

# ¿Cuál es la respuesta correcta?

## Ilustrando las dificultades de calificar en matemáticas

por

ALBERTO ARNAL-BAILERA, JOSÉ MARÍA MUÑOZ-ESCOLANO Y ANTONIO M. OLLER

(Universidad de Zaragoza, Universidad de Zaragoza y Centro Universitario de la Defensa de Zaragoza)

Las PAU constituyen un marco interesante en el que analizar las actuaciones de distintos correctores a la hora de calificar las respuestas de los alumnos con el objetivo de mejorar la fiabilidad de dichas pruebas lo que es deseable, dada su gran importancia administrativa, institucional y social. En este sentido, diversos trabajos de investigación (Gairín, Muñoz y Oller, 2012; 2013) ponen de manifiesto la variabilidad existente cuando varios correctores actúan sobre un mismo examen.

En esta línea, nos planteamos abordar los siguientes objetivos:

- Estudiar la existencia de una (aceptable) fiabilidad global de un buen número de correctores (entendida por una baja variabilidad entre correctores) en sus calificaciones al corregir exactamente una misma respuesta.
- Analizar el papel que juega la experiencia docente o la formación recibida por el corrector respecto a la calificación de las respuestas.

### Diseño del cuestionario

Para alcanzar los objetivos enunciados anteriormente, se diseña un cuestionario donde se solicitaba a 91 profesores de matemáticas (26 en formación, 45 de secundaria en ejercicio y 20 de universidad) calificar respuestas de estudiantes a un mismo problema de localización de puntos críticos análogo a los que aparecen en varios de los exámenes de PAU de Matemáticas II del Bachillerato de Ciencias y Tecnología en Aragón:

Dada la función  $f(x)=x^2/(4-x)$ , calcula sus extremos relativos.

Las tres respuestas de alumnos que se proponían para calificar poseen los siguientes rasgos:

- Los métodos de resolución de las tres respuestas aparecen en libros de texto revisados y los dos primeros (Figuras 1 y 2) son frecuentemente utilizados por estudiantes en las PAU revisadas.
- No presentan ninguna incorrección matemática manifiesta.
- En las tres respuestas se obtiene la solución correcta.
- El nivel de argumentación de las tres respuestas es similar. De hecho, es equiparable al nivel medio de argumentación observado en las PAU revisadas.

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

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

$$x(8-x) = 0$$

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

Derivada segunda

$$f''(x) = \frac{(8-2x)(4-x)^2 - (8x-x^2)2(4-x)(-1)}{(4-x)^4} = \frac{32 - \cancel{8x} - \cancel{8x} + \cancel{8x^2} + \cancel{16x} - \cancel{2x^2}}{(4-x)^3}$$

$$= \frac{32}{(4-x)^3}$$

$$f''(0) = \frac{32}{4^3} > 0 \rightarrow \boxed{(0,0) \text{ m\u00ednimo relativo}}$$

$$f''(8) = \frac{32}{(-4)^3} < 0 \rightarrow \boxed{(8,-16) \text{ m\u00e1ximo relativo}}$$

$$f(8) = \frac{64}{-4} = -16$$

Figura 1. Respuesta según el Método 1

En las figuras 1, 2 y 3 se muestran las tres respuestas incluidas en la versión final del cuestionario.

El método 1 (detección de extremos usando la segunda derivada) y el método 2 (estudio del crecimiento y decrecimiento cerca del punto crítico) son mucho más comunes que el método 3 (aplicación de la definición de máximo y mínimo relativo para clasificar el punto crítico). De hecho, desde el punto de vista de su utilización por parte de los alumnos en sus respuestas en las P.A.U. revisadas, el método 3 está prácticamente ausente y el método 1 es más común que el método 2.

