

# VALORACIÓN DE LA INCLUSIÓN DE REVISIONES ADICIONALES EN LA CONFECCIÓN DE PRUEBAS EVALUATORIAS CON PREGUNTAS DE OPCIÓN MÚLTIPLE

García Sánchez, Pablo César<sup>1</sup>, Álvarez Harris, Sara<sup>2</sup>

1: Departamento de Fisioterapia  
Facultad de Ciencias de la Salud  
UEM  
C/Tajo, s/n 28670 Villaviciosa de Odón, Madrid  
e-mail: pablocesar.garcia@uem.es, web: <http://www.uem.es/titulacion/grado-en-fisioterapia>

2: Departamento de Fisioterapia  
Facultad de Ciencias de la Salud  
UEM  
C/Tajo, s/n 28670 Villaviciosa de Odón, Madrid  
e-mail: sara.alvarez@uem.es, web: <http://www.uem.es/titulacion/grado-en-fisioterapia>

**Resumen.** *En el presente estudio se valora la eficacia de añadir una revisión simple o una doble revisión en la confección exámenes de Preguntas de Opción Múltiple (POM) en dos asignaturas del Grado de Fisioterapia. El objetivo es mejorar la calidad del examen que le llega al alumno como herramienta de valoración y plantear acciones de mejora futuras en la confección de este tipo de prueba.*

*Para ello se ha seleccionado dos asignaturas que emplean pruebas con POM en su evaluación. En una asignatura se ha realizado una revisión simple por parte de un profesor ajeno a la asignatura. En la segunda, paralelamente a esta revisión se ha añadido otra revisión por parte de un profesor de la asignatura que no había confeccionado previamente ninguna pregunta. Posteriormente los autores originales del examen han valorado cuáles de los cambios propuestos por los revisores se incorporaban al documento definitivo*

*Tras la realización de este estudio se detectó que había una ratio elevada de fallos en la confección inicial de las pruebas siendo el tipo de fallos detectados por el revisor externo e interno diferentes. Ambas revisiones RS y DR han mejorado la calidad del examen final. De todo esto se puede concluir que hace falta establecer un plan de mejora en la confección inicial de las POM.*

**Palabras clave:** Preguntas de Opción Múltiple, Tests, Revisión, Evaluación

## 1. INTRODUCCIÓN

Existe una necesidad de evaluar conocimientos, competencia y habilidades de distintas maneras en el entorno universitario. Los exámenes tipo test o exámenes de Preguntas de Opción Múltiple (POM) son uno de los métodos más utilizados sobre todo en grupos muy numerosos, y calificados con asiduidad como “neutrales”, “justos” y “eficaces” (Haladyna 1993). La preparación de POM es una de las tareas habituales a las que se

enfrenta el profesor universitario durante su trabajo como docente (Palmer 2007; Martínez 2009)

Por otro lado, desde hace muchos años este método de evaluación ha sido considerado muy controvertido entre los pedagogos universitarios (Haladyna 2002). De la misma manera, algunos autores consideran que puede que no sea la mejor manera de medir algunas competencias y que la forma de confeccionar las mismas puede encerrar en sí misma muchas dificultades para la resolución de las tareas que se plantean encontrándose en la literatura decenas de guías sobre cómo confeccionar las mismas (McCoubrie 2004; Roediguer 2005; Palmer 2007).

Sin embargo los autores de este estudio no hemos encontrado ninguna recomendación sobre la revisión de estas preguntas una vez han sido confeccionadas por los autores.

Con el objetivo de minimizar nuestros errores en el examen al que se enfrentan los alumnos decidimos llevar a cabo un estudio con dos asignaturas en el que probamos dos tipos de revisiones adicionales diferentes en la elaboración POM del tipo "convencional" según la clasificación de Haladyna (2002).

## **2. MÉTODOS**

Hasta el momento de iniciarse este estudio, la confección de preguntas se realizaba a través de una serie de criterios no formales empleados a través de la experiencia de los profesores y recomendados genéricamente en la literatura y en cursos de formación.

