

Índice de dificultad y discriminación de ítems para la evaluación en asignaturas básicas de medicina

Difficulty and discrimination index of items for evaluation in basic medical
subjects

Víctor Manuel Gómez López^{1*} <https://orcid.org/0000-0002-2423-0973>

Sandra Rosales Gracia² <https://orcid.org/0000-0003-4646-8111>

José Luis García Galaviz¹ <https://orcid.org/0000-0001-9234-7949>

Karla Isabel Berrones Sánchez¹ <https://orcid.org/0000-0002-8513-4950>

Constanza Margarita Berrones Sánchez¹ <https://orcid.org/0000-0002-7228-5135>

¹Universidad del Noreste, Escuela de Medicina “Dr. José Sierra Flores”. Tampico, Tamaulipas, México.

²Universidad del Noreste, Área de Ciencias de la Salud. Tampico, Tamaulipas, México.

*Autor para la correspondencia: vgomez@une.edu.mx

RESUMEN

Introducción: La determinación de la calidad de un instrumento de medición implica la evaluación tanto de la exactitud como de la estabilidad de dicha medición o el resultado.

Objetivo: Evaluar el índice de dificultad y discriminación de los ítems de los instrumentos de medición de algunas asignaturas del eje curricular de Fundamentos de la medicina de una escuela de medicina particular.

Métodos: Mediante estudio observacional, comparativo, transversal y prolectivo, se analizaron los ítems del examen ordinario de las asignaturas Morfología, Bioquímica y Biología Molecular, y Microbiología y Virología Médica, los cuales fueron contestados por 112 alumnos de Morfología, 101 de Bioquímica y Biología Molecular y 89 de Microbiología y Virología Médica. El número de reactivos resultó 100 para Morfología, 80 para Bioquímica y Biología Molecular y 75 para Microbiología y Virología Médica. Se utilizó estadística descriptiva, y se calculó el índice de dificultad y el índice de

discriminación de los ítems y de todo el examen. Además, se utilizó ANOVA, considerando estadísticamente significativo un resultado igual o menor a 0,05.

Resultados: Los ítems resultaron más difíciles en el examen de Morfología, ya que el 19 % correspondió a la clasificación de relativamente difíciles y el 50 % a la de difíciles, con un total del 69 % entre relativamente difíciles y difíciles. El índice de discriminación resultó mayor en la asignatura de Microbiología y Virología Médica, mientras que en Bioquímica y Biología Molecular se presentó una buena capacidad de discriminación.

Conclusiones: Existió variabilidad en el grado de dificultad y el poder de discriminación de los ítems de los exámenes analizados.

Palabras clave: índice de dificultad; poder de discriminación; exámenes departamentales.

ABSTRACT

Introduction: Determining the quality of a measuring instrument implies the evaluation of both the accuracy and the stability of such measurement or result.

Objective: To evaluate the index of difficulty and discrimination of the items of the measuring instruments of some subjects of the curricular axis of Fundamentals of Medicine of a particular medical school.

Methods: By means of an observational, comparative, cross-sectional and prolective study, the items of the ordinary examination of the subjects Morphology, Biochemistry and Molecular Biology, and Microbiology and Medical Virology were analyzed, which were answered by 112 students of Morphology, 101 of Biochemistry and Biology Molecular, and 89 of Microbiology and Medical Virology. The number of reagents was 100 for Morphology, 80 for Biochemistry and Molecular Biology, and 75 for Microbiology and Medical Virology. Descriptive statistics were used, and the difficulty index and the discrimination index of the items and of the entire exam were calculated. In addition, ANOVA was used, considering a statistically significant result equal to or less than 0.05.

Results: The items were more difficult in the Morphology test, since 19% corresponded to the classification of relatively difficulty and 50% corresponded to those of difficulty, with a total of 69% between relative difficulty and difficulty. The discrimination index was higher in the subject of Microbiology and Medical Virology, while in Biochemistry and Molecular Biology there was a good capacity for discrimination.

