

Estudio psicométrico de las pruebas MIR de 2020 y 2021

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

Introducción. En España, el ejercicio de cualquier especialidad médica exige disponer del título de especialista. El acceso a la formación como especialista se realiza a través de la prueba MIR. Esta prueba es obligatoria para poder optar a una plaza de formación en cualquier especialidad médica. El presente artículo realiza un estudio psicométrico tanto a través de la teoría clásica de los test como de la teoría de la respuesta al ítem de las pruebas MIR de 2020 y 2021 (de 175 y 200 preguntas, respectivamente).

Material y métodos. La base de datos utilizada en el presente trabajo contiene las respuestas de un total de 4.810 opositores que se examinaron en 2020 y 3.888 que lo hicieron en 2021. La información disponible se analizó haciendo uso de indicadores como el alfa de Cronbach, la fórmula número 21 de Kuder y Richardson, el índice de dificultad, el índice de dificultad con corrección de los efectos del azar, el índice de discriminación, el índice de correlación biserial puntual, así como los índices de dificultad y discriminación según la teoría de respuesta al ítem.

Resultados. Las pruebas MIR de 2020 y 2021 presentan una alta fiabilidad, con variaciones pequeñas de un año a otro en el comportamiento de las distintas asignaturas y bloques de preguntas en lo relativo a los índices utilizados en el presente estudio.

Conclusiones. Los resultados obtenidos en el presente estudio permiten afirmar que el examen MIR es un examen objetivo, de dificultad y discriminación adecuadas, así como estable de una convocatoria a otra.

Palabras clave. Estadísticas. Estudiantes de medicina. Mediciones educativas. Prueba MIR. Psicometría. Rendimiento académico.

Psychometric study of the MIR tests of 2020 and 2021

Introduction. In Spain, in order to practise any medical speciality, it is necessary to have qualified as a specialist. Access to training as a specialist is gained through the MIR test. This test is mandatory to qualify for a training position in any medical specialty. This article carries out a psychometric study of the 2020 and 2021 MIR tests (with 175 and 200 questions, respectively) through both the classical test theory and the item response theory.

Material and methods. The database used in this study contains the responses from a total of 4,810 individuals who took the exam in 2020 and 3,888 who did so in 2021. The information available was analysed using indicators such as Cronbach's alpha, the formula number 21 of Kuder and Richardson, the difficulty index, the difficulty index corrected for random effects, the discrimination index and the point biserial correlation index. The difficulty and discrimination indices were also used, according to item response theory.

Results. The MIR tests of 2020 and 2021 present high reliability, with only slight variations from one year to another in the behaviour of the different subjects and question blocks in terms of the indices used in the present study.

Conclusions. The results obtained in the present study allow us to affirm that the MIR exam is objective, of appropriate difficulty and discrimination, as well as stable from one year to another.

Key words. Academic performance. Educational measurements. Medicine students. MIR exam. Psychometrics. Statistics.

Introducción

El ejercicio de cualquier especialidad médica en España requiere que el profesional que la desempeñe tenga el título de médico especialista. El acceso a la formación especializada no es libre, sino que, dado

que se dispone de un número limitado de plazas, todos los graduados interesados deben realizar el examen MIR. Esta prueba la convocan anualmente desde 1978 de manera conjunta los Ministerios de Sanidad, y Educación y Formación Profesional. La normativa aplicable a dicha prueba se recoge en el

Cursos Intensivos MIR Asturias (J. Baladrón-Romero, A. García-Guerrero, C. Peñalver-San Cristóbal). Departamento de Matemáticas. Facultad de Ciencias. Universidad de Oviedo (F. Sánchez-Lasheras). Servicio de Oncología Médica. ISPA. Hospital Universitario Central de Asturias. Oviedo (P. Jiménez-Fonseca). Facultad de Medicina (J. Curbelo). Facultad de Psicología. Universidad Francisco de Vitoria. Pozuelo de Alarcón (C. Baladrón-Labordá). Clínica Oftalmológica Villacampa. Avilés, España (T. Villacampa).

Correspondencia:

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

E-mail:

sanchezfernando@uniovi.es

Recibido:

24.11.22.

Aceptado:

01.02.23.

Conflicto de intereses:

No declarado.

Cómo citar este artículo:

Baladrón-Romero J, Sánchez-Lasheras F, Curbelo J, García-Guerrero A, Peñalver-San Cristóbal C, Villacampa T, et al. Estudio psicométrico de las pruebas MIR de 2020 y 2021. FEM 2023; 26: 19-28. doi: 10.33588/fem.261.1256.

© 2023 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

Real Decreto 589/2022 [1]. Es este real decreto el que desarrolla lo dispuesto en el artículo 25 de la Ley 44/2003, de 21 de noviembre, de Ordenación de las Profesiones Sanitarias [2].

A pesar de la existencia de críticas razonables al actual formato de la prueba MIR [3-5], así como a su capacidad para medir los conocimientos en medicina de los médicos opositores [5], no es menos cierto que el uso de las pruebas de tipo test como criterio de evaluación para el ejercicio de la profesión médica o el acceso a la especialización se lleva a cabo en muchos países [6,7]. Además, también cabe señalar la existencia de trabajos previos [8-13] en los que se argumenta que, desde el punto de vista psicométrico y del contenido de las preguntas, la prueba MIR presenta un buen rendimiento.