La programación sobre el proceso de confección y revisión de las POM era el siguiente:

- 1º. Distribución de las preguntas del temario entre el profesorado
- 2º. Elaboración individual de las preguntas
- 3º. Revisión global del examen en conjunto por los autores de las mismas

Para la realización de este estudio, se llevó a cabo una revisión de la literatura, adoptando las 31 recomendaciones de consenso de la taxonomía que aportan Haladyna et al. en una revisión de guías y libros de texto publicada en 2002; y añadiendo los requerimientos formales de maquetación y formato utilizados en nuestros exámenes de manera habitual. (Anexo I)

En cuanto a la programación del proceso de confección y revisión de las POM en el estudio propusimos las siguientes fases que se añadirían al finalizar las 3 habituales:

- Grupo A "Revisión Simple" (RS):
  - 4º. Revisión por parte de un profesor "externo" que no imparte la asignatura
  - 5º. Revisión de las recomendaciones/no conformidades indicadas por el profesor "externo" que no imparte la asignatura
- Grupo B "Doble Revisión" (DR):
  - 4º. Revisión en paralelo por parte de un profesor "externo" que no imparte la asignatura
  - 5º. Revisión en paralelo por parte de un profesor que imparte la asignatura pero que no ha confeccionado las preguntas
  - 6º. Revisión de las recomendaciones/no conformidades indicadas por el profesor "externo" que no imparte la asignatura

### **2.1. Participantes**

Los participantes en la elaboración de las POM son 15 profesores del Departamento de Fisioterapia, 9 mujeres y 6 hombres, con una edad media de 35,7 años, una experiencia

media como fisioterapeuta de 12,3 años y 4,1 años de media como profesores de fisioterapia a nivel universitario. Los datos demográficos de los profesores que confeccionaron las preguntas están reflejados en la tabla 1 junto con la desviación estándar. Los revisores “externos” y “no externos” fueron informados sobre la forma de realizar las correcciones y rellenar los formularios.

El grupo de la RS estuvo compuesto por 9 profesores y 2 revisores “externos”. Se realizaron y revisaron 20 preguntas en cada uno de los dos exámenes de la asignatura de primer curso Terapia Manual Básica. (n=40)

El grupo de la DR estuvo compuesto por 7 profesores, de los cuáles 4 fueron a su vez revisores de otros tantos exámenes que ellos no habían confeccionado y 4 revisores “externos”. Se realizaron y revisaron 23 preguntas en cada uno de los cuatro exámenes de la asignatura de segundo curso Técnicas Especiales en Fisioterapia II(n=92)

	Media	SD	Rango
Edad (a)	35,7	7,7	58-28
Experiencia como fisioterapeuta (a)	12,3	6,1	31-7
Experiencia como profesor universitario (a)	4,1	5,7	17-0,5

*Tabla 1.*Datos profesores participantes

## 2.2. Revisiones

Durante las revisiones, se apuntan las no conformidades en la hoja del examen y se anotan su cantidad bajo 15 apartados agrupados en 3 categorías que pueden observarse en la tabla 1.

Redacción y formulación	1	Incomprensible. No quede clara la intención de lo que se pide
	2	Precisión semántica, necesidad de ser más específico
	3	Faltas de ortografía
	4	Dobles negaciones
	5	Utilización excesiva de acrónimos
Contenido	6	No adecuado al temario, explicaciones, asignatura
	7	No adecuado al nivel de la pregunta
	8	Una respuesta/pregunta responda a otras del examen
	9	Pregunta mal contestada
	10	Más de una opción o ninguna
	11	Respuestas no útiles tipo: “todas anteriores son correctas...”
Artefactos	12	Maquetación
	13	Aparición de un asterisco
	14	Preguntas repetidas
	15	Faltan o sobran preguntas

*Tabla 2.* Formulario de “no conformidades” agrupadas arbitrariamente en 3 categorías

### **3. RESULTADOS**

Las tablas 3, 4 y 5 ilustran los resultados más relevantes encontrados en el estudio. En el conjunto de las revisiones que se realizaron se encontraron un total de 215 errores en 123 preguntas. Su distribución fue de un 47,7% en “Precisión semántica, necesidad de ser específico”. Seguido de un 31,8% de “Faltas de ortografía”. El porcentaje de propuestas de corrección se agrupa con un 90,2% en el conjunto definido como cambios de “Redacción y Formulación”. Esta distribución se repite en ambas revisiones. Con la DR se detectaron una ratio de 20.47 frente a los 9.75 de la RS.

