

# ANÁLISIS DE LAS ANOTACIONES REALIZADAS POR PROFESORES AL CALIFICAR PRUEBAS ESCRITAS DE MATEMÁTICAS

## Analysis of the annotations made by teachers when grading mathematics written exams

Arnal-Bailera, A.<sup>a</sup>, Muñoz-Escolano, J. M.<sup>a</sup> y Oller-Marcén, A. M.<sup>b</sup>

<sup>a</sup>Universidad de Zaragoza, <sup>b</sup>Centro Universitario de la Defensa de Zaragoza

### Resumen

*Pese a la existencia de múltiples modelos, métodos e instrumentos para la evaluación, la calificación de exámenes sigue siendo una de las actividades cotidianas de los profesores de matemáticas. Diversos trabajos han puesto de manifiesto la complejidad de esta tarea, así como los diversos factores que influyen en los profesores a la hora de abordarla. En este trabajo, bajo el paradigma de la teoría fundamentada, ampliamos estudios anteriores analizando las anotaciones de 21 profesores universitarios de matemáticas cuando califican diez respuestas (algunas correctas y otras con errores de distinto tipo) a una tarea de tipo procedimental en un contexto de análisis matemático. A partir de este análisis, se han identificado cinco grandes categorías (con diversas subcategorías) en torno a las que se organizan los comentarios de los profesores.*

**Palabras clave:** calificación, evaluación, pruebas escritas, análisis temático, teoría fundamentada.

### Abstract

*In spite of the many models, methods and instruments to the assessment, the grading of written exams is still a daily task of mathematics teachers. Several works point out the complexity of this task, as well as the many factors affecting the teachers when they carry it out. In this work, based on the grounded theory paradigm, we extend previous work by analysing the annotations of 21 university mathematics teachers when grading 10 answers (some correct and some containing different types of errors) to a procedural task in a context of mathematical analysis. From this analysis, we have identified five main categories (with several subcategories) which organise the teachers' comments.*

**Keywords:** grading, assessment, written exams, thematic analysis, grounded theory.

### INTRODUCCIÓN Y OBJETIVOS

En el trabajo de Arnal-Bailera, Muñoz-Escolano y Oller-Marcén (2016), se estudiaron las actuaciones de un grupo de profesores de secundaria al calificar tres producciones correctas de una misma tarea, típica en las P.A.U., de marcado carácter procedimental. Además de constatar una gran variabilidad en las calificaciones otorgadas, un estudio cualitativo mostró que, durante el proceso de calificación, los profesores analizados realizaban anotaciones y comentarios sobre diversos aspectos. En concreto se identificaron tres grandes temas en torno a los que giraban dichas anotaciones: el método de resolución seguido por el alumno, la corrección matemática de la respuesta y el grado de argumentación de la misma.

Resulta de interés entonces ampliar el estudio cualitativo anterior, incluyendo en el mismo no sólo producciones correctas, sino también otras que incluyan diferentes errores. De este modo, además de identificar eventualmente nuevas temáticas asociadas a la calificación de producciones correctas,

también aparecerán temáticas asociadas a la calificación de producciones que contienen errores de distinta naturaleza.

Así, el objetivo fundamental de este trabajo consiste en identificar y caracterizar los diferentes temas que emergen en los comentarios de profesores cuando éstos califican las respuestas, tanto correctas como erróneas, dadas por estudiantes a una tarea de tipo procedimental en un contexto de análisis matemático.

## **MARCO TEÓRICO**

La evaluación afecta de manera determinante a los procesos de enseñanza y aprendizaje de cualquier disciplina. La evaluación en matemáticas tiene distintas funciones (Giménez, 1997) y atendiendo a su propósito, a cuando se realiza y a los agentes implicados en ellas, podemos clasificar distintos tipos de evaluación: formativa, sumativa, de diagnóstico, inicial, continua, final, interna, externa, etc. (Castillo, 1999). Pese a esta diversidad, los exámenes siguen siendo un instrumento muy usado por parte de los profesores de matemáticas (Cárdenas, Blanco, Guerrero y Caballero, 2016).

Webb (1993) señala cinco aspectos comunes a todo tipo de evaluación, ya sea la evaluación realizada por un profesor (o persona que evalúa) en un contexto de aula o en un estudio internacional: la tarea que se evalúa, la respuesta que da el estudiante, la interpretación que hace el profesor de esta respuesta, la valoración de dicha interpretación ubicándola dentro de una escala y, finalmente, el registro e informe del resultado de la evaluación.

