

Falso o verdadero, ¿es esa la pregunta?

María Claudia Abaúnza Chagín¹, Elga Johanna Vargas Carreño², Carlos Javier Rincón Rodríguez³

RESUMEN

Objetivos: en los exámenes para estudiantes de medicina se emplean preguntas objetivas de la modalidad Falso o Verdadero (F/V), que solo tienen dos opciones de respuesta con alta probabilidad de ser contestadas correctamente por azar. El propósito de este trabajo fue demostrar que solicitar la argumentación de las respuestas disminuye el número de las correctas mejorando la capacidad de las preguntas para verificar los aprendizajes.

Métodos: se revisaron 8.188 preguntas de F/V que requirieron sustentación de la respuesta, aplicadas en 28 evaluaciones de Patología de la carrera de Medicina de la Universidad de la Sabana durante cuatro semestres. Las respuestas se clasificaron en correctas e incorrectas y las sustentaciones, en correctas, parcialmente correctas o incorrectas. Se calculó el coeficiente de concordancia kappa entre las preguntas F/V y las preguntas F/V con sustentación, así como por período y por evaluación.

Resultados: de un total de 8.188 preguntas, 6.112 (74,6%) fueron contestadas correctamente, 3.655 (44,6%) tuvieron sustentación correcta y 2.336 (28,5%) no fueron bien sustentadas, con un índice de concordancia global de 0,378 e índices bajos de concordancia por período.

Conclusión: las preguntas de F/V pueden acompañarse de una sustentación que argumente la respuesta con el objetivo de reducir la probabilidad de responder simplemente por azar, desarrollar procesos cognitivos superiores y competencias comunicativas.

PALABRAS CLAVE

Educación Médica; Educación de Pregrado en Medicina; Evaluación Educacional

SUMMARY

True or false: is that the question?

Objectives: Standard tests designed for medical students frequently include objective True/False (T/F) questions, in which there are only two possible answers that can be successfully

¹ Jefa de Área y Profesora Asociada, Área de Patología, Facultad de Medicina, Universidad de la Sabana, Chía, Colombia.

² Profesora Asistente, Área de Patología, Facultad de Medicina, Universidad de la Sabana, Chía, Colombia.

³ Profesor Asistente, Departamento Integrado de Investigación, Facultad de Medicina, Universidad de la Sabana, Chía, Colombia.

Correspondencia: María Claudia Abaúnza Chagín; maria.abauza@unisabana.edu.co

Recibido: marzo 07 de 2014

Aceptado: mayo 15 de 2014

Cómo citar: Abaúnza Chagín MC, Vargas Carreño EJ, Rincón Rodríguez CJ. Falso o verdadero, ¿es esa la pregunta? Iatreia. 2015 Abr-Jun;28(2):120-127. DOI 10.17533/udea.iatreia.v28n2a02.

reached at random. The purpose of this work was to demonstrate that asking for support of the answers diminishes the number of randomly correct ones, so improving the ability of questions to verify learnings.

Methods: 8.188 T/F questions whose answers had to be supported; they were applied in 28 written tests of Pathology, at the undergraduate medical program (La Sabana University, Chía, Colombia). Answers were classified as correct or incorrect, and their support, as correct, partially correct and incorrect. In order to analyze the data, the kappa agreement coefficient between T/F questions and T/F questions with support was calculated, as well as per term and per test.

Results: Out of 8.188 T/F questions, 6.112 (74.6%) were correctly answered as true or false, 3.655 (44.6%) had a correct support, and 2.336 (28.5%) were not correctly supported, exhibiting a global agreement index of 0.378 and low agreement indexes per term.

Conclusion: T/F questions could be accompanied by support of the answers in order to reduce the probability of randomly correct ones, to enhance the development of superior cognitive processes, and communicative essential competences.

KEY WORDS

Education, Medical; Educational Measurement; Undergraduate Medical Education

RESUMO

Falso ou verdadeiro, é essa a pergunta?

Objetivos: nos exames para estudantes de medicina se empregam perguntas objetivas da modalidade Falso ou Verdadeiro (F/V), que só têm duas opções de resposta com alta probabilidade de ser contestadas corretamente por casualidade. O propósito deste trabalho foi demonstrar que solicitar a argumentação das respostas diminui o número das corretas melhorando a capacidade das perguntas para verificar as aprendizagens.

