



Investigación en
Educación Médica

<http://riem.facmed.unam.mx>



ARTÍCULO DE REVISIÓN

Reflexiones sobre adivinar en preguntas de opción múltiple y cómo afecta el resultado del examen



Alma Jurado-Núñez^a e Iwin Leenen^{b,*}

^a Programa de Apoyo y Fomento a la Investigación Estudiantil (AFINES), Facultad de Medicina, Universidad Nacional Autónoma de México, México D.F., México

^b Departamento de Evaluación, Secretaría de Educación Médica, Facultad de Medicina, Universidad Nacional Autónoma de México, México D.F., México

Recibido el 18 de enero de 2015; aceptado el 27 de julio de 2015

Disponible en Internet el 9 de septiembre de 2015

PALABRAS CLAVE

Exámenes de opción múltiple;
Adivinar;
Teoría Clásica de los Tests;
Teoría de Respuesta al Ítem;
México

KEYWORDS

Multiple-choice tests;
Guessing;
Classical Test Theory;

Resumen Los exámenes de opción múltiple (EOM) son la herramienta más difundida en educación médica, pero su utilidad está supeditada a la confiabilidad del instrumento y la validez de las inferencias que emanan de la medición. La posibilidad de adivinar, inherente al formato de evaluación, puede introducir varianza irrelevante a la medición y reducir la representación del rasgo latente en la calificación del examen por diferencias individuales respecto a *educated guessing*, *testwiseness* y la tendencia a adivinar. En este artículo se presentan brevemente las características generales de la Teoría Clásica de los Tests (TCT) y la Teoría de Respuesta al Ítem (TRI) y su abordaje al problema de adivinar. Asimismo, se propone un modelo teórico dentro de la TCT que integra los mecanismos que afectan la adivinación y se determina la variación de la probabilidad de aprobar un EOM, en función de ciertos supuestos respecto a adivinar a través de un análisis teórico dentro de un modelo TRI. Es posible concluir que algunas características de los ítems propician la adivinación, y cuando ésta ocurre se encuentran inmersas diversas variables, relacionadas o independientes, del rasgo que se pretende medir, que determinan la magnitud de su efecto.

Derechos Reservados © 2015 Universidad Nacional Autónoma de México, Facultad de Medicina. Este es un artículo de acceso abierto distribuido bajo los términos de la Licencia Creative Commons CC BY-NC-ND 4.0.

Insights into guessing in multiple choice questions and its effect in the assessment outcome

Abstract Multiple-choice tests (MCT) are the most employed assessment tool in medical education; however, its use is limited to the instrument reliability and validity of the inferences made

* Autor para correspondencia: Departamento de Evaluación, Secretaría de Educación Médica. Av. Universidad N° 3000, Edif. B, 3er piso, C.U., C.P. 04510, México D.F., México. Tel.: +56-23-23-00 Ext. 43034.

Correo electrónico: iwin.leenen@gmail.com (I. Leenen).

La revisión por pares es responsabilidad de la Universidad Nacional Autónoma de México.

Item Response Theory; Mexico

upon the measurement. Guessing, an inherent element of this evaluating tool, may introduce construct-irrelevant variance and reduce the load of the latent trait in the score of the exam due to individual differences regarding educated guessing, testwiseness and guessing tendency. This article presents an overview of Classical Test Theory (CTT) and Item Response Theory (IRT), including a discussion of how each theory addresses the guessing phenomenon. With respect to the latter, we propose a theoretical model that integrates factors related to guessing within the CTT framework. We further include a theoretical analysis, which displays the variation of the probability of passing a MCT reliant on certain assumptions regarding guessing that are akin to a particular IRT model. In conclusion, various features of the items increase the likelihood of guessing, and, when guessing takes place, the magnitude of its effect is determined by some variables that can be dependent or independent of the latent trait.

All Rights Reserved © 2015 Universidad Nacional Autónoma de México, Facultad de Medicina. This is an open access item distributed under the Creative Commons CC License BY-NC-ND 4.0.