En concreto, durante las fases de interpretación y valoración, el evaluador debe identificar aspectos importantes de la tarea, establecer criterios para valorar la respuesta e identificar y ponderar los posibles errores en la producción. En este punto, juega un papel importante la formación matemática y la experiencia profesional de los correctores (Meier, Rich y Cady, 2006; Wang y Cai, 2006). A este respecto existen estudios sobre prácticas evaluativas en la Universidad tanto con profesores de la Licenciatura de Matemáticas (Jarero, Aparicio y Sosa, 2013), como del Grado o Diplomatura de Magisterio (Palacios y López-Pastor, 2013). Además, también son relevantes las concepciones del evaluador sobre los contenidos involucrados y sobre los posibles modos correctos de resolución de la tarea, así como sus concepciones y creencias acerca de la enseñanza y el aprendizaje de las matemáticas escolares (Cárdenas, Blanco, Guerrero y Caballero, 2016; Gil y Rico, 2003, Carrillo y Contreras, 1995).

Según muchos de los autores anteriores, el diferente papel que tiene el error es uno de los indicadores de las concepciones del profesorado. Por ejemplo, Huitrado y Climent (2013) establecen dimensiones para la caracterización de saberes en la comprensión de los errores mediante el estudio de las argumentaciones dadas por un docente mientras califica producciones de estudiantes. González, Gómez y Restrepo (2015) caracterizan distintos usos del error en la enseñanza de las matemáticas en el contexto de un curso de formación continua del profesorado donde los profesores debían analizar, diseñar e implementar la enseñanza de un tema. Estudiando los propósitos, las acciones docentes vinculadas y los resultados finales de dichas acciones, los autores establecen 16 usos del error por parte de los docentes. Un propósito concreto asociado a evaluar el estado cognitivo del estudiante es calificar o emitir una valoración final sobre dicho estado tras un determinado periodo formativo. Para conseguir este propósito concreto, se usan los errores para llevar a cabo las acciones de enunciar los criterios de evaluación correspondientes al tema y de diseñar los instrumentos de evaluación. Por otro lado, en cuanto a la presencia de errores en las respuestas de los estudiantes, Wang y Cai (2006) señalan que éste es uno de los factores que influyen en la variabilidad de las calificaciones otorgadas por distintos correctores.

En la literatura, existen múltiples propuestas para evaluar las producciones de los estudiantes cuando resuelven problemas de matemáticas. Charles, Lester y O'Daffer (1987) señalan tres

aspectos a evaluar: comprensión del problema, planificación de la solución y respuesta obtenida. Lane (1993), en su diseño de rúbrica general para evaluar tareas dentro del proyecto QUASAR, distingue entre conocimiento matemático, conocimiento de estrategias y comunicación. Szetela y Nicol (1992) señalan como categorías a evaluar: respuesta, expresión de la respuesta, estrategia seleccionada e implementación de la estrategia. Más recientemente, Cáceres y Chamoso (2015) proponen una matriz de valoración de resolución de problemas donde destacan cinco aspectos: la comprensión del problema, planificación y ejecución de la estrategia, solución del problema, análisis del proceso y la solución y la presentación.

La calificación de una producción solo está en el punto final de todo el proceso de evaluación. Para llegar a ella, como hemos visto, es necesario un proceso de interpretación y valoración. Así, es necesario prestar atención a cómo los correctores argumentan y justifican la calificación otorgada a cada producción. Fernández, Callejo y Márquez (2014) recogen los argumentos explicitados por maestros en formación cuando califican problemas de división-medida; mientras que Sakonidis y Klohou (2007) estudian las justificaciones dadas por maestros en ejercicio al calificar cuatro respuestas al mismo problema indicando los recursos y posiciones adoptados frente a la evaluación (Morgan, Tsatsaroni y Lerman, 2002; Morgan y Watson, 2002).

Las pruebas externas y anónimas, como los estudios de evaluación a gran escala, pueden estar menos afectadas por los efectos de las expectativas del profesor hacia cada estudiante (Morgan y Watson, 2002). Además, pueden ser un buen contexto para analizar y caracterizar las distintas dificultades de los estudiantes (Nortes y Nortes, 2010; Cai, Mok, Reddy y Stacey, 2016) así como para analizar las actuaciones de distintos correctores a la hora de valorar las respuestas de los estudiantes y proponer modelos de calificación (Gairín, Muñoz y Oller, 2012; 2013; Mengual, Gorgorió y Albarracín, 2013).

