PROPUESTA DE UN MODELO PARA LA CALIFICACIÓN DE EXÁMENES DE MATEMÁTICAS

A PROPOSAL OF A MODEL FOR MARKING MATHEMATICS EXAMS

Gairín, J.M. (1), **Muñoz, J.M.** (1), **Oller, A.M.** (2)

Universidad de Zaragoza (1); *Centro Universitario de la Defensa* (2)

Resumen

A partir de un estudio de tipo indagatorio sobre la corrección de exámenes de matemáticas de las Pruebas de Acceso a la Universidad, llegamos a relacionar la importancia de los errores cometidos por los alumnos con la tipología de las tareas en que las que se producen y construimos un modelo de penalización de errores que introduce criterios de calificación más objetivos que los observados.

Abstract

We use an exploratory study about the corrections of mathematics exams from a University Entrance Examination to relate the importance of the mistakes made by the students with the typology of the tasks where they appear. Thus, we propose a model for mistake penalization which introduces marking criteria which are more objective than the observed ones.

Palabras clave: *Evaluación. Calificación. Errores. Matemáticas. Pruebas de Acceso a la Universidad.*

Key words: *Evaluation. Marking. Mistakes. Mathematics. University Entrance Examination.*

Introducción

En el proceso de enseñanza y aprendizaje de las materias, la evaluación juega un papel de gran relevancia porque es la única manera de saber si lo que se ha enseñado ha sido aprendido por el alumno. Además de conocer el grado de dominio alcanzado por el alumno en relación con los objetivos propuestos, la evaluación también sirve para determinar si el proceso de enseñanza ha sido adecuado para alcanzar dichos objetivos (Cantón y Pino-Juste, 2011).

En el ámbito de la enseñanza y aprendizaje de las matemáticas se han realizado estudios específicos encaminados a adaptar métodos e instrumentos a las especificidades de la materia. Existen interesantes monografías dedicadas a las muy diversas facetas del proceso de evaluación (Giménez, 1997; Kaur y Wong, 2011). También encontramos trabajos centrados en el diseño de herramientas y métodos para la evaluación en distintos ámbitos de la matemática y en diferentes niveles educativos (Lian et al., 2010; Torres-Skoumal, 2001; Carrillo y Guevara, 1996; Gallardo y González, 2005; Gutiérrez y Jaime, 1995); y otros trabajos (Boesen et al., 2010) que muestran el modo en que el propio proceso de evaluación influye en el modo de trabajar de los alumnos.

Los exámenes de matemáticas de las Pruebas de Acceso a la Universidad[1] son un ejemplo de prueba de evaluación externa con gran influencia en la enseñanza y aprendizaje de las matemáticas de Bachillerato (Ordóñez y Contreras, 2011; Ruiz de Gauna, 2011). Sobre las calificaciones de los exámenes de estas pruebas conocemos algunos estudios estadísticos (Escudero y Bueno, 1994; Grau et al., 2002), pero no hemos encontrado ningún trabajo que estudie cómo desempeñan los profesores la tarea de calificar los exámenes de matemáticas en las PAU.

No es habitual que el trabajo de corregir exámenes de matemáticas figure en los planes de formación de los profesores. Las experiencias como alumnos, las conversaciones con compañeros, los debates con miembros de departamentos de matemáticas, la lectura sobre la forma de actuar de otros profesionales, etc., constituyen en la práctica la formación real de quienes han de corregir exámenes de matemáticas.

En el caso especial de las PAU, los armonizadores de cada asignatura elaboran unos criterios específicos de corrección que indican la calificación que corresponde a los alumnos que responden correctamente a las distintas partes que componen las preguntas. No obstante, se constata la existencia de diferencias significativas entre las calificaciones de distintos correctores cuando califican un mismo examen. Así lo pone de manifiesto el análisis estadístico, realizado por uno de los autores de este artículo, sobre la actuación de los dos correctores de la asignatura Matemáticas Aplicadas a las Ciencias Sociales II del Tribunal nº 8 de la convocatoria de junio de 2010 de la Universidad de Zaragoza:

	Media	Desviación Típica	Notas en [8,10]	Notas en [0,4]
Corrector A	5,74	2,03	11/74 (15%)	9/74 (12%)
Corrector B	3,86	1,98	5/84 (6%)	42/84 (50%)

Téngase en cuenta que a los correctores se les asignan de forma aleatoria los cerca de 90 exámenes que tienen que calificar y que los correctores tienen que aplicar los

[1] PAU en lo sucesivo.

mismos criterios de corrección proporcionados por el armonizador de la prueba. En estas condiciones pensamos que las diferencias entre las calificaciones de los dos correctores son achacables, en buena medida, a que existe una fuerte componente de subjetividad en la valoración de las respuestas de los alumnos y que, en consecuencia, se producen diferencias en las calificaciones otorgadas por distintos correctores.

Así pues, el objetivo de este trabajo es presentar un marco para la calificación de exámenes de matemáticas que reduzca el grado de subjetividad propio de esta actividad.

