

Evaluación de la calidad de las preguntas de selección múltiple utilizadas en los exámenes de Certificación y Recertificación en Cardiología en el año 2009

Evaluation of the Quality of Multiple-Choice Questions Used in Cardiology Certification and Recertification Exams

AMANDA GALLI, HÉCTOR ROITER[§], DANIELA DE MOLLEIN, SANDRA SWIESZKOWSKI^{MTSAC}, NICOLÁS ATAMAÑUK, ANDRÉS AHUAD GUERRERO^{MTSAC}, HUGO GRANCELLI^{MTSAC}, CARLOS BARRERO^{MTSAC}

Recibido: 01/12/2009
Aceptado: 07/04/2010

Dirección para separatas:

Sociedad Argentina de Cardiología
Amanda Galli
Centro de Educación Permanente
(CEP). Área de Docencia
Azcuénaga 980
(1115) CABA
e-mail:
amandagalli@ciudad.com.ar

RESUMEN

La Sociedad Argentina de Cardiología, en los procesos de Certificación y Recertificación de Especialistas, implementa exámenes escritos de selección múltiple. Dos observadores independientes revisaron las 200 preguntas utilizadas en dos exámenes (A y B) realizados durante 2009. Se usó el Índice de Calidad de Galofré; este instrumento toma en cuenta 10 criterios que se deben considerar en la redacción de las preguntas de selección múltiple y establece una escala de 1 a 5 puntos según la cantidad de defectos de construcción que tenga la pregunta. El valor máximo de calidad posible es 5. La media aritmética de los valores de calidad de las preguntas expresa el Índice de Calidad del examen en su totalidad.

Se encontró que el 30% de las preguntas tenían muy buena calidad técnica (valores de calidad 4 y 5); cerca del 40% eran preguntas aceptables (valor de calidad 3) que se deberían mejorar y el 30% eran inaceptables (valor de calidad 1 y 2).

El examen A tiene un Índice de Calidad de 2,15 y el examen B, de 3,21. En ambos exámenes se encontró que los defectos más frecuentes en la redacción de las preguntas eran la falta de viñeta (caso clínico o problema) y la exploración de conocimientos sólo a nivel de memoria o recordación de datos aislados. Se concluyó que sería conveniente constituir en la Sociedad Argentina de Cardiología un grupo de trabajo permanente para revisar y mejorar las preguntas y armar un banco de ítems.

REV ARGENT CARDIOL 2011;79:419-422.

Palabras clave > Educación médica - Preguntas de selección múltiple - Evaluación de la calidad de los exámenes

Abreviaturas > **NBME** National Board of Medical Examiners | **SAC** Sociedad Argentina de Cardiología

INTRODUCCIÓN

La Sociedad Argentina de Cardiología (SAC) implementa anualmente dos exámenes de selección múltiple como parte de los procesos de Certificación y de Recertificación de Especialistas en Cardiología. Tanto la Certificación como la Recertificación son procesos que pretenden evaluar las aptitudes de los profesionales para ejercer la especialidad. En cada caso se analizan los antecedentes académicos y laborales del sujeto y, además, se implementa un examen escrito.

En el proceso de Certificación, el examen escrito se complementa o amplía con un segundo examen teórico-práctico que se rinde en un Servicio de Cardiología con pacientes reales.

En la Recertificación, el examen escrito es opcional, se ofrece como una oportunidad más para sumar créditos en caso de que las actividades de educación continua

y/o las publicaciones y/o participación en congresos y jornadas no sean suficientes para alcanzar el nivel de exigencia establecido.

Los exámenes escritos, con preguntas estructuradas de tipo selección múltiple, permiten una evaluación válida y confiable siempre que se hayan planificado convenientemente y que las preguntas estén bien redactadas.