En este trabajo se analiza sobre qué aspectos o a qué temas hacen referencia distintos profesores de universidad cuando califican producciones variadas de estudiantes de 2º de Bachillerato ante una tarea de tipo procedimental y con un rol de evaluadores externos.

## **MÉTODO Y MUESTRA**

El estudio realizado es de carácter cualitativo. En particular, se aborda un análisis temático de tipo inductivo (Braun y Clarke, 2006) que se enmarca en el paradigma de la teoría fundamentada. En este tipo de diseños de investigación se “utiliza un procedimiento sistemático cualitativo para generar una teoría que explique en un nivel conceptual una acción, una interacción o un área específica” (Hernández, Fernández y Baptista, 2010, p. 492).

En el marco de la teoría fundamentada (Charmaz, 2006) se suele optar por un muestreo teórico, entendido éste como la búsqueda de fuentes de datos potencialmente ricas desde el punto de vista de la investigación que se pretende realizar. En nuestro caso, se diseñó una situación de práctica virtual (Carrillo y Contreras, 1995) consistente en un cuestionario formado por 10 respuestas distintas (creadas por el equipo investigador) a un mismo problema de optimización (Figura 1) similar a los propuestos habitualmente en las P.A.U. (Zamora-Pérez, 2014). Se trató de que las 10 respuestas diseñadas presentaran la mayor diversidad posible (Tabla 1). Para ello, tres de ellas son respuestas correctas que involucran distintos métodos de resolución (Arnal-Bailera, Muñoz-Escolano y Oller-Marcén, 2016), una de ellas presenta un procedimiento inacabado y las seis restantes, presentan errores en tareas de diversa índole según la clasificación de Gairín, Muñoz y Oller (2012). De estas seis, tres presentan errores relacionados con aspectos procedimentales algebraicos, dos con aspectos procedimentales analíticos y una con aspectos conceptuales; en algunos casos el mismo error aparece repetido en la misma respuesta.

Tabla 1. Detalles principales de las 10 respuestas

| Respuesta | Resultado  | Error                   | Tipo de tarea en la que sucede |
|-----------|------------|-------------------------|--------------------------------|
| A1        | Incorrecto | Manipulación algebraica | Tarea auxiliar general         |
| A2        | Incorrecto | Derivación              | Tarea auxiliar específica      |
| A3        | Ninguno    | Método incompleto       | Tarea principal                |
| A4        | Correcto   | -                       | -                              |
| A5        | Correcto   | -                       | -                              |
| A6        | Correcto   | -                       | -                              |
| A7        | Incorrecto | Método incorrecto       | Tarea principal                |
| A8        | Incorrecto | Derivación              | Tarea auxiliar específica      |
| A9        | Correcto   | Derivación              | Tarea auxiliar específica      |
| A10       | Incorrecto | Manipulación algebraica | Tarea auxiliar general         |

Finalmente, las respuestas calificadas contienen distintos sistemas de representación y presentan grados diversos de justificación, orden y limpieza en la presentación. Pese a no tratarse de respuestas reales de alumno, todas las respuestas presentan características observadas en el análisis de respuestas reales de alumnos a problemas similares en la P.A.U. (Gairín, Muñoz y Oller, 2013).

Dada la función  $f(x) = \frac{x^2}{4-x}$ , calcula sus extremos relativos

Figura 1. Enunciado del problema utilizado en el cuestionario.

Una vez diseñado el cuestionario, se solicitó a 21 profesores universitarios de matemáticas, todos ellos doctores en matemáticas, y en ejercicio en diversos centros de la Universidad de Zaragoza que calificaran cada una de las 10 respuestas que lo formaban. Al tratarse de profesores de matemáticas a nivel universitario se minimiza la influencia del factor de la formación matemática. Por otra parte, de los 21 participantes, 5 habían impartido clase a nivel de Educación Secundaria y 16 de ellos habían actuado como correctores en los tribunales de las P.A.U. en la Universidad de Zaragoza por lo que el factor de la experiencia profesional queda también controlado.

Puesto que cada uno de los 21 correctores actuó sobre 10 respuestas, se dispuso finalmente de una muestra de 210 correcciones realizadas cada una de ellas sobre un documento como el mostrado en la Figura 2. Como se observa en la figura, además de la calificación otorgada a la pregunta, se dedicaba un espacio en el que los correctores debían aportar justificaciones acerca de la calificación emitida. Las unidades de análisis fijadas para el estudio fueron las distintas anotaciones y comentarios realizados por los correctores en cada una de ellas. Estas anotaciones podían ser realizadas tanto en la zona dedicada a las justificaciones, como sobre la respuesta del alumno.