Método y muestra

Este es un estudio de tipo indagatorio organizado, de manera secuencial, en las siguientes fases:

Fase 1: Identificar algunos fenómenos no deseables en el proceso de corrección de exámenes de matemáticas de las PAU. La fiabilidad interna de esta fase se mejora con la presencia de tres investigadores que actúan sobre los mismos registros observacionales (Goetz y Lecompte, 1988).

Fase 2: Elaborar propuestas de mejora de dicho proceso encaminadas a evitar los fenómenos detectados.

Para llevar a cabo nuestro trabajo se ha utilizado información procedente de los exámenes correspondientes a las PAU de la convocatoria de Septiembre de 2009 en Comunidad de Aragón. En concreto se han analizado las actuaciones de 6 correctores, de los cuales:

- Tres correctores han calificado la asignatura Matemáticas II, de la que hemos analizado los 203 exámenes correspondientes a la opción A.
- Tres correctores han calificado la asignatura Matemáticas Aplicadas a las Ciencias Sociales II, de la que hemos analizado los 206 exámenes de la opción A.

Fase 1

En nuestra opinión, las decisiones y actuaciones que conlleva la calificación de la respuesta que ofrece un alumno al enunciado de un problema determinado deben seguir el siguiente gráfico:

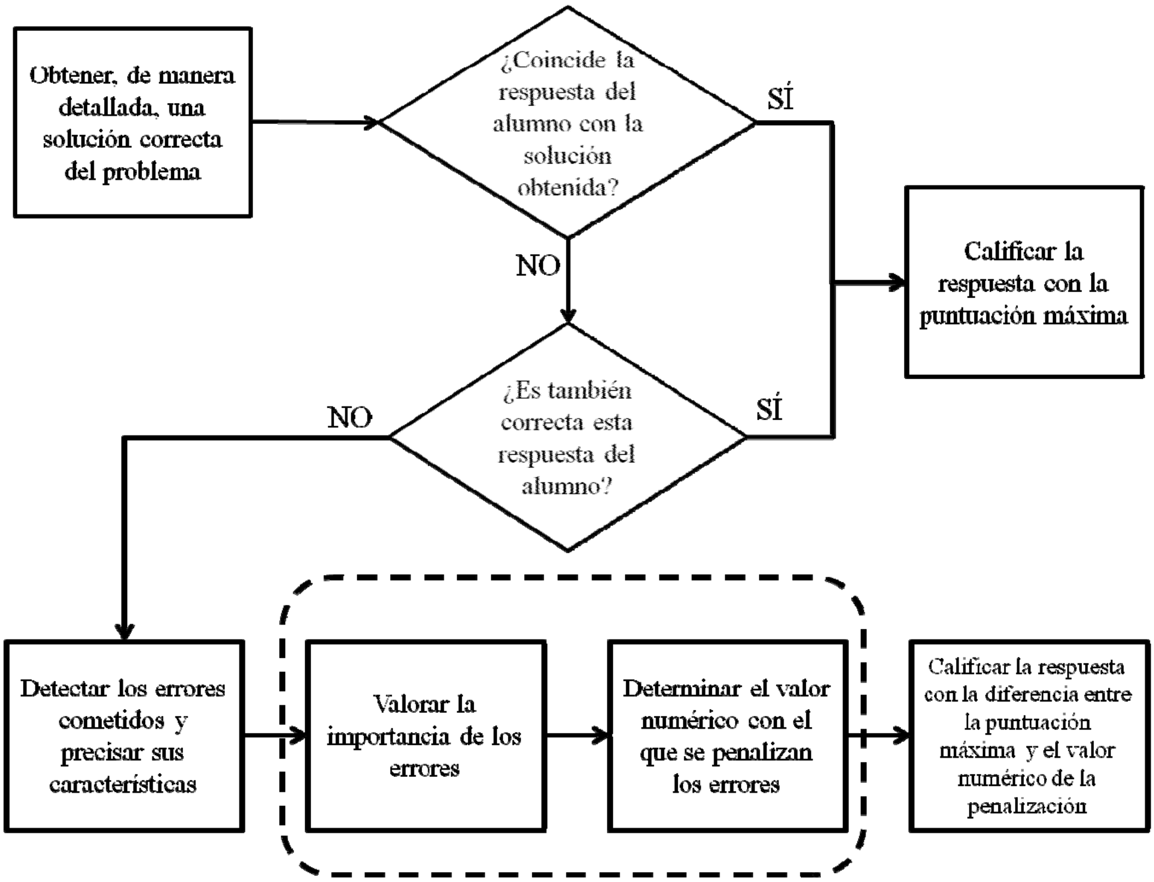


Gráfico 1

En este proceso, los primeros pasos no presentan dificultades reseñables para los correctores. Sin embargo, hemos detectado varios fenómenos anómalos en la corrección de los exámenes asociados a los pasos señalados con línea discontinua en el gráfico 1.

Fenómenos reseñables

Ante la presencia de un error en la respuesta de un alumno el corrector tiene que tomar una decisión sobre la "gravedad" o "levedad" que clasifique dicho error y debe asignar un valor numérico para penalizarlo. En el desempeño de esta tarea de hemos observado cuatro fenómenos destacables:

Fenómeno 1: Existen posiciones claramente diferenciadas entre distintos correctores al valorar la importancia de un mismo tipo de error.