Dependiendo de la fase de la revisión se detectaron claras diferencias. En la ratio durante la corrección simple se pasa de 6 durante la corrección por parte de profesores de la asignatura, a 3.25 por profesores ajenos a la asignatura a 0.5 en la prueba final presentado a los alumno. En la DR la tendencia también es decreciente pero con distinta ratio, en concreto 9.07 por profesores de la asignatura, 8.6 en la siguiente fase (1.51 profesor ajeno a la asignatura, 7.09 otro profesor de la asignatura no participante en la elaboración inicial de las preguntas) en la revisión final la ratio de cambios es del 2.56 con un 0.2 en la prueba final presentada a los alumnos.

En la DR, de las modificaciones propuestas por los profesores ajenos a la asignatura el 92.3% pertenecen al grupo de preguntas denominadas de “Redacción y Formulación”. El número de cambios propuestos suponen un 7.43% del total de cambios realizados. En la corrección simple los profesores ajenos a la asignatura propusieron el 33.3% de las modificaciones, siendo el 92.3% también pertenecientes al grupo de preguntas denominadas de “Redacción y Formulación”.

### **4. DISCUSIÓN**

Para que una evaluación sea efectiva, ésta ha de reflejar distintos aspectos de las competencias de los alumnos y no ser la herramienta elegida una dificultad en sí misma. Las Preguntas de Opción Múltiple son consideradas en la literatura como una de las herramientas más utilizadas y variadas (Haladyna, 2002).

Un estudio llevado a cabo por Roediguer y Marsh (2005) realizado con pre-graduados universitarios, refleja que esta manera de evaluar puede tener consecuencias positivas y negativas en el aprendizaje de los alumnos. Sin embargo como señala Palés-Agullós (2010), “Respecto a su validez, son adecuados para medir un espectro amplio del conocimiento y se pueden usar con seguridad para medir los componentes relacionados con los conocimientos y, hasta cierto punto, la capacidad de aplicarlos, siempre y cuando se formulen adecuadamente. Otras capacidades que pueden evaluarse con los exámenes de POM son el razonamiento clínico y la toma de decisiones. Por otra parte, tienen otra ventaja y es que son fáciles de administrar y corregir, especialmente con números altos de alumnos”

Haladyna et al. (2002) realizan una revisión sistemática de 46 guías relacionados con la elaboración de preguntas tipo test de respuesta múltiple. En esta revisión, los autores constatan que hay varios formatos de POM y que unos pueden ser más deseables que otros. Proponen una serie de recomendaciones sobre cómo elaborarlas para mejorar su utilidad. Nos parece el trabajo más completo que hemos encontrado hasta ahora, aunque

carece de algunas recomendaciones sobre distribución de POM según el tipo de destreza mental que estemos evaluando como recomiendan otros estudios (Palmer 2007)

A priori, el ratio número de fallos por pregunta nos parece moderado, aunque no hemos encontrado ningún otro estudio con el que comparar nuestros resultados.

Martínez et al. (2009) realizan un estudio sobre la influencia de 5 directrices específicas recogidas en la taxonomía utilizada por nosotros, en las que apuntan que sólo las diferencias de las opciones en cuanto a contenido y la utilización de respuestas tipo “Ninguna de las anteriores” pueden tener un leve efecto perjudicial sobre la discriminación del ítem.

Aunque no aportan ningún resultado experimental Galofré y Wright (2010) proponen un sistema interesante de valoración de la calidad de las preguntas basados en un trabajo anterior de Josefowic et al. en el que se valora de 1 a 5 la calidad de las mismas según cumplan 10 criterios, la mayoría recogidos en la taxonomía que nosotros hemos elegido.

De las categorías analizadas, hemos comprobado que hay seis en las que no se producen ningún error, siendo las 4 últimas especialmente sensibles en cuanto a la anulación de preguntas el día del examen:

- Dobles negaciones
- Utilización excesiva de acrónimos
- Respuestas no útiles tipo: “todas anteriores son correctas...”
- Aparición de un asterisco
- Preguntas repetidas
- Faltan o sobran preguntas

No tenemos una explicación clara para justificar que el mayor número de no conformidades se encontraran en la meta-categoría de “Redacción y Formulación”, pero es cierto que es la parte más vital y complicada del proceso

A pesar de las revisiones introducidas, 4 preguntas aparecieron con fallos en la copia final del examen, teniéndose que anular dos de ellas.

