





Influencia de plantear unas preguntas de respuesta breve en papel o en la pantalla sobre la nota obtenida

Effect of asking short-answer questions in paper or on-screen on the achieved score

Fidel Salas Vicente^a, Rut Benavente Martínez^b

^a Dpto. Ingeniería Mecánica y de Materiales. Universitat Politècnica de València. fisavi@dimm.upv.es 

^a Dpto. Ingeniería Mecánica y de Materiales. Universitat Politècnica de València. rutbmr@upvnet.upv.es 

How to cite: Salas Vicente, F. y Benavente Martínez, R. (2024). Influencia de plantear unas preguntas de respuesta breve en papel o en la pantalla sobre la nota obtenida. En libro de actas: *X Congreso de Innovación Educativa y Docencia en Red*. Valencia, 11 - 12 de julio de 2024. Doi: <https://doi.org/10.4995/INRED2024.2024.18392>

Abstract

The evaluation method known as short-answer questions is based on questions about important concepts of the studied matter that can be answered in a short time. Its use, that can be both summative and formative, combines a duration similar to that of a test with the requirement for the student to elaborate the answer.

In this study two ways of posing this kind of questions have been compared during a course on materials technology: On the screen, one after another, and without the possibility of going backwards or on paper. Although the two ways of asking the questions are clearly different, the removal of the influence of other factors indicates that either of them can be chosen without noticeably affecting the grades

Keywords: *short-answer questions; assessment; score*

Resumen

El método de evaluación basado en preguntas de respuesta breve consiste en cuestiones sobre conceptos importantes de una materia que pueden responderse en un tiempo breve. Su uso, que puede ser tanto en evaluación formativa como sumativa, combina una rapidez similar a la de un test con la necesidad de que sea el propio alumno quien elabore la respuesta.

En este trabajo se han comparado dos formas distintas de plantear este tipo de preguntas en una asignatura sobre tecnología de materiales: En la pantalla de clase, una tras otra y sin posibilidad de retroceder, o en papel. Aunque ambas formas de plantear las preguntas son claramente distintas, la eliminación de la influencia de otros factores indica que se puede optar por cualquiera de ellas sin que las calificaciones se vean prácticamente afectadas

Palabras clave: *Preguntas del minuto; evaluación; notas*

1. Introducción

Las preguntas de respuesta breve constituyen un tipo de evaluación basada en preguntas cuya respuesta consiste en unas pocas líneas e incluso en una sola palabra. Este tipo de preguntas se suele emplear como refuerzo y comprobación de la comprensión de la materia explicada a lo largo de una clase si se plantean al acabar ésta, e incluso como sistema para obtener información sobre la visión del alumno de la materia impartida (Stowe, 2010), dando origen en este caso a la técnica conocida en inglés como “one-minute paper” (Angelo & Cross, 1993), consistente en 2 preguntas destinadas a conocer si los alumnos habían asimilado los conceptos más importantes de la clase y cuales eran los puntos que les resultaban de más difícil comprensión. Estas 2 preguntas son “¿Qué es lo más importante que has aprendido hoy?” y “¿Qué cuestiones han quedado más confusas?”

El uso de las preguntas de respuesta corta, aunque asociado en ocasiones a la memorización, ha dado evidencias de que mejora las habilidades cognitivas de orden superior (Chamberlain et al., 2011) cuando se usa para dividir conceptos complejos en otros más manejables. Además, el uso de pequeñas evaluaciones usando estas preguntas fuerza a los alumnos a no dejar el estudio de la asignatura para los días previos al examen (Cole & Spence, 2012), aunque hay que tener en cuenta que no está claro que una evaluación continua estricta vaya a mejorar el aprendizaje (Reboredo, 2017).

Naturalmente, éste tipo de evaluación no es único ni el más usado, correspondiendo ese puesto a los tests de respuesta múltiple, con los que podrían intercambiarse en muchas ocasiones. Comparado con los tests, las preguntas de respuesta corta presentan tanto ventajas como desventajas, sobre todo relacionadas con el tiempo que requieren para su calificación (Stead, 2005), ya que la de los tests puede ser automatizada con facilidad. Como ventaja se puede citar una mayor fiabilidad y capacidad de discriminación (van Wijk et al., 2023) o una mayor mejora del aprendizaje (Greving & Richter, 2022) ya que, al no facilitarse la respuesta, se requiere un mayor esfuerzo mental.