Reproducimos las actuaciones de tres correctores al calificar las respuestas al siguiente problema:

a) Utilizar el cambio de variable $t^3 = 1 - x$ para calcular el siguiente límite

$$\lim_{x \to 0} \frac{(1-x)^{1/3} - 1}{x}. \quad \text{(1 punto)}$$

$$\lim_{x\to 0} \frac{(1-x)^{1/3}-1}{x} \overset{\text{cambio}}{=} \lim_{\cancel{t\to 0}} \frac{(t^3)^{1/3}-1}{1-t^3} =$$

$$= \lim_{\cancel{t\to 0}} \frac{t-1}{1-t^3} = \lim_{\cancel{x\to 0}} \frac{0-1}{1-0} = \frac{-1}{1} = -1$$

0.5

2) $\lim_{x\to 0} \frac{(1-x)^{1/3}-1}{x} \Rightarrow \lim_{t\to 0} \frac{t-1}{1-t^3} = \frac{0-1}{1-0} = \frac{-1}{1} = \boxed{-1}$

para $t^3 = 1-x$

$x = 1-t^3$

0'3

2.

$\lim_{x\to 0} \frac{(1-x)^{1/3}-1}{x} \Rightarrow \lim_{t\to 0} \frac{t-1}{1-t^3} \Rightarrow \lim_{t\to 0} \frac{0-1}{1-0} = -1$

$t^3 = 1-x \Rightarrow x = 1-t^3$

$t = (1-x)^{1/3}$

0 ptos

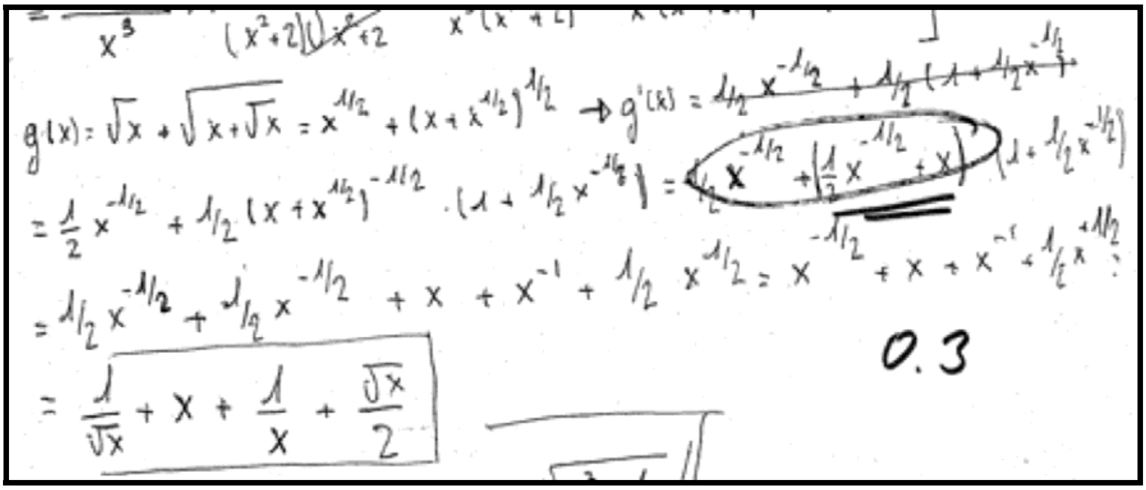
Puede observarse que los tres correctores señalan el mismo tipo de error y que cada uno de los correctores otorga calificaciones significativamente diferentes. Este hecho pone de manifiesto que cada uno de los correctores valora la importancia del error de acuerdo con las concepciones establecidas sobre evaluación (Gil, Rico y Fernández, 2002).

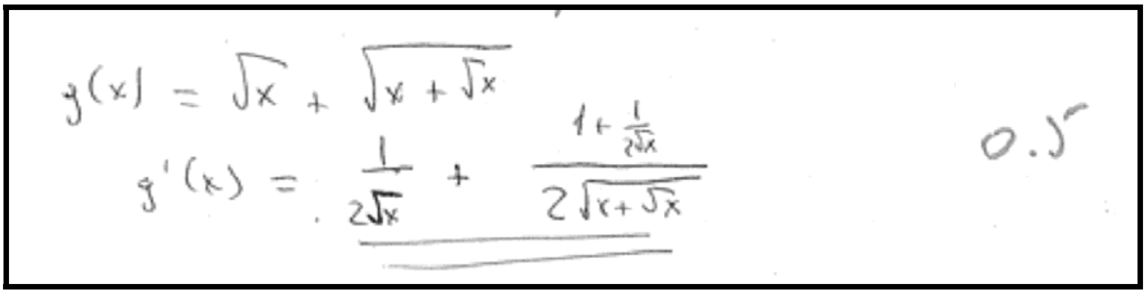
Fenómeno 2: La importancia de los errores se establece con independencia de las exigencias del enunciado de los problemas.