Métodos: revisaram-se 8.188 perguntas de F/V que requereram sustentação da resposta, aplicadas em 28 avaliações de Patologia da carreira de Medicina da Universidade da Sabana durante quatro semestres. As respostas se classificaram em corretas e incorretas e as sustentações, em corretas, parcialmente corretas

ou incorretas. Calculou-se o coeficiente de concordância kappa entre as perguntas F/V e as perguntas F/V com sustentação, bem como por período e por avaliação.

Resultados: de um total de 8.188 perguntas, 6.112 (74,6%) foram respondidas corretamente, 3.655 (44,6%) tiveram sustentação correta e 2.336 (28,5%) não foram bem sustentadas, com um índice de concordância global de 0,378 e índices baixos de concordância por período. **Conclusão:** as perguntas de F/V podem acompanhar-se de uma sustentação que argumente a resposta com o objetivo de reduzir a probabilidade de responder simplesmente por casualidade, desenvolver processos cognitivos superiores e concorrências comunicativas.

PALAVRAS CHAVES

Avaliação Educacional; Educação Médica; Educação de Graduação em Medicina

INTRODUCCIÓN

La evaluación es una de las partes más importantes del proceso enseñanza-aprendizaje y su valor ha sido reconocido desde el siglo IV a. C., cuando Sócrates preguntaba a sus discípulos para comprobar sus habilidades. De acuerdo con Ianfrancesco (1), es un acto educativo, un proceso sistemático y permanente que comprende la búsqueda y el análisis de la información sobre la calidad del desempeño, el rendimiento y los logros del estudiante, así como de la calidad de los procesos, procedimientos y estrategias empleados por los educadores para hacer un diagnóstico conforme a los objetivos de formación y tomar decisiones que orienten y aseguren el aprendizaje (2). La evaluación debe ser dinámica, multidimensional, reflexiva, autónoma, responsable, continua, cooperativa, flexible, integral y sistemática (3,4).

La evaluación puede ser sumativa o formativa. La primera de ellas ha tenido una aceptación amplia, se utiliza para asignar una calificación y dar valores estadísticos, y puede ocurrir en distintos momentos o al final del período académico. La evaluación formativa se emplea para mejorar la calidad de la enseñanza, utiliza la autoevaluación y la retroalimentación, facilita a los estudiantes la adquisición de destrezas de

aprendizaje y el dominio del material aprendido, y permite correcciones continuas del proceso educativo (1,4,5,6).

Después de que el docente conoce qué es evaluación, por qué y para qué evaluar, a quién y cuándo, se debe resolver la pregunta de cómo evaluar y es aquí donde aparecen las teorías de la evaluación y el abanico de instrumentos que deben ser conocidos en profundidad, diseñados y elaborados por los profesores. Estos instrumentos deben ser planeados, validados, variados, claros, precisos, novedosos y adecuados a los contenidos de los saberes de cada elemento de la competencia, aplicados por diferentes docentes y construidos con enunciados neutrales que no condicionen las respuestas, procurando mejorar las habilidades de pensamiento de los estudiantes (1,3,7).

En las facultades de Medicina ha sido tradicional el uso de exámenes de tipo objetivo con predominio de las preguntas convergentes como selección múltiple, apareamiento, complementar y falso o verdadero (F/V) (1,5,8), las cuales se emplean para evaluar aprendizajes específicos o factuales con el fin de comprobar si el estudiante posee los conocimientos requeridos. En las pruebas objetivas los estudiantes tienen poca oportunidad para argumentar o ser originales (3,9,10).

En las preguntas de F/V el enunciado solo tiene dos posibilidades dando una información dicotómica que permite evaluar niveles cognoscitivos como conocimiento y comprensión (11,12). En este tipo de preguntas hay tres posibles causas de respuestas incorrectas: el no conocimiento, el conocimiento parcial y el anti-conocimiento (13,14).

El propósito de este trabajo fue demostrar que solicitar la argumentación de las respuestas disminuye el número de las correctas comparadas con la simple opción de contestar F o V.