Es en base a estas características, y a la obligatoriedad de evaluar al alumno usando distintos métodos, que las evaluaciones mediante el uso de preguntas de respuesta corta se introdujo hace años en varias asignaturas del área de materiales del Departamento de Ingeniería Mecánica y de Materiales, pero siempre usando una única forma de plantear las preguntas según la preferencia del responsable de la asignatura: O se presentaban una tras otra en la pantalla del aula o se daban escritas en una hoja. Es en base a la duda que surgió durante el primer cuatrimestre del curso 23/24 respecto a cual sería la mejor manera de plantear las preguntas que se decidió comparar ambas formas y evaluarlas desde el punto de vista del resultado académico.

En un primer momento se puede pensar que dar las preguntas en papel supondrá una ventaja para el alumno puesto que tiene todas las preguntas disponibles desde el primer momento y puede empezar por contestar aquellas que le resultan más sencillas, con lo que reduce el nerviosismo asociado a toda evaluación. Además, puede volver a las primeras preguntas en cualquier momento, cosa más complicada si los enunciados se muestran en la pantalla. No obstante, es algo sobre lo que es preciso tener datos reales para poder decidir. En este trabajo se presenta como se llevó a cabo el estudio y cuales fueron los resultados.

2. Objetivos

El objetivo de este estudio fue comparar, en cuanto a las calificaciones obtenidas por los alumnos, dos formas de plantear las preguntas de respuesta corta a lo largo de curso, para lo cual se llevó a cabo un

análisis de la influencia sobre las notas de otros factores con el fin de aislar y cuantificar la influencia de la manera de plantear las preguntas.

3. Desarrollo de la innovación

El estudio se llevó a cabo en la asignatura “Ciencia de materiales II” de la titulación de Ingeniería Mecánica de la Escuela Técnica Superior de Diseño de la Politécnica de València. Esta asignatura contó en el curso 23/24 con un total de 198 matriculados, divididos en 3 grupos de teoría (1, 2 y 3).

Durante el curso se plantearon 4 evaluaciones sumativas basadas en preguntas del minuto, con 10 preguntas cada vez, dos actos de evaluación antes del primer examen parcial y otras dos después de éste. Las evaluaciones se hicieron en el aula donde se impartió clase a cada grupo, justo a su comienzo y se avisó de las pruebas con un mínimo de una semana de antelación en todos los casos.

Teniendo en cuenta que cada grupo tenía que hacer cuatro pruebas a lo largo del curso se disponía de 12 pruebas en total con las que plantear un diseño estadístico de experimentos 2^k . Se podría haber incrementado este número, pero se estimó que una cantidad mayor de evaluaciones sería excesiva para los alumnos, que tienen otras materias y trabajos que realizar. Incluso aunque fuese la única asignatura, si se emplea al finalizar cada clase o cada uno de los temas en los que se divide el curso los alumnos llegarían con facilidad al hastío y a un desinterés por la actividad (Stead, 2005).

12 pruebas son insuficientes para estudiar la influencia de 4 variables puesto que $2^4=16$. Así pues, se decidió estudiar la influencia del modo en el que se plantean las preguntas: en papel o en la pantalla, y la de otros dos factores que también pueden hacer variar la nota obtenida y cuya influencia debe tenerse en cuenta si se quiere aislar la del modo de presentar las preguntas. Estas 2 variables fueron:

- La clase, ya que se sabe que los alumnos con mejor curriculum tienen tendencia a elegir primero los horarios de mañana con mejor horario (grupos 1 y 2), quedando los de tarde con mayor proporción de alumnos con calificaciones bajas (grupo 3).
- Pruebas realizadas antes o después del primer examen parcial, ya que los resultados del primer examen pueden influir en la dedicación de los alumnos al estudio de la asignatura.