Reproducimos la actuación de un mismo corrector ante dos respuestas al segundo apartado del problema:

3. a) Derive las siguientes funciones: *(1,5 puntos)*

$$f(x) = \ln\frac{x^3}{\sqrt{x^2+2}}, \qquad g(x) = \sqrt{x} + \sqrt{x+\sqrt{x}}, \qquad h(x) = e^{\sqrt{x^2+1}}$$



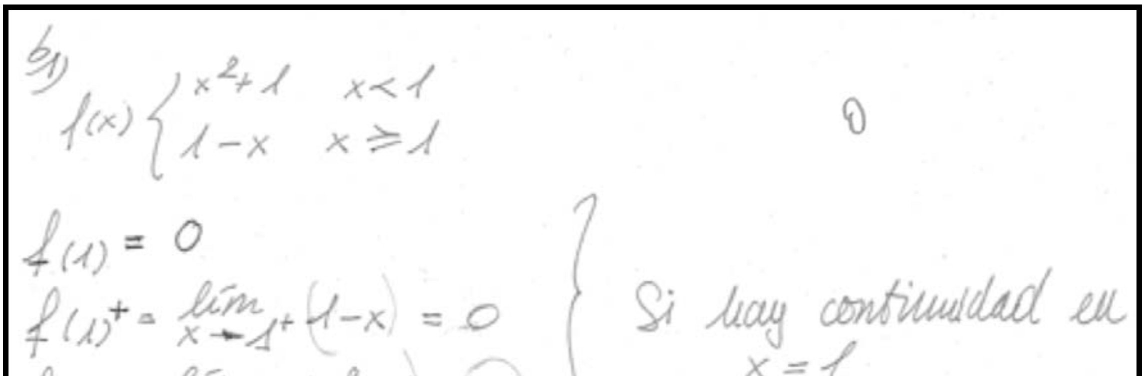


En las dos imágenes se calcula de forma correcta la función derivada, por lo que ambas deberían calificarse con 0,5 puntos. Pero el corrector penaliza la primera respuesta con un 40% de la calificación debido a un error algebraico.

Fenómeno 3: La importancia de un error se valora con independencia del conocimiento matemático que se pretende evaluar con el problema.

Reproducimos la actuación de un corrector al calificar la respuesta de un alumno al problema:

b) Estudiar la continuidad de $f(x)=\begin{cases} x^2+1 & x<1 \\ 1-x & x\geq 1 \end{cases}$ *(1 punto)*



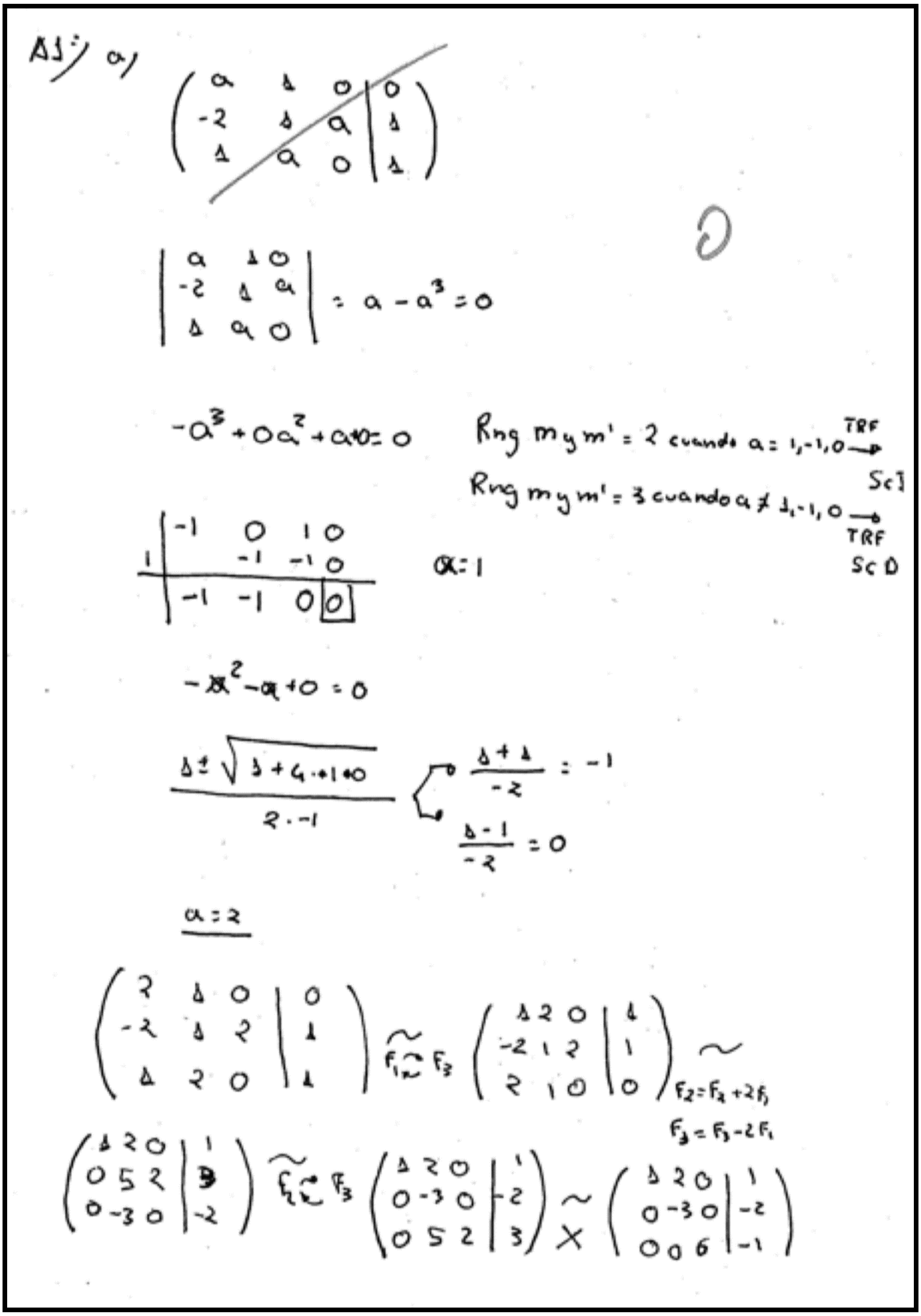
El corrector señala que el alumno ha obtenido como límite 0. Entendemos que con esta actuación el corrector deja constancia de que el alumno tiene unos conocimientos muy escasos de operar con números enteros; mientras que, atendiendo a las exigencias del problema, el corrector debería calificar los conocimientos que tiene el alumno sobre las funciones continuas.