Las preguntas de un examen pueden explorar distintos niveles taxonómicos o categorías de conocimiento, ya que al leer una pregunta y al pensar la respuesta o solución se ponen en juego distintos procesos intelectuales. (1) Para responder una pregunta del nivel memoria se requiere simplemente la recordación de un dato aislado, mientras que para responder las preguntas de nivel superior hay que interpretar datos, comprender y aplicar información, sacar conclusiones y/o tomar decisiones.

^{MTSAC} Miembro Titular de la Sociedad Argentina de Cardiología

[§] Para optar a Miembro Adherente de la Sociedad Argentina de Cardiología

El National Board of Medical Examiners (NBME) (2) señala que todas las preguntas de un examen deben ser relevantes, es decir, vinculadas con las tareas del médico y que para evaluar la capacidad de usar la información se deben construir preguntas de nivel de aplicación o superior. Recomienda que los ítems de selección múltiple tengan la siguiente estructura:

1. Viñeta.
2. Pregunta.
3. Opciones.

La **viñeta** es la presentación de un paciente; siempre debe incluir edad, sexo, lugar de atención (consultorio externo, unidad coronaria u otros), motivo de la consulta, duración de los síntomas, antecedentes, hallazgos en el examen físico. Eventualmente se pueden mencionar los resultados de estudios de diagnóstico y/o la evolución del paciente.

La **pregunta** se ubica a continuación de la presentación del paciente y/o problema, que puede referirse a ciencias básicas, diagnóstico o tratamiento. Por ejemplo:

- ¿Cuál sería la ubicación más probable de la lesión que presenta el paciente?
- ¿Cuál de las siguientes sería la explicación más probable de estos hallazgos?
- ¿Cuál de los siguientes datos confirmaría el diagnóstico?

Las **opciones** incluyen un listado de tres o cuatro diagnósticos, medicamentos, medidas terapéuticas, hallazgos de laboratorio, sitios anatómicos, patógenos, mecanismos fisiopatológicos. Una de las opciones es la respuesta correcta (clave) y las otras son distractores.

Todas las opciones deben ser homogéneas en forma y en contenido.

Alberto Galofré, (3) después de traducir el manual del NBME y de revisar otras publicaciones, diseñó un Índice de Calidad de los exámenes que se basa en la calidad de cada pregunta. Cada pregunta de selección múltiple se evalúa teniendo en cuenta 10 criterios y se considera que la calidad de la pregunta será óptima si cumple los 10 criterios.

La mayoría de las publicaciones sobre exámenes en educación médica (4-7) han analizado los resultados obtenidos y han sacado conclusiones, pero no revisaron la calidad de las preguntas incluidas en el examen. Cabe señalar que el Índice de Calidad de las preguntas recién estuvo disponible a partir de 2006.

El propósito de este trabajo fue evaluar la calidad de las preguntas utilizadas en dos exámenes implementados en la SAC aplicando el Índice de Calidad de Galofré.

MATERIAL Y MÉTODOS

Dos observadores independientes (un médico y una pedagoga) revisaron dos exámenes (A y B) de 100 preguntas cada uno, implementados durante 2009 en la SAC en el marco de los procesos de Certificación y Recertificación de Especialistas. La evaluación se realizó utilizando el Índice de Calidad de Galofré, que es una sistematización de las recomendaciones publicadas sobre la construcción

de preguntas de selección múltiple. Los 10 criterios que se tomaron en cuenta son:

Presencia de viñeta: se refiere a la presencia de un caso clínico o procedimiento de laboratorio o problema.

Enunciado completo: incluye toda la información necesaria para comprender la pregunta. Se puede pensar la respuesta antes de leer las opciones.

Evita uso de negaciones: como Excepto, NO es, Incorrecto.

Concordancia gramatical entre enunciado y opciones: la pregunta del enunciado y las opciones respetan las reglas gramaticales en cuanto a género y número. Por ejemplo, si se pregunta por "el", no poner "los" en una opción.

Distractores verosímiles: todas las opciones deben parecer verdaderas, igualmente atractivas; no se aceptan opciones "absurdas".