El análisis de la información obtenida en el cuestionario se realizó en dos fases (Charmaz, 2006). En una fase inicial, se generó y codificó una serie de categorías provisionales a partir de los datos. A continuación, en una segunda fase, se refinaron las categorías conceptuales identificadas. Puesto que el equipo investigador consideró que se había alcanzado la saturación de las categorías (Charmaz, 2006), el proceso concluyó con el análisis de la muestra considerada dando como resultado un instrumento que permite analizar las anotaciones realizadas por profesores al calificar pruebas escritas de matemáticas.

**ALUMNO 4**

1. Dada la función  $f(x) = \frac{x^2}{4-x}$ , calcula sus extremos relativos.

$$f(x) = \frac{x^2}{4-x}$$

$$f'(x) = \frac{2x(4-x) - x^2(-1)}{(4-x)^2} = \frac{8x - 2x^2 + x^2}{(4-x)^2} = \frac{8x - x^2}{(4-x)^2}$$

$$\frac{8x - x^2}{(4-x)^2} = 0$$

$$8x - x^2 = 0$$

$$\begin{cases} x = 0 \\ x = 8 \end{cases}$$

|            |      |      |      |           |
|------------|------|------|------|-----------|
| $-\infty$  | 0    | 4    | 8    | $+\infty$ |
| $8x - x^2$ | -    | +    | +    | -         |
| $(4-x)^2$  | +    | +    | +    | +         |
| $f'(x)$    | -    | +    | +    | -         |
|            | dec. | cre. | cre. | dec.      |

$f(0) = 0$        $f(8) = -16$        $f(4) = \text{no existe}$   
 Mínimo en  $(0, 0)$  y máximo en  $(8, -16)$

CALIFICACIÓN (DE 0 A 10 PUNTOS):

JUSTIFICACIÓN:

Figura 2. Enunciado del problema utilizado en el cuestionario.

La validez y fiabilidad internas de la investigación se mejoran mediante la “triangulación de investigadores” (Flick, 2007, p. 42); en concreto con la presencia de tres investigadores que actúan sobre los mismos registros observacionales para reducir los posibles sesgos individuales. Para ello, los tres investigadores analizaron por separado todas las producciones y posteriormente se compararon y discutieron las diferentes categorías que emergían de los datos analizados en sucesivas reuniones y puestas en común.

Cada uno de los correctores realizó entre 3 y 5 unidades de análisis o comentarios a clasificar para justificar la calificación de cada una de las 10 respuestas. Esto da un total por corrector de más de 30 comentarios cuyo análisis detallado excede del espacio disponible para este trabajo.

## RESULTADOS

Dentro del conjunto de correctores se distinguen 3 casos en los que, previamente a la calificación de las respuestas, establecen escalas de puntuaciones analíticas en las que dividen la tarea en varios aspectos como la corrección del proceso de resolución, la justificación de detalles técnicos o la comprobación de errores de cálculo, entre otros. Uno de estos correctores no realiza comentarios propiamente dichos en cada tarea, limitándose a asignar una calificación numérica en cada apartado de la escala y a proporcionar la calificación final de la respuesta en el cuadro reservado para ello. Los apartados en los que se divide cada una de estas escalas no han sido considerados comentarios como tales ya que no se pueden relacionar con una respuesta en particular. Los otros dos correctores, al margen de asignar la puntuación atendiendo a los criterios antes enunciados, sí que realizan comentarios durante la corrección a cada una de las respuestas y son analizados en el estudio.

Durante la primera fase del análisis se generaron y codificaron una serie de categorías provisionales a partir de los datos analizados. En la Tabla 2 mostramos los resultados relativos a esta fase del estudio.