Fenómeno 4: La detección de un error considerado como muy grave da por finalizado el proceso de corrección.

Reproducimos la actuación de un corrector al calificar la respuesta de un alumno al problema:

a) Discutir y resolver cuando sea posible el siguiente sistema lineal: *(1,75 puntos)*

$$\begin{cases} ax + y = 0 \\ -2x + y + az = 1 \\ y + az = 1 \end{cases}$$



A1) a)

$$\cancel{\left(\begin{array}{ccc|c} a & 1 & 0 & 0 \\ -2 & 1 & a & 1 \\ 1 & a & 0 & 1 \end{array}\right)}$$

0

$$\begin{vmatrix} a & 1 & 0 \\ -2 & 1 & a \\ 1 & a & 0 \end{vmatrix} = a - a^3 = 0$$

$$-a^3 + 0a^2 + a + 0 = 0$$

Rng m y m' = 2 cuando a = 1, -1, 0 $\xrightarrow{TRF}$ ScI

Rng m y m' = 3 cuando a ≠ 1, -1, 0 $\xrightarrow{TRF}$ ScD

$$\begin{array}{r|rrrr} & -1 & 0 & 1 & 0 \\ 1 & & -1 & -1 & 0 \\ \hline & -1 & -1 & 0 & \boxed{0} \end{array} \qquad a = 1$$

$$-a^2 - a + 0 = 0$$

$$\frac{1 \pm \sqrt{1 + 4 \cdot -1 \cdot 0}}{2 \cdot -1} \begin{cases} \frac{1+1}{-2} = -1 \\ \frac{1-1}{-2} = 0 \end{cases}$$

a = 2

$$\left(\begin{array}{ccc|c} 2 & 1 & 0 & 0 \\ -2 & 1 & 2 & 1 \\ 1 & 2 & 0 & 1 \end{array}\right) \underset{F_1 \leftrightarrow F_3}{\sim} \left(\begin{array}{ccc|c} 1 & 2 & 0 & 1 \\ -2 & 1 & 2 & 1 \\ 2 & 1 & 0 & 0 \end{array}\right) \underset{\substack{F_2 = F_2 + 2F_1 \\ F_3 = F_3 - 2F_1}}{\sim}$$

$$\left(\begin{array}{ccc|c} 1 & 2 & 0 & 1 \\ 0 & 5 & 2 & 3 \\ 0 & -3 & 0 & -2 \end{array}\right) \underset{F_2 \leftrightarrow F_3}{\sim} \left(\begin{array}{ccc|c} 1 & 2 & 0 & 1 \\ 0 & -3 & 0 & -2 \\ 0 & 5 & 2 & 3 \end{array}\right) \underset{X}{\sim} \left(\begin{array}{ccc|c} 1 & 2 & 0 & 1 \\ 0 & -3 & 0 & -2 \\ 0 & 0 & 6 & -1 \end{array}\right)$$

Pensamos que el corrector entiende que el alumno ha cometido un error al representar el sistema de ecuaciones en forma matricial. Dicho corrector parece considerar muy grave el error del alumno y decide que no tiene sentido proseguir con el proceso de calificación; aunque, como puede constatarse, el alumno demuestra conocer el contenido principal que se evalúa.

Comentarios relativos a los fenómenos

1. Los correctores valoran la importancia de los errores más por creencias personales (Gil, Rico y Fernández, 2002), que por la aplicación de criterios universalmente establecidos y aceptados.

2. No se deben valorar como importantes los errores cometidos al realizar tareas que no se solicitan explícitamente en el enunciado de la pregunta.

3. Para valorar la importancia de los errores así como para determinar un valor numérico de la penalización, es necesario clarificar el tipo de tareas que conlleva la resolución de un problema de matemáticas y jerarquizarlas atendiendo al papel que juegan en la resolución del problema.

4. Ante las respuestas que contienen errores el corrector debe tomar dos decisiones: valorar y cuantificar la penalización que corresponde al error detectado y si debe finalizar o proseguir el proceso de calificación del alumno.

