

ARTÍCULO ORIGINAL

Validación de la calidad de las preguntas en un examen final de la asignatura Morfofisiología Humana

Validation of the quality of questions in a final examination of the subject Human Morphophysiology

MsC. Luis Ulpiano Pérez Marqués,^I MsC. Gisela Murillo Jorge,^I MsC. Juan Luis Cobas Pérez,^I Dra. Rosa Navarro Alemán,^I Dra. Vilma García Díaz^I y Lic. Gisela Alina Pérez Murillo^{II}

^I Universidad de Ciencias Médicas, Santiago de Cuba, Cuba.

^{II} Universidad de Oriente, Santiago de Cuba, Cuba.

RESUMEN

Introducción: El índice de dificultad y el poder de discriminación son indicadores fáciles de calcular y útiles para el análisis de la correspondencia entre los resultados esperados y los obtenidos de un instrumento evaluativo.

Objetivo: Evaluar la calidad de las preguntas del examen final ordinario de Morfofisiología Humana V.

Métodos: Fueron incluidos en esta investigación los 265 exámenes teóricos realizados por los estudiantes del segundo año en la Facultad de Medicina No. 2 de la Universidad de Ciencias Médicas de Santiago de Cuba, durante el curso 2011-2012, a los que se les calculó el índice de dificultad y el poder de discriminación en cada una de las 7 preguntas aplicadas.

Resultados: Las preguntas de respuesta alternativa, que evaluaban los contenidos sobre la sangre y el corazón, mostraron un índice de dificultad por debajo de 0,1 y un poder de discriminación inferior a 0,2, lo que hace necesaria su reformulación en próximos instrumentos evaluativos. Los valores más altos para ambos indicadores fueron 0,34 y 0,86, respectivamente, y correspondieron a una pregunta de selección múltiple sobre vasos sanguíneos y linfáticos, siguiéndoles en orden las preguntas de respuesta abierta.

Conclusiones: Se demostró la pertinencia de la mayoría de las preguntas, destacándose la capacidad de 5 de ellas para distinguir estudiantes de alto y bajo rendimientos.

Palabras clave: Morfofisiología Humana V, evaluación del aprendizaje, índice de dificultad, poder de discriminación, validación de calidad, universidad de ciencias médicas.

ABSTRACT

Introduction: The difficulty index and the discrimination power are easy indicators to calculate and useful for the analysis of the correspondence between the expected and obtained results of an evaluative instrument.

Objective: To evaluate the quality of questions of the regular final examination of Human Morphophysiology V.

Methods: The 265 theoretical examinations carried out by the second year students in the Medical Faculty No. 2 of the Medical University in Santiago de Cuba, during the course 2011-2012 were included in this investigation, to which the difficulty index and the discriminating power were calculated in each of the 7 applied questions.

Results: The questions with alternative answer which evaluated the contents on blood and heart, showed a difficulty index below 0.1 and a discriminating power lower than 0.2, making necessary their reformulation in future evaluative instruments. The highest values for both indicators were 0.34 and 0.86, respectively, and they corresponded to a question of multiple selections on blood and lymphatic vessels, followed in order by the questions of open answer.

Conclusions: The relevancy of most of the questions was demonstrated, and the capacity of 5 of them to distinguish students of high and low performances was emphasized.

Key words: Human Morphophysiology V, evaluation of learning, difficulty index, discrimination power, validation of quality, university of medical sciences.

INTRODUCCIÓN

La evaluación del aprendizaje implica el control y la valoración de los conocimientos, las habilidades, los hábitos y modos de actuación que los estudiantes van adquiriendo, a través del proceso docente educativo, en correspondencia con los objetivos de cada disciplina y el plan de estudio en general.¹

En los últimos años, las preguntas con respuesta objetivamente estructurada están siendo usadas ampliamente en las evaluaciones escritas, de manera que la persona examinada elige la respuesta entre varias alternativas proporcionadas por el examinador.