Conclusions: There was variability in the degree of difficulty and the power of discrimination of the items of the analyzed exams.

Keywords: difficulty index; discrimination power; department tests.

Recibido: 19/10/2018

Aceptado: 19/01/2020

Introducción

Determinar la calidad de un instrumento de medición implica plantearse preguntas, tanto de la exactitud con que este mide el fenómeno en estudio, la estabilidad de la medición y el resultado como de su pertinencia para los sujetos o la realidad en la que se aplican las mediciones. Este conjunto de interrogantes alude directamente a los conceptos de “confiabilidad” y “validez de un instrumento”. Por otro lado, la estimación de la calidad de un cuestionario o escala es requisito esencial no solo en su proceso de construcción, sino en la utilización de aquellos instrumentos obtenidos en la literatura médica internacional.⁽¹⁾

Una meta del proceso de enseñanza aprendizaje consiste en que los alumnos logren desarrollar las competencias de cada una de las asignaturas cursadas. Para conseguir lo anterior, se debe supervisar que los instrumentos de medición utilizados para la evaluación sumativa tengan la validez adecuada; además, verificar algunos indicadores acerca de las respuestas realizadas a ese instrumento. Desde esta perspectiva existen algunos indicadores que describen cómo ha funcionado una pregunta en una situación dada en exámenes objetivos empleados en la evaluación del aprendizaje; por eso decimos que estos índices describen lo que ha sucedido.⁽²⁾

La función principal de un instrumento de medición en el ámbito educativo, cuando se crea como medida para inferir las capacidades de las personas, es ofrecer información para la correcta toma de decisiones.⁽³⁾

La medición resulta, simplemente, el proceso de asignar valores a ciertos eventos de la realidad.⁽⁴⁾ Esta constituye la etapa más importante de la práctica docente, ya que forma parte del proceso de evaluación, la cual es fundamental para la toma de decisiones acerca del aprendizaje de los alumnos.

Hay que precisar que la confiabilidad (o consistencia) de un test radica en su propiedad de mostrar resultados similares en repetidas mediciones, o en la precisión con que este realiza la medición en una población determinada y en las condiciones normales de aplicación.^(5,6)

Sin embargo, la consistencia del instrumento de medición puede afectarse por otros factores inherentes al propio instrumento. Para evaluar la calidad de un examen, generalmente se utilizan el índice de dificultad y la discriminación obtenida de los reactivos, ya que estos se relacionan con la estructura de la pregunta y con los procesos cognitivos que se demandan del alumno.⁽⁷⁾

El índice de dificultad mide el nivel de calibración de la prueba y responde a la pregunta ¿cuántos de los evaluados respondieron correctamente a la prueba? Por su parte, el índice de discriminación contesta ¿cuáles son las preguntas que han permitido diferenciar a los examinados con altos puntajes de los de bajos puntajes?, y permite caracterizar la calidad de la prueba como elemento de discriminación del desempeño de los estudiantes.^(8,9,10) Es decir, en cuanto al índice de discriminación, un buen ítem debe discriminar entre los que obtuvieron buenas calificaciones en la prueba y los de bajas calificaciones. Generalmente, se calcula como la diferencia entre dos proporciones: proporción de aciertos en el grupo superior (AS/N) menos proporción de aciertos en el grupo inferior (AI/N). Es suficiente con utilizar el 27 % de las personas con las puntuaciones más altas en el test y el 27 % de las más bajas. Mientras mayor resulte el índice de discriminación, el reactivo diferenciará mejor a las personas con altas y bajas calificaciones. De acuerdo con el manual del EXHCOBA (examen de habilidades y conocimientos básicos), el nivel medio de dificultad del examen debe oscilar entre 0,5 y 0,6, al distribuirse los valores de p de la manera siguiente: 5 % de reactivos fáciles, 20 % medianamente fáciles, 50 % con una dificultad media, 20 % medianamente difíciles y 5 % difíciles.⁽³⁾

El objetivo de este estudio es evaluar el índice de dificultad y discriminación de los ítems de los instrumentos de medición de algunas asignaturas del eje curricular de fundamentos de la medicina de una escuela de medicina particular.

