

Análisis psicométrico de las preguntas repetidas en los simulacros de la prueba MIR

Jaime BALADRÓN-ROMERO, Fernando SÁNCHEZ-LASHERAS, Carmen PEÑALVER-SAN CRISTÓBAL, Alberto GARCÍA-GUERRERO, José M. ROMEO-LADRERO, Tomás VILLACAMPA, José CURBELO, Paula JIMÉNEZ-FONSECA

Introducción. El presente trabajo analiza 13 preguntas que se repitieron dos veces en distintas pruebas que simulan el examen MIR, separadas en el tiempo entre dos y 15 semanas. A partir de los resultados obtenidos, se ha analizado el proceso de aprendizaje de los estudiantes que se preparan para la prueba MIR.

Material y métodos. Para el presente estudio se hizo uso de una base de datos formada por un total de 2.600 preguntas respondidas en promedio por 1.585 alumnos de características relativamente homogéneas.

Resultados. En general, resulta posible afirmar que entre la primera y la segunda exposición de los alumnos a la misma pregunta se produjo una mejora de los índices analizados. De los resultados obtenidos, se observa que, en promedio, a lo largo de la preparación, la dificultad de las preguntas para el grupo de estudiantes se reduce a la mitad. En lo referente al índice de discriminación, éste se incrementa cerca del 60%, lo que supone que, según transcurre la preparación, aumenta la proporción de personas que aciertan la pregunta en el grupo fuerte con respecto a las que la aciertan en el grupo débil. Finalmente, también se produce una mejora en el valor del coeficiente de correlación biserial puntual, lo que supone una mejora de la capacidad discriminativa de las preguntas.

Conclusiones. Según transcurre la preparación, las preguntas analizadas se vuelven más fáciles y más discriminativas. Además, se observa cómo no todos los alumnos consiguen el mismo aprovechamiento del proceso de aprendizaje.

Palabras clave. Educación médica. Encuestas y cuestionarios. Estudiantes de medicina. Preguntas de examen. Psicometría. Teoría de respuesta al ítem.

Psychometric analysis of the questions repeated in the MIR test mockups

Introduction. The present work analyses 13 questions that were repeated twice in different tests that simulate the MIR exam, separated in time between 2 and 15 weeks. From the results obtained, the learning process of the students preparing for the MIR test has been analysed.

Material and methods. For the present study, a database made up of a total of 2,600 questions which were answered on average by 1,585 students with relatively homogeneous characteristics was employed.

Results. In general, it is possible to affirm that between the first and second exposure of the students to the same question there was an improvement in the indexes analysed. From the results obtained, it is observed that, on average, throughout the preparation, the difficulty of the questions for the group of students is reduced by half. Regarding the discrimination index, this increases by about 60%, which means that as the preparation progresses, the proportion of people who get the question right in the strong group increases with respect to those who get it right in the weak group. Finally, there is also an improvement in the value of the biserial point correlation coefficient, which implies an improvement in the discriminative capacity of the questions.

Conclusions. As the preparation progresses, the analysed questions become easier and discriminating. In addition, it is observed how not all students obtained the same performance.

Key words. Examination question. Item response theory. Medical education. Medical students. Psychometrics. Surveys and questionnaires.

Curso Intensivo MIR Asturias (J. Baladrón-Romero, C. Peñalver-San Cristóbal, A. García-Guerrero). Departamento de Matemáticas. Universidad de Oviedo (F. Sánchez-Lasheras). Servicio de Oncología Médica. Hospital Universitario Central de Asturias. Oviedo (P. Jiménez-Fonseca). Blog MIRentrelazados. Zaragoza (J.M. Romeo-Ladrero). Clínica Oftalmológica Villacampa. Avilés (T. Villacampa). Servicio de Medicina Interna. Hospital Universitario La Princesa. Madrid, España (J. Curbelo).

Correspondencia:

Dr. Fernando Sánchez-Lasheras. Departamento de Matemáticas. Facultad de Ciencias. Universidad de Oviedo. c/ Federico García Lorca, 18. E-33007 Oviedo.

E-mail:

sanchezfernando@uniovi.es

Recibido:

30.07.21.

Aceptado:

14.01.22.

Conflicto de intereses:

No declarado.

Competing interests:

None declared.