Tabla 2. Resumen de los resultados de la primera fase.

| Nombre (código)                    | Definición   | Ejemplos  |
|------------------------------------|--|---|
| Procedimiento esperado (PE)        | Comentarios acerca del ajuste del proceso general seguido por el alumno al esperado o preferido por el corrector.                    | <i>“No se aprecia si conoce el estudio a través de la segunda derivada y/o crecimiento/ decrecimiento.”</i><br><i>“La manera de ver si hay máximo o mínimo no es la más adecuada.”</i>                                      |
| Procedimiento correcto (PC)        | Comentarios acerca de la corrección del proceso general seguido por el alumno.   | <i>“Se utilizan las mínimas herramientas necesarias...”</i><br><i>“Conoce la mecánica.”</i>   |
| Procedimiento incorrecto (PI)      | Comentarios acerca de la incorrección o incompletitud del proceso general seguido por el alumno.                                     | <i>“El estudiante trabaja mecánicamente, aplicando mal determinadas reglas. No relaciona derivadas con crecimiento/ decrecimiento.”</i><br><i>“No sabe cómo se calculan los puntos críticos ni los extremos relativos.”</i> |
| Tarea algebraica correcta (TalC)   | Comentarios acerca de la corrección de una tarea algebraica llevada a cabo por el alumno en el proceso de resolución del problema.   | <i>“No se equivoca en las operaciones.”</i><br><i>“Sabe operar.”</i>  |
| Tarea algebraica incorrecta (TalI) | Comentarios acerca de la incorrección de una tarea algebraica llevada a cabo por el alumno en el proceso de resolución del problema. | <i>“... pone que <math>(ab+c)/b^2=(a+c)/b</math>.”</i><br><i>“Errores [...] de manipulación algebraica.”</i>  |
| Tarea analítica correcta (TanC)    | Comentarios acerca de la corrección de una tarea analítica llevada a cabo por el alumno en el proceso de resolución del problema.    | <i>“Calcula bien <math>f'</math> y <math>f''</math>.”</i><br><i>“Realiza la 1ª derivada bien.”</i>  |
| Tarea analítica incorrecta (TanI)  | Comentarios acerca de la incorrección de una tarea analítica llevada a cabo por el alumno en el proceso de resolución del problema.  | <i>“Fallo al derivar...”</i><br><i>“Error en la fórmula de derivación.”</i>   |
| Gravedad absoluta (GA)             | Comentarios acerca de la gravedad de un error cometido por el alumno en términos absolutos.  | <i>“...tiene un error de simplificación MUY grave...”</i><br><i>“Despiste.”</i>   |
| Gravedad relativa (GR)             | Comentarios acerca de la gravedad de un error cometido por el alumno en términos relativos.  | <i>“Creo que al nivel de las PAU no se puede aprobar un ejercicio en el que se comete un error de este tipo.”</i>   |
| Reiteración (R)                    | Comentarios respecto a la repetición a lo largo del problema de un mismo error cometido por el alumno.                               | <i>“Comete un error grave 2 veces...”</i><br><i>“Mismo fallo repetido 2 veces.”</i>   |
| Limpieza (LI)                      | Comentarios acerca del orden y limpieza en la presentación de la respuesta del alumno.   | <i>“Exposición esquemáticamente muy clara.”</i><br><i>“Claro y muy bien.”</i>   |
| Argumentación (AR)                 | Comentarios acerca del grado de argumentación presente en la respuesta del alumno.   | <i>“No justifica suficientemente las monotonías.”</i><br><i>“Quizás falte algo de letra comentando la teoría.”</i>  |
| Sistemas de representación (SR)    | Comentarios acerca del uso por parte del alumno de distintos sistemas de representación.   | <i>“No hubiera estado mal añadir una pequeña gráfica de la función.”</i><br><i>“Aproxima <math>8/3</math> con lo que podría generar un error.”</i>  |
| Resultado final (RF)               | Comentarios acerca de la corrección del resultado final presentado por el alumno.  | <i>“¡El “acierto” en el mínimo es casualidad!”</i><br><i>“El resultado es correcto.”</i>  |
| Verificación del resultado (VR)    | Comentarios acerca del uso de estrategias por parte del alumno para la comprobación de la  | <i>“No analiza el hecho de que salgan dos mínimos.”</i>   |

validez o corrección del resultado final.

*“Con los resultados obtenidos debería darse cuenta de que algo va mal.”*

Las 15 categorías identificadas en esta primera fase recogen la práctica totalidad de las unidades de análisis estudiadas. De hecho, los escasos comentarios que no corresponden a ninguna de las categorías anteriores se refieren a aspectos que no están directamente relacionados con la calificación de la pregunta. Un ejemplo de este tipo de comentarios aparece recogido en la Figura 3.