En la convocatoria de 2021, la prueba consistió en un total de 200 preguntas de tipo test con cuatro opciones de respuesta de las que únicamente una era correcta. Dado que existe la posibilidad de que las preguntas sean anuladas, los candidatos que se presentan a esta prueba reciben un cuadernillo con un total de 210 preguntas de las que las últimas 10 se consideran preguntas de reserva y su respuesta se tiene en cuenta únicamente si alguna de las 200 primeras se anula, pues dichas preguntas se emplean para sustituir a las anuladas. En la prueba de 2020, el número total de preguntas fue de 185, de las que 10 eran de reserva.

La puntuación total obtenida en la prueba MIR, ponderada con el expediente académico, determina el orden de prelación a la hora de escoger una plaza de formación especializada. Para el cálculo de la puntuación total, el resultado obtenido en el examen tiene un peso del 90%, y el baremo académico obtenido en el grado en medicina, del 10%. Además, se establece una nota de corte que es igual a un 35% del promedio de la puntuación de los 10 mejores exámenes de cada convocatoria.

En la prueba de 2020 se inscribieron 15.176 médicos tanto españoles como extranjeros, de los que fueron admitidos 14.425, y se presentaron al examen 13.332 aspirantes. En la prueba de 2021, estos números fueron inferiores. Así, se inscribieron 13.924 médicos, de los que fueron admitidos 13.080, y se presentaron al examen 11.827 aspirantes. En la prueba de 2020, los candidatos a especialista compitieron por un total de 7.899 plazas, mientras que, en la prueba de 2021, el número de plazas objeto de concurso fue de 8.188.

El objetivo de este trabajo es analizar desde el punto de vista psicométrico los resultados de las pruebas MIR de 2020 y 2021 comparándolas entre sí, por ser las más cercanas en el tiempo, a pesar de

tener distinto número de preguntas. Para ello se dispone de una muestra de las respuestas a las preguntas MIR de los opositores de ambas convocatorias.

Material y métodos

Bases de datos

Se ha hecho uso de la información disponible a través de una aplicación *ad hoc* desarrollada por el Curso Intensivo MIR Asturias. Este sistema de recogida de información ya se había empleado en estudios anteriores [10-14]. Así, en el caso de la prueba MIR de 2020, se dispone de las respuestas de 4.810 opositores, mientras que, en el caso de la de 2021, se tiene la de 3.888. Seguidamente, se definen las métricas fundamentales que se han empleado en esta investigación.

El alfa de Cronbach y la fórmula 21 de Kuder y Richardson

La fiabilidad del test en su conjunto se analizó con la ayuda de las fórmulas del alfa de Cronbach y la fórmula 21 de Kuder y Richardson [15].

La fórmula del alfa de Cronbach viene definida por la ecuación:

$$\alpha = \frac{n}{n-1} \left(1 - \frac{\sum_{j=1}^n \sigma_j^2}{\sigma_x^2} \right) \quad (1)$$

Donde:

n representa el número de ítems del test.

σ_j^2 es la varianza del ítem j -ésimo.

σ_x^2 es el valor de la varianza de las puntuaciones obtenidas por todos los opositores en el test.

La fórmula número 21 de Kuder y Richardson se expresa como:

$$KR_{21} = \frac{n}{n-1} \left(1 - \frac{n \cdot p \cdot (1-p)}{\sigma_x^2} \right) \quad (2)$$

En esta ecuación, n y σ_x^2 tienen el mismo significado que en la ecuación anterior, mientras que p representa la probabilidad promedio de acierto de todas las preguntas del test.

Índice de dificultad

Se define el índice de dificultad como [15,16]:

$$DF = \frac{F+D}{N_1+N_2} \quad (3)$$

En esta fórmula, F representa el número de personas que aciertan la pregunta de entre los que pertenecen al grupo fuerte, el cual se define como el que está formado por el 27% de los opositores que obtuvieron las puntuaciones más altas en la prueba [10], mientras que D representa el número de respuestas correctas en el grupo débil, y dicho grupo es el formado por el 27% de los estudiantes que obtuvieron las peores puntuaciones en el test [10]. De igual manera, N_1 y N_2 representan el número de estudiantes que respondieron la pregunta, la acertaron o no, en el grupo fuerte y débil, respectivamente. A mayor dificultad de la pregunta, menor será el valor de este índice.

Según la bibliografía, las preguntas se pueden clasificar atendiendo a su índice de dificultad en las siguientes categorías [10,16]:

- Fácil: $>0,8$.
- Aceptable: $>0,6-0,8$.
- Óptimo: $>0,5-0,6$.
- Aceptable: $>0,3-0,5$.
- Difícil: $0-0,3$.

Índice de dificultad con corrección de los efectos del azar

El índice de dificultad con corrección de los efectos del azar viene dado por la fórmula [10,16]:

$$ID = \frac{A - \frac{E}{K-1}}{N} \quad (4)$$

En esta fórmula, A representa el número de opositores que aciertan la pregunta objeto de análisis, mientras que el número de opositores que o fallan la pregunta o la dejan en blanco viene dado por E . Nótese que K representa el número de alternativas de respuesta y, en el caso de la prueba MIR objeto de análisis, es igual a 4 para todas sus preguntas. N representa el número total de individuos presentados a la prueba.

De manera similar a lo que se hace con el índice de dificultad, según la bibliografía [10,16], las preguntas se pueden clasificar atendiendo a su índice de dificultad con corrección de los efectos del azar en las siguientes categorías:

- Muy fácil: $>0,8$.
- Fácil: $>0,66-0,8$.
- Óptimo: $>0,33-0,66$.
- Difícil: $>0-0,33$.
- Muy difícil: $-0,33-0$.