Cómo citar este artículo:

Baladrón-Romero J, Sánchez-Lasheras F, Peñalver-San Cristóbal C, García-Guerrero A, Romeo-Ladrero JM, Villacampa T, et al. Análisis psicométrico de las preguntas repetidas en los simulacros de la prueba MIR. FEM 2022; 25: 19-23. doi: 10.33588/fem.251.1164.

© 2022 FEM



Artículo *open access* bajo la licencia CC BY-NC-ND (<https://creativecommons.org/licenses/by-nc-nd/4.0/>).

ISSN: 2014-9832

ISSN (ed. digital): 2014-9840

Tabla. Distancia en semanas entre las dos exposiciones a una misma pregunta, asignatura a la que corresponde, índice de dificultad con corrección de los efectos del azar en la primera y la segunda exposición a la pregunta (Dif. 1 y Dif. 2), índices de discriminación en la primera y la segunda exposición a la pregunta (Disc. 1 y Disc. 2) e índice de correlación biserial puntual en la primera y la segunda exposición a la pregunta (Bis. 1 y Bis. 2).

N.º	Semanas	Asignatura	Dif. 1	Disc. 1	Bis. 1	Dif. 2	Disc. 2	Bis. 2
1	2	Inmunología	0,6458	0,9616	0,3625	0,768	0,9725	0,3352
2	2	Traumatología	0,2753	0,1308	0,0903	0,5153	0,5724	0,2731
3	3	Oftalmología	0,0225	0,5709	0,2546	0,4077	0,8581	0,385
4	3	Nefrología	0,8561	0,5404	0,1632	0,9369	0,5719	0,1182
5	4	Digestivo	-0,0471	0,4226	0,1867	0,2996	0,7653	0,3539
6	6	Traumatología	0,2115	0,2483	0,1406	0,4862	0,4047	0,2115
7	8	Anestesiología	-0,0934	0,3552	0,162	0,5362	0,6579	0,3175
8	8	Urgencias Cirugía Plástica	0,7923	0,5804	0,223	0,8348	0,9943	0,3197
9	8	Traumatología y Urgencias	0,6527	0,66	0,278	0,8441	0,8423	0,2614
10	12	Traumatología	0,2431	0,5055	0,2556	0,706	1,1459	0,4091
11	13	Preventiva	0,2111	0,6386	0,298	0,6683	1,216	0,4239
12	14	Oncología y Endocrinología	0,0533	0,5624	0,2648	0,53	0,8935	0,3906
13	15	Enfermedades Infecciosas	0,4996	0,8145	0,3556	0,7938	1,1934	0,3735
		Promedio	0,3325	0,5378	0,2335	0,6405	0,8529	0,321

Introducción

En la actualidad, el acceso a la formación médica especializada en España se realiza a través de una prueba objetiva y de carácter nacional que se celebra con una periodicidad anual. Esta prueba se viene celebrando desde 1978 y en la actualidad se convoca por medio de una orden del Ministerio de Sanidad [1].

Una forma de medir la evolución del conocimiento de los estudiantes que se preparan para la prueba MIR podría consistir en someterlos en distintos momentos de su preparación a un mismo test. Sin embargo, el proceder de esta forma se encontraría con la limitación de que los alumnos podrían haber memorizado las respuestas correctas, dado que ya fueron expuestos a esas preguntas, y

que sus respuestas no representen su verdadero nivel de conocimiento, sino un mero ejercicio de memoria. Una posible alternativa sería introducir dentro de un test algunas preguntas que ya hayan sido contestadas previamente.

El presente trabajo analiza 13 preguntas que se repitieron dos veces en distintas pruebas que simulan el examen MIR, separadas en el tiempo entre dos y 15 semanas. Esta repetición se produjo de forma accidental. Es decir, el examinador no era consciente de estar proponiendo una pregunta que ya había sido utilizada en una prueba anterior y, por tanto, los sujetos analizados no fueron advertidos en forma alguna. En este análisis se hace uso de este hecho para observar el proceso de aprendizaje de los estudiantes. Los datos analizados corresponden a un grupo de estudiantes españoles, recién graduados y que se preparan por primera vez para esta prueba.

El objetivo de este trabajo es comparar los resultados que obtienen los alumnos en estas preguntas y emplearlas como forma de medir la evolución del nivel de conocimiento de un grupo de estudiantes que se preparan para la prueba MIR.