Evidentemente, hay otros factores que fueron considerados a la hora de plantear la investigación y que fueron desestimados, aunque no olvidados pues dentro de lo posible se buscó minimizar su influencia. Estos fueron:

- La dificultad de la materia evaluada en cada caso. Este es el primer año que los autores de la investigación se han hecho cargo de la asignatura, en la que se han introducido importantes cambios, tanto en la forma de impartirla y de evaluarla como en la materia impartida, por lo que no se dispone de resultados válidos de años anteriores (tampoco se han considerado los de asignaturas similares de otras titulaciones). Evidentemente, los profesores disponen de una opinión al respecto y se ha dedicado más tiempo y esfuerzo a aquellos apartados del temario que se han considerado más extensos y profundos para tratar de equilibrar este aspecto.

- El profesor. En la asignatura han impartido clases dos docentes y se sabe que la calidad de un profesor es un factor importante en el proceso de aprendizaje. No obstante, en este caso ambos usan las mismas transparencias y documentación y ambos tienen una buena valoración en las encuestas de los alumnos y un estilo similar a la hora de impartir la clase, por lo que se estimó que éste no sería un factor tan importante como cabría esperar en otros casos.

También se podría considerar que, como cada profesor era el encargado de poner las preguntas a su grupo, y también de corregirlas, éste fuese un factor también de gran importancia. Se decidió compensar este problema mediante la revisión de todas las preguntas propuestas por parte de los 2 profesores, de forma que se asegurase, dentro de lo posible, que todas las preguntas tuviesen un nivel similar. En cuanto a la calificación de las preguntas, cada profesor corrigió una muestra aleatoria de los exámenes del otro sin saber qué calificación habían recibido. La comparación de las notas indicó que ambos usaban criterios muy similares a la hora de la corrección.

Un último factor dificultaba el estudio de la influencia del docente y es que no se consideró pertinente que el profesor de un grupo cambiase a mitad de curso.

Teniendo en cuenta el rango de parámetros que se pretendía estudiar y el número de pruebas que se harían en total a lo largo del curso se optó por el diseño de experimentos mostrado en la tabla 1:

Tabla 1. *Pruebas realizadas*

Prueba nº	Grupo	Parcial	Modo
1	1	1º	Pantalla
2	1	1º	Papel
3	1	2º	Pantalla
4	1	2º	Papel
5	2 y 3	1º	Pantalla
6	2 y 3	1º	Papel
7	2 y 3	2º	Pantalla
8	2 y 3	2º	Papel

En cuanto a como se realizaron las pruebas de evaluación, su desarrollo se describe a continuación:

- Preguntas en pantalla: Se presentaron una tras otra usando el proyector de clase. Una vez se había leído la pregunta se esperaba hasta que se observaba que la mayor parte de los alumnos hubiese terminado de escribir. En ese momento se pasaba a la siguiente. Es preciso remarcar que todas las preguntas podían contestarse con una o dos frases y, en ocasiones, hasta con una sola palabra. Al comienzo de la prueba se avisaba de esto y se indicaba que no se extendiesen en la respuesta, indicaciones que no siempre se seguían.
- Preguntas en papel: Se entregaba un hoja boca abajo con las 10 preguntas y espacio para las respuestas. El terminar el reparto los alumnos podían dar la vuelta a la hoja y comenzar. En estos casos se indicaba que el tiempo estaba limitado a 10 minutos y se visualizaba la hora en la pantalla del aula. No obstante, si el profesor veía que los alumnos seguían escribiendo se dejaba algún minuto más.

El control del tiempo que se deja para responder las preguntas reviste gran importancia, tanto para el alumno como para el profesor. Un tiempo insuficiente hará que algunos alumnos no sean capaces de contestar a todas las preguntas con el nivel de desarrollo que desean mientras que un tiempo excesivo podría hacer que llegasen a dudar y a cambiar la respuesta dada inicialmente por otra incorrecta, perdiendo puntos (Brothen & Wambach, 2004; Salas & Vicente, 2019). Además, una vez los alumnos han terminado de contestar las preguntas cuya respuesta conocen se hacen evidentes las miradas hacia el compañero de al lado, en un intento de encontrar la inspiración que no llega.