Índice de discriminación

Se define el índice de discriminación como [17]:

$$DS = 2 \cdot \frac{F-D}{N_1+N_2} \quad (5)$$

En esta fórmula, F , D , N_1 y N_2 tienen el mismo significado que en la fórmula (1). Atendiendo al resultado de este índice, las preguntas se clasifican en las siguientes categorías [10,15,17,18]:

- Excelente: $>0,34$.
- Bueno: $>0,24-0,34$.
- Revisable: $>0,14-0,24$.
- Malo: $0-0,14$.
- Muy malo: <0 .

Índice de correlación biserial puntual

La fórmula del índice de correlación biserial puntual viene dada por [15]:

$$\rho_{bis} = \frac{\mu_p - \mu_q}{\sigma_x} \sqrt{\frac{ID}{1-ID}} \quad (6)$$

En esta fórmula, μ_p representa la puntuación media en el test de los sujetos que aciertan el ítem, mientras que μ_q es la puntuación media de los que lo fallan. Además, σ_x representa la desviación típica de las puntuaciones totales en el test de todos los médicos que se examinaron e ID es el índice de dificultad con corrección de los efectos del azar definido según la fórmula 4.

Al igual que para las fórmulas anteriores, también en el caso del índice de correlación biserial puntual se hará uso de una clasificación categórica de las puntuaciones obtenidas [10,12]. Dicha clasificación es la que se relaciona a continuación:

- Excelente: $>0,39$.
- Buena: $>0,3-0,39$.
- Regular: $>0,2-0,3$.
- Pobre: $0-0,2$.
- Pésima: <0 .

Índices de dificultad y discriminación según el modelo de dos parámetros de la teoría de respuesta al ítem

La teoría de respuesta al ítem se centra en el desarrollo de modelos capaces de prever la forma en la que los individuos responden a las preguntas que se les plantean en función de su nivel de conocimientos sobre una disciplina [11]. A través de los resultados obtenidos en estudios previos [11,13], es conocido que el modelo que mejor refleja el comportamiento de los opositores que se presentan a la prueba MIR es el modelo de dos parámetros [19]. Dicho modelo expresa la probabilidad que cada opositor tiene de acertar la pregunta j -ésima por medio de la fórmula:

$$P(\mu_j | a_j, b_j) = \frac{\exp[-1,7 \cdot a_j \cdot (\theta_i - b_j)]}{1 + \exp[-1,7 \cdot a_j \cdot (\theta_i - b_j)]} \quad (7)$$

En esta ecuación, θ_i se refiere al nivel de conocimiento del opositor i -ésimo, mientras que a_j y b_j representan, respectivamente, los valores de discriminación y dificultad de la pregunta j -ésima. A la vista de esta fórmula, es posible afirmar que la probabilidad que tiene un individuo de acertar una pregunta viene dada tanto por el nivel de conocimientos que tiene dicho individuo, como por las características de la pregunta, entendiendo como tales su dificultad y discriminación.

Resultados

La media del baremo académico de todos los médicos presentados a la prueba MIR de 2020 fue de 7,32, con una mediana de 7,46, mientras que, en el caso del subconjunto de los que decidieron introducir sus respuestas a las preguntas de la prueba, su media de baremo fue de 7,58, con una mediana de 7,6. En lo referente a la prueba de 2021, el conjunto de todos los presentados a la prueba tuvo una media de baremo de 7,34, con una mediana de 7,45, mientras que el subgrupo de los que introdujeron sus respuestas en la aplicación informática presentó una media de baremo de 7,65, con una mediana de 7,68. Es decir, en ambos casos, los baremos de los que introdujeron sus respuestas en la aplicación resultaron ligeramente superiores a los de toda la población MIR.

Haciendo uso de la información relativa a los opositores que introdujeron su examen en la herramienta informática utilizada en la presente investigación, se determinó para la prueba de 2020 un valor según la fórmula 21 de Kuder y Richardson de 0,9423, mientras que el resultado en 2021 fue de 0,9485. En lo que se refiere al alfa de Cronbach, los resultados obtenidos fueron de 0,932 y 0,9365 para las pruebas MIR de 2020 y 2021, respectivamente. Por tanto, se puede afirmar que ambas pruebas presentan una buena fiabilidad.

En la prueba de 2020 se anularon dos preguntas, mientras que en la de 2021 se anularon tres. Los análisis efectuados en este trabajo incluyen todas las preguntas propuestas, a excepción de las anuladas. Además, las preguntas propuestas en la prueba MIR de 2020 correspondían a 32 asignaturas distintas, mientras que las propuestas en la prueba MIR de 2021 se agrupaban en 35 asignaturas, las mismas 32 que la convocatoria de 2020, más las asignaturas de Alergología, Anestesiología y Oncología.

De las 183 preguntas no anuladas propuestas en la prueba de 2020, las asignaturas con más preguntas fueron Digestivo, con 13, seguida por Cardiol-

gía y Reumatología, con 12. En la prueba de 2021, de las 207 preguntas no anuladas, la asignatura con mayor número de preguntas fue nuevamente Digestivo, con 13, y también Neumología y Pediatría, con el mismo número de preguntas. Si bien no hubo ninguna especialidad con 12 preguntas en esta prueba, hubo tres especialidades de las que se propusieron 11 preguntas. Estas fueron Cardiología, Ginecología y Obstetricia y Traumatología.