Comparando los resultados de las revisiones, con el proceso habitual sólo se detectan alrededor de la mitad de las no conformidades totales encontradas. No tenemos elementos cuantitativos para juzgar la influencia del sesgo de expectativa y el empeño de los revisores en encontrar las mismas. Las RS ha llevado aproximadamente un 33% más de tiempo que una revisión normal y la DR un aproximadamente un 66% más también. Debido a la “sensibilidad” de la evaluación en la formación del alumno creemos que puede ser un tiempo muy bien empleado.

Con todo y con eso creemos que el foco de la mejora debería venir en el establecimiento de un plan de intervención en el aumento de la calidad durante la elaboración de las POM. Suponemos que esto hará disminuir el número de no conformidades en las revisiones.

## **5. CONCLUSIÓN**

A la luz de los resultados obtenidos en el estudio sacamos las siguientes conclusiones:

- La confección inicial del examen presenta unos ratios elevados de fallos, aunque carecemos de datos para compararlos
- El revisor externo e interno detectan diferentes tipos de errores
- La utilización de una RS o una DR ha mejorado la calidad del examen final
- La necesidad de establecer un plan de mejora en la confección inicial de las POM

## **6. DECLARACIÓN DE INTERESES**

Todas PCG y SAH son profesores de las asignaturas utilizadas para el estudio y participaron parcialmente en la elaboración de preguntas y revisiones de las mismas

## **7. CONTRIBUCIÓN DE LOS AUTORES**

PCG y SAH contribuyeron por igual en el diseño, realización del proyecto y elaboración del manuscrito

			TOT	%	ratio	1X	2X
Redacción y formulación	1	Incomprensible. No quede clara la intención	23	10,7	1,07	4	19
	2	Precisión semántica, necesidad de ser específico	102	47,7	4,77	24	78
	3	Faltas de ortografía	68	31,8	3,18	5	63
Contenido	6	No adecuado al temario, explicaciones, asignatura	3	1,4	0,14	0	3
	7	No adecuado al nivel de la pregunta	1	0,47	0,05	0	1
	8	Una respuesta/pregunta responda a otras	3	1,4	0,14	0	3
	9	Pregunta mal contestada, asterisco mal colocado	7	3,27	0,33	1	6
	10	Más de una opción o ninguna	6	2,8	0,28	5	1
Artefactos	12	Maquetación	1	0,47	0,05	0	1

*Tabla 3.* Resultados totales generales en porcentajes, ratios de cambios por cada 10 preguntas y resultados por grupo de estudio

			1°	2°	3°	TOT	ratio	%
Redacción y formulación	1	Incomprensible. No quede clara la intención	4	0	0	4	1	10,26
	2	Precisión semántica, necesidad de ser más específico	14	10	0	24	6	61,54
	3	Faltas de ortografía	3	2	0	5	1,25	12,82
Contenido	6	No adecuado al temario, explicaciones, asignatura	0	0	0	0	0	0,00
	7	No adecuado al nivel de la pregunta	0	0	0	0	0	0,00
	8	Una respuesta/pregunta responda a otras del examen	0	0	0	0	0	0,00
	9	Pregunta mal contestada, asterisco mal colocado	1	0	0	1	0,25	2,56
	10	Más de una opción o ninguna	2	1	2	5	1,25	12,82
Artefactos	12	Maquetación	0	0	0	0	0	0,00
TOT			24	13	2	39		
ratio			6	3,25	0,5	9,75		
%			61,6	33,3	5,1	100		

*Tabla 4.* Resultados totales por fase en el grupo de RS en porcentajes, ratios de cambios por cada 10 preguntas

			1°	2°	3°	4°	5°	TOT	ratio	%	
Redacción y formulación	1	Incomprensible. No quede clara la intención	4	4	8	3	0	19	2,21	10,86	
	2	Precisión semántica, necesidad de ser específico	37	5	24	12	0	78	9,07	44,57	
	3	Faltas de ortografía	34	3	21	5	0	63	7,33	36,00	
Contenido	6	No adecuado al temario, explicaciones, asignatura	0	0	3	0	0	3	0,35	1,71	
	7	No adecuado al nivel de la pregunta	0	0	0	1	0	1	0,12	0,57	
	8	Una respuesta/pregunta responde a otras	1	0	2	0	0	3	0,35	1,71	
	9	Pregunta mal contestada, asterisco mal colocado	2	1	3	1	0	7	0,81	4,00	
	10	Más de una opción o ninguna	0	0	0	0	1	1	0,12	0,57	
Artefactos	12	Maquetación	0	0	0	0	1	1	0,12	0,57	
			78	13	61	22	2	176			
			ratio	9,07	1,51	7,09	2,56	0,2	20,47		
			%	44,6	7,43	34,9	12,6	1,1	100		

Tabla 5 .Resultados totales por fase en el grupo de RD en porcentajes, ratios de cambios por cada 10 pregunta



## **8. BIBLIOGRAFÍA**

Galofré, A., & Wright, A. Índice de calidad para evaluar preguntas de opción múltiple. *LA SALUD*, , 141.

