de técnicas	4 (0,26)	3 (0,2)	7 (0,23)	2 (0,03)	14 (0,23)	16 (0,13)	11 (0,73)	4 (0,26)	15 (0,5)
B. Aspectos cognitivos	B1 Razonamiento matemático	15 (1)	14 (0,93)	29 (0,96)	15 (0,25)	35 (0,58)	50 (0,41)	14 (0,93)	14 (0,93)	28 (0,93)
	B2 Comprensión del problema	6 (0,4)	10 (0,66)	16 (0,53)	14 (0,23)	29 (0,48)	43 (0,35)	6 (0,4)	5 (0,33)	11 (0,36)
	B3 Creatividad	2 (0,13)	4 (0,26)	6 (0,2)	1 (0,01)	3 (0,05)	4 (0,03)	0	1 (0,06)	1 (0,03)
	B4 Generalización	2 (0,13)	1 (0,06)	3 (0,1)	0	0	0	1 (0,06)	1 (0,06)	2 (0,06)
	B5 Selección de la operación	2 (0,13)	2 (0,13)	4 (0,13)	3 (0,05)	7 (0,11%)	10 (0,08)	2 (0,13)	1 (0,06)	3 (0,1)
	B6 Capacidad de análisis	1 (0,06)	1 (0,06)	2 (0,06)	0	2 (0,03)	2 (0,01)	0	0	0
	B7 Capacidad de atención	0	1 (0,06)	1 (0,03)	10 (0,16)	41 (0,68)	51 (0,42)	1 (0,06)	1 (0,06)	2 (0,06)
C. Aspectos formales	C1 Presentación	0	1 (0,06)	1 (0,03)	8 (0,13)	10 (0,16)	18 (0,15)	2 (0,13)	0	2 (0,06)
	C2 Notación/Indicación	0	1 (0,06)	1 (0,03)	0	1 (0,01)	1 (0,008)	1 (0,06)	1 (0,06)	2 (0,06)
D. Resultados	D1 Resultados finales	7 (0,46)	8 (0,53)	15 (0,5)	29 (0,48)	42 (0,7)	71 (0,59)	12 (0,8)	15 (1)	27 (0,9)
E. Auto-regulación	E1 Tiempo de resolución	0	0	0	3 (0,05)	6 (0,1)	9 (0,07)	0	0	0
	E2 Autonomía	3 (0,2)	1 (0,06)	4 (0,13)	0	2 (0,03)	2 (0,01)	1 (0,06)	0	1 (0,03)
	E3 Esfuerzo/atención	0	1 (0,06)	1 (0,03)	17 (0,28)	36 (0,6)	53 (0,44)	0	0	0
F. Otros	F1 Otros	1 (0,06)	1 (0,06)	2 (0,06)	20 (0,33)	19 (0,31)	39 (0,32)	4 (0,26)	4 (0,26)	8 (0,26)

Nota: Elaboración propia

Como puede apreciarse en el siguiente gráfico, el razonamiento matemático es el criterio más frecuentemente mencionado y utilizado. Sin embargo, se observa una clara discrepancia (de entre 40 y 50 puntos) entre el porcentaje de profesores que refiere al menos una vez en las entrevistas los criterios relativos a la corrección de las operaciones, la escritura de números o el resultado final, y el porcentaje de profesores que realmente los tienen en cuenta en la evaluación ciega.

El porcentaje de los alumnos que perciben la corrección de las operaciones como uno de los criterios más relevantes, en cambio, se aproxima mucho al porcentaje de utilización en la evaluación ciega (Figura 1). No ocurre lo mismo en el resto de los criterios percibidos como importantes por una parte significativa de los estudiantes, cuyos porcentajes difieren bastante de los criterios utilizados.

**Figura 1.** Porcentaje de profesores que refirieron y utilizaron los principales criterios de evaluación, así como de los estudiantes que declararon percibirlos.