El reparto de preguntas por bloques de asignaturas fue el siguiente. En el MIR de 2020, el 45,36% correspondió a Aparatos; el 9,85%, a Básicas; y el 44,8%, a otras. En el caso de la prueba MIR de 2021, el 41,06% de las preguntas pertenecía a la categoría de Aparatos; el 9,66%, a Básicas; y el 49,28%, a otras, sin que se aprecien, por tanto, diferencias estadísticamente significativas en la proporción de preguntas entre bloques en ambas convocatorias ($\chi^2 = 0,8292$; $p = 0,660604$).

Análisis de las preguntas de las pruebas MIR de 2020 y 2021 por asignaturas

Las tablas I y II muestran el número de preguntas por asignatura junto con los valores medios y desviaciones estándar de sus índices de dificultad, índice de dificultad con corrección de los efectos del azar, índice de discriminación, índice de correlación biserial puntual, índice de dificultad del modelo de dos parámetros de la teoría de respuesta al ítem e índice de discriminación del mismo modelo correspondiente a las pruebas MIR de 2020 y 2021, respectivamente. Tal y como ya se ha indicado con anterioridad, el número de preguntas propuestas en dichas pruebas no ha sido el mismo. A pesar de esto, resulta de interés comparar los valores obtenidos en los promedios de los distintos índices, sobre todo en las asignaturas con un mayor número de preguntas.

En lo que se refiere al índice de dificultad de todas las preguntas del estudio, existe una gran similitud tanto en el valor promedio como en la desviación estándar de las dos pruebas. Así, el promedio de 2020 fue de 0,6853, con una desviación estándar de 0,2249, y el de 2021, de 0,6895, con una desviación estándar de 0,2128. Ambos valores colocan el promedio de la dificultad de las preguntas en la categoría 'aceptable'. Si se realiza el análisis de las asignaturas que en la prueba de 2020 estuvieron presentes en 10 o más preguntas (Digestivo, Cardiología, Reumatología, Ginecología y Obstetricia, Pediatría, Endocrinología, Neurología, Preventiva y Traumatología), se observa cómo la diferencia de valores medios del índice de dificultad está en to-

Tabla I. Prueba MIR de 2020. Número de preguntas por asignatura junto con su valor medio y desviación estándar de sus índices de dificultad, índice de dificultad con corrección de los efectos del azar, índice de discriminación, índice de correlación biserial puntual, índice de dificultad del modelo de dos parámetros de la teoría de respuesta al ítem (TRI) e índice de discriminación del modelo de dos parámetros de la TRI.