Métodos

A través de un estudio observacional, comparativo y transversal, se llevó a cabo el análisis de los ítems del examen ordinario del ciclo académico 2017-02 de las asignaturas de

Morfología, Bioquímica y Biología Molecular, y Microbiología y Virología Médica, los cuales fueron contestados por 112 alumnos de Morfología, 101 de Bioquímica y Biología Molecular, y 89 de Microbiología y Virología Médica; estas disciplinas se cursan en los primeros tres semestres de la licenciatura de medicina de una escuela particular. En todos los casos se utilizaron instrumentos de medición elaborados *ex profeso* para tal finalidad (exámenes departamentales) por los profesores de dichas asignaturas, los cuales quedaron integrados por 100 reactivos para Morfología, 80 para Bioquímica y Biología Molecular, y 75 para Microbiología y Virología Médica. El instrumento se aplicó simultáneamente a los participantes de las tres materias.

Para calcular el índice de dificultad por ítem, se dividió el número de alumnos que contestó correctamente entre el total de alumnos implicados. Generalmente, a esta proporción se le denota con una p ; se utilizó la siguiente fórmula $p = A1/N1$, donde p = índice de dificultad, A = número de aciertos y N = total de alumnos que contestaron los ítems.

En la distribución de la clasificación de dificultad esperada en los ítems se consideraron la publicada por *Backhoff, Larrazolo y Rosas*,⁽³⁾ y los intervalos del índice de dificultad, de acuerdo con *Ortiz y otros*⁽¹¹⁾ (Tabla 1).

Tabla 1- Clasificación de la dificultad de los ítems y su interpretación

Clasificación del ítem	Valor del índice de dificultad del ítem	Distribución porcentual
Fácil	0,91-1	5
Relativamente fácil	0,81-0,90	20
Dificultad adecuada (Media)	0,51-0,80	50
Relativamente difícil	0,40-0,50	20
Difícil	0-0,39	5

El índice de discriminación consistió en la diferencia entre dos proporciones: proporción de aciertos en el grupo superior (AS/N) menos proporción de aciertos en el grupo inferior (AI/N). Este expresó, por lo tanto, hasta qué punto la pregunta discrimina y contribuye a

situar a un sujeto en el grupo superior o inferior. A mayor diferencia en número de acertantes entre los grupos superior e inferior, el ítem es más discriminante, y mejor sitúa a un sujeto entre los primeros o los últimos.⁽²⁾

La discriminación de los ítems se calculó a través del índice de discriminación (D), de acuerdo con la siguiente fórmula: $D = GA - GB/N$, donde D = índice de discriminación del reactivo i, $GA_{aciertos}$ = número de aciertos en el reactivo i del 27 % de personas con las puntuaciones más altas en el test, $GB_{aciertos}$ = número de aciertos en el reactivo i del 27 % de personas con las puntuaciones más bajas en el test, y N = número de personas en uno de los grupos.

En la tabla 2 se consignan los valores y la interpretación del índice de discriminación.⁽¹²⁾

Tabla 2 - Poder de discriminación de los ítems de acuerdo con el valor D

Índice de discriminación (D)	Calidad	Recomendación
< 0,01	Pésimo	Descartar definitivamente
0 a 0,19	Pobre	Descartar o revisar a profundidad
0,20 a 0,29	Regular	Necesidad de revisar
0,30 a 0,39	Buena	Posibilidad de mejorar
0,40 a 1	Excelente	Conservar

Para analizar los resultados se utilizó tanto la estadística descriptiva como el promedio, y la desviación estándar y los porcentajes. Además, se calcularon los índices de dificultad y discriminación de los ítems y de todo el examen. Para la comparación de los índices calculados en los tres exámenes se utilizó ANOVA, al considerar estadísticamente significativo un resultado igual o menor a 0,05.

Toda la información recabada con motivo de este protocolo se manejó con estricto apego al anonimato y la confidencialidad.