De hecho, esta forma de evaluación ofrece numerosas ventajas que les han aportado una gran aceptación, a saber: aseguramiento de la objetividad en la calificación, posibilidad de ampliar sensiblemente el número de áreas a explorar, calificación en un tiempo relativamente breve, facilidad de buena retroalimentación a los estudiantes y docentes, ventajas para los estudiantes cuya lengua materna no es el español, automatización de la calificación (donde existan o se creen las condiciones) y sencillez relativa del análisis estadístico. Además, con este tipo de preguntas no solo son valorados los conocimientos, sino también las habilidades en la solución de problemas, incluso complejos, por lo que se han estado utilizando de preferencia en la disciplina Morfofisiología, desde que esta se introdujo en el currículo de la carrera de medicina.²⁻⁶

Por otra parte, las preguntas de respuesta abierta o tradicionales ofrecen la posibilidad de evaluar determinados aspectos de la competencia profesional, así como algunas habilidades, lo que no puede efectuarse con las de respuesta objetivamente estructurada, entre ellas las de explicar y describir; también facilitan al estudiante que se examina ordenar sus ideas, exponer sus conocimientos y demostrar su capacidad de expresión, pues son ideales para que este formule hipótesis, halle relaciones y aplique procedimientos conocidos a situaciones nuevas.¹ Debe agregarse a tales consideraciones la utilidad para evaluar la ortografía de los educandos.

En efecto, la necesidad de conocer la efectividad, pertinencia, validez y confiabilidad, además de otros atributos de los instrumentos evaluativos empleados durante el proceso

docente, como parte de la función de retroalimentación que ellos cumplen; conduce a la aplicación de métodos y procedimientos, cuya función es satisfacer dicha necesidad.

Para el grupo de profesores de cualquier asignatura resulta importante conocer si los resultados alcanzados al aplicar un instrumento evaluativo se corresponden con los teóricamente esperados. El índice de dificultad (ID) y el poder de discriminación (PD) son indicadores útiles para ese propósito, y fáciles de calcular además; el primero expresa la proporción de preguntas desaprobadas, mientras que el segundo define la manera en que una pregunta es respondida por estudiantes de diferente rendimiento.^{7,8}

De igual forma, se ha establecido una escala de diversas categorías, que permiten determinar la validez de las preguntas realizadas. Sin embargo, a pesar de su utilidad, el uso de estos indicadores no es sistemático y habitualmente los análisis del claustro se fundamentan casi por completo en los resultados cuantitativos y cualitativos de promoción, por lo que se desconoce la manifestación de otros indicadores de calidad.

Aunque en un instrumento evaluativo puede existir diversidad de tipos de preguntas, se les pueden calcular el ID y el PD a todas por igual, ya sean de respuesta abierta o de respuesta objetivamente estructurada. Conocer el ID y el PD de las preguntas de un examen, retroalimenta a los profesores acerca de los contenidos, los tipos de pregunta y las habilidades que ofrecen mayor grado de dificultad a los estudiantes, lo cual da la posibilidad no solo de saber lo ocurrido, sino de intervenir con medidas concretas en el perfeccionamiento del proceso docente.

En una investigación anterior de los autores de este trabajo se obtuvo que las preguntas de ensayo tuvieron índices de dificultad más elevados.⁹ Ese resultado motivó a profundizar en la evaluación de la calidad de las preguntas de este tipo empleadas con posterioridad, lo cual conformó el propósito del presente artículo.

MÉTODOS

Fueron incluidas en el estudio las 7 preguntas del examen teórico de la asignatura Morfofisiología Humana V, aplicado a 265 estudiantes del segundo año del curso 2011-2012 durante la primera convocatoria (examen final ordinario), en la Facultad de Medicina No. 2 de la Universidad de Ciencias Médicas de Santiago de Cuba, a fin de evaluar la calidad de cada una de las preguntas por medio de la obtención y el análisis del índice de dificultad y el poder de discriminación.

Para medir el ID, se halló el cociente resultante de la división del número de respuestas incorrectas entre el total de respuestas. Esta proporción se expresó como fracción de la unidad y se consideró⁷ normal un ID de 0,3 a 0,7.

El cálculo del PD se efectuó mediante el ordenamiento de los estudiantes, según la calificación obtenida por cada uno. Así, se dividió el total en 4 partes iguales y en ambos extremos de la lista quedaron definidos: un cuartil que contenía a los estudiantes de mayor rendimiento, con las calificaciones más altas (cuartil superior: 66 estudiantes, para 25 % del total) y otro cuartil con los de menor rendimiento, o sea, las calificaciones más bajas (cuartil inferior: 66 estudiantes, para 25 % del total).