MATERIALES Y MÉTODOS

Se revisaron 28 evaluaciones escritas aplicadas a estudiantes de Patología del programa de Medicina de la Universidad de La Sabana durante cuatro semestres (períodos) en los años 2006 y 2007. Cada período incluyó los mismos temas divididos en dos secciones: Patología Básica y Patología Sistémica.

De estas evaluaciones se seleccionaron todas las preguntas de F/V que requirieron sustentación de la respuesta. Las tres primeras pruebas incluyeron temas de Patología Básica y las restantes, de Patología Sistémica. Los docentes que dictaron las clases fueron los encargados de plantear y evaluar las preguntas.

Se encontraron 8.188 preguntas de F/V con sustentación, que fueron reevaluadas de manera conjunta por dos docentes de Patología de la Universidad de La Sabana y los datos se incluyeron en una base de Excel con las siguientes convenciones: cero (0) si la pregunta estaba correctamente clasificada como falsa o verdadera, uno (1) si estaba incorrecta o nueve (9) si no había sido contestada. La sustentación o argumentación se evaluó como cero (0) si estaba correcta, uno (1) si estaba parcialmente correcta, dos (2) si estaba incorrecta y nueve (9) si no había sido contestada.

Para considerar una sustentación como correcta debía ser congruente con el enunciado, completa, no corresponder a una paráfrasis y estar de acuerdo con la literatura pertinente para el nivel de los estudiantes. Se consideró que estaba parcialmente correcta cuando era congruente, pero estaba incompleta y se la consideró incorrecta cuando la argumentación era incongruente, incomprensible o consistía en una paráfrasis. Cuando hubo discrepancia entre las evaluadoras, se revisó la literatura correspondiente llegando siempre a un consenso.

Para el análisis de los datos, se obtuvo el número de estudiantes y preguntas por cada período. Posteriormente, se calculó el coeficiente kappa de concordancia entre el enunciado (F o V) y el enunciado con sustentación. Para ello se excluyeron las preguntas no sustentadas y se agruparon bajo la categoría incorrecta las sustentaciones incorrectas y parcialmente correctas. Adicionalmente, se calculó el coeficiente kappa por período y por prueba. Se hizo la interpretación del coeficiente a partir de los siguientes puntos de corte: < 0,2 pobre; 0,21 a 0,4 baja; 0,41 a 0,6 moderada; 0,61 a 0,8 buena y 0,81 a 1 muy buena.

RESULTADOS

En total se evaluaron 8.188 preguntas de F/V con sustentación aplicadas a 267 estudiantes en cuatro

períodos académicos. La distribución por período se presenta en la tabla 1.

Tabla 1. Distribución de las preguntas por período

Período	Número de preguntas	Número de estudiantes
1	1.735	62
2	2.373	78
3	2.321	64
4	1.759	63
Total	8.188	267

Del total de las preguntas (8.188), se evidenció que 3.655 (44,6%) tenían sustentación correcta, 1.688 (20,6%) tenían sustentación incorrecta, 648 (7,9%) tenían sustentación parcialmente correcta y 121 (1,5%) no tenían sustentación; por lo tanto, 2.336 (28,5%) respuestas no fueron bien sustentadas. La concordancia global fue de 0,378, es decir, baja. Por otra parte, en 1.743 (21,3%) preguntas el enunciado fue calificado como incorrecto y no respondieron 333 (4,1%) preguntas (tabla 2).

Tabla 2. Concordancia global de las respuestas F/V con la sustentación

F-V/Sustentación	n	%
Correcta - Correcta	3.655	44,6
Correcta - Incorrecta	1.688	20,6
Correcta - Parcialmente correcta	648	7,9
Correcta - NR	121	1,5
Incorrecta - Correcta	79	1,0
Incorrecta - Incorrecta	1.553	18,7
Incorrecta - Parcialmente correcta	33	0,4
Incorrecta - NR	78	0,9
NR- Correcta	8	0,1
NR- Incorrecta	20	0,2
NR- Parcialmente correcta	6	0,07
NR- NR	299	3,7
Total	8.188	100,00