Resultados

En la tabla 3 se describen los estadígrafos básicos de los índices de dificultad y discriminación de los exámenes analizados, donde se observa que, en general, el examen de Morfología resultó con mayor grado de dificultad, ya que su promedio del índice de dificultad fue de 0,44.

Tabla 3 - Promedio y desviación estándar del índice de dificultad y discriminación de los exámenes analizados

Asignatura	Ítems (n)	Promedio del índice		Desviación estándar	
		<i>Dificultad*</i> (p)	<i>Discriminación*</i> (D)	p	D
Morfología	100	0,44	0,18	0,19	0,17
Bioquímica y Biología molecular	80	0,57	0,37	0,16	0,22
Microbiología y Virología Médica	75	0,58	0,27	0,22	0,17

Nota: *Valor p: 0,001 (ANOVA).

Los ítems resultaron más difíciles en el examen de Morfología (Tabla 4), ya que un 19 % de estos correspondió a la clasificación de relativamente difíciles y el 50 % a difíciles, lo que hizo un total de 69 % entre ambos. Bioquímica y Biología Molecular apareció después de Morfología, la asignatura con mayor proporción de ítems en la clasificación de relativamente difíciles y difíciles, con un porcentaje global de 37,5 %, cuando se esperaba un 25 %.

Por otro lado, la calidad de los ítems, desde el punto de vista de la discriminación, resultó mayor en la asignatura Microbiología y Virología Médica, como se puede observar en la tabla 5, donde el índice de discriminación llegó hasta el 48 % en el nivel de buena y excelente discriminación, y el mayor porcentaje de ítems con pobre discriminación correspondió a Morfología.

Tabla 4 - Distribución porcentual del índice de dificultad de los ítems en los exámenes analizados

Clasificación del ítem	Porcentaje esperado	Porcentaje observado		
		<i>Morfología</i>	<i>Bioquímica y Biología Molecular</i>	<i>Microbiología y Virología Médica</i>
Fácil	5	0	2,5	8
Relativamente fácil	20	4	5	8
Dificultad media	50	27	55	50,7
Relativamente difícil	20	19	23,75	10,7
Difícil	5	50	13,75	22,6

Tabla 5 - Porcentaje de ítems de acuerdo con el índice de discriminación en los exámenes analizados

Calidad del ítem	Porcentaje observado		
	<i>Morfología</i>	<i>Bioquímica y Biología Molecular</i>	<i>Microbiología y Virología Médica</i>
Pésima	7	7,5	5,3
Pobre	45	22,5	32
Regular	15	32,5	14,7
Buena	18	33,8	21,3
Excelente	15	3,7	26,7

Discusión

La evaluación del aprendizaje constituye el elemento fundamental del proceso educativo, ya que redirecciona las acciones a seguir, tanto por parte del alumno como del profesor. Por ello es indispensable que los instrumentos de medición utilizados para tal propósito reúnan los criterios mínimos de calidad, a efecto de que la evaluación realizada realmente refleje el aprendizaje desarrollado por cada uno de los alumnos. Por otra parte, resulta necesario que los instrumentos de medición utilizados en la evaluación del aprendizaje en programas de licenciatura sean válidos y confiables, así como estandarizados, para que el resultado realmente refleje el estado actual del aprendizaje en las asignaturas evaluadas y, de esta manera, se puedan implementar las acciones de mejora en el proceso educativo. De acuerdo con el objetivo del presente estudio, el índice de dificultad más bajo correspondió a Morfología, lo cual difiere con lo reportado en la misma asignatura por *Argudín, Díaz y Leyva*,⁽¹³⁾ quienes establecieron un índice de dificultad de 0,78. En este mismo estudio, la asignatura de Bioquímica obtuvo el valor más alto del índice de dificultad (0,97); y en el presente artículo, el valor fue de 0,57, el cual es muy similar al conseguido por *Menéndez, Somontes y Capote*,⁽¹⁴⁾ los cuales obtuvieron un índice de dificultad en el examen de Bioquímica de 0,51. Esto se debió a que, en el estudio de comparación, el examen se confeccionó con incisos muy fáciles y con poca profundidad, que no exigieron el razonamiento del estudiante para ser respondidos, situación que contrasta con el instrumento utilizado en este estudio, donde predominaron los ítems de aplicación y análisis, y el alumno tuvo que realizar un proceso de razonamiento para emitir su respuesta.