**Nota:** Elaboración propia

El índice de Jaccard, que mide la convergencia entre los criterios referidos por cada profesor y los que utilizó posteriormente en la actividad de evaluación ciega varía entre 0 y 0,75 ( $X=0,38$ ;  $DT=0,18$ ). Sólo 9 de los 30 profesores obtuvieron valores superiores a 0,50, lo que sugiere una cierta discrepancia entre los criterios de evaluación referidos y utilizados en la mayoría de la muestra de profesores.

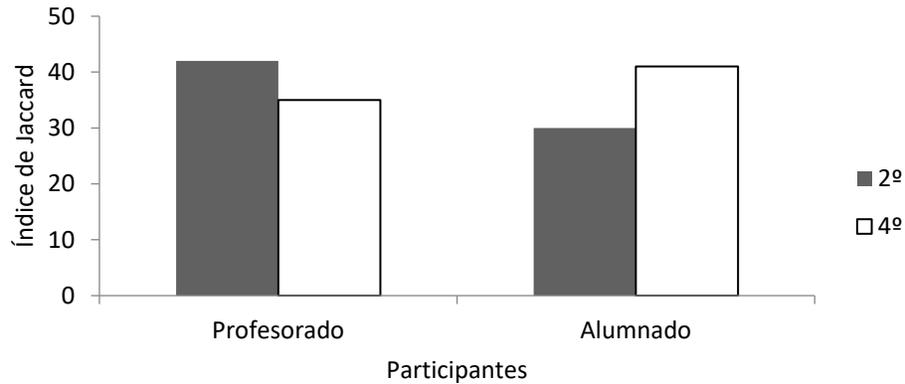
El índice que cuantifica la convergencia entre los criterios percibidos por los alumnos y los utilizados por los profesores en la evaluación ciega varía entre un mínimo de 0 y un máximo de 0,8. El promedio fue, en este caso, 0,35 y la desviación típica 0,15; lo que refleja también una clara discrepancia.

### **Diferencias en función del curso académico**

En cuanto al curso académico en que impartía docencia el profesorado, no se encontraron diferencias significativas en cuanto a los criterios referidos en las entrevistas por los profesores de 2º y de 4º curso. Tampoco se encontraron diferencias significativas en cuanto a los criterios utilizados en la evaluación ciega, salvo en las categorías *Reducciones* ( $X^2=7,77$ ;  $p<0,01$ ), más frecuentemente utilizado en 4º curso; y *Automatización técnicas* ( $X^2=6,53$ ;  $p=0,01$ ), más utilizado en 2º. En cuanto al grado de convergencia no hay una diferencia significativa entre el promedio de los índices de Jaccard obtenidos en la muestra de profesorado de 2º ( $X=0,42$ ;  $DT=0,21$ ) y la de 4º ( $X=0,35$ ;  $DT=0,17$ ).

En cambio, el porcentaje de alumnos de 2º curso que aludió a cada criterio fue mucho menor que el porcentaje de alumnado de 4º (salvo en escritura de números, donde no hay una diferencia significativa). También resultó más bajo el promedio de los resultados del índice de Jaccard ( $X=0,30$ ;  $DT=0,17$ ), en comparación con el de 4º ( $X=0,41$ ;  $DT=0,24$ ). Dicho de otro modo, la convergencia de la percepción de los alumnos de 2º con respecto a los criterios realmente utilizados por sus profesores es menor que la del alumnado de 4º (Figura 2). La prueba t de Student confirmó que se trata de una diferencia significativa ( $t=2,87$ ;  $p=0,005$ ).

**Figura 2:** Promedios de los índices de convergencia de criterios (índice Jaccard) obtenidos en la muestra de profesorado y alumnado de 2º y 4º curso.



**Nota:** Elaboración propia

### **Diferencias de percepción del alumnado en función del rendimiento académico**

En cuanto al rendimiento académico, por último, la Tabla 3 muestra una coincidencia en la percepción del alumnado con medio y bajo rendimiento en considerar las operaciones aritméticas y los resultados finales como los criterios de evaluación más importantes para conseguir una buena calificación en matemáticas.

**Tabla 3.** Frecuencia y porcentaje de aparición de las categorías de criterios de evaluación en la entrevista a los alumnos (criterios percibidos) en función del rendimiento académico (normal o bajo).