NR: no responde

Al discriminar por período, se observó que el porcentaje de preguntas con respuesta F o V correcta y sustentaciones incorrectas –incluyendo las parcialmente correctas– fue de 26,6%; 28,1%; 28,4% y 31% para cada uno de acuerdo con el orden cronológico. Los índices kappa mostraron que la concordancia fue baja, con los siguientes valores en los períodos 1, 2, 3 y 4, respectivamente: 0,3925; 0,4125; 0,3672 y 0,3257 (tabla 3)

Finalmente, al discriminar por cada una de las evaluaciones o pruebas se observó que el porcentaje de preguntas con respuesta F o V correcta y sustentaciones incorrectas –incluyendo las parcialmente correctas– osciló entre 35,7% de la prueba 6 y 22,3% de la prueba 4. Los índices kappa muestran que la concordancia fue baja a moderada, con los siguientes valores para cada una de las 12 pruebas aplicadas: 0,4007; 0,4382; 0,3376; 0,4228; 0,4101; 0,288; 0,446; 0,3096; 0,318; 0,3862; 0,3953 y 0,351 (tabla 4).

DISCUSIÓN

Las pruebas de tipo objetivo tienen la ventaja de poderse aplicar a grupos grandes y obtener una calificación rápida y sencilla, aunque requieren mucho tiempo para su adecuada elaboración. Estas pruebas pueden medir niveles superiores de pensamiento y solución de problemas (1). Para la *National Board of Medical Examiners*, de los Estados Unidos de América, las preguntas de selección de la mejor respuesta son las más apropiadas para evaluar la aplicación de conocimientos, la integración, la síntesis y el juicio (15).

Las preguntas de F/V solo tienen dos posibilidades de respuesta y, por lo tanto, una alta probabilidad (50%) de ser contestadas correctamente por azar (11,14,16). En este tipo de preguntas, los estudiantes se limitan a contestar si un enunciado es o no verdadero, no hay oportunidad de respuestas diferentes y permiten evaluar los niveles más simples de la taxonomía de Bloom (5,9,11,12,17). Además, se ha demostrado que las preguntas de F/V pueden producir un efecto de sugestión negativa que favorece el aprendizaje de información incorrecta (18). En un estudio de Burton, se encontró que al hacer un test de 10 preguntas con dos opciones de respuesta, el índice 50/60 PIP (*Probability of inappropriately passing*) es igual al 52,6%; es decir, que la probabilidad de que un estudiante con el 50% del conocimiento alcance el resultado esperado de un estudiante con el 60% del conocimiento, es igual al 52,6% (19).

Tabla 3. Concordancia por período de las respuestas F/V con la sustentación

Período	Respuesta F/V	Sustentación								Total General	
		Correcta		Incorrecta		Regular		NR			
		n	%	n	%	n	%	n	%	n	%
1	Correcta	828	47,7	307	17,7	155	8,9	27	1,6	1.317	75,9
	Incorrecta	24	1,4	321	18,5	3	0,2	12	0,7	360	20,7
	NR	2	0,1	3	0,2	1	0,1	52	3,0	58	3,3
	Total 1	854	49,2	631	36,4	159	9,2	91	5,2	1.735	100,0
2	Correcta	996	42,0	468	19,7	199	8,4	42	1,8	1.705	71,8
	Incorrecta	10	0,4	514	21,7	19	0,8	23	1,0	566	23,9
	NR	2	0,1	5	0,2	3	0,1	92	3,9	102	4,3
	Total 2	1.008	42,5	987	41,6	221	9,3	157	6,6	2.373	100,0
3	Correcta	1.091	47,0	490	21,1	170	7,3	26	1,1	1.777	76,6
	Incorrecta	25	1,1	406	17,5	8	0,3	23	1,0	462	19,9
	NR	1	0,0	2	0,1	0	0,0	79	3,4	82	3,5
	Total 3	1.117	48,1	898	38,7	178	7,7	128	5,5	2.321	100,0
4	Correcta	740	42,1	423	24,0	124	7,0	26	1,5	1.313	74,6
	Incorrecta	20	1,1	312	17,7	3	0,2	20	1,1	355	20,2
	NR	3	0,2	10	0,6	2	0,1	76	4,3	91	5,2
	Total 4	763	43,4	745	42,4	129	7,3	122	6,9	1.759	100,0