Fase 2.

Apoyándonos en las consideraciones anteriores, presentamos a continuación dos propuestas:

I. Una tipología y jerarquización de las tareas,

II. Un modelo de penalización de errores.

I. Tipología y jerarquía de tareas

Los problemas que se proponen en los exámenes de matemáticas exigen del alumno la realización de tareas de distinta naturaleza:

a) Tareas principales

Son aquellas tareas que claramente constituyen el objetivo principal de la calificación: valorar la comprensión del alumno sobre los contenidos matemáticos propios de los temarios de matemáticas de un curso determinado. A su vez, estas tareas también pueden distinguirse atendiendo a su carácter conceptual o procedimental.

b) Tareas auxiliares

En el proceso de resolución del problema también hay que realizar otro tipo de tareas, que denominamos *tareas auxiliares*, con la finalidad de obtener las informaciones necesarias para dar la respuesta al problema. Identificamos dos grandes grupos de tareas auxiliares: las específicas y las generales.

b1) Tareas auxiliares específicas

Son aquellas tareas que juegan un papel instrumental para alcanzar la solución de un problema en el que aparecen tareas principales sobre contenidos específicos. Por ejemplo, para calcular los extremos relativos de una función, la obtención de derivadas es una tarea auxiliar específica.

b2) Tareas auxiliares generales

Consideramos como tareas auxiliares generales de un determinado curso de matemáticas a todo tipo de tareas matemáticas que ha realizado el alumno a lo largo de su formación matemática anterior. Atendiendo a su naturaleza, estas tareas podrían dividirse en tareas de tipo algebraico, de tipo aritmético, de tipo geométrico, de tipo gráfico y de representación.

Una vez establecida la tipología de las tareas, para determinar una jerarquía de las mismas es necesario estudiar el papel que juegan en las distintas categorías de problemas propuestos en los exámenes de las PAU. Hemos detectado dos categorías de problemas:

Problemas de categoría 1.

Incluye los problemas cuya tarea principal consiste en la aplicación directa de una técnica de cálculo, como ocurre en el siguiente ejemplo:

Calcule la siguiente integral $\int_1^2 (x^2 + 3x + 1)dx$

En este problema el cálculo de primitivas y la regla de Barrow juegan el papel de tareas principales mientras que las tareas de simplificación y los cálculos aritméticos constituyen las tareas auxiliares generales.

En esta categoría de problemas la jerarquía de tareas es:

1. Tareas principales
2. Tareas auxiliares generales

Problemas de categoría 2.

Son problemas en los que debe manejarse algún concepto matemático, como muestra el siguiente ejemplo:

Calcule el área de la superficie limitada por las funciones $f(x) = x^2 + 1$ y $g(x) = 2x + 3$

En este problema, la delimitación del área pedida, el concepto de la integral definida de cada una de las funciones y la interpretación que tiene la integral de la resta de las dos funciones constituyen las tareas principales. El cálculo de integrales definidas juega el papel de tarea auxiliar específica. Las representaciones gráficas, la resolución de ecuaciones, la simplificación de expresiones algebraicas y el cálculo aritmético suponen las tareas auxiliares generales.

En esta categoría de problemas la jerarquía de tareas es:

1. Tareas principales
2. Tareas auxiliares específicas
3. Tareas auxiliares generales

<u>II. Modelo de penalización de errores</u>

Es posible y deseable establecer un marco que oriente las penalizaciones de todos los correctores de exámenes de matemáticas de manera que se atenúen las diferencias entre las actuaciones de distintos correctores. En este sentido, proponemos un modelo de penalización de errores sustentado en la teoría curricular interpretativa (Sierra y Pérez, 2007), del que hacemos, en primer lugar, una representación gráfica y, en segundo lugar, una descripción de su sentido y alcance.