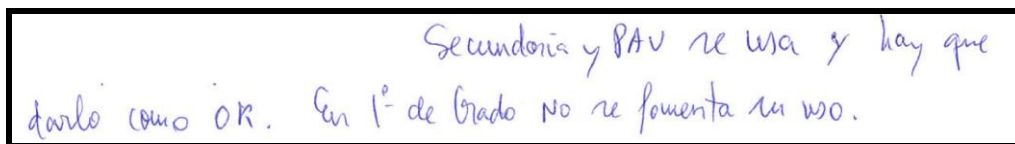


Figura 3. Comentario ajeno al proceso de calificación: “En 1º de Grado no se fomenta su uso”.

En la segunda fase, las categorías anteriormente identificadas se refinaron y agruparon en otras más generales tal y como se recoge en la Tabla 3.

Tabla 3. Resumen de los resultados de la primera fase.

| Nombre        | Código | Categorías de la primera fase que la componen | Definición  |
|---------------|--------|---|---|
| Procedimiento | P      | PE, PC, PI                                    | Comentarios acerca del proceso general seguido por el alumno.   |
| Cálculos      | C      | TalC, TalI, TanC, TanI                        | Comentarios acerca de una tarea concreta llevada a cabo por el alumno dentro del proceso general de resolución. |
| Errores       | E      | GA, GR, RE                                    | Comentarios acerca de un error cometido por el alumno.  |
| Exposición    | EX     | LI, AR, SR                                    | Comentarios acerca de la calidad expositiva de la respuesta del alumno.   |
| Resultado     | R      | RF, VR  | Comentarios acerca del resultado final presentado por el alumno.  |

## DISCUSIÓN Y CONCLUSIONES

Los resultados del estudio se centran fundamentalmente en la tercera característica apuntada por Webb (1993) sobre la interpretación de la respuesta por parte del corrector y, de manera tangencial, la cuarta característica sobre la valoración de la misma. Del análisis de los comentarios de los profesores, han surgido 15 categorías temáticas que hemos agrupado en 5 categorías más generales: procedimiento, cálculos, errores, exposición y resultado. Ampliar el estudio a un número mayor de producciones con distintas características ha supuesto que hayan aparecido nuevos temas en los comentarios de los correctores. Las temáticas emergidas en Arnal-Bailera, Muñoz-Escolano y Oller-Marcén (2016) (método esperado, argumentación y corrección matemática) han sido refinadas y ampliadas con estas nuevas categorías que hacen referencia a comentarios sobre los cálculos realizados para llegar a la solución, el resultado final y los errores cometidos por el alumno.

Las categorías presentadas en la Tabla 3 pueden ser útiles para analizar cualitativamente el modo en que profesores de matemáticas califican pruebas escritas. Este análisis permitiría una descripción más detallada de los perfiles que eventualmente puedan identificarse a partir de un estudio estadístico de carácter cuantitativo. Un trabajo de este tipo, pero circunscrito únicamente a la calificación de respuestas correctas, ya fue llevado a cabo por Arnal-Bailera, Muñoz-Escolano y Oller-Marcén (2016). La consideración de todo tipo de respuestas permitiría considerar aspectos como la relación entre las calificaciones y el tipo de comentarios realizados, así como la realización de un análisis del contenido de dichos comentarios.

Por otro lado, la identificación de estas temáticas en las justificaciones que emplean los correctores puede permitir el establecimiento de criterios de evaluación más completos, tal y cómo señalan González, Gomez y Restrepo (2015). Así, el análisis de los comentarios que aparecen durante la evaluación de tareas de escolares y la reflexión sobre las mismas es un valioso recurso didáctico para la formación inicial del profesorado, tanto para abordar el tratamiento didáctico de los contenidos (Fernández, Callejo y Márquez, 2014) como para trabajar contenidos sobre evaluación en matemáticas (Arnal-Bailera, Cid, Muñoz-Escolano y Oller-Marcén, 2018).

De hecho, a pesar de que la mayor parte de los correctores analizados no establecieron ninguna escala de evaluación a priori y puntuaron mediante la penalización por error, todas las categorías emergentes que aparecen son comparables con las escalas de puntuación analítica y de criterios de evaluación de problemas y tareas presentes en la literatura. Por ejemplo, las consideraciones acerca de los procedimientos, cálculos y exposición están presentes en todos los indicadores para la evaluación de problemas mencionados anteriormente (Charles, Lester y O'Daffer, 1987; Szetela y Nicol, 1992; Lane, 1993; Cáceres y Chamoso, 2015). Comentarios acerca de la importancia del error (grave o leve) aparecen explícitamente tanto en la rúbrica general de Lane (1993) como en las categorías de Szetela y Nicol (1992). De hecho, estos últimos autores también advierten sobre la reiteración de un mismo error. Cáceres y Chamoso (2015) valoran el hecho de que el estudiante obtenga la respuesta correcta al problema, y que, además, sea capaz de valorar si la solución que obtiene de ejecutar un determinado plan es correcta y tiene sentido en el contexto del problema.