Extensión similar entre las opciones: todas las opciones deben tener una longitud similar, más o menos el mismo número de palabras.

Evita las opciones "ninguna" y "todas las anteriores": bastaría saber que dos opciones son correctas para saber que si existe una opción "todas las anteriores" ésta es la respuesta.

Opciones ordenadas: ya sea lógica o numéricamente. Opciones con números en orden creciente o decreciente.

Opciones homogéneas: similares en cuanto al contenido; todas las opciones son medicamentos, patógenos, diagnósticos, hallazgos u otros.

Nivel de aplicación de conocimiento o superior: que la pregunta promueva un proceso intelectual superior a la memoria o recuerdo. Pensar la respuesta requiere razonamiento, interpretar y relacionar datos, sacar conclusiones, tomar decisiones.

La ausencia de alguno de los criterios se considera un "defecto" de construcción; según la cantidad de defectos que tiene la pregunta le corresponde un valor de calidad que varía entre 1 y 5; el valor 5 es el máximo nivel de calidad posible.

Los cinco valores de calidad posibles son:

- 5 = pregunta con viñeta, sin defectos de construcción.
- 4 = pregunta sin viñeta, sin defectos de construcción.
- 3 = pregunta con o sin viñeta, con un defecto.
- 2 = pregunta con o sin viñeta, con dos defectos.
- 1 = pregunta con o sin viñeta, con tres o más defectos.

Se consideró que la calidad de una pregunta era aceptable cuando tenía un valor de 3 como mínimo.

El promedio de los valores de calidad de las preguntas expresa el Índice de Calidad del examen en su totalidad.

RESULTADOS

Los valores asignados a cada pregunta por ambos observadores fueron coincidentes.

En la Tabla 1 se presentan los resultados de la evaluación de la calidad de las preguntas de cada examen.

En el examen A, el 39% (IC 95% 29,40-49,27) de las preguntas fueron aceptables (valor de calidad 3 y más) y en el examen B, el 69% (IC 95% 58,97-77,87) de las preguntas correspondía a dicha categoría.

El Índice de Calidad del examen A fue de 2,15 y el del examen B fue de 3,21.

En la Tabla 2 se presentan los defectos de construcción que se registraron con mayor frecuencia.

Tabla 1. Cantidad de preguntas según valor de calidad. Exámenes A y B

Valor de calidad	Examen A (n = 100 preguntas)		Examen B (n = 100 preguntas)	
	Nº de preguntas	% acumulado	Nº de preguntas	% acumulado
5	3	3	24	24
4	1	4	8	32
3	35	39	37	69
2	28	67	27	96
1	33	100	4	100

Tabla 2. Cantidad de preguntas según defecto de construcción. Exámenes A y B

Defecto de construcción	Examen A (n = 100 preguntas)	Examen B (n = 100 preguntas)
Falta de viñeta	91	64
Explora conocimientos sólo a nivel de memoria	87	59
Enunciados incompletos	42	16
Falta de concordancia gramatical entre enunciado y opciones	22	2
Uso de negaciones	9	21

La falta de viñeta y el bajo nivel taxonómico de las preguntas resultaron los defectos más frecuentes.

DISCUSIÓN

De las 200 preguntas evaluadas, el 30% tiene muy buena calidad técnica (puntajes 4 y 5); cerca del 40% son preguntas aceptables que se deberían mejorar (puntaje 3) y el 30% son inaceptables (puntajes 1 y 2).