<i>Dimensiones</i>	<i>Categorías (criterios)</i>	<i>Normal</i>	<i>Bajo</i>
A. Aspectos procedimentales	A1 Operaciones	47 (0,78)	41 (0,68)
	A2 Tablas	15 (0,25)	14 (0,23)
	A3 Cálculo mental	2 (0,03)	2 (0,03)
	A4 Escritura de números	24 (0,4)	17 (0,28)
	A5 Reducciones	7 (0,11)	3 (0,05)
	A6 Geometría	1 (0,01)	2 (0,03)
	A7 Automatización de técnicas	8 (0,13)	8 (0,13)
B. Aspectos cognitivos	B1 Razonamiento	29 (0,48)	21 (0,35)
	B2 Comprensión	21 (0,35)	22 (0,36)
	B3 Creatividad	3 (0,05)	2 (0,03)
	B4 Generalización	0	0
	B5 Selección da operación	6 (0,1)	4 (0,06)
	B6 Capacidad de análisis	1 (0,01)	1 (0,01)
	B7 Capacidad de atención	28 (0,46)	23 (0,38)
C. Aspectos formales	C1 Presentación	8 (0,13)	10 (0,16)
	C2 Notación/Indicación	1 (0,01)	0
D. Resultados	D1 Resultados finales	31 (0,51)	40 (0,66)
E. Auto-regulación	E1 Tiempo de resolución	3 (0,05)	6 (0,01)
	E2 Autonomía	1 (0,01)	1 (0,01)
	E3 Esfuerzo	24 (0,4)	29 (0,48)
F. Otros	F 1 Otros	17 (0,28)	22 (0,36)

**Nota:** Elaboración propia

Las diferencias entre el porcentaje de estudiantes con rendimiento normal vs. bajo que mencionaron (al menos una vez) cada criterio de evaluación no resultó estadísticamente significativa.

## Discusión

Los resultados anteriores nos aportan información útil acerca de la diversidad del pensamiento del profesorado de Educación Primaria sobre los criterios de evaluación de tareas matemáticas. Las categorías criterioales extraídas en la Tabla 1 reflejan claramente 4 de los 6 elementos competenciales propuestos por la OCDE,

que desglosábamos en la introducción de este trabajo.

Apenas se detectan diferencias significativas entre los maestros de 2º y 4º curso de Educación Primaria. Cabe destacar que casi todo el profesorado otorga mucha importancia, no sólo a la corrección de las operaciones aritméticas y al resultado final, sino también a criterios de índole cognitiva-procesual, relacionados principalmente con el *razonamiento matemático*, es decir, con la realización de inferencias lógicas basadas en conectar elementos de problemas para justificar las operaciones o conclusiones matemáticas (Turner et al., 2015). Este criterio de evaluación fue, de hecho, el único que los profesores refirieron explícitamente en las entrevistas y utilizaron realmente en la actividad de evaluación. Sin embargo, de acuerdo con las medidas obtenidas en el índice de Jaccard, casi un tercio de los maestros muestran una significativa discrepancia entre el resto de criterios que refirieron en las entrevistas y lo que realmente utilizaron en la actividad de evaluación ciega. En concreto, más de un 40% de los participantes que no mencionaron la escritura de números, los algoritmos aritméticos y la corrección del resultado final de los problemas como criterios de evaluación relevantes, los utilizaron posteriormente en la actividad de evaluación ciega (Figura 1).

En cuanto al alumnado, no hemos encontrado una percepción significativamente diferente en función del rendimiento académico. En cambio, en función del curso, los alumnos más inmaduros (de 2º) mencionan en las entrevistas un número significativamente menor de criterios que los de 4º; lo que sugiere una escasa conciencia de la diversidad de elementos que se pueden tener en cuenta en la evaluación. Esto explicaría por qué, de acuerdo con el índice de Jaccard obtenido como promedio en la muestra de 2º, la discrepancia entre los criterios percibidos y utilizados en este nivel educativo es mayor que en 4º curso (Figura 2).