Modelo de penalización de errores

Modelo de tercios

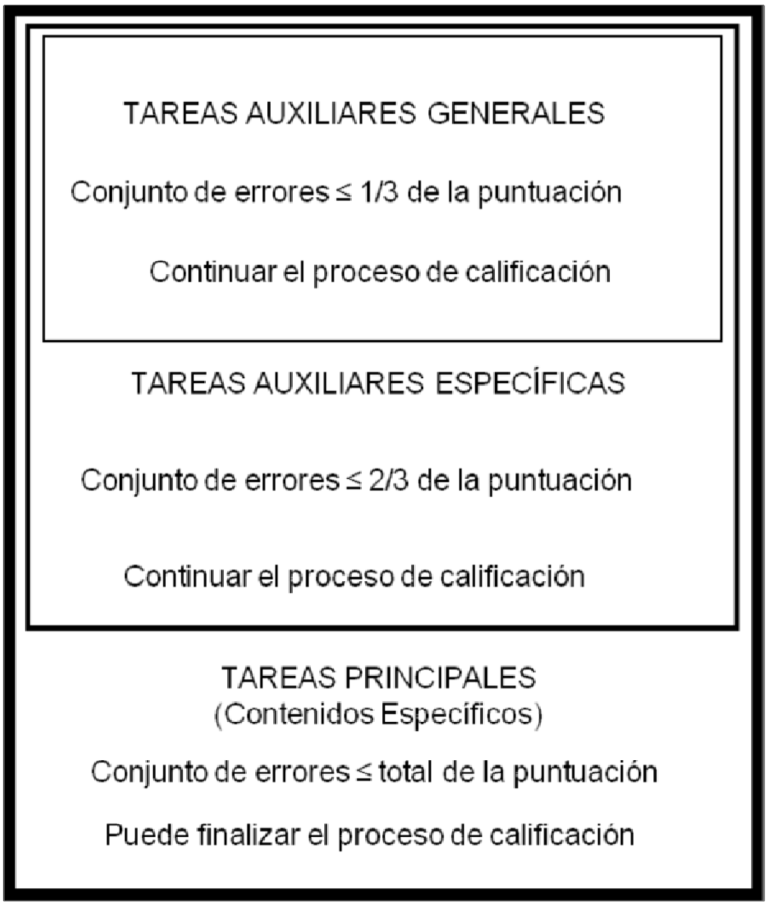


Gráfico 2

a) El conjunto de todos los errores que comete un alumno al realizar tareas auxiliares generales no podrá penalizarse con un valor superior a 1/3 de la puntuación asignada al problema, porque el conjunto de estas tareas no debe ser determinante al valorar la comprensión de los contenidos específicos contemplados en los currículos oficiales. Además, el proceso de calificación debe continuar porque quedan 2/3 de la puntuación por asignar a la respuesta que da el alumno, considerando como correctos los resultados que se hayan obtenido en las tareas auxiliares generales.

b) Los errores cometidos en las tareas auxiliares específicas podrán penalizarse con un máximo de 2/3 de la calificación asignada a la respuesta correcta, porque el objetivo fundamental de la calificación son los contenidos específicos y no las tareas auxiliares específicas. Es más, en este límite de penalización han de contabilizarse tanto los errores en las tareas auxiliares específicas como en las tareas auxiliares generales. Además, el corrector debe continuar calificando el ejercicio tomando como punto de partida los resultados producidos al calcular erróneamente.

c) En este modelo los errores localizados en los contenidos específicos, tanto de conceptos como de procedimientos, pueden penalizarse hasta con el 100% de la puntuación total, porque entendemos que este tipo de contenidos constituye el objetivo principal de la calificación de los alumnos. En el caso de que en la resolución de un problema estén implicados dos o más contenidos específicos, deberían asignarse puntuaciones independientes a cada uno de ellos. Además, el corrector, teniendo en cuenta la importancia de estos errores determinará si continúa con el proceso de calificación o si lo da por concluido.

d) En este modelo se han establecido tres intervalos de igual amplitud para marcar los valores entre los que ubicar las penalizaciones por los errores detectados en las respuestas. Entendemos que nuestras motivaciones no se sustentan en razones científicas irrefutables y en consecuencia, admitimos que, por otras razones diferentes, los intervalos de penalización de errores se modifiquen.

e) En este modelo se deja un amplio margen de actuación a los correctores por cuanto la penalización de los errores se ubica en intervalos de amplitud 1/3 de la puntuación máxima. Se han establecido intervalos de penalización porque, entre otras razones, resulta muy complejo establecer penalizaciones concretas para cada tipo de error, pues el número de casos diferentes que se puede presentar es muy amplio.

Conclusiones

Entendemos la necesidad de investigar sobre las creencias de los profesores acerca de la enseñanza y la evaluación, pues ello permitirá poner de manifiesto las disonancias, las contradicciones y las repercusiones al calificar exámenes de matemáticas (Marcia, 2008). En ese camino aportamos un modelo de penalización de errores que constituye el objetivo de la investigación propuesto en la introducción del trabajo.

La calificación de un problema de matemáticas tiene una doble interpretación: una privada, que solamente conoce el corrector, pues solamente éste sabe las razones que le han llevado a determinar cuál es la puntuación que ha asignado, y otra pública, que asocia la calificación de la respuesta con el grado de conocimiento que tiene el autor de dicha respuesta respecto al contenido matemático que figura explícitamente en el enunciado de la pregunta. Con este modelo pretendemos avanzar hacia una interpretación única de la calificación como la medida del grado de conocimiento que tiene un alumno sobre los contenidos matemáticos que figuran explícitamente en el enunciado del problema.

La utilización de este modelo de penalización de errores limita la subjetividad de la calificación de los exámenes de matemáticas evitando la valoración absoluta de los errores. Los errores no son “graves” o “leves” en sí mismos; su importancia queda asociada a la tarea en la que aparecen y a la relación de ésta tarea con el objetivo principal de la evaluación.

El modelo que hemos propuesto, a diferencia de la práctica actual, limita la penalización de los errores cometidos en tareas auxiliares a una parte de la puntuación máxima. Por tanto, la calificación de un problema informa con mayor fiabilidad sobre los conocimientos que tiene el alumno acerca de los contenidos que figuran en el enunciado del problema.

Este modelo persigue establecer un marco general en el que puedan moverse los correctores de un mismo colectivo de alumnos. Este marco obliga a respetar sus límites, aunque éstos entren en contradicción con las creencias de los correctores. La aceptación de estos límites es condición necesaria para ir reduciendo las diferencias entre las calificaciones de distintos correctores.