4. Resultados

En la Tabla 2 se muestra el número de alumnos que participó en cada una de las pruebas y la nota media obtenida en cada una de ellas.

Tabla 2. Resultados de las pruebas evaluativas. Entre paréntesis, el número de alumnos presentado a cada prueba

	Prueba nº 1	Prueba nº 2	Prueba nº 3	Prueba nº 4
Grupo 1	Pn=5,801 (52)	Pp=7,170 (53)	Pp=6,417 (52)	Pn=6,611 (52)
Grupo 2	Pn=5,574 (53)	Pp=6,622 (52)	Pp=6,164 (51)	Pn=6,788 (52)
Grupo 3	Pp=4,381 (70)	Pn=5,549 (71)	Pp=4,759 (70)	Pn=5,531 (75)

* Pn = Pantalla; Pp = Papel

Antes de pasar a hacer una análisis de la influencia de los distintos factores considerados en el diseño 2³ se puede extraer información mediante el uso de los cuadrados latinos suponiendo que las variables consideradas son independientes. Aunque el uso de cuadrados latinos de tamaño 2x2 entraña muchas limitaciones debido a que el grado de libertad de los residuos es 0 y eso impide tener en cuenta la variabilidad de los datos, sigue siendo una herramienta que puede aportar datos sobre el efecto de como se plantean las preguntas. Por desgracia, en este caso no fue posible plantear un tercer modo de presentar las preguntas para diseñar un cuadrado 3x3.

Con los datos disponibles se pueden plantear 6 cuadros 2x2, tal y como muestra la Figura 1. La media de las notas obtenidas en cada caso se muestra en la Figura 2.

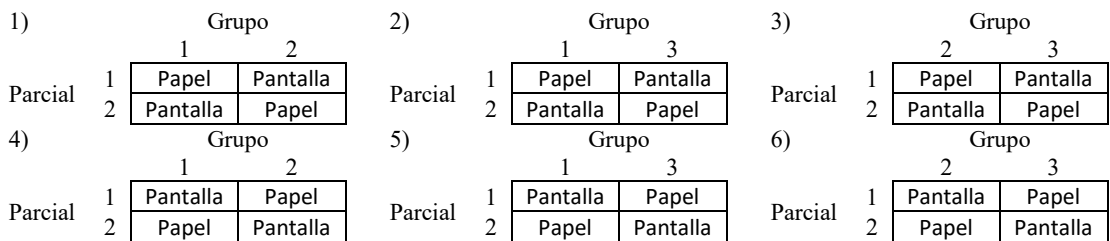


Figura 1. Cuadrados latinos planteados

Figura 2. Resultados obtenidos para los cuadrados latinos planteados

El modelo aditivo de los cuadrados latinos supone que el valor obtenido es

$$y_{ijk} = \eta + \beta_i + \gamma_j + \zeta_k + \varepsilon_{ijk} \quad (1)$$

Donde η es el valor medio, β_i son los efectos de fila (parcial 1 o 2), γ_j son los efectos de columna (uno u otro grupo), ζ_k es el efecto de dar las preguntas en papel o en pantalla y ε_{ijk} son los residuos o diferencias entre el valor medido y el calculado con la fórmula. Como en este caso se trabaja con cuadrados 2x2, los residuos son siempre cero. La media y el efecto de las distintas variables sobre la media son los siguientes:

Media cuadrado 1: 6,037

Parcial		Grupo		Modo	
1º	2º	1	2	Papel	Pantalla
-0,350	0,350	0,169	-0,169	-0,055	0,055

Media cuadrado 2: 5,388

Parcial		Grupo		Modo	
1º	2º	1	3	Papel	Pantalla
-0,297	0,297	0,818	-0,818	-0,108	0,108

Media cuadrado 3: 5,638

Parcial		Grupo		Modo	
1º	2º	2	3	Papel	Pantalla
0,136	-0,136	1,067	-1,067	0,053	-0,053

Media cuadrado 4: 6,749

Parcial		Grupo		Modo	
1º	2º	1	2	Papel	Pantalla
0,147	-0,147	-0,044	0,044	-0,229	0,229