En conjunto, a diferencia de sus profesores, sólo una minoría de los estudiantes coinciden en percibir que el razonamiento matemático tiene un peso significativo en las calificaciones de sus ejercicios. Para la gran mayoría lo más

importante es la corrección de las operaciones y algoritmos aritméticos. Casi la mitad de los estudiantes entrevistados considera también que la concentración o el esfuerzo tienen un peso muy relevante en las calificaciones, y un 15% le otorga mucha importancia a la presentación formal de los ejercicios.

## **Conclusiones**

En definitiva, podemos extraer dos conclusiones principales de estos resultados. Por un lado, los datos obtenidos en las entrevistas y en la actividad de evaluación ciega ponen de manifiesto que la mayoría de los profesores que participaron en el estudio consideran importante evaluar las tareas matemáticas, no sólo en función de la corrección de las operaciones aritméticas y los resultados de los problemas, sino también según criterios procesuales relacionados con la comprensión de los enunciados y el razonamiento matemático. Esta conclusión contrasta con lo encontrado por [Rosales et al. \(2012\)](#) en actividades de evaluación de problemas aritméticos en Educación Primaria. El análisis de la interacción verbal en dichas actividades mostró que la mayoría de los profesores tendían a afrontar este tipo de tarea de un modo “paradigmático” y “superficial”, es decir, centrándose más en la resolución mecánica de los algoritmos que en la comprensión situacional de los problemas y en los procesos de autorregulación. Nuestra investigación evidencia la preocupación del profesorado por evaluar la comprensión y la toma de decisiones sobre las operaciones para resolver un problema matemático, aunque no tanto por evaluar y apoyar los procesos de autorregulación. La información más útil para potenciar la autonomía de los estudiantes es aquella, de carácter cualitativo, que facilita que los estudiantes tomen consciencia de lo que necesitan afianzar o corregir: lo que hacen bien y lo que puede mejorar. En este sentido, la consideración de criterios relacionados con valorar la capacidad de los alumnos de co-evaluar y autorregular sus propias tareas es fundamental para conseguir una auténtica evaluación formativa ([Clark, 2012](#)). Aunque la actividad de evaluación ciega no facilita que los profesores utilicen este tipo de criterios, tampoco fueron mencionados por los profesores o los alumnos en las entrevistas.

Por otro lado, la discrepancia encontrada entre los criterios referidos por los maestros en las entrevistas, los percibidos por sus alumnos y los realmente utilizados en la actividad de evaluación ciega sugiere una insuficiente conciencia de buena parte del profesorado acerca de los criterios que realmente tienen en cuenta en sus actividades de evaluación, así como la necesidad de clarificarlos mejor a los alumnos. La explicitación y la comprensión de los criterios por parte de los estudiantes tiene una importante influencia en el impacto que la evaluación tiene sobre el aprendizaje (Amigues & Guignard-Andreucci, 1981; Barbosa & Alaíz, 1994; Bonniol, 1986). No se trata simplemente de comunicar con precisión qué y cómo se va a evaluar a los estudiantes. Más allá de la evaluación *formativa* se ha reclamado la importancia de promover una evaluación *formadora* (Sanmarti, 2007), haciendo al estudiante participe de la concreción y aplicación de los criterios de evaluación; no únicamente como objeto de la evaluación, sino también como sujeto activo del proceso de evaluación desde su inicio.

### **Limitaciones y prospectiva**

Cabe señalar, sin embargo, algunas cuestiones, relativas a la naturaleza de las tareas y el procedimiento de evaluación empleado en la presente investigación, que obligan a considerar con cautela estas conclusiones. Como en muchas otras investigaciones similares, los procedimientos de entrevista y autoinforme del profesorado están fuertemente amenazados por un sesgo de *deseabilidad social*, que podría haber influido en la preponderancia de los criterios relacionados con la comprensión y el razonamiento matemático.

No cabe duda, además, de la artificialidad de la actividad de *evaluación ciega*. El hecho de que al docente no se le permitiera elaborar la prueba de evaluación, y que el estudiante que la realizó no estuviera presente, ha podido influir en desatención de ciertos criterios de evaluación *formativa*, principalmente los relativos a la autorregulación. La actividad que se recrea en esta situación experimental se parece a los exámenes de orientación *sumativa*, que muchos profesores emplean

en combinación con otro tipo de estrategias de enfoque formativo; razón por la que sería arriesgado concluir que los profesores no tienen en cuenta otros tipos de criterios de evaluación en el conjunto del proceso de enseñanza y aprendizaje.