Por otro lado, en Morfología, el mayor porcentaje de la distribución porcentual de los ítems, según el grado de dificultad, correspondió a los niveles de relativamente difícil y difícil, con un 69 %; en Bioquímica este porcentaje fue de 37 % y en Microbiología tan solo del 33 %, lo que se distribuyó entre el 50-55 % de los ítems de estas dos asignaturas en el nivel de dificultad media. En Morfología solo el 27 % correspondió a este grado de dificultad, lo que evidenció un mayor rigor en su examen.

Saldaña, Delgadillo y Méndez⁽¹⁵⁾ evaluaron el índice de dificultad y la discriminación de un examen parcial de Bioquímica, y encontraron que el 4 % de los reactivos fueron difíciles, el 17 % medianamente difíciles, el 43 % de dificultad media, el 16 % medianamente fáciles y el 20 % fáciles, lo cual difiere con nuestros hallazgos, donde el 14 % resultaron difíciles, el 24 % relativamente difíciles, el 55 % de dificultad media, el

5 % relativamente fáciles y solamente el 2,5 % correspondió a fáciles. Lo anterior manifiesta un grado de dificultad alto, comparado con la distribución porcentual esperada, de acuerdo con el índice de dificultad, ya que el 38 % de los ítems correspondían a la clasificación de relativamente difícil y difícil, en relación con el porcentaje esperado, que entre estos dos niveles debería de ser del 25 %.

En relación con el índice de discriminación del total de los exámenes, el que corresponde a la asignatura de Morfología resultó con el más bajo poder discriminante y el de Bioquímica con el valor más alto (0,37), lo que distribuyó el porcentaje de ítems según el poder de discriminación en excelente (3,7 %) y buena (33,8 %); esto difiere de los resultados aportados por *Saldaña, Delgadillo y Méndez*,⁽¹⁵⁾ donde el 73 % de los ítems se clasificaron en el nivel excelente y el 11 % tuvo buena discriminación.

Aunque la variabilidad de los resultados de estos índices es predecible, también existen coincidencias como, por ejemplo, el hecho de que los hallazgos de *Pérez, Acuña y Arratia*⁽¹⁶⁾ reportan un índice de dificultad promedio de 0,44 en un examen de Citohistología de la carrera de medicina, lo cual coincide con el valor promedio del índice de dificultad del examen final de Morfología obtenido en este estudio.

Finalmente, se puede decir que existe variabilidad en el grado de dificultad y el poder de discriminación de los ítems de los exámenes analizados.

Referencias bibliográficas

1. Alarcón A, Muñoz S. Medición en salud: Algunas consideraciones metodológicas. Rev Méd Chile. 2008 [acceso 12/02/2018];136:125-30. Disponible en: https://scielo.conicyt.cl/scielo.php?pid=S0034-98872008000100016&script=sci_arttext
2. Morales P. Análisis de ítems en las pruebas objetivas. Madrid: Universidad Pontificia Comillas. 2012 [acceso 08/09/2018]. Disponible en: <http://www.upcomillas.es/personal/peter/otrosdocumentos/analisisitemspruebasobjetivas.pdf>
3. Backhoff E, Larrazolo N, Rosas M. Nivel de dificultad y poder de discriminación del examen de habilidades y conocimientos básicos (EXHCOBA). Revista Electrónica de Investigación Educativa. 2000 [acceso 13/02/2018];2(1):11-29. Disponible en: <https://redie.uabc.mx/redie/article/view/15/26>