Media cuadrado 5: 6,167

Parcial		Grupo		Modo	
1º	2º	1	3	Papel	Pantalla
0,193	-0,193	0,627	-0,627	0,184	-0,184

Media cuadrado 6: 5,704

Parcial		Grupo		Modo	
1º	2º	2	3	Papel	Pantalla
-0,143	0,143	0,164	-0,164	-0,152	0,152

Aunque no sea muy ortodoxo, ya que las diferencias en cuanto al rendimiento académico de cada grupo no son las mismas entre los distintos grupos (en los parciales las medias de cada grupo fueron de 6,06 para el grupo 1, de 5,00 para el grupo 2 y de 4,30 para el grupo 3), con el fin de obtener datos orientativos, se puede suponer que así es y usar la distribución t de Student para estimar si la influencia media de cada uno de los parámetros considerados sobre la nota media se puede considerar distinta de 0 con una confianza aceptable.

Calculando las medias y desviaciones típicas muestrales y planteando la hipótesis de que el efecto de cada parámetro es nulo se llega a que:

- El que la prueba se haga antes o después del segundo parcial tiene una influencia media sobre la nota media de -0,098 con un nivel de significación para Pr ($t > 0,975$) de 0,147, por lo que no se puede afirmar con total seguridad que el parcial influya sobre la nota.
- El grupo al que pertenece el alumno tiene una influencia media sobre la nota media de 0,467 con un nivel de significación para Pr ($t > 2,393$) de 0,042, por lo que su influencia queda demostrada con una confianza de más del 95%.
- El que las preguntas se planteen en pantalla o en papel tiene una influencia media sobre la nota media de 0,015 con un nivel de significación para Pr ($t > 0,207$) superior a 0,4, por lo que no se puede afirmar

que influya sobre la nota. Es más, dado el bajo nivel de influencia media se podría descartar esa influencia, pues, de existir, sería de 1,5 centésimas como media.

Hecho este análisis usando los cuadrados latinos y la t de Student se puede pasar a hacer un análisis factorial 2³. Dada la cantidad de datos disponibles se ha decidido combinarlos para realizar 3 estudios distintos igual que se ha hecho con los cuadrados latinos, seleccionando como niveles para la variable “grupo” las combinaciones grupo 1-grupo 2, grupo 1-grupo 3 y grupo 2-grupo 3. En las Tabla 3 se muestran los resultados obtenidos de los análisis factoriales (Grupo A indica que es el de mayor nota media en la totalidad del curso).

Tabla 3. Efectos estimados en los 3 diseños factoriales extraídos de los datos

	Grupos 1 y 2	Grupos 1 y 3	Grupos 2 y 3	Media
Media:	6,393	5,778	5,671	5,947
Grupo A:	0,213	1,444	1,231	0,963
Papel:	-0,284	-0,292	0,205	-0,124
Parcial 1:	-0,203	-0,104	-0,279	-0,195
Grupo A x Papel:	-0,497	-0,489	0,007	-0,326
Grupo A x Parcial 1:	0,175	0,076	-0,099	0,050
Papel x Parcial 1:	0,124	0,191	0,903	0,406
Grupo A x Papel x Parcial 1:	-0,712	-0,779	-0,067	-0,519

Como puede verse, igual que en el caso anterior, destaca la influencia de la concentración de buenos alumnos en algunos de los grupos, en este caso los grupos 1 y 2, algo que no debe extrañar en ningún caso. A partir de aquí conviene destacar la importancia de los efectos de las interacciones entre los factores, incluso las de tercer nivel, algo que puede invalidar los resultados obtenidos mediante el análisis de los cuadrados latinos. No obstante, dada la magnitud de los valores obtenidos y la dificultad existente a la hora de explicar estas interacciones, los autores han decidido no descartar los resultados obtenidos mediante los cuadrados y presentarlos junto con los del análisis factorial, ya que los resultados del análisis podrían deberse también a alguno de los factores no considerados y que no se ha podido minimizar los suficiente.