Sin embargo, entre los comentarios analizados no aparece mención alguna a la comprensión del enunciado del problema, aspecto considerado por todos los autores anteriores en sus trabajos. Esto puede deberse al carácter estándar del enunciado de la tarea propuesta, que aparece consistentemente en las P.A.U. y se trabaja de forma exhaustiva en las aulas de bachillerato. Cabría ampliar el estudio para analizar los efectos en las temáticas identificadas al cambiar la tarea a evaluar, tanto en lo relativo al contenido matemático (en este caso, cálculo de extremos relativos) como al tipo de demanda cognitiva (en este caso, baja demanda cognitiva) (Smith y Stein, 1998). Por ejemplo, nuestra categoría 'Argumentación' incluye comentarios con diferente grado de exigencia argumentativa, desde la mera descripción de los pasos efectuados hasta la demanda de la explicitación de los resultados teóricos que avalan el procedimiento seguido por el alumno.

### **Agradecimientos**

Este trabajo ha sido parcialmente financiado por el Gobierno de Aragón y el Fondo Social Europeo (S36\_17D Investigación en Educación Matemática) así como por el Ministerio de Economía y Competitividad (Proyecto EDU2015-65378-P).

### **Referencias**

- Arnal-Bailera, A., Cid, E., Muñoz-Escolano, J. M. y Oller-Marcén, A. M. (2018). Marking mathematics exams as a tool for secondary teacher training. En M. E. Strutchens, R. Huang, D. Potari y L. Losano (Eds.), *Educating Prospective Secondary Mathematics Teachers*. New York, NY: Springer. En prensa.
- Arnal-Bailera, A., Muñoz-Escolano, J. M. y Oller-Marcén, A. M. (2016). Caracterización de las actuaciones de correctores al calificar pruebas escritas de matemáticas. *Revista de Educación*, 371, 35-60.
- Braun, V. y Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology* 3(2), 77-101.
- Cai, J., Mok, I., Reddy, V. y Stacey, K. (2016). *International comparative studies in mathematics: Lessons for improving students' learning*. New York, NY: Springer.
- Cáceres, M. J. y Chamoso, J. M. (2015). La evaluación sobre la resolución de problemas de matemáticas. En L. Blanco, J. A. Cárdenas y A. Caballero (Eds.), *La resolución de problemas de Matemáticas en la formación inicial de profesores de Primaria* (pp. 225-241). Cáceres: Universidad de Extremadura.