NR: no responde

Tabla 4. Concordancia por evaluación de las respuestas F/V correctas con la sustentación

Prueba	Respuesta	Sustentación								Total General	
		Correcta		Incorrecta		Regular		NR			
		n	%	n	%	n	%	n	%	n	%
1	Correcta	530	45,0	226	19,2	98	8,3	14	1,2	868	73,6
	Incorrecta	16	1,4	246	20,9	4	0,3	9	0,8	275	23,3
	NR	0	0,0	3	0,3	0	0,0	33	2,8	36	3,1
	Total 1	546	46,3	475	40,3	102	8,7	56	4,7	1.179	100,0
2	Correcta	284	51,3	76	13,7	57	10,3	7	1,3	424	76,5
	Incorrecta	7	1,3	104	18,8	1	0,2	6	1,1	118	21,3
	NR	0	0,0	0	0,0	0	0,0	12	2,2	12	2,2
	Total 2	291	52,5	180	32,5	58	10,5	25	4,5	554	100,0
3	Correcta	522	41,4	274	21,7	135	10,7	14	1,1	945	74,9
	Incorrecta	6	0,5	240	19,0	4	0,3	13	1,0	263	20,8
	NR	2	0,2	3	0,2	0	0,0	49	3,9	54	4,3
	Total 3	530	42,0	517	41,0	139	11,0	76	6,0	1.262	100,0
4	Correcta	165	55,7	34	11,5	32	10,8	6	2,0	237	80,1
	Incorrecta	1	0,3	34	11,5	7	2,4	2	0,7	44	14,9
	NR	1	0,3	2	0,7	0	0,0	12	4,1	15	5,1
	Total 4	167	56,4	70	23,6	39	13,2	20	6,8	296	100,0

Tabla 4. Concordancia por evaluación de las respuestas F/V correctas con la sustentación (continuación)

Prueba	Respuesta	Sustentación								Total General	
		Correcta		Incorrecta		Regular		NR			
		n	%	n	%	n	%	n	%	n	%
5	Correcta	268	40,0	144	21,5	47	7,0	2	0,3	461	68,8
	Incorrecta	10	1,5	166	24,8	9	1,3	3	0,4	188	28,1
	NR	1	0,1	2	0,3	3	0,4	15	2,2	21	3,1
	Total 5	279	41,6	312	46,6	59	8,8	20	3,0	670	100,0
6	Correcta	171	30,8	132	23,8	66	11,9	9	1,6	378	68,1
	Incorrecta	10	1,8	138	24,9	1	0,2	6	1,1	155	27,9
	NR	0	0,0	1	0,2	0	0,0	21	3,8	22	4,0
	Total 6	181	32,6	271	48,8	67	12,1	36	6,5	555	100,0
7	Correcta	660	51,4	262	20,4	44	3,4	15	1,2	981	76,3
	Incorrecta	9	0,7	233	18,1	4	0,3	15	1,2	261	20,3
	NR	1	0,1	0	0,0	2	0,2	40	3,1	43	3,3
	Total 7	670	52,1	495	38,5	50	3,9	70	5,4	1.285	100,0
8	Correcta	336	45,5	167	22,6	55	7,4	19	2,6	577	78,1
	Incorrecta	2	0,3	100	13,5	0	0,0	8	1,1	110	14,9
	NR	1	0,1	1	0,1	0	0,0	50	6,8	52	7,0
	Total 8	339	45,9	268	36,3	55	7,4	77	10,4	739	100,0
9	Correcta	359	42,5	190	22,5	73	8,6	9	1,1	631	74,8
	Incorrecta	14	1,7	150	17,8	2	0,2	8	0,9	174	20,6
	NR	1	0,1	3	0,4	0	0,0	35	4,1	39	4,6
	Total 9	374	44,3	343	40,6	75	8,9	52	6,2	844	100,0
10	Correcta	111	57,8	35	18,2	11	5,7	0	0,0	157	81,8
	Incorrecta	2	1,0	26	13,5	0	0,0	1	0,5	29	15,1
	NR	0	0,0	0	0,0	0	0,0	6	3,1	6	3,1
	Total 10	113	58,9	61	31,8	11	5,7	7	3,6	192	100,0
11	Correcta	111	44,0	61	24,2	10	4,0	4	1,6	186	73,8
	Incorrecta	2	0,8	52	20,6	0	0,0	0	0,0	54	21,4
	NR	1	0,4	5	2,0	1	0,4	5	2,0	12	4,8
	Total 11	114	45,2	118	46,8	11	4,4	9	3,6	252	100,0
12	Correcta	138	38,3	87	24,2	20	5,6	22	6,1	267	74,2
	Incorrecta	0	0,0	64	17,8	1	0,3	7	1,9	72	20,0
	NR	0	0,0	0	0,0	0	0,0	21	5,8	21	5,8
	Total 12	138	38,3	151	41,9	21	5,8	50	13,9	360	100,0