	Asignaturas	Número de preguntas	Índice de dificultad	Índice de dificultad con corrección de los efectos del azar	Índice de discriminación	Correlación biserial puntual	Dificultad según la TRI	Discriminación según la TRI
Aparatos	Cardiología	12	0,7989 (0,1232)	0,7308 (0,1712)	0,2652 (0,11)	0,3829 (0,1089)	-1,613 (0,511)	1,156 (0,5819)
	Digestivo	13	0,6149 (0,226)	0,465 (0,3233)	0,2731 (0,0909)	0,3041 (0,1345)	-0,003 (2,454)	0,823 (0,5103)
	Endocrinología	10	0,74 (0,2553)	0,6345 (0,3341)	0,2418 (0,1143)	0,3305 (0,1232)	-0,426 (3,906)	0,9107 (0,4434)
	Enfermedades Infecciosas	6	0,6814 (0,2377)	0,5663 (0,3281)	0,234 (0,0896)	0,2868 (0,1016)	-1,028 (1,909)	0,7356 (0,3899)
	Hematología	5	0,8163 (0,1146)	0,7135 (0,1952)	0,3096 (0,0983)	0,3806 (0,0393)	-1,512 (0,89)	1,0464 (0,2115)
	Nefrología	6	0,7206 (0,2987)	0,6376 (0,4209)	0,2273 (0,1318)	0,3516 (0,2211)	-3,927 (5,309)	1,1845 (0,8741)
	Neumología	9	0,7323 (0,1752)	0,6061 (0,2398)	0,2351 (0,0744)	0,2754 (0,0959)	-1,373 (1,34)	0,6818 (0,3264)
	Neurología	10	0,6146 (0,1995)	0,4655 (0,2564)	0,2661 (0,1598)	0,2783 (0,1419)	-0,023 (2,17)	0,6664 (0,4201)
Básicas	Reumatología	12	0,7408 (0,1528)	0,6513 (0,2403)	0,3015 (0,1512)	0,3881 (0,1665)	-1,345 (1,393)	1,2658 (0,7999)
	Anatomía patológica	6	0,6317 (0,2659)	0,4615 (0,3658)	0,1935 (0,1704)	0,239 (0,1351)	-13,681 (32,342)	0,62 (0,4043)
	Bioquímica	2	0,4179 (0,1995)	0,1491 (0,2293)	0,34 (0,171)	0,2545 (0,0858)	1,51 (1,735)	0,6124 (0,1797)
	Farmacología	5	0,5789 (0,3644)	0,4436 (0,5037)	0,2051 (0,2115)	0,2903 (0,2354)	21,803 (43,748)	0,9481 (0,841)
	Genética	2	0,6454 (0,2012)	0,4597 (0,2213)	0,2319 (0,0731)	0,2468 (0,0112)	-0,728 (1,433)	0,4999 (0,0102)
Otras	Inmunología	3	0,7459 (0,1984)	0,6439 (0,2492)	0,3485 (0,1558)	0,4215 (0,1005)	-1,087 (0,854)	1,3272 (0,731)
	Bioética	1	0,8552	0,7844	0,179	0,2698	-2,702	0,6432
	Cirugía maxilofacial	2	0,877 (0,0195)	0,8291 (0,0342)	0,1877 (0,0724)	0,3348 (0,0672)	-2,453 (0,404)	0,9146 (0,3247)
	Cirugía plástica	1	0,4325	0,1904	0,2966	0,2261	1,283	0,5007
	Cirugía vascular	5	0,8041 (0,1649)	0,7287 (0,2513)	0,2005 (0,1518)	0,3424 (0,1803)	-1,804 (1,269)	1,1376 (0,7893)
	Dermatología	2	0,8196 (0,1835)	0,7418 (0,2642)	0,1822 (0,1278)	0,2598 (0,0149)	-2,453 (1,475)	0,7046 (0,3137)
	Fisiología	4	0,5737 (0,2002)	0,4203 (0,2942)	0,3409 (0,1462)	0,3283 (0,1471)	-0,085 (1,376)	0,8461 (0,5069)
	Geriatría	2	0,9107 (0,111)	0,8633 (0,1459)	0,1145 (0,0856)	0,3339 (0,1016)	-2,623 (0,077)	1,4037 (1,1472)
	Gestión clínica	2	0,8953 (0,055)	0,8647 (0,0718)	0,173 (0,0585)	0,3282 (0,0094)	-2,565 (0,321)	1,0104 (0,1982)
	Ginecología y Obstetricia	11	0,6741 (0,1285)	0,5621 (0,1752)	0,2743 (0,1126)	0,2948 (0,1038)	-1,146 (0,667)	0,675 (0,3519)
	Medicina legal	1	0,8886	0,8354	0,1144	0,2483	-3,461	0,5955
	Oftalmología	2	0,4316 (0,3465)	0,2228 (0,4811)	0,3068 (0,1316)	0,2691 (0,1223)	1,197 (2,959)	0,6969 (0,2086)
	Otorrinolaringología	3	0,7668 (0,1752)	0,6845 (0,2653)	0,213 (0,0196)	0,2887 (0,0892)	-1,692 (1,228)	0,7573 (0,3581)
	Paliativos	1	0,3545	0,1025	0,2555	0,2148	1,751	0,4531
	Pediatría	11	0,588 (0,279)	0,4457 (0,388)	0,2722 (0,1326)	0,3126 (0,147)	1,582 (5,67)	0,9598 (0,7468)
	Preventiva	10	0,6872 (0,262)	0,5703 (0,3604)	0,2183 (0,1413)	0,319 (0,1702)	0,696 (4,666)	0,9202 (0,6057)
	Psiquiatría	7	0,7604 (0,1863)	0,6726 (0,2623)	0,2594 (0,0981)	0,345 (0,0958)	-1,624 (1,264)	0,9624 (0,4792)
Traumatología	10	0,6347 (0,2727)	0,4909 (0,3698)	0,2323 (0,0967)	0,2696 (0,1134)	0,188 (3,428)	0,662 (0,3319)	
Urgencias	7	0,5454 (0,2757)	0,361 (0,3418)	0,1472 (0,0999)	0,1561 (0,096)	-7,914 (17,422)	0,3166 (0,2139)	
Total		183	0,6853 (0,2249)	0,5644 (0,3106)	0,248 (0,1225)	0,3095 (0,1338)	-0,836 (10,483)	0,868 (0,5555)

Tabla II. Prueba MIR de 2021. Número de preguntas por asignatura junto con su valor medio y desviación estándar de sus índices de dificultad, índice de dificultad con corrección de los efectos del azar, índice de discriminación, índice de correlación biserial puntual, índice de dificultad del modelo de dos parámetros de la teoría de respuesta al ítem (TRI) e índice de discriminación del modelo de dos parámetros de la TRI.