- Cárdenas, J. A., Blanco, L. J., Guerrero, E. y Caballero, A. (2016). Manifestaciones de los profesores de matemáticas sobre sus prácticas de evaluación de la resolución de problemas. *Bolema*, 30(55), 649-669.
- Carrillo, J. y Contreras, L. C. (1995). Un modelo de categorías e indicadores para el análisis de las concepciones del profesor sobre la matemática y su enseñanza. *Educación matemática*, 7(3), 79-92.
- Castillo, S. (1999). Sentido educativo de la evaluación en la Educación Secundaria. *Educación XXI*, 2, 65-96.
- Charles, R., Lester, F. y O'Daffer, P. (1987). *How to evaluate progress in problem-solving*. Reston, VA: National Council of Teachers of Mathematics.
- Charmaz, K. (2006). *Constructing Grounded Theory. A Practical Guide through Qualitative Analysis*. London: SAGE.
- Fernández, C., Callejo, M. L. y Márquez, M. (2014) Conocimiento de los estudiantes para maestro cuando interpretan respuestas de estudiantes de primaria a problemas de división-medida. *Enseñanza de las Ciencias*, 32(3), 407-424.
- Flick, U. (2007). *Managing Quality in Qualitative Research*. London: SAGE.
- Gairín, J. M., Muñoz, J. M. y Oller, A. M. (2012). Propuesta de un modelo para la calificación de exámenes de matemáticas. En A. Estepa, Á. Contreras, J. Deulofeu, M. C. Penalva, F. J. García y L. Ordóñez (Eds.), *Investigación en Educación Matemática XVI* (pp. 261-274). Jaén: SEIEM.
- Gairín, J. M., Muñoz, J. M. y Oller, A. M. (2013). Anomalías en los procesos de identificación de errores en las pruebas escritas de matemáticas de las P.A.U. *Campo abierto: Revista de Educación* 32(2), 27-50.
- Gil, F. y Rico, L. (2003). Concepciones y creencias del profesorado de secundaria sobre enseñanza y aprendizaje de las matemáticas. *Enseñanza de las ciencias*, 21(1), 27-47.
- Giménez, J. (1997). *Evaluación en matemáticas. Una integración de perspectivas*. Madrid: Síntesis.
- González, M. J., Gómez, P. y Restrepo, Á. M. (2015). Usos del error en la enseñanza de las matemáticas. *Revista de Educación*, 370, 71-95.
- Hernández, R., Fernández, C. y Baptista, M. P. (2010) *Metodología de la Investigación*. México: McGraw Hill Educación.
- Huitrado, J. L. y Climent, N. (2013). Conocimiento profesional del profesor ante errores relativos al álgebra de los alumnos de secundaria. En A. Berciano, G. Gutiérrez, A. Estepa y N. Climent (Eds.), *Investigación en Educación Matemática XVII* (pp. 327-336). Bilbao: SEIEM.
- Jarero, M., Aparicio, E. y Sosa, L. (2013). Pruebas escritas como estrategias de evaluación de aprendizajes matemáticos: un estudio de caso a nivel superior. *RELIME*, 16(2), 213-243.
- Lane, S. (1993). The conceptual framework for the development of a mathematics performance assessment instrument. *Educational Measurement: Issues and Practice*, 12(2), 16-23.
- Meier, S. L., Rich, B. S. y Cady, J. (2006). Teachers' use of rubrics to score non-traditional tasks: factors related to discrepancies in scoring. *Assessment in Education*, 13(01), 69-95.
- Mengual, E., Gorgorió, N. y Albarracín, L. (2013). Validación de un instrumento para la calificación de exámenes de Matemáticas. En A. Berciano, G. Gutiérrez, A. Estepa y N. Climent (Eds.), *Investigación en Educación Matemática XVII* (pp. 367-381). Bilbao: SEIEM.
- Morgan, C. y Watson, A. (2002). The Interpretative Nature of Teachers' Assessment of Students' Mathematics: Issues for Equity. *Journal for Research in Mathematics Education*, 33, 78-110.
- Morgan, C., Tsatsaroni, A. y Lerman, S. (2002). Mathematics teachers' positions and practices in discourses of assessment. *British Journal of Sociology of Education*, 23(3), 445-461.
- Nortes, A. y Nortes, R. (2010). Resolución de problemas de matemáticas en las pruebas de acceso a la universidad. Errores significativos. *Educatio Siglo XXI*, 28(1), 317-341.

- Palacios, A. y López-Pastor, V. M. (2013). Haz lo que yo digo, pero no lo que yo hago: Sistemas de evaluación del alumnado en la formación inicial del profesorado. *Revista de Educación*, 361, 279-305.
- Sakonidis, H. y Klothou, A. (2007). On Primary teachers' assessment of pupils' written work in mathematics. En J-H. Woo, H-C. Lew, K-S. Park y D-Y. Seo (Eds.), *Proceedings of the 31st conference of the International Group for the Psychology of Mathematics Education*, vol 4, (pp. 153-160). Seoul: PME.
- Smith, M. S. y Stein, M. K. (1998). Selecting and creating mathematical tasks: From research to practice. *Mathematics Teaching in the Middle School*, 3, 344-350.
- Szetela, W. y Nicol, C. (1992). Evaluating Problem Solving in Mathematics. *Educational Leadership*, 49(8), 42-45.
- Wang, N. y Cai, J. (2006). An investigation of factors influencing teachers' scoring student responses to mathematics constructed-response assessment tasks. En J. Novotná, H. Moraová, M. Krátká y N. Stehliková (Eds.), *Proceedings 30th Conference of the International Group for the Psychology of Mathematics Education*, vol. 5, (pp. 369-376). Prague: PME.
- Webb, N. L. (1993). Assessment for the Mathematics Classroom. En N. L. Webb y A. F. Coxford (Eds.), *Assessment in the Mathematics Classroom. 1993 Yearbook*. (pp. 1-7). Reston, VA: National Council of Teachers of Mathematics.
- Zamora-Pérez, R. F. (2014). *Análisis de las pruebas de acceso a las universidades de Castilla y León (Matemáticas II)*. (Tesis doctoral inédita). Universidad de Valladolid, Valladolid.