Si se usa la t de Student para estudiar la significación del efecto de cada factor e interacción planteando la hipótesis de una influencia nula se obtiene (ver Tabla 4) que tienen influencia con una confianza de más del 90% son el grupo, el parcial y la interacción “Grupo Axpapelxparcial”

Tabla 4. Significación de los factores y sus interacciones en la nota obtenida

	t ₀	Significación
Grupo A:	2,534	0,069
Papel:	-0,752	0,268
Parcial 1:	-3,866	0,033
Grupo A x Papel:	-1,956	0,100
Grupo A x Parcial 1:	0,631	0,302
Papel x Parcial 1:	1,629	0,136
Grupo A x Papel x Parcial 1:	-2,287	0,081

En definitiva, el análisis factorial parece confirmar la escasa influencia que tiene el que las preguntas se planteen en papel o en la pantalla comparando con otros factores y que, en cualquier caso parece que plantear las preguntas en papel lleva a una ligera disminución de la nota.

5. Conclusiones

Los resultados obtenidos en este estudio, aunque con notables limitaciones, han servido como evidencia de que plantear las preguntas de respuesta corta en la pantalla o en papel no tiene una gran influencia sobre la calificación obtenida por los alumnos y de que, en todo caso, plantearlas en papel lleva a una ligera disminución de la nota.

Con el esfuerzo por aislar la influencia de la forma de presentar las preguntas se ha podido ver también que el factor más influyente de lejos sobre la nota es, lógicamente, la nota media del grupo. También se detecta una ligera mejora de la nota en las pruebas del segundo parcial, debido tanto a que los alumnos conocen mejor las pruebas evaluativas como, posiblemente, a que aumenta su esfuerzo.

Aún así, estos resultados deben tomarse con prudencia y confirmarse con nuevos datos.

Referencias

- Angelo, T. A., & Cross, K. P. (1993). *Classroom Assessment Techniques: A Handbook for College Teachers* (2ª). Jossey Bass Publisher.
- Brothen, T., & Wambach, C. (2004). The Value of Time Limits on Internet Quizzes. *Teaching of Psychology*, 31(1), 4–4. https://doi.org/10.1207/s15328023top3101_1
- Chamberlain, C., Button, A., Dison, L., Grenville, S., & Delmont, E. (2011). The role of short answer questions in developing higher-order thinking. *Per Linguam*, 20(2). <https://doi.org/10.5785/20-2-83>
- Cole, J. S., & Spence, S. W. T. (2012). Using continuous assessment to promote student engagement in a large class. *European Journal of Engineering Education*, 37(5), 508–525. <https://doi.org/10.1080/03043797.2012.719002>
- Greving, S., & Richter, T. (2022). Practicing retrieval in university teaching: short-answer questions are beneficial, whereas multiple-choice questions are not. *Journal of Cognitive Psychology*, 34(5), 657–674. <https://doi.org/10.1080/20445911.2022.2085281>
- Reboredo, J. C. (2017). Do continuous assessment results affect final exam outcomes? Evidence from a microeconomics course. *Multidisciplinary Journal for Education, Social and Technological Sciences*, 4(1), 88. <https://doi.org/10.4995/muse.2017.6548>
- Salas, F., & Vicente, A. (2019). Influencia del tiempo disponible y usado en un examen online sobre la nota obtenida. *XXVII Congreso Universitario de Innovación Educativa En Las Enseñanzas Técnicas*, 1–9.
- Stead, D. R. (2005). A review of the one-minute paper. *Active Learning in Higher Education*, 6(2), 118–131. <https://doi.org/10.1177/1469787405054237>
- Stowe, K. (2010). A quick argument for active learning: The effectiveness of one-minute papers. *Journal for Economic Educators*, 10(1), 33–39.
- van Wijk, E. V., Janse, R. J., Ruijter, B. N., Rohling, J. H. T., van der Kraan, J., Crobach, S., Jonge, M. de, Beaufort, A. J. de, Dekker, F. W., & Langers, A. M. J. (2023). Use of very short answer questions compared to multiple choice questions in undergraduate medical students: An external validation study. *PLOS ONE*, 18(7), e0288558. <https://doi.org/10.1371/journal.pone.0288558>