	Asignaturas	Número de preguntas	Índice de dificultad	Índice de dificultad con corrección de los efectos del azar	Índice de discriminación	Correlación biserial puntual	Dificultad según la TRI	Discriminación según la TRI
Aparatos	Cardiología	11	0,7508 (0,1746)	0,6447 (0,239)	0,3451 (0,15225)	0,3549 (0,1032)	-1,419 (1,461)	0,9283 (0,3615)
	Digestivo	13	0,7312 (0,1913)	0,5975 (0,2527)	0,2978 (0,09475)	0,3056 (0,093)	-1,056 (1,503)	0,7704 (0,3388)
	Endocrinología	10	0,6625 (0,1431)	0,5314 (0,2174)	0,3026 (0,21407)	0,2941 (0,1766)	-0,306 (2,13)	0,7783 (0,6448)
	Enfermedades Infecciosas	7	0,7181 (0,124)	0,6127 (0,1543)	0,2978 (0,07958)	0,304 (0,0437)	-1,456 (0,805)	0,6752 (0,154)
	Hematología	4	0,8485 (0,0724)	0,7951 (0,0765)	0,4036 (0,08794)	0,5059 (0,0243)	-1,406 (0,268)	1,6875 (0,2424)
	Nefrología	9	0,7792 (0,167)	0,6862 (0,231)	0,302 (0,12145)	0,381 (0,1331)	-1,252 (1,146)	1,1567 (0,6478)
	Neumología	10	0,6826 (0,2853)	0,5367 (0,3738)	0,2871 (0,16292)	0,322 (0,1422)	0,092 (3,368)	0,948 (0,572)
	Neurología	13	0,6333 (0,2025)	0,4767 (0,2915)	0,3956 (0,17677)	0,3534 (0,1343)	1,871 (8,871)	1,0077 (0,5359)
	Reumatología	8	0,6866 (0,1719)	0,5646 (0,2536)	0,3247 (0,16485)	0,3391 (0,162)	-0,618 (1,743)	0,9691 (0,6573)
Básicas	Anatomía Patológica	4	0,8669 (0,1419)	0,788 (0,193)	0,2137 (0,14399)	0,3017 (0,031)	-2,504 (1,149)	0,8793 (0,2941)
	Bioquímica	2	0,3803 (0,1688)	0,1256 (0,2078)	0,239 (0,07257)	0,1724 (0,0379)	2,09 (1,959)	0,4008 (0,0221)
	Farmacología	7	0,6442 (0,2393)	0,4896 (0,3442)	0,3398 (0,11624)	0,34 (0,1408)	-0,054 (2,094)	0,9671 (0,5543)
	Genética	2	0,9057 (0,0149)	0,8293 (0,0823)	0,2652 (0,03341)	0,3878 (0,1305)	-2,098 (0,169)	1,1863 (0,6198)
	Inmunología	5	0,5906 (0,2352)	0,386 (0,3274)	0,2762 (0,05901)	0,2806 (0,1395)	0,409 (1,791)	0,7805 (0,735)
Otras	Alergología	3	0,5939 (0,2728)	0,4201 (0,3836)	0,1789 (0,0853)	0,1873 (0,1252)	0,658 (4,947)	0,4042 (0,3018)
	Anestesiología	2	0,7765 (0,1557)	0,6861 (0,2327)	0,2547 (0,00494)	0,3313 (0,1299)	-1,683 (0,611)	0,8152 (0,4954)
	Bioética	2	0,8769 (0,0551)	0,8077 (0,1016)	0,1932 (0,00099)	0,2665 (0,0738)	-2,924 (0,179)	0,7016 (0,3323)
	Cirugía maxilofacial	2	0,7915 (0,0712)	0,7029 (0,0836)	0,2438 (0,01563)	0,2724 (0,0041)	-2,255 (0,456)	0,5734 (0,0587)
	Cirugía plástica	2	0,502 (0,2372)	0,2895 (0,311)	0,1397 (0,03641)	0,1393 (0,0745)	2,077 (5,198)	0,2534 (0,1055)
	Cirugía vascular	4	0,8472 (0,1492)	0,7257 (0,1516)	0,2686 (0,0907)	0,298 (0,0569)	-2,076 (0,853)	0,7342 (0,2323)
	Dermatología	2	0,6635 (0,0183)	0,5378 (0,0631)	0,4445 (0,14807)	0,3371 (0,0559)	-0,905 (0,544)	0,7978 (0,2284)
	Fisiología	4	0,5548 (0,2953)	0,3696 (0,4031)	0,3163 (0,26152)	0,2929 (0,2327)	-17,5 (34,391)	0,8097 (0,714)
	Geriatría	6	0,66 (0,3028)	0,5064 (0,4197)	0,1728 (0,09131)	0,23 (0,1272)	6,086 (17,586)	0,5907 (0,3885)
	Gestión clínica	1	0,8916	0,8457	0,2463	0,3964	-2,134	1,1252
Otras	Ginecología y Obstetricia	11	0,6517 (0,2601)	0,5203 (0,3577)	0,1976 (0,07073)	0,2581 (0,1245)	-0,203 (2,911)	0,6696 (0,452)
	Medicina legal	1	0,7924	0,6729	0,2101	0,1979	-2,638	0,4302
	Oftalmología	4	0,7341 (0,0995)	0,6034 (0,1619)	0,2474 (0,11001)	0,2467 (0,1151)	-1,823 (1,411)	0,5647 (0,3679)
	Oncología	4	0,6208 (0,1395)	0,4522 (0,1976)	0,2184 (0,08147)	0,2042 (0,0885)	-0,564 (1,513)	0,415 (0,2223)
	Otorrinolaringología	4	0,6699 (0,2625)	0,5312 (0,3665)	0,2525 (0,10592)	0,2814 (0,1285)	-0,593 (2,578)	0,7289 (0,5117)
	Paliativos	2	0,7889 (0,2036)	0,6733 (0,288)	0,2186 (0,06976)	0,2653 (0,0465)	-1,914 (1,565)	0,6437 (0,2686)
	Pediatría	13	0,6166 (0,2416)	0,4774 (0,3434)	0,2996 (0,12961)	0,3101 (0,1377)	0,212 (2,724)	0,8449 (0,4943)
	Preventiva	9	0,6333 (0,2614)	0,4889 (0,3557)	0,2187 (0,14946)	0,2635 (0,174)	0,972 (8,116)	0,7736 (0,7667)
	Psiquiatría	9	0,6866 (0,2527)	0,5681 (0,3323)	0,2919 (0,17451)	0,3258 (0,113)	-0,78 (2,508)	0,874 (0,3935)
	Traumatología	11	0,7476 (0,2364)	0,6308 (0,3)	0,2754 (0,13283)	0,3101 (0,1214)	7,367 (30,408)	0,7856 (0,3819)
Urgencias	6	0,6047 (0,194)	0,4309 (0,2699)	0,3092 (0,13808)	0,2783 (0,1193)	0,226 (2,34)	0,6566 (0,346)	
Total		207	0,6895 (0,2128)	0,5566 (0,2913)	0,2856 (0,13894)	0,306 (0,1291)	-0,167 (9,545)	0,8224 (0,4973)

Tabla III. Prueba MIR de 2020. Número de preguntas por bloque de asignaturas junto con su valor medio y desviación estándar de sus índices de dificultad, índice de dificultad con corrección de los efectos del azar, índice de discriminación, índice de correlación biserial puntual, índice de dificultad del modelo de dos parámetros de la teoría de respuesta al ítem (TRI) e índice de discriminación del modelo de dos parámetros de la TRI.