Haladyna, T. M., Downing, S. M., & Rodriguez, M. C. (2002). A review of multiple-choice item-writing guidelines for classroom assessment. *Applied Measurement in Education*, 15(3), 309-333.

Haladyna, T. M., Haladyna, R., & Soto, C. M. (1993). Preparación de preguntas de opciones múltiples para medir el aprendizaje de los estudiantes. *OEI-Revista Iberoamericana De Educación*,

Martínez, R. J., Moreno, R., Martín, I., & Trigo, E. (2009). Evaluation of five guidelines for option development in multiple-choice item-writing. *Psicothema*, 21(2), 326-330.

McCoubrie, P. (2004). Improving the fairness of multiple-choice questions: A literature review. *Medical Teacher*, 26(8), 709-712.

Palés-Argullós, J. (2010). ¿ Cómo elaborar correctamente preguntas de elección múltiple? *Educación Médica*, 13(3), 149-155.

Palmer, E., & Devitt, P. (2007). Assessment of higher order cognitive skills in undergraduate education: Modified essay or multiple choice questions? research paper. *BMC Medical Education*, 7(1), 49.

Roediger III, H. L., & Marsh, E. J. (2005). The positive and negative consequences of multiple-choice testing. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 31(5), 1155.

## **9. ANEXOS**

Anexo I: Recomendaciones de consenso de la taxonomía de Haladyna et al., 2002 incluyendo específicas del formato de nuestros exámenes

### **A. Contenidos**

1. Cada una de las preguntas deberá responder a un **CONTENIDO ESPECÍFICO** y deberá utilizar solo una **HABILIDAD MENTAL** específica
2. Evitar contenidos triviales o intrascendentes
3. Utilizar un lenguaje nuevo pero adecuado, no parafrasear texto o poner exactamente las mismas palabras de los contenidos teóricos
4. Mantén el contenido de una pregunta independiente del resto de las preguntas. Evita que unas preguntas respondan a otras
5. Evita preguntas muy generales o muy específicas que hacen difícil su respuesta
6. Evita preguntas con truco

7. Utiliza un vocabulario simple para el tipo de grupo que está siendo evaluado
- B. Formato
1. Redacta las preguntas y sus opciones de manera vertical y no horizontal
  2. Fuente: Calibri 11
- C. Estilo: Utiliza un corrector ortográfico.
1. Usa una correcta gramática, puntuación, capitalización y deletreado
  2. Sin 2 puntos al final de cada pregunta
  3. Sin interrogaciones en la pregunta
  4. Sin puntos al final de cada respuesta
- D. Escribiendo la pregunta
1. Asegúrate de que las directrices de la pregunta sean claras, que el alumno sepa qué quieres de él
  2. Incluye la idea central en la pregunta no en las respuestas
  3. Evita adornos excesivos y/o accesorios que distraigan la atención
  4. Evita las negaciones en la medida de lo posible (no, excepto) y nunca utilices las dobles negaciones. Si lo usas hazlo con cuidado y pon la palabra en negrita o en mayúsculas (NO, NUNCA, EXCEPTO)
- E. Escribiendo las respuestas
1. Estate seguro de que solo una es la adecuada
  2. Varía la localización de la respuesta adecuada entre las 3 opciones
  3. Las opciones no deberían *pisarse*, sobreponerse las unas a las otras
  4. Mantén una homogeneidad en cuanto a la estructura gramatical y longitud de las respuestas
  5. Evita el uso de “ninguna de las de arriba/las anteriores”
  6. Evita el uso de “todas las de arriba/las anteriores”
  7. Evita frases negativas
  8. Evita dar pistas
  9. Haz todas las opciones plausibles, ninguna imposible o ridícula
  10. Utiliza como respuesta errores típicos de los alumnos como respuestas falsas
  11. Evita/Limita el uso del humor; circunscríbelo a un ambiente evaluativo relajado no a un examen importante