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

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

SUMMARY

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 vignettes (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 task force in the Argentine Society of Cardiology.

REV ARGENT CARDIOL 2011;79:419-422.

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

Abbreviations >	NBM	National Board of Medical Examiners	SAC	Argentine Society of Cardiology (Sociedad Argentina de Cardiología)

BACKGROUND

The Argentine Society of Cardiology (SAC) uses multiple-choice questions for Cardiology Certification and Recertification exams. Certification and Recertification are evaluation processes designed to assess the knowledge and skills required for competent medical practice required by professionals. The academic and working backgrounds of each candidate are evaluated and candidates undergo a written examination.

The Certification process consists of a written examination and a second theory and practice test taken at a Department of Cardiology with real patients.

The written examination is optional in the Recertification process. Candidates who do not achieve the necessary credits earned by continuous medical education and/or publications and/or attending congresses and conferences can take the written exam to earn more credits and be eligible for Recertification.

Written examination with structured multiple-choice questions conveniently planned and properly formulated offer a valid and reliable evaluation.

The questions of an examination may explore different taxonomic levels or categories of knowledge, as several intellectual processes are brought into play at the moment of reading a question and thinking the answer or solution. (1) At the lowest level of cognitive processing of knowledge, a person remembers isolated information, while higher-order processing require interpretation, comprehension and application of the information, as well as the ability to draw conclusions and make decisions.

The National Board of Medical Examiners (NBME) (2) states that the application of knowledge questions should be used as they are relevant, related with physician's tasks and require an examinee to use the information. The NBME recommends the following structure for multiple-choice questions:

1. Vignette.
2. Question.
3. Options.

The vignette commonly provides the patient's age, gender, site of care (physician's office or coronary care unit, among others), presenting complaint, duration of symptoms, followed by personal history, family history and physical examination information. Results of diagnostic studies and/or patient's outcomes may eventually be provided.

Questions follow patient's presentation or problem, and may refer to basic science, diagnosis or treatment. For example:
– Which of the following is the most likely location of the patient's lesion?
– Which of the following is the most likely explanation for these findings?
– Which of the following data is most likely to confirm the diagnosis?

The **option** list includes three or four diagnoses, medications, therapeutic measures, laboratory findings, anatomic sites, pathogens or pathophysiological mechanisms. One of the options is the correct (key) answer and the others are distractors.

All the options should be homogeneous in shape and content.

After translating the NBME manual and reviewing other publications, Alberto Galofré (3) designed a Quality Index of examinations based on the quality of each question. Each multiple-choice question is evaluated taking into account 10 criteria; an optimal question should fulfill the 10 criteria.

Most of the publications about examinations in medical education (4-7) have analyzed the results obtained and drawn conclusions; yet, they did not evaluate the quality of the questions included in the examination. It should be mentioned that the Quality Index of questions has been available since 2006.

The goal of this study was to evaluate the quality of questions used in two examinations implemented in the SAC applying Galofré's Quality Index.

MATERIAL AND METHODS

Two independent observers (a physician and a woman educationist) revised two examinations (A and B) with 100 questions each, implemented by the SAC during 2009 in the setting of the Certification and Recertification processes of the specialty. The evaluation was performed using the Galofré Quality Index based on selected guidelines recommended for well constructed multiple-choice questions. The following 10 criteria were considered:

Presence of vignette: presence of a clinical case, laboratory procedure or problem.

Complete stem: including the necessary information to understand the question. The answer can be predicted before looking at the options.

Lack of negatively phrased items: as those with Except or Not or Incorrect in the lead-in.

Written options that are grammatically consistent and logically compatible with the stem: the stem and the options should be consistent in terms of gender and number. For example, if the stem asks for "an" item, do not include items in "plural" in an option.

Plausible distractors: all the options must seem true and equally attractive; "absurd" options are not accepted.

The options are similar in length: all the options must have the same length and approximately the same number of words.

Lack of using the options "none of the above" and "all of the above": if "all of the above" is an option and students know two of the options are correct, the answer must be "all of the above".

The choices are placed in some meaningful order: place the choices in numerical, chronological or conceptual order. Arrange the options with numbers in increasing or decreasing order.