4. Kerlinger F, Lee H. Investigación del comportamiento. Métodos de investigación en ciencias sociales. 4 ed. México: McGraw-Hill;2002.
5. Kaplan RM, Saccuzzo DP. Pruebas psicológicas. 6 ed. México: Thomson; 2006.
6. Anastasi A, Urbina S. Tests Psicológicos. 7 ed. México: Prentice Hall; 1998.
7. Ram P, Van Der Vleuten C, Tethans JJ, Schouten B, Hobma S. Assessment in general practice: the predictive value of griten-knowledge test and multiple-station examination for actual medical performance in daily practice. Med Educ. 1999 [acceso 13/02/2018];33:197-203. Disponible en: <https://www.ncbi.nlm.nih.gov/pubmed/10211240>
8. Torres J. Los exámenes nacionales de medicina (ENAM) en el Perú. Rev Perú Med Exp Salud Pública. 2008 [acceso 10/01/2018];25(3):316-18. Disponible en: <http://www.scielo.org.pe/pdf/rins/v25n3/a11v25n3.pdf>
9. Abad F, Garrido J, Olea J, Ponsoda V. Introducción a la psicometría: teoría clásica de los test y teoría de la respuesta al ítem. España: Universidad Autónoma de Madrid; 2006.
10. Soubirón E, Camarano S. Diseño de pruebas objetivas. Montevideo: Unidad académica de Educación Química. 2006 Jun [acceso 10/01/2018]. Disponible en: <http://web.ua.es/es/ice/documentos/recursos/materiales/ev-pruebas-objetivas.pdf>
11. Ortiz G, Díaz P, Llanos O, Pérez S, González K. Dificultad y discriminación de los ítems del examen de Metodología de la Investigación y Estadística. EDUMECENTRO. 2015 [acceso 15/02/2018];7(2):19-35. Disponible en: http://scielo.sld.cu/scielo.php?script=sci_arttext&pid=S2077-28742015000200003
12. Ebel L, Frisbie A. Essentials of Education Measurement. 5 ed. Englewood Cliffs, NJ: Prentice Hall; 1991.
13. Argudín E, Díaz P, Leyva E. Índice de Dificultad del examen de Morfofisiología Humana I. Educ Méd Sup. 2011 [acceso 20/03/2018];25(2):97-106. Disponible en: <https://pdfs.semanticscholar.org/a61c/e86ae4acf466f2b4d5a435740c07d1729729.pdf>
14. Menéndez A, Somontes D, Capote A. Análisis de las preguntas de selección múltiple de exámenes profesionales de Bioquímica. AMC. 2010 [acceso 24/09/2018];14(2):[Aprox. 8p]. Disponible en: http://scielo.sld.cu/scielo.php?script=sci_arttext&pid=S1025-02552010000200007&lng=es

15. Saldaña Y, Delgadillo HJ, Méndez I. Evaluación de un examen parcial de Bioquímica. REB. 2014 [acceso 13/01/2018];33(4):104-10. Disponible en: <http://www.scielo.org.mx/pdf/reb/v33n4/1665-1995-reb-33-04-00104.pdf>
16. Pérez J, Acuña N, Arratia E. Nivel de dificultad y poder de discriminación del tercer y quinto examen parcial de la cátedra de cito-histología 2007 de la carrera de medicina de la UMSA. Cuadernos. 2008 [acceso 15/01/2018];53(2):16-22. Disponible en: <http://www.revistasbolivianas.org.bo/pdf/chc/v53n2/v53n2a03.pdf>

Conflicto de intereses

El autor declara que no existe conflicto de intereses.

Contribución de los autores

Víctor Manuel Gómez López: Idea original de la investigación, revisión bibliográfica y documental, análisis e interpretación de los datos, redacción del manuscrito y aprobación de su versión final.

Sandra Rosales Gracia: Revisión bibliográfica y documental, Análisis e interpretación de los datos, participación en la redacción del manuscrito y aprobación de su versión final.

José Luis García Galaviz, Karla Isabel Berrones Sánchez y Constanza Margarita Berrones Sánchez: Recopilación y procesamiento de la información, análisis y discusión de los datos y aprobación de la versión final.