NR: no responde

Algunos autores han diseñado variantes de las preguntas de F/V para optimizar este tipo de evaluación como seleccionar ítems erróneos dentro de párrafos de lecturas (20) o penalizar las respuestas erradas (14). Nosotros empleamos otra alternativa a este tipo de pregunta que consiste en solicitar la justificación de la respuesta F o V.

Consideramos que las preguntas de F/V clásicas son de difícil elaboración en Medicina, dado que muy pocas temáticas de este campo permiten aplicar el concepto de 100% falso o verdadero y por el contrario existen numerosas "zonas grises"; además, nos exponemos a inexactitudes y a preguntar datos superfluos o temas irrelevantes. Por el contrario, si se solicita sustentar las respuestas se promueven en los estudiantes el análisis, la argumentación y el juicio, además de incentivar el adecuado uso de las normas gramaticales y ortográficas. De este modo, no solo evaluamos conocimientos de Patología, sino que mejoramos las competencias comunicativas esenciales en el ejercicio médico.

En nuestra experiencia docente hemos tratado de mantener un proceso permanente de reflexión y autoevaluación tanto en la elaboración como en la implementación de nuestras evaluaciones. Consideramos que los profesores no debemos limitarnos a emplear un solo tipo de pregunta, aunque existen tendencias que consideran que las pruebas con preguntas de selección múltiple son el tipo "ideal" de evaluación escrita.

Los docentes debemos evaluar constantemente los instrumentos de valoración que aplicamos a nuestros estudiantes, de acuerdo con los niveles cognitivos y competencias esperadas; estos instrumentos no se deben concebir como simples elementos rutinarios de modelos pedagógicos clásicos y deben ser objeto de renovación, combinación e innovación continuas, tal como lo exige un proceso evaluativo dinámico.

En este trabajo se encontró que 6.112 (74,6%) de un total de 8.188 preguntas estaban contestadas correctamente como falsas o verdaderas, pero solo 3.655 (44,6%) tenían una sustentación correcta con bajos índices de concordancia global y de concordancia por período. Estos valores de coeficiente kappa nos indican que es valioso incorporar la sustentación en este tipo de preguntas, ya que si los valores se hubieran

acercado a uno, no habría diferencia entre realizar preguntas F/V o preguntas F/V con sustentación.

Cuando se analizaron los resultados de concordancia entre las evaluaciones, los índices kappa mostraron que ella fue de baja a moderada; los porcentajes más bajos (por debajo del 60%) se detectaron en las pruebas 6, 8 y 9. Las pruebas 8 y 9 correspondieron principalmente a las temáticas de patología hematolinfoide y endocrino, que en nuestra opinión son de mayor grado de dificultad, incluso para médicos patólogos, por su complejidad y extensión.

Además, no se observó una mejoría en la concordancia a medida que se aplicaban las pruebas en cada uno de los períodos académicos. Finalmente, solo 8 (0,1%) y 6 (0,07%) preguntas que no fueron clasificadas como falsas o verdaderas poseían una sustentación correcta o parcialmente correcta, respectivamente, lo que se podría explicar por falta de atención durante el desarrollo de la prueba y la presencia de algún grado de conocimiento del estudiante; sin embargo, estas fueron situaciones infrecuentes.

CONCLUSIONES

La formación docente requiere un conocimiento amplio sobre el proceso evaluativo en el que es pertinente seleccionar los instrumentos de acuerdo con las competencias que se espera alcanzar. Así como otros autores han diseñado variantes de las preguntas de F/V, nosotros proponemos que en este tipo de preguntas se le solicite al estudiante una corta sustentación en la que argumente su respuesta, con el objetivo de reducir la probabilidad de responder simplemente por azar y desarrollar procesos cognitivos superiores y competencias comunicativas esenciales en la formación médica.