El modelo que hemos definido se ha mostrado más discrepante con la actuación de los correctores al penalizar los errores en las tareas auxiliares; y también ha mostrado diferencias, aunque más pequeñas, en la penalización de errores en las tareas auxiliares específicas. Sin embargo, no muestra diferencias significativas en la valoración de errores que surgen en las tareas sobre contenidos específicos.

Este modelo no pretende zanjar la problemática de la calificación de los exámenes de matemáticas. Antes bien, se ha construido como un primer paso en el camino de ajustar las calificaciones de respuestas iguales que son evaluadas por diferentes correctores. Además, la presentación de este modelo pretende abrir un debate entre el profesorado de matemáticas sobre la búsqueda de criterios de corrección universales que hagan lo más objetivo posible el proceso de calificación de los exámenes.

Este modelo es válido para corregir exámenes de matemáticas de cualquier nivel educativo, porque el modelo se sustenta en relacionar los errores con el tipo de tareas en que se producen y porque en todo tipo de problemas se realizan tareas que están más o menos relacionadas con el grado de conocimiento que se quiere calificar.

Agradecimientos

El primer autor recibe apoyo del Proyecto I+D+i EDU2009-12063 del Ministerio de Ciencia e Innovación (MICINN) de España. El segundo autor recibe apoyo del Proyecto MTM2010-19938-C03-03 de la Dirección General de Investigación del MICINN (España).

Bibliografía

BOESEN, J.; LITHNER, J. y PALM, T. (2010). The relation between types of assessment tasks and the mathematical reasoning students use. *Educ. Stud. Math.*, 7(1), pp. 89-105.

CANTÓN, I. y PINO-JUSTE, M. (2011). *Diseño y desarrollo del currículum.* Madrid: Alianza Editorial.

CARRILLO, J. y GUEVARA, F. (1996). Un instrumento para evaluar la resolución de problemas. *Uno*, 3(8), pp. 65-81.

ESCUDERO, T. y BUENO, C. (1994). Examen de selectividad: el estudio de un tribunal paralelo. *Revista de educación*, 304, pp. 281-298.

GALLARDO, J. y GONZÁLEZ, J.L. (2005). Una aproximación operativa al diagnóstico y la evaluación de la comprensión del conocimiento matemático. En Maz, A. et al. (Eds.), *Investigación en educación matemática. Actas del IX Simposio de la Sociedad Española de Investigación en Educación Matemática* (pp. 197-204). Córdoba: SEIEM.

GIL, F.; RICO, L. y FERNÁNDEZ, A. (2002). Concepciones y creencias del profesorado de secundaria sobre evaluación en matemáticas. *Revista de Investigación Educativa*, 20(1), pp. 47-75

GIMÉNEZ, J. (1997). *Evaluación en matemáticas. Una integración de perspectivas.* Madrid: Síntesis.

GOETZ, J.P. Y LECOMPTE, M.D. (1988). *Etnografía y diseño cualitativo en investigación educativa*. Madrid: Morata

GRAU, R.; CUXART, A. y MARTÍ-RECOBER, M. (2002). La calidad en el proceso de corrección de las Pruebas de Acceso a la Universidad: variabilidad y factores. *Revista de Investigación Educativa*, 20(1), pp. 209-223.

GUTIÉRREZ, A. y JAIME, A. (1995). Towards the design of a standard test for the assessment of the students' reasoning in geometry. En Meira, L. et al*., 19th Annual conference of the International Group for the Psychology of Mathematics Education (PME 19)*, 3, pp.11-18.

KAUR, B. y WONG, K.Y. (eds.) (2011). *Assessment in the mathematics classroom. Yearbook 2011, Association of Mathematics Educators.* Hackensack, NJ: World Scientific.

LIAN, L.H.; YEW, W.T. e IDRIS, N. (2010). Superitem test: An alternative assessment tool to assess students' algebraic solving ability. *Int. J. Math. Teach. Learn.*, 5, pp. 1-15.

MARCIA, P. (2008). Creencias de los profesores sobre evaluación y efectos incidentes. *Revista de Pedagogía*, 84(29), pp. 123-144.

ORDÓÑEZ, L. y CONTRERAS, A. (2011). La integral definida en Bachillerato. Restricciones institucionales de las Pruebas de Acceso a la Universidad. En Marín, M. et al. (Eds.), *Investigación en educación matemática XV* (pp. 461-470). Ciudad Real: SEIEM.

RUIZ DE GAUNA, J. (2011). Tesis: La enseñanza de las matemáticas del bachillerato, los libros de texto y las pruebas de acceso a la UPV-EHU (1970-2008). *Enseñanza de las Ciencias*, 29(2), pp. 305-306.

SIERRA, B. y PÉREZ, M. (2007). La relación teoría-práctica: una clave epistemológica de la didáctica. *Revista de Educación*, 342, pp. 553-576.

TORRES-SKOUMAL, M. (2001). Alternative assessment models: assessment through group work and the use of CAS as a self-assessment tool. *Int. J. Comput. Algebra Math. Educ.*, 8(1), pp. 61-83.