El defecto más frecuente es el bajo nivel taxonómico que exploran las preguntas, es decir que sólo demandan la recordación de datos específicos. Entre los profesionales cardiólogos que colaboran y/o participan activamente en la docencia en la SAC se manifiesta cierta inquietud en relación con la dirección o sentido que está teniendo la práctica profesional y la orientación que están recibiendo los jóvenes en formación: justamente refieren el énfasis puesto en la memorización de datos y conclusiones de los estudios publicados y la falta de capacidad para “razonar” clínicamente. Pareciera que las preguntas de los exámenes analizados están reforzando la dirección no deseada. Mejorar la calidad de las preguntas para llegar a evaluar el razonamiento médico y la capacidad para tomar decisiones es un desafío. (8)

Uno de los exámenes implementados en la SAC es un examen “a libro abierto”, que se responde en domicilio. Se considera que esta modalidad de examen puede constituir una experiencia de aprendizaje muy significativa si estimula a leer en profundidad los temas importantes. Este examen en particular se debería mejorar, ya que “*en los exámenes a libro abierto no tiene sentido realizar preguntas sobre hechos aislados que pueden encontrarse en una sola página del libro de texto... Deben concentrarse en la comprensión de principios y conceptos fundamentales de situaciones problemáticas*”. (9)

Otro reto es la construcción de preguntas relevantes que exploren contenidos pertinentes para la práctica pro-

fesional, preguntas que se refieran a las conductas médicas más habituales en los distintos contextos asistenciales. (10)

Según estudios realizados, (11) la mayoría de los exámenes, orales y escritos, incluyen preguntas que sólo exploran memoria; por lo tanto, no es un problema particular de los exámenes escritos estructurados como se suele mencionar con frecuencia. Desde fines de la década de los setenta se vienen desarrollando otros instrumentos para evaluar las competencias profesionales de los médicos, (12-16) pero los exámenes escritos con preguntas de selección múltiple se siguen utilizando mucho en las carreras de grado, en la selección de residentes y en la certificación de especialistas. (17)

El Índice Calidad del examen A fue de 2,15 y el del examen B resultó de 3,31. Galofré informa que aplicó el Índice de Calidad a distintos exámenes, de ciencias básicas y clínicas, utilizados en Chile y en otros países y encontró que el Índice de Calidad de esos exámenes variaba entre 1,6 y 4,6. Los exámenes de Cardiología analizados se ubican dentro del amplio rango mencionado por Galofré y muy por debajo del índice de calidad (4, 8) de un examen de selección de residentes implementado en un hospital de comunidad. (18)

Como estrategia para mejorar la redacción de preguntas para los exámenes, se propone constituir en la SAC un grupo de trabajo permanente que sea responsable de la revisión de las preguntas antes de diagramar e implementar los exámenes. También resultaría muy útil formar un “banco de ítems”, un archivo de preguntas clasificadas por tema y según sus propiedades psicométricas, que se podrían utilizar nuevamente, previa actualización, en futuros exámenes. (19)

CONCLUSIONES

La evaluación realizada permitió identificar un porcentaje elevado de preguntas deficientes; si bien uno

de los exámenes presentó una proporción mayor de preguntas de calidad aceptable, se advirtió que el margen de oportunidad para el mejoramiento era amplio en ambas pruebas.

SUMMARY

Evaluation of the Quality of Multiple-Choice Questions Used in Cardiology Certification and Recertification Exams

The Argentine Society of Cardiology uses multiple-choice questions for Cardiology Certification and Recertification exams. Two independent observers revised 200 questions used in two examination tests (A and B) during 2009. We used the Galofré Quality Index, which considers 10 criteria for writing multiple-choice questions and establishes a scale ranging from 1 to 5 points according to the number of construction defects the question has. The maximum value of quality is 5. Mean value of the quality of questions represents the Quality Index of the whole test.

We found that 30% of the questions had a very good technical quality (quality value 4 and 5); about 40% were acceptable questions (quality value 3) that should be improved, and 30% were unacceptable (quality value 1 and 2).

The Quality Index of test A was 2.15 and of test B was 3.21. The most frequent writing defects in both exams were the lack of bullet points (case report or problem) and exploring knowledge only in terms of memory or recalling isolated information. In conclusion, it would be convenient to constitute a permanent work group in the Argentine Society of Cardiology to revise and improve the questions and create a bank of items.