Use of homogeneous options in content: all the options are medications, pathogens, diagnoses, findings, or others.

Use of application of knowledge questions: the question promotes an intellectual process instead of requiring little more than rote memory and recall. The question requires reasoning and comprehension; the examinee has to relate the information, reach a conclusion or select a course of action.

The absence of any of the criteria is considered a construction "defect". The index 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.

The five possible values of quality are:
5 = question with a vignette and flawless construction.
4 = question without a vignette and flawless construction.
3 = question with or without a vignette, with one defect.
2 = question with or without a vignette, with two defects.
1 = with or without a vignette and three defects or greater.

A minimum score of 3 was necessary to consider that the quality of the question was acceptable.

Mean value of the quality of questions represents the Quality Index of the whole test.

RESULTS

Both observers agreed with the values assigned for each question.

Table 1 presents the results of the evaluation of the quality of the questions of each exam.

The cumulative percent of acceptable questions with a quality value of 3 or greater was 39% (95% CI 29.40-49.27) for exam A and 69% (95% CI 58.97-77.87) for exam B.

The Quality Index was 2.15 and 3.21 for test A and test B, respectively.

Table 2 shows the most common construction defects.

The lack of a vignette and the low taxonomic level of the questions were the most common defects.

DISCUSSION

We found that 30% of the 200 questions evaluated 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 low taxonomic level of the questions exploring only recall memory of isolated facts was the most common defect. The cardiologists with active participation in the educational area of the SAC are concerned about the direction and orientation of medical practice among young in-training

Table 1. Number of question by quality value. Exams A and B

Quality value	Exam A (n = 100 questions)		Exam B (n = 100 questions)	
	Number of questions	Cumulative %	Number of questions	Cumulative %
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

Table 2. Number of questions by construction defects. Exams A and B

Construction defect	Exam A (n = 100 questions)	Exam B (n = 100 questions)
Absence of a vignette	91	64
Explores only rote memory of isolated facts	87	59
Incomplete stems	42	16
The options are nor grammatically consistent with the stem	22	2
Use of negatively phrased items	9	21

professionals, particularly in the emphasize trainees put in recalling the information and conclusions of the published trials and the lack of capability of clinical reasoning. It seems that the questions of the tests analyzed are stimulating the unwanted direction. The challenge is to improve the questions in order to evaluate medical reasoning and the ability to make decisions. (8)

.The SAC implements an "open-book test" to be taken at home. This modality is considered a significant learning tool if it stimulates deep reading of the relevant topics. This exam should be particularly improved as " for open-book tests, it is pointless to ask questions about isolated facts that can be looked up quickly on a single page of a text book, so test material developed for these tests tends to focus more on understanding of key concepts and principles in problem situations". (9)

Another challenge is to construct relevant questions exploring the appropriate contents for professional practice. The questions should focus more on daily practice in the different settings of medical care. (10)

Most oral and written exams (11) include questions that measure simple recall of isolated information; thus this problem is not exclusive of structured written examinations as frequently mentioned. Since the end of the seventies, several instruments are being developed to measure the professional competencies of physicians. (12-16) Yet, written tests with multiple-choice questions are still used for undergraduate students, selection of medical residents and certification of specialists. (17)

The Quality Index was 2.15 and 3.21 for test A and test B, respectively. Galofré reports that he has applied the Quality Index to different exams on basic and clinical sciences used in Chile and other countries, with scores ranging from 1.6 to 4.6. The cardiology tests analyzed are within the range mentioned by Galofré and well below the quality index (4, 8) of a resident selection examination implemented in a community hospital. (18)

It would be convenient to constitute a permanent task force in the SAC as a strategy to revise and improve the construction of questions before designing and implementing the examinations. The creation of a bank of items - a file with questions classified by subjects and psychometric properties - would also be very useful. Updated items could be used again in future tests. (19)

CONCLUSIONS

This analysis allowed the identification of a high percent of deficient questions. Although the quality index of one of the tests was acceptable, both tests should be improved.

RESUMEN

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

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.

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

BIBLIOGRAPHY

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 N° 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 N° 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).