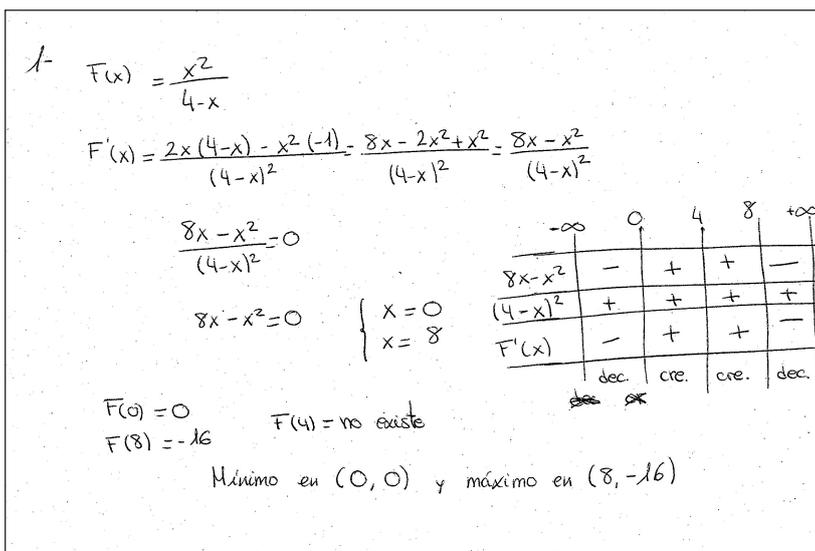


Figura 2. Respuesta según el Método 2

### Resultados

En la tabla 1 se muestran las medias, las desviaciones típicas, las medianas, el intervalo entre el primer cuartil ( $Q_1$ ) y el tercer cuartil ( $Q_3$ ) y el rango estadístico de las calificaciones otorgadas por los 91 profesores que constituyen la muestra para cada uno de los métodos de resolución.

Resulta interesante observar que, conforme la nota media disminuye, aumenta la desviación típica, el rango estadístico total y en el último caso, el rango semi-intercuartílico,  $Q_3 - Q_1$ . Además estos fenómenos se producen conforme el método se vuelve menos estándar.

Desde el punto de vista estadístico, realizando los test correspondientes, se puede afirmar con un nivel de confianza del 99% que la calificación media del método 1 es mayor que la calificación media de los métodos 2 y 3 y que, a su vez, la calificación media del método 2 es mayor que la del método 3.

Si analizamos los datos por colectivos (Tabla 2), observamos que tanto entre los profesores en formación, como entre los profesores de Secundaria se dan fenómenos similares a los señalados en el análisis global. Las medias decrecen y aumenta la desviación típica. Las diferencias entre las medias siguen siendo significativas al 99% en el caso de los profesores de Secundaria y al 95% en el caso de los profesores en formación. Sin embargo, entre el profesorado de Universidad se observa que los métodos 1 y 2 reciben un tratamiento similar. De hecho no hay diferencias

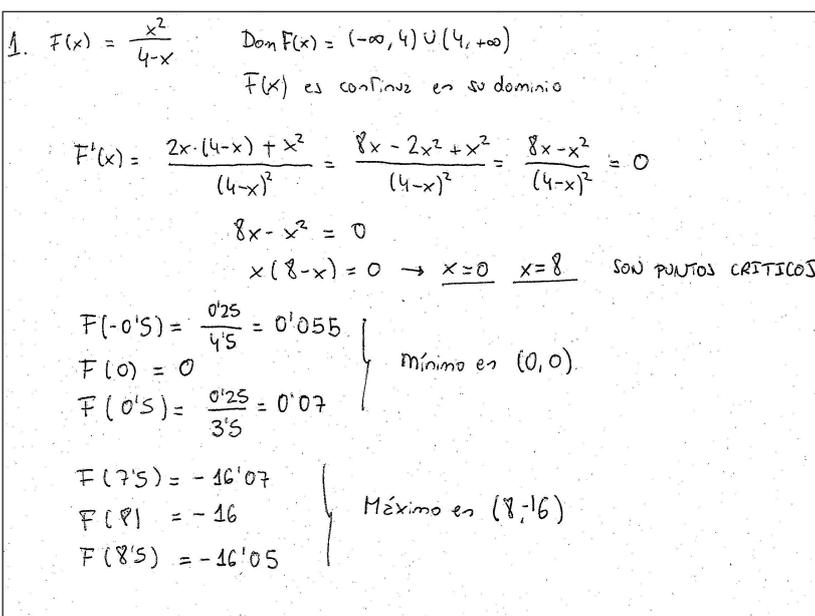


Figura 3. Respuesta según el Método 3

	Media	Desviación típica	Mediana	$[Q_1, Q_3]$	Rango
Método 1	9,53	0,84	10	[9,10]	4
Método 2	9,06	1,53	10	[9,10]	6,5
Método 3	7,87	2,14	8,5	[6,10]	8

Tabla 1. Resultados de las calificaciones de las tres respuestas

estadísticamente significativas entre las medias y la desviación típica es casi idéntica. Sí que se observa (con un nivel de confianza del 99%) que las medias de los métodos 1 y 2 son mayores que la del método 3.