REFERENCIAS BIBLIOGRÁFICAS

1. Ianfrancesco G. La evaluación integral y del aprendizaje: Fundamentos y estrategias. Bogotá: Cooperativa Editorial Magisterio; 2005. p. 17-40.
2. Henson KT, Eller BF. Psicología Educativa para la enseñanza eficaz. México: Thomson; 2000.

3. Tobón S. De la evaluación a la valoración de competencias. En: *Formación Integral y Competencias: Pensamiento complejo, currículo, didáctica y evaluación*. 3ª ed. Bogotá: ECOE; 2010. p. 279-309.
4. Rodríguez J. *Módulo de Evaluación, Especialización en Educación Médica*. Chía: Universidad de La Sabana; 2004. p. 1-33.
5. Orlich D, Harder R, Callahan R, Kauchak D, Pendergrass RA, Keogh A, et al. *Técnicas de enseñanza*. México: Limusa-Noriega; 1995. p. 50-2, 231-9.
6. Venturelli J. La evaluación en el proceso de formación de los profesionales médicos. *Principios de Evaluación Formativa*. En: *Educación médica: Nuevos enfoques, metas y métodos*. Washington: Serie PAL-TEX Salud y Sociedad OPS/OMS; 2003. p. 164-88.
7. de Camilloni ARW, Celman S, Litwin E, Palou de Maté MC. La evaluación de los aprendizajes en el debate didáctico contemporáneo. Buenos Aires: Paidós; 1998. p. 3, 71-6.
8. Pamphlett R. It takes only 100 true-false items to test medical students: true or false? *Med Teach*. 2005 Aug;27(5):468-72.
9. McInerney DM, McInerney V. *Educational Psychology: Constructing Learning*. 2º ed. Sydney: Prentice Hall; 1998. p. 297-304.
10. Schwartz PL, Loten EG. Brief problem-solving questions in medical school examinations: is it necessary for students to explain their answers? *Med Educ*. 1999 Nov;33(11):823-7.
11. University of Waterloo. Exam Question: Types, Characteristics and suggestions [Internet]. [cited 2014 Mar 3]. Available from: [https://uwaterloo.ca/centre-for-teaching-excellence/teaching-tips/developing-assignments/exams/questions-types-characteristics-suggestions](https://uwaterloo.ca/centre-for-teaching-excellence/teaching-resources/teaching-tips/developing-assignments/exams/questions-types-characteristics-suggestions)
12. Richardson R. The multiple choice true/false question: what does it measure and what could it measure? *Med Teach*. 1992;14(2-3):201-4.
13. Koeslag JH, Melzer CW. The incorrect response in multiple-choice examinations. *S Afr Med J*. 1981 Oct;60(15):591-2.
14. Koeslag JH, Melzer CW, Schach SR. Penalties in multiple-choice and true-false questions. *S Afr Med J*. 1983 Jan;63(1):20-2.
15. Case S, Swanson D. Formatos de ítems de opción múltiple. En: *Cómo elaborar preguntas para evaluaciones escritas en el área de ciencias básicas y clínicas*. Philadelphia: National Board of Medical Examiners; 2006. p. 13-8.
16. Frisbie D, Becker D. An Analysis of Textbook Advice About True-False Tests. *Appl Meas Educ*. 1991;4(1):67-83.
17. Rath L, Wassermann S. Las operaciones del pensamiento. En: *Como enseñar a pensar*. Buenos Aires: Paidós; 1997. p. 27-50.
18. Toppino T, Brochin H. Learning from test: The case of True/ False examinations. *J Educ Res*. 1989 Nov-Dec;83(2):119-24.
19. Burton R. Quantifying the Effects of Chance in Multiple Choice and True/False Tests: question selection and guessing of answers. *Assess eval high educ*. 2001;26(1):41-50.
20. Goldwater B, Grabavac D, Acker L. The distorted-item test: a true-false test with a difference. *IJT*. 2005;5(1):75-80.