Introducción

Los exámenes de opción múltiple (EOM) son la herramienta más utilizada en educación médica al hacer posible una evaluación objetiva, estandarizada, costo-efectiva y eficiente¹⁻³. Un ítem de opción múltiple (IOM) desarrollado de forma correcta puede evaluar desde memoria hasta procesos cognitivos superiores como razonamiento clínico y toma de decisiones. Otras ventajas incluyen: *a*) un muestreo más extenso del contenido por evaluar, y *b*) una administración y calificación en poco tiempo, aun en grandes poblaciones de estudiantes⁴. Esta forma de evaluación se ha utilizado a gran escala internacionalmente y ha sido un tema central en investigación en educación desde el siglo pasado. Lo que se espera de cualquier método de evaluación educativa es que proporcione una medición lo más certera posible del rasgo latente que se pretende medir. La validez del método es especialmente trascendente cuando se trata de exámenes sumativos, donde el puntaje se utiliza para tomar decisiones sobre los sustentantes (en educación médica, concretamente, de éste puede depender la emisión de un título o el acceso a un curso de especialidad). Por ello, es indispensable que los EOM, como instrumento de medición, sean lo más confiables, válidos y justos posible.

Respecto a estas tres características, surge un elemento inherente a este formato de evaluación que siempre se ha cuestionado: la posibilidad de adivinar. Existe la posibilidad de que los sustentantes sumen puntos que no representen dominio sobre el tema e incluso, al menos teóricamente, que aprueben un examen a expensas de una ventaja del propio formato. Esta posibilidad se presenta aunque se adivine a ciegas. Sin embargo, pocas veces el sustentante tiene que adivinar a ciegas. Por un lado, los EOM suelen incluir ítems con deficiencias en su construcción, que permiten descartar una o más opciones o inferir la respuesta correcta, incluso con un bajo nivel de conocimiento⁵⁻⁷. Este fenómeno se conoce como *testwiseness*⁸. Por otro lado, es común que, con base en un conocimiento parcial, el sustentante pueda descartar uno o más distractores de un IOM, lo cual se ha denominado *informed guessing* o *educated guessing*⁹⁻¹¹. En general, adivinar constituye una fuente de error de medición que puede convertirse en sesgo estadístico¹². Por lo tanto, la decisión sobre el punto de corte entre

aprobar y reprobar un EOM debería tomar en cuenta que un porcentaje de los aciertos no refleja el dominio del tema o contenido evaluado por la pregunta, sino que se debe al azar.

Para el análisis de pruebas en general (y EOM en específico), existen dos enfoques psicométricos principales: la Teoría Clásica de los Tests (TCT) y la Teoría de Respuesta al Ítem (TRI). Ambos modelos tienen como objetivo realizar inferencias sobre uno o más rasgos latentes, imposibles de observarse directamente, a partir de las respuestas a las preguntas. Por otro lado, sus supuestos subyacentes difieren de forma notable, incluyendo la forma en que consideran y enfrentan el problema de acertar preguntas por adivinación.

El presente artículo tiene dos objetivos: primero, revisamos brevemente las características de la TCT y la TRI, en general (para una revisión más extensa, véase Leenen¹³) y cómo cada teoría ha abordado el problema de adivinar. Al respecto, incluimos *a*) un modelo que integra los mecanismos que afectan la adivinación con los conceptos claves de la TCT, y *b*) una breve descripción de algunos modelos TRI específicos para EOM. Segundo, investigamos, a partir de un análisis teórico enmarcado dentro del modelo Nedelsky, cómo la probabilidad de aprobar un EOM varía en función de ciertos supuestos o estrategias respecto a adivinar. Los resultados de esta investigación aportan evidencia sobre la (baja) plausibilidad de acreditar un examen a expensas de adivinar a ciegas o con base en *testwiseness* o *educated guessing* exclusivamente.

Dos marcos psicométricos y cómo tratan al problema de la adivinación

Conceptos generales de la Teoría Clásica de los Tests

Los supuestos y procedimientos de la TCT son relativamente simples y consideran la prueba en su totalidad. Distingue entre dos factores que componen la puntuación observada (comúnmente representada por X) en el examen: la puntuación verdadera (V) y el error de medición (E):

$$X = V + E \quad (1)$$

Download English Version:

<https://daneshyari.com/en/article/3474597>

Download Persian Version:

<https://daneshyari.com/article/3474597>

[Daneshyari.com](https://daneshyari.com)