Una última limitación tiene que ver con el sistema de categorías empleado para clasificar los criterios de evaluación. Algunos criterios, como la escritura de números o las tablas de multiplicar, están claramente condicionados por los contenidos curriculares propios de estos niveles educativos. El *razonamiento matemático*, por su parte, es un criterio difícil de diferenciar de otros conceptos adyacentes. Los estudiantes, e incluso los propios profesores, tienen dificultades para conceptualizar con precisión este tipo de criterios; lo que podría explicar la discrepancia de su percepción con respecto al profesorado. En futuras investigaciones convendría por tanto revisar y operativizar mejor este tipo de categorías, con muestras de profesorado y alumnado de otros niveles educativos de la Educación Primaria.

## Referencias

- Alves Martins, M., Matta, I., Mata, L., & Nunes, C. (1991). Avaliação em língua materna e matemática do 1º ciclo do ensino básico: critérios referidos e utilizados por professores e percebidos por alunos com sucesso e insucesso escolar. *Análise Psicológica*, 3-4 (9), 453-466. [http://repositorio.ispa.pt/bitstream/10400.12/5213/1/AP-1991\\_34\\_453.pdf](http://repositorio.ispa.pt/bitstream/10400.12/5213/1/AP-1991_34_453.pdf)
- Amigues R., & Guignard-Andreucci, C. (1981). A propos d'une recherche sur l'évaluation formative en situation éducative: prise en compte et modifications des données de la situation. *Bulletin de Psychologie*, 353, 167-172.
- Barbosa, J., & Alaíz, V. (1994). Explicitação de critérios - exigência fundamental de uma avaliação ao serviço da aprendizagem. *Pensar a avaliação, melhorar a aprendizagem*. Instituto de Inovação Educacional.
- Black, P., & Wiliam, D. (1998). Assessment and classroom learning. *Assessment in Education: Principles, Policy & Practice*, 5(1), 7-73.

<https://www.gla.ac.uk/t4/learningandteaching/files/PGCTHE/BlackandWilliam1998.pdf>

- Black, P., & William, D. (2009). Developing the theory of formative assessment. *Educational Assessment, Evaluation and Accountability*, 1 (1), 5-31. [https://www.researchgate.net/publication/225590759\\_Developing\\_the\\_theory\\_of\\_formative\\_assessment](https://www.researchgate.net/publication/225590759_Developing_the_theory_of_formative_assessment).
- Bonniol, J. J. (1986). Recherches et formations : pour une problématique de l'évaluation formative. Em J. M. De Ketele (Éd). *L'évaluation: approche descriptive ou prescriptive?* (pp. 119-133). De Boeck.
- Clark, I. (2012). Formative Assessment: Assessment Is for Self-regulated Learning. *Educational Psychology Review*, 24 (2), 205–249. <https://doi.org/10.1007/s10648-011-9191-6>
- Coll, C., Barberá, E., & Onrubia, J. (2000). La atención a la diversidad en las prácticas de evaluación. *Infancia y Aprendizaje*, 23 (90), 111-132 <https://doi.org/10.1174/021037000760087991>
- De Ory Azcárate, M., & Ruiz Suárez, V. M. (2011). La evaluación en el aula de primaria. Factor clave para el aprendizaje de las ciencias y las matemáticas. *Revista Eureka sobre Enseñanza y Divulgación de las Ciencias* 8(2), 212-220. [https://doi.org/10.25267/Rev\\_Eureka\\_ensen\\_divulg\\_cienc.2011.v8.i2.08http://reuredc.uca.es](https://doi.org/10.25267/Rev_Eureka_ensen_divulg_cienc.2011.v8.i2.08http://reuredc.uca.es)
- Hadji, C. (2001). *Avaliação desmistificada*. Artmed Editora.
- Jaccard, P. (1901). Distribution de la flore alpine dans le bassin des Dranses et dans quelques régions voisines. *Bulletin de la Société Vaudoise des Sciences Naturelles*, 37, 241-272. [https://www.researchgate.net/publication/243457811\\_Distribution\\_de\\_la\\_Flore\\_Alpine\\_dans\\_le\\_Bassin\\_des\\_Dranses\\_et\\_dans\\_quelques\\_regions\\_voisines](https://www.researchgate.net/publication/243457811_Distribution_de_la_Flore_Alpine_dans_le_Bassin_des_Dranses_et_dans_quelques_regions_voisines)
- Melo, M., & Veiga, F. (2013). Aprendizagem: perspectivas socioconstructivistas. En F. Veiga, (ed.), *Psicologia da educação: Teoria, investigação e aplicação: Envolvimento dos alunos na escola* (pp. 263-296). Climepsi Editores.