	Número de preguntas	Índice de dificultad	Índice de dificultad con corrección de los efectos del azar	Índice de discriminación	Índice de correlación biserial puntual	Dificultad según la TRI	Discriminación según la TRI
Aparatos	83	0,7121 (0,2036)	0,6009 (0,2835)	0,2634 (0,1156)	0,3314 (0,1357)	-1,080 (2,561)	0,9448 (0,5726)
Básicas	18	0,6138 (0,2638)	0,4520 (0,3601)	0,2431 (0,1683)	0,2863 (0,1561)	-1,402 (31,027)	0,8148 (0,6006)
Otras	82	0,6739 (0,2347)	0,5521 (0,3222)	0,2334 (0,1173)	0,2924 (0,1248)	-1,079 (6,082)	0,8019 (0,5239)
Todo	183	0,6853 (0,2249)	0,5644 (0,3106)	0,2480 (0,1225)	0,3095 (0,1338)	-0,836 (10,483)	0,8680 (0,5555)

dos los casos por debajo de la décima, salvo en el de Digestivo y Traumatología.

El índice de dificultad con corrección de los efectos del azar tomó en la prueba de 2020 un valor promedio de 0,5644, con una desviación estándar de 0,3106; de igual manera, en la prueba de 2021, su valor promedio fue de 0,5566, con una desviación estándar de 0,2913. Ambos valores colocan el promedio de la dificultad de las preguntas tras la corrección del azar en la categoría 'óptima'. De las nueve asignaturas que tuvieron 10 o más preguntas en la prueba de 2020, únicamente en tres de ellas se encontró una diferencia superior a la décima en los valores medios del índice de dificultad con corrección de los efectos del azar. Estas fueron Digestivo, Endocrinología y Traumatología.

Al observar los valores del índice de discriminación, se aprecia que el valor promedio de 2020 fue de 0,248, con una desviación estándar de 0,1225; mientras que en 2021 el valor promedio fue de 0,2856, con una desviación estándar de 0,13894. Ambos valores colocan el promedio de la discriminación de las preguntas en la categoría 'buena'. En el caso de este indicador, solamente en una de las nueve asignaturas con 10 o más preguntas, Neurología, se produjo una diferencia superior a la décima entre los valores promedios de ambas convocatorias, y la siguiente asignatura con mayores diferencias fue Cardiología, con una diferencia de 0,0799.

El valor promedio del índice de correlación biserial puntual de las preguntas correspondientes a la prueba de 2020 fue de 0,3095, con una desviación estándar de 0,1338, y estos valores fueron muy similares a los de 2021, con una media de 0,306 y una desviación estándar de 0,1291. Ambos valores colocan el promedio de la discriminación de las preguntas medida con el índice de correlación biserial puntual en la categoría 'buena'. En este caso, no

existe asignatura alguna que presente una diferencia entre los promedios de ambos años superior a la décima, y la asignatura con una mayor diferencia es Neurología con 0,0751, seguida de Preventiva, con 0,0555.

La dificultad promedio de las preguntas de la prueba MIR 2020 calculada según la teoría de respuesta al ítem fue de -0,836, con una desviación estándar de 10,483, mientras que, en 2021, dicha dificultad promedio fue de -0,167, con una desviación estándar de 9,545. Si se realiza un análisis detallado de las nueve asignaturas que tuvieron 10 o más preguntas en la prueba de 2020, la que tiene mayor diferencia en el valor de dificultad promedio calculado según la teoría de respuesta al ítem fue Traumatología, seguida de Neurología, Pediatría, Digestivo y Ginecología y Obstetricia.

Finalmente, los valores de discriminación según la teoría de respuesta al ítem de las pruebas de 2020 y 2021 fueron muy similares. De esta forma, el valor promedio de 2020 fue de 0,868, con una desviación estándar de 0,5555, mientras que en 2021 se obtuvo 0,8224, con una desviación estándar de 0,4973. En lo relativo al análisis de las asignaturas con mayor número de preguntas, la asignatura que presentó una mayor diferencia entre las dos convocatorias objeto de estudio fue Neurología, seguida de Reumatología, Cardiología, Preventiva y Endocrinología.

Análisis de las preguntas de las pruebas MIR de 2020 y 2021 por bloques

La tabla III muestra el número de preguntas por bloque de las asignaturas de la prueba MIR de 2020, junto con el valor medio y la desviación estándar de sus índices de dificultad, índice de dificultad con corrección de los efectos del azar, índice de discrimi-

Tabla IV. Prueba MIR de 2021. Número de preguntas por bloque de asignaturas junto con su valor medio y desviación estándar de sus índices de dificultad, índice de dificultad con corrección de los efectos del azar, índice de discriminación, índice de correlación biserial puntual, índice de dificultad del modelo de dos parámetros de la teoría de respuesta al ítem (TRI) e índice de discriminación del modelo de dos parámetros de la TRI.