La dispersión de las distribuciones de las muestras es considerable. No obstante, respecto a los métodos 1 y 2, esta apreciación se puede matizar en cierta manera al observar que los rangos intercuartílicos en todos los colectivos son menores o iguales que 1. Sin embargo, respecto al método 3, sí que se aprecia gran dispersión ya que los rangos intercuartílicos son mayores o iguales que 3 en todos los colectivos.

También se observa en algunos colectivos y métodos (principalmente en los estudiantes del Máster para los métodos 2 y 3) que las distribuciones de los datos son muy poco simétricas presentando un sesgo a la izquierda ya que existe una gran diferencia entre media y mediana y que, en el caso del método 2, llega incluso a situarse fuera del intervalo  $[Q_1, Q_3]$ .

No se observan diferencias estadísticamente significativas entre las calificaciones otorgadas por los distintos colectivos para el método 1. Respecto al método 2, se puede afirmar (nivel de confianza del 99%) que el profesorado de Universidad otorga una calificación mayor que los profesores en formación. Aunque la nota media otorgada por el profesorado de Secundaria también es mayor que la de los profesores en formación, esta diferencia no resulta ser significativa. No hay diferencias entre el profesorado de Secundaria y el de Universidad. En cuanto al método 3, el profesorado de Universidad otorga calificaciones mayores (nivel de confianza del 95%) que los otros dos colectivos. Los profesores en formación otorgan mayores calificaciones (nivel de confianza del 90%) que el profesorado de Secundaria.

## Conclusiones

La tarea de corregir y calificar problemas dista mucho de ser sencilla y requiere de una profunda reflexión de ámbito matemático sobre los contenidos (conceptos, procedimientos) que se preguntan y sobre las distintas estrategias que pueden emplearse para su resolución. En este estudio, comprobamos que incluso para aquellas tareas sobre las que, en principio, existe un amplio consenso de que están correctamente resueltas, éste no es absoluto. La respuesta correspondiente al método 1 sí que presenta un amplio consenso en los tres colectivos que, sin embargo, no es tan mayoritario en el caso de las otras dos respuestas.

Los resultados por colectivos parecen indicar que la pertenencia a un colectivo u otro influye a la hora de calificar las tareas siendo el colectivo de profesores en formación el que presenta una mayor dispersión de calificaciones, por encima de los años de experiencia. Esta conclusión apunta al hecho de que es necesaria formación específica sobre este aspecto.

Los resultados del estudio indican que la calificación otorgada por gran parte de los correctores es mayor cuando se utilizan métodos más cercanos al modo de hacer y a la experiencia docente del corrector.

## Referencias

- GAIRÍN, J. M., J. M. MUÑOZ y A. M. OLLER (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*, SEIEM, Jaén, SEIEM
- (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*, n.º 32(2), 27-50.

		Media	Desviación típica	Mediana	$[Q_1, Q_3]$	Rango
Máster	Método 1	9,58	0,98	10	[9,63; 10]	4
	Método 2	8,65	2,15	9,75	[9, 10]	6,5
	Método 3	8,02	2,46	9,25	[6, 10]	8
Secundaria	Método 1	9,51	0,86	10	[9, 10]	4
	Método 2	9,03	1,33	9,5	[9, 10]	5
	Método 3	7,48	1,98	8	[6, 9]	6
Universidad	Método 1	9,51	0,59	9,88	[9, 10]	2
	Método 2	9,64	0,60	10	[9,38; 10]	2
	Método 3	8,55	1,93	9,5	[7, 10]	6

Tabla 2. Resultados de las calificaciones de las tres respuestas por colectivo