- Noizet, G., & Caverni, J. P. (1985). *Psicologia da avaliação escolar*. Coimbra Editora.
- NCTM. (2000). *Principles and standards for school mathematics*. National Council of Teachers of Mathematics.
- Nunes, C. (1990). *A Avaliação de textos escritos: Critérios referidos e utilizados pelos professores e a sua percepção por alunos com diferente estatuto escolar*. [Tese de mestrado em Psicologia Educacional]. Instituto Superior de Psicologia Aplicada, Lisboa.
- OECD. (2013). *PISA 2012 assessment and analytical framework: Mathematics, reading, science, problem solving and financial literacy*. OECD. <https://doi.org/10.1787/9789264190511-en>
- Onrubia, J., & Lago, J. R (2008). Asesoramiento psicopedagógico y mejora de las prácticas de evaluación. *Infancia y Aprendizaje*, 31 (3), 363-383. <https://doi.org/10.1174/021037008785702929>
- Pacheco, J. (2002). Critérios de avaliação na escola. Em Ministério da Educação (ed.), *Avaliação das aprendizagens: das concepções às práticas* (pp.53-64). Departamento da Educação Básica.
- Pinto, J., & Santos, L. (2006). *Modelos de avaliação das aprendizagens*. Universidade Aberta.
- Riggan, M., & Oláh, L. N. (2011). Locating interim assessments within teachers' assessment practice. *Educational Assessment*, 16(1), 1–14. <https://doi.org/10.1080/10627197.2011.551085>
- Rosales, J., Vicente, S., Chamoso, J. M., Muñoz, D., & Orrantia, J. (2012). Teacher-student interaction in joint word problem solving. The role of situational and mathematical knowledge in mainstream classrooms. *Teaching and Teacher Education*, 28, 1185-1195. <https://doi.org/10.1016/j.tate.2012.07.007>
- Sanmarti, N. (2007). *10 ideas clave: evaluar para aprender*. Graó.
- Suurtamm, C., Koch, M., & Arden, A. (2010). Teachers' assessment practices in mathematics: Classrooms in the context of reform. *Assessment in Education: Principles, Policy & Practice*, 17(4), 399–417. <https://doi.org/10.1080/0969594X.2010.497469>

- Struyven, K., Dochy, F., & Janessens, S. (2004). Students' perceptions about evaluation and assessment in higher education: a review. *Assessment & Evaluation in Higher Education*, 30(4), 331-347. <https://doi.org/10.1080/0260293042000318091>
- Teledahl, A. (2015). Different modes in teachers' discussions of students' mathematical texts. *Teaching and Teacher Education*, 51, 68-76. <https://doi.org/10.1016/j.tate.2015.06.002>
- Turner, R., Blum, W., & Niss, M. (2015). Using competencies to explain mathematical item demand: A work in progress. In K. Stacey & R. Turner (Eds.), *Assessing mathematical literacy: The PISA experience* (pp. 85–115). Springer.
- Veldhuis, M., Van den Heuvel-Panhuizen, M., Vermeulen J. A., & Eggen, T. J. H. M. (2013). *Teachers' use of classroom assessment in primary school mathematics education in the Netherlands*. *CADMO*, 21(2), 35–53. <https://doi.org/10.1080/03055698.2016.1193475>