Posteriormente se calculó el PD de cada pregunta mediante la ecuación:

$$PD = (Ca - Cb) / N$$

En la ecuación C_a es el número de respuestas correctas de los estudiantes con rendimiento alto, C_b es el número de respuestas correctas del grupo con rendimiento bajo y N es la cifra de integrantes de uno de los grupos. Los resultados pueden encontrarse en los valores de -1 a $+1$, pero la calidad de la pregunta en cuanto a la diferenciación de estudiantes de rendimientos alto y bajo, se clasificó de la forma siguiente:⁸

- Excelente: 0,35 y más
- Buena: de 0,25 a 0,34
- Dudosa: de 0,15 a 0,24
- Mala: menos de 0,15

RESULTADOS

Aprobaron 224 estudiantes, lo que representó 84,5 % del total y de ellos, 58,1 % obtuvo notas de 3 (83 educandos, para 31,3 %) y 4 (71, para 26,8 %); en tanto 70 (26,4 %) calificaron con 5 puntos y 41 desaprobaron (15,5 %).

Las preguntas enumeradas como 1 y 3 fueron desaprobadas solo por 5 y 15 estudiantes, respectivamente; mientras que las preguntas 4, 6 y 7 tuvieron una cantidad mayor de desaprobados (con 90, 55 y 82 estudiantes, en ese orden), por lo cual mostraron valores de ID más elevados, al igual que mayor PD (tabla).

Tabla. Contenidos evaluados, índice de dificultad y poder de discriminación según los tipos de preguntas

Pregunta	Tipo	Contenido	ID	PD
1	Respuesta alternativa	Sangre	0,02	0,08
2	Selección múltiple	Sistema hemolinfopoyético	0,09	0,36
3	Respuesta alternativa	Corazón	0,06	0,18
4	Selección múltiple	Vasos sanguíneos y linfáticos	0,34	0,86
5	Completamiento	Vasos sanguíneos	0,10	0,35
6	Respuesta corta	Sistema cardiovascular	0,21	0,55
7	Ensayo	Presión arterial	0,31	0,68

DISCUSIÓN

Previo a la confección del examen se elaboró una tabla de contenidos, que incluyó los distintos objetivos, la proporción en que estos serían explorados y el tipo de preguntas para cada caso; de este modo se logró un balance adecuado de dichos componentes. Tanto el diseño del instrumento evaluativo como su versión final fueron revisados y aprobados anticipadamente por el grupo de expertos del Departamento Metodológico de la Dirección del Área Básica de la Facultad. Igualmente, la clave de calificación se confeccionó con antelación, de forma completa y detallada.

Al respecto, el colectivo de profesores habría realizado una exploración muy limitada si hubiera considerado suficientes y satisfactorios los indicadores de promoción

cuantitativos y cualitativos, obtenidos a través del instrumento evaluativo. Este criterio condujo a la búsqueda de un método que permitiera efectuar un análisis de calidad más profundo del instrumento, lo que se materializó finalmente con el cálculo del ID y el PD de las distintas preguntas del examen.

Entre los estudiantes desaprobados se apreció un predominio de malas calificaciones en una de las preguntas de selección múltiple y en aquellas de respuesta abierta (corta y de ensayo). No obstante, debe ser aclarado si estos resultados se debieron al tipo de pregunta, al poco dominio de los estudiantes sobre los contenidos u objetivos explorados en ellas, o a su escasa capacidad para demostrar la habilidad evaluada.^{10,11}

Como ya fue señalado⁷ en otros escenarios y, en general, internacionalmente, se considera normal un ID de 0,3 a 0,7. Aunque no existen estudios al respecto, la experiencia empírica indica que en la educación médica superior del territorio donde se desarrolló este estudio, dicho indicador habitualmente muestra valores inferiores a los referidos, lo que pudiera estar en relación con el rigor en el cumplimiento de los aspectos metodológicos establecidos para confeccionar los instrumentos evaluativos y las claves de calificación.

Respecto a la investigación⁹ en la que se mostraron resultados semejantes, la única diferencia entre las preguntas con los mayores ID y las restantes fue el tipo de pregunta y la habilidad a explorar que la relacionaba (habilidad de explicar). Carrazana *et al*¹² obtuvieron para la pregunta de ensayo aplicada en un examen de Morfofisiología I, un nivel de dificultad de 0,23 y una capacidad de discriminación de 0,40, lo que determinó que se clasificara como adecuada y de excelente discriminación. Ellos exploraron las habilidades: identificar y clasificar. En el presente trabajo las preguntas implican la habilidad de explicar, lo que puede haber incidido en el mayor ID, en particular en la pregunta 7.