	Número de preguntas	Índice de dificultad	Índice de dificultad con corrección de los efectos del azar	Índice de discriminación	Índice de correlación biserial puntual	Dificultad según la TRI	Discriminación según la TRI
Aparatos	85	0,7103 (0,1873)	0,587 (0,2582)	0,2808 (0,1486)	0,3403 (0,1294)	-0,461 (3,898)	0,9439 (0,5275)
Básicas	20	0,6751 (0,2432)	0,521 (0,3393)	0,2331 (0,1034)	0,3055 (0,1211)	-0,4184 (2,139)	0,8682 (0,5329)
Otras	102	0,675 (0,2268)	0,5383 (0,3075)	0,214 (0,1261)	0,2775 (0,1245)	0,1269 (13,121)	0,7123 (0,441)
Todo	207	0,6895 (0,2128)	0,5566 (0,2913)	0,2433 (0,1371)	0,306 (0,1291)	-0,1672 (9,545)	0,8224 (0,4973)

Tabla V. Resultados obtenidos de la aplicación del test de Mann-Whitney a la comparación de los valores que toman los índices de dificultad, dificultad con corrección de los efectos del azar, discriminación, correlación biserial puntual, dificultad según la teoría de respuesta al ítem (TRI) y discriminación según la TRI en las pruebas MIR de 2020 y 2021.

	W	Valor de p
Índice de dificultad	40.366	0,927
Índice de dificultad con corrección de los efectos del azar	39.763,5	0,526
Índice de discriminación	40.629	0,886
Correlación biserial puntual	40.120	0,754
Dificultad según la TRI	41.334,5	0,436
Discriminación según la TRI	39.825	0,563

minación, índice de correlación biserial puntual, índice de dificultad del modelo de dos parámetros de la teoría de respuesta al ítem e índice de discriminación del modelo de dos parámetros de la teoría de respuesta al ítem. La tabla IV muestra la misma información relativa a la prueba MIR de 2021.

Al igual que ocurría cuando se realizaba el análisis del conjunto de la prueba o de las asignaturas con mayor número de preguntas, al comparar por bloques los valores promedios de los índices, se observan pocas diferencias. Así, en la tabla V se presentan los resultados obtenidos de la aplicación del test de Mann-Whitney a la comparación de los valores que toman los índices de dificultad, dificultad con corrección de los efectos del azar, discriminación, correlación biserial puntual, dificultad según la teoría de respuesta al ítem y discriminación según la teoría de respuesta al ítem en las pruebas

MIR de 2020 y 2021. En el caso de todas estas variables, no se observan diferencias estadísticamente significativas entre ambas convocatorias. De manera similar, la tabla VI estudia, a través del test de Kruskal-Wallis, si existen diferencias estadísticamente significativas en las medias de esas mismas variables cuando se realizan comparaciones entre bloques de preguntas en las convocatorias de 2020 y 2021. En el caso de la prueba de 2020, no se encontró diferencia alguna, mientras que en la de 2021 sí se encontraron diferencias de significación estadísticas en las medias de bloques en las variables relacionadas con la capacidad discriminativa de las preguntas; éstas son el índice de discriminación, el índice de correlación biserial puntual y el índice de discriminación según la teoría de respuesta al ítem. En todos los casos, los valores más discriminativos de estos índices correspondieron al bloque de aparatos.

Discusión

Los resultados obtenidos permiten afirmar que, en primer lugar, las pruebas MIR de 2020 y 2021 presentaron una alta fiabilidad como escala para la medición de los conocimientos de los presentados a éstas. Este hallazgo concuerda con los resultados obtenidos en estudios previos [10,13].

En lo referente a las limitaciones propias de este estudio, en opinión de los autores, la más importante es que no se dispone de información relativa a las respuestas a las preguntas de todos los médicos que se examinaron en las convocatorias de la prueba MIR de 2020 y 2021, sino de dos muestras de más de 3.000 personas sesgadas hacia aspirantes con mayor mediana de conocimiento medido en preguntas netas y baremo académico que las respecti-

vas poblaciones MIR. Sin embargo, a favor de la utilidad del presente estudio y de la conveniencia de realizarlo, cabe señalar que no existe estudio alguno en los últimos años que haya podido emplear datos relativos a toda la población, y el último ha sido el realizado por Bonillo [8] para las convocatorias de las pruebas MIR de 2005 y 2006.

Otra posible limitación de este trabajo es que únicamente realiza un análisis cuantitativo de la prueba sin tener en cuenta los aspectos cualitativos de las preguntas que, sin duda alguna, también resultan de interés para una buena comprensión de esta. Un trabajo de esta índole relativo a las convocatorias de 2009 a 2013 ya se ha publicado [9].

Finalmente, como fortaleza destacable de este estudio se puede señalar que se presenta un análisis que combina indicadores propios de la teoría clásica de los test con otros de la teoría de respuesta al ítem.

Conclusiones

Para poder acceder a la formación como especialista en España, los médicos necesitan superar la prueba MIR. La preparación suele estar dirigida hacia las tendencias detectadas en el análisis de los exámenes oficiales y el tiempo dedicado a cada una de las asignaturas es proporcional al número de preguntas de cada asignatura en las últimas convocatorias del examen MIR.

La prueba MIR se caracteriza por presentar una alta fiabilidad, así como una gran estabilidad entre convocatorias. También se mantiene en un rango de pequeña variación el reparto de las preguntas de la prueba por bloques.

En opinión de los autores, es importante mantener una estabilidad en el criterio escogido sobre la importancia de cada asignatura, entendiendo como tal la proporción de