Key words > Medical Education - Multiple-choice Questions - Evaluation of Quality of Examinations

BIBLIOGRAFÍA

1. Bloom B. Taxonomía de los objetivos de la educación. La clasificación de las metas educacionales. Buenos Aires: El Ateneo; 1971.
2. Case SM, Swanson DB. Cómo elaborar preguntas para evaluaciones escritas en el área de ciencias básicas y clínicas. Philadelphia: National Board of Medical Examiners; 2006. Disponible en www.saidem.org.ar (accedido 18 diciembre 2009).
3. Galofré A. Instrucciones para calcular un índice de calidad para preguntas de selección múltiple. Universidad Católica del Norte, Chile. 2007. Disponible en <http://www.ucn.cl/oem/?id=3&subid=611> (accedido 15 de octubre de 2009).
4. Iglesias R, Sosa Liprandi A, Darú V, González M, Almirón N, Galli A. Selección de Médicos Residentes. Evaluación de los conocimientos cardiológicos y de las experiencias del pregrado. Medicina 1992;52:549-54.
5. Neuman M, Questa U, Kaufman R. Concurso de residencias médicas en la ciudad de Buenos Aires. Educación Médica 2004;7:90-6.
6. Borracci R, Arribalzaga E. Aplicación de análisis de conglomerados y redes neuronales artificiales para la clasificación y selección de candidatos a residencias médicas. Educación Médica 2005;8:22-30.
7. Alves de Lima A, Botto F, Azzari F, Galli A, Guzmán L, Trivi M y col. Exámenes de selección para postulantes a residencias en cardiología. Variables que se asocian con los resultados. Rev Argent Cardiol 2001;69:327-31.
8. McCoubrie P. Improving the fairness of multiple-choice questions: a literature review. Med Teach 2004;26:709-12.
9. Case SM, Swanson DB. Cómo elaborar preguntas para evaluaciones escritas en el área de ciencias básicas y clínicas. Philadelphia: National Board of Medical Examiners; 2006. Disponible en www.saidem.org.ar (accedido 12 enero 2010).
10. Guilbert JJ. Guía Pedagógica para el personal de salud. OPS/OMS. Ginebra, 1981.
11. Charvat J, McGuire C, Parsons V. Características y aplicaciones de los exámenes en la enseñanza de la medicina. Cuadernos de Salud Pública Nº 36. Organización Mundial de la Salud. Ginebra, 1969. Disponible en http://whqlibdoc.who.int/php/WHO_PHP_36_spa.pdf (accedido 8 de agosto 2009).
12. Harden R, Gleeson F. Asme Medical Education Booklet Nº 8 Assessment of medical competence using an objective structured clinical examination (OSCE). Medical Education 1997;13:41-45.
13. Collins J, Harden R. Real patients, simulates patients and simulators in clinical examinations. Med Teach 1998;20:508-21.
14. Norcini JJ, Blank LL, Duffy D, Fortna GS. The Mini-CEX: a method for assessing clinical skills. Ann Intern Med 2003;138:476-81.
15. Fornells Vallés JM. El ABC del Mini-Cex. Educ Med 2009;9:83-9.
16. Martínez Carretero JM. Los métodos de evaluación de la competencia profesional: la evaluación clínica objetiva estructurada (ECOE). Educación Médica 2005;8:S18-S22.
17. Shumway JM, Harden RM; Association for Medical Education in Europe. AMEE Guide No. 25: The assessment of learning outcomes for the competent and reflective physician. Med Teach 2003;25:569-84.
18. Galli A, Häbich D, Yulitta H. Selección de residentes en un hospital de comunidad: resultados obtenidos por los egresados de distintas universidades y análisis de la calidad del examen administrado. Revista Argentina de Educación Médica 2008;2:8-14.
19. Martínez Cervantes RJ, Moreno Rodríguez R. Construcción de un banco de ítems informatizado para la evaluación de conocimientos sobre una materia universitaria. Universidad de Sevilla. 2002. Disponible en www.aloj.us.es/gmbytdlc/fevaluacionenfmt.pdf (accedido 10 de marzo 2010).