La escasa capacidad de expresión de numerosos estudiantes y sus limitaciones para demostrar el cumplimiento de objetivos en un nivel productivo del conocimiento, se relacionaron con las insuficiencias de muchos en la habilidad de explicar, y demostró la necesidad de incrementar el trabajo pedagógico en el desarrollo de dicha habilidad. Un estudio preliminar sobre evaluaciones frecuentes posibilitó identificar el problema descrito en este artículo, de modo que el colectivo de la asignatura aplicó algunas medidas, entre las que se incluyeron el incremento del número de preguntas abiertas en dichas evaluaciones y una atención diferenciada a los estudiantes con mayores dificultades, lo cual influyó positivamente en el logro de la promoción final.

Excepto los bajos índice de dificultad y poder de discriminación de las preguntas 1 y 3 -- que indicó la necesidad de su reelaboración en evaluaciones venideras --, los resultados de ambos indicadores en la casuística permitieron ubicarlos entre los esperados, en correspondencia con lo señalado en la bibliografía sobre el tema. En particular se destacó la eficacia de la mayoría de las preguntas para distinguir estudiantes de alto y bajo rendimiento; por tanto, pudo concluirse que los resultados demostraron la validez de las preguntas utilizadas en el instrumento evaluativo analizado.

REFERENCIAS BIBLIOGRÁFICAS

1. Salas Perea RS. La evaluación en la educación superior contemporánea. 2 ed. San Francisco de Macorís: Universidad Católica Nordestana; 2005.
2. Palmer EJ, Devitt PG. Assessment of higher order cognitive skills in undergraduate education: modified essay or multiple choice questions? Research Paper. BMC Med Educ. 2007; 7: 49-55.
3. Stupans I. Multiple choice questions: Can they examine application of knowledge? Pharmacy Education. 2006; 6(1): 59-63.
4. Cañizares Luna O, Sarasa Muñoz N. Algunas reflexiones acerca de los recursos para el aprendizaje de la disciplina Morfofisiología Humana. Educ Med Super. 2007 [citado 12 Feb 2012]; 21(2). Disponible en: http://scielo.sld.cu/scielo.php?script=sci_arttext&pid=S0864-21412007000200001&lng=es
5. Rodríguez Cortés V, Del Pino Sánchez E. Proceso docente educativo: una visión para el profesor de la asignatura de Morfofisiología Humana (I). Educ Med Super. 2009 [citado 12 Feb 2012]; 23(3). Disponible en: http://scielo.sld.cu/scielo.php?script=sci_arttext&pid=S0864-21412009000300010&lng=es
6. Pérez Marqués LU, Murillo Jorge G. Experiencias en el uso de preguntas tipo test objetivo en Morfofisiología Humana [monografía en CD-ROM]. Santiago de Cuba: FACMED/FRH; 2010.
7. Brodin U, Fors U, Laksov KB. The application of Item Response Theory on a teaching strategy profile questionnaire. BMC Med Educ. 2010 [citado 3 Mar 2012]; 10(14): 1-30. Disponible en: <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2830224/>
8. Carneson J, Delpierre G, Masters K. Scorings and statistics. Question attributes. En: Randall V. Designing and managing of multiple choice questions. Dayton : The University of Dayton School of Law [citado 3 Mar 2012]; 2008. Disponible en: <http://academic.udayton.edu/legaled/ProfResp/00Syllabus/extracredit02e.html>
9. Pérez Marqués LU, Murillo Jorge G, Cobas Pérez JL, Navarro Alemán R, Poll Feris E. Índice de dificultad de las preguntas del examen final ordinario de Morfofisiología Humana VI [monografía en CD-ROM]. Santiago de Cuba: Pedagogía y Salud; 2011.
10. Schuwirth LWT, Van der Vleuten CPM. ABC of learning and teaching in medicine written assessment. BMJ. 2003; 326(7390): 643-5.
11. Epstein RM. Assessment in Medical Education. N Engl J Med. 2007; 356(4): 387-96.
12. Carrazana Lee A, Salas Perea RS, Ruiz Salvador AK. Nivel de dificultad y poder de discriminación del examen diagnóstico de la asignatura Morfofisiología Humana I. Educ Med Super. 2011 [citado 4 Dic 2011]; 25(1): 103-14. Disponible en: http://scielo.sld.cu/scielo.php?script=sci_arttext&pid=S0864-21412011000100010&lng=es

Recibido: 5 de julio de 2012.

Aprobado: 30 de noviembre de 2012.

Luis Ulpiano Pérez Marqués. Universidad de Ciencias Médicas, avenida de las Américas, entre calles I y E, reparto Sueño, Santiago de Cuba, Cuba. Correo electrónico: ulpgis@medired.scu.sld.cu