Material y métodos

La base de datos

La base de datos a la que pertenecen las preguntas analizadas está constituida por un total de 2.600 preguntas que corresponden a 13 simulacros de examen distintos a los que se somete a médicos que preparan la prueba MIR con Cursos Intensivos MIR Asturias. Dichos simulacros forman parte de los 33 simulacros de la prueba MIR que se realizan durante los últimos siete meses de preparación. Por tanto, las 13 parejas de preguntas repetidas suponen el 1% del total de las preguntas propuestas. En promedio, cada uno de estos simulacros fue cumplimentado por 1.585 alumnos, con una desviación estándar de 45,6.

Herramientas de comparación

Para efectuar la comparación de estas parejas de preguntas, se hizo uso de las siguientes herramientas

- *Histograma*. En el que se indican el número de alumnos que ha escogido cada una de las opciones de respuesta.
- *Localización de la pregunta dentro de la curva de dificultad del examen*. En esta curva se presentan, ordenadas de izquierda a derecha y de

mayor a menor dificultad, todas las preguntas que constituyen la prueba a la que pertenecen. Con esta disposición se puede conocer la dificultad de cada pregunta en relación con todas las demás que le acompañan en cada simulacro.

- *Curva de probabilidad.* En ella se representa la probabilidad que tiene un alumno de responder correctamente a una pregunta en función de su nivel de conocimiento relativo al resto de individuos de la población objeto de estudio. La curva representada corresponde al modelo de dos parámetros de la teoría de respuesta al ítem (TRI) [2].

Además, también se hace uso de los tres indicadores numéricos que se explican a continuación:

- En primer lugar, se encuentra el índice de dificultad [3] con corrección de los efectos del azar (*ID*). Esta fórmula pondera el número de individuos que responden de forma correcta cierta pregunta (*A*), teniendo en cuenta el número de los que la fallan (*E*) y las opciones de respuesta disponible (*K*), así como el total de individuos que realizan la prueba [3,4]. Se representa por medio de la fórmula:

$$ID = \frac{A - \frac{E}{K-1}}{N}$$

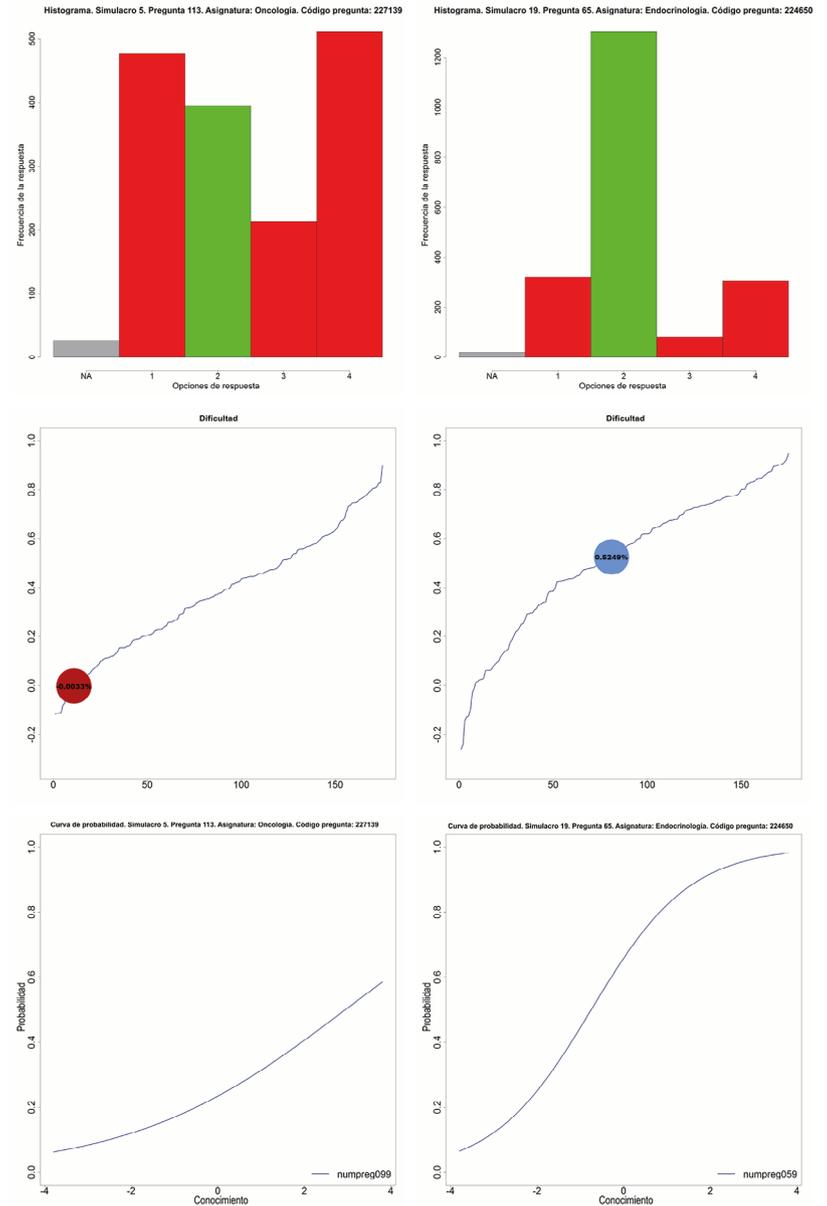
- Seguidamente, el índice de discriminación (*DS*) relaciona el fallo o acierto de los individuos en cada uno de los ítems con la puntuación que estos obtienen en la prueba [5,6]. Para ello, es necesario definir primeramente los conceptos de grupo fuerte y grupo débil. El grupo fuerte es el formado por el 27% de los estudiantes que obtuvieron la mejor puntuación en el test, mientras que el grupo débil es el del 27% de los estudiantes que obtuvieron las peores puntuaciones en la misma prueba. Así, el índice de discriminación viene dado por la ecuación [6]:

$$DS = 2 \frac{F-D}{N_1+N_2},$$

donde *F* es el número de respuestas correctas entre los estudiantes que forman parte del grupo fuerte y *D* es el número de respuestas correctas del grupo débil. De igual manera, *N₁* es el número de estudiantes que respondió a la pregunta en el grupo fuerte y *N₂* el que respondió a la pregunta en el grupo débil.

- Por último, y con el fin de estudiar la validez dis-

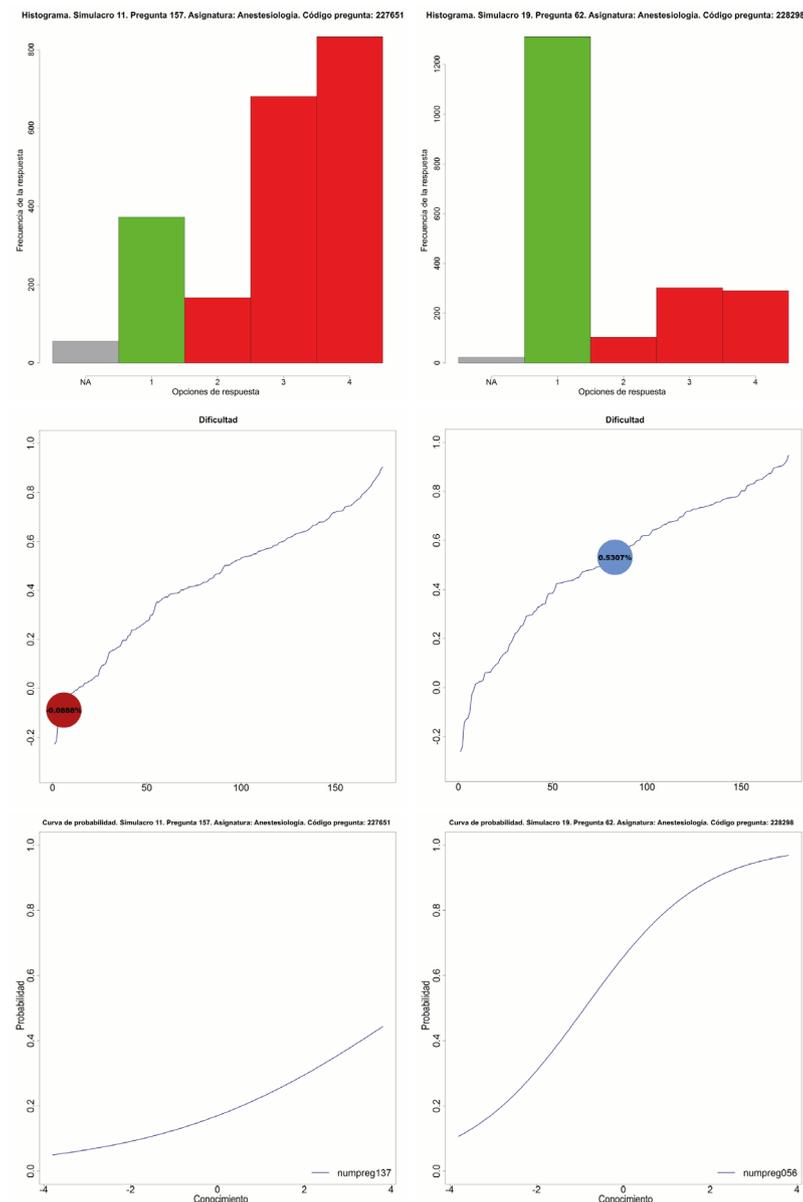
Figura 1. Pregunta caracterizada como perteneciente a las asignaturas de Oncología y Endocrinología presentada con 14 semanas de diferencia. Histograma, localización de la pregunta dentro de la curva de dificultad del examen y curva de probabilidad.



criminativa de las preguntas, se define el índice de correlación biserial puntual (ρ_{bp}) a partir de la fórmula [7]:

$$\rho_{bp} = \frac{\mu_p - \mu_q}{\sigma_x} \sqrt{\frac{ID}{1-ID}},$$

Figura 2. Pregunta caracterizada como perteneciente a la asignatura de Anestesiología presentada con ocho semanas de diferencia. Histograma, localización de la pregunta dentro de la curva de dificultad del examen y curva de probabilidad.



donde μ_p es la puntuación media en el test de los sujetos que aciertan el ítem, mientras que μ_q es la puntuación media de los que fallan dicho ítem; σ_x es la desviación estándar de la puntuación total del test y representa la proporción de

sujetos que aciertan el ítem. En el caso de este índice [7], cuanto mayor sea su valor, mayor será la relación entre obtener una puntuación alta en el test y el hecho de haber contestado correctamente a la pregunta en cuestión.

Resultados y discusión

La tabla muestra los índices psicométricos correspondientes a las parejas de preguntas ordenadas de menor a mayor número de semanas de diferencia entre las dos veces que se preguntaron. En dicha tabla se incluye la asignatura a la que corresponden junto con los valores de los índices de dificultad con corrección de los efectos del azar, índices de discriminación e índices de correlación biserial puntual, tanto para la primera como para la segunda vez que los alumnos responden a la pregunta. Nótese que la primera y la segunda vez que se preguntan, tanto el enunciado como las opciones de respuesta son exactamente iguales.

De los resultados obtenidos, se observa que, en promedio, el valor del índice de dificultad de las preguntas prácticamente se multiplica por dos y se incrementa un 92,6%, lo que se debe interpretar como un aumento considerable del número de alumnos que la aciertan. En lo referente al índice de discriminación, éste se incrementa un 58,6%, lo que supone que, según transcurre la preparación, aumenta la diferencia en la proporción de personas que aciertan la pregunta en el grupo fuerte con respecto a las que la aciertan en el grupo débil. Es decir, el incremento de los conocimientos del grupo fuerte es superior a la del grupo débil.

Finalmente, también se produce una mejora en el valor del coeficiente de correlación biserial puntual, lo que se traduce en su aumento en un 37,5%. Este aumento supone que, en promedio, en lo relativo a su capacidad discriminativa, las preguntas pasen de estar clasificadas como aceptables a buenas. Esto va en la línea de que, según mejora la preparación, los efectos de ésta se van haciendo más notables y existen mayores diferencias entre estudiantes como consecuencia del distinto aprovechamiento que hacen de ella.

En general, resulta posible afirmar que entre la primera y la segunda exposición de los alumnos a la pregunta se produce una mejora de los tres índices analizados, a excepción del coeficiente de correlación punto biserial de las preguntas con números de orden 1, 4 y 9, dado que el acierto de éstas pasa a tener una menor relación con el resultado global del test, cosa lógica, dado que son tres

de las preguntas más fáciles de entre las 13 objeto de análisis.

La figura 1 muestra el histograma, la localización de la pregunta dentro de la curva de dificultad del examen y la curva de probabilidad correspondiente a una misma pregunta que se propuso en dos pruebas separadas 14 semanas entre sí. El histograma muestra que, mientras que en la primera exposición existen dos opciones que los alumnos prefieren a la opción correcta, en la segunda exposición la opción correcta es claramente la preferida por la mayoría de los alumnos. Por tanto, a lo largo de la preparación se fijan conceptos que ayudan a discriminar la respuesta correcta. En lo referente a la localización de la pregunta dentro de la curva de dificultad del examen, ésta evoluciona, y la pregunta pasa de ser una de las más difíciles a una de dificultad media. Por último, en lo referente a la curva de probabilidad, se observa cómo una pregunta cuya probabilidad de acierto en alumnos con niveles de conocimientos altos dentro del conjunto de los que hicieron el test era ligeramente superior al 50% se convierte en una pregunta fácil para los alumnos con conocimientos ligeramente superiores a la media, ya que la acierta prácticamente el 100% de los mejores alumnos que realizan el segundo test. Nótese que tener la mayor pendiente para niveles de conocimiento intermedios (*ability* alrededor de cero) indica que ésa es la zona de la curva en la que la pregunta resulta más discriminativa, es decir, en la que un pequeño incremento de los conocimientos relativos al resto del grupo del individuo consigue un mayor incremento de sus probabilidades de acertar la pregunta.

En la pregunta que se presenta en la figura 2 también se observa cómo el transcurso de la preparación consigue que los alumnos se inclinen mayoritariamente por la opción correcta, disminuyendo también el número de los que no contestan la pregunta. Como era esperable, la pregunta se vuelve más fácil en la clasificación relativa dentro de la prueba. La interpretación gráfica de las probabilidades resulta muy relevante, dado que se aprecia cómo una pregunta en cuya primera exposición la probabilidad de acierto era muy baja, la segunda vez que se propone presenta una curva clásica de pregunta discriminativa, pues la probabilidad de acertar varía con el nivel de conocimiento y pasa a ser fácil para los alumnos de elevado nivel de conocimiento.

Conclusiones

En general, los resultados que obtienen los alumnos la segunda vez que son expuestos a la misma pregunta son mejores que la primera. Este resultado se produce con independencia del número de semanas transcurridas en el intervalo de este estudio, comprendido entre las dos y las 15 semanas.

En opinión de los autores, el resultado más interesante que presenta este trabajo es que se aprecia cómo, en general, según transcurre la preparación, las preguntas analizadas se vuelven más discriminativas. Dado que el grupo de alumnos que realizan todas las pruebas objeto de análisis es superior al 90% de los participantes, se puede considerar que se trata del mismo conjunto y que, por tanto, lo que se está reflejando en la variación de la discriminación y en la gráfica de probabilidad de la pregunta es cómo no todos los alumnos consiguen el mismo aprovechamiento de ésta. Por último, desde nuestro punto de vista, dado el pequeño porcentaje que suponen las preguntas repetidas con respecto al total de las preguntas de esos simulacros (el 1% del total) y que los alumnos no esperan que ninguna pregunta lo sea, se puede suponer que en la respuesta a las preguntas interviene mayoritariamente el nivel de conocimiento real de los opositores.

Bibliografía

1. Baladrón J, Sánchez-Lasheras F, Villacampa T, Romeo-Ladrero JM, Jiménez-Fonseca P, Curbelo J, et al. Propuesta metodológica para la detección de preguntas susceptibles de anulación en la prueba MIR. Aplicación a las convocatorias 2010 a 2015. FEM 2017; 20: 161-75.
2. Baladrón J, Sánchez-Lasheras F, Villacampa T, Romeo-Ladrero JM, Jiménez-Fonseca P, Curbelo J, et al. El examen MIR 2015 desde el punto de vista de la teoría de respuesta al ítem. FEM 2017; 20: 29-38.
3. Baladrón J, Curbelo J, Sánchez-Lasheras F, Romeo-Ladrero JM, Villacampa T, Fernández-Somoano A. El examen al examen MIR 2015: aproximación a la validez estructural a través de la teoría clásica de los tests. FEM 2016; 19: 217-26.
4. Baladrón J, Sánchez Lasheras F, Romeo Ladrero JM, Curbelo J, Villacampa Menéndez P, Jiménez Fonseca P. Evolución de los parámetros dificultad y discriminación en el ejercicio de examen MIR. Análisis de las convocatorias 2009 a 2017. FEM 2018; 21: 181-93.
5. Guilbert JJ. Educational handbook for health personnel (offset publication, n.º 35). 1 ed. Geneva: WHO; 1977.
6. Ministerio de Sanidad y Consumo. Pruebas selectivas para el acceso a plazas de formación de médicos especialistas (1982-1992). Madrid: Ministerio de Sanidad y Consumo; 1993.
7. Muñiz J. Teoría clásica de los test. 2 ed. Madrid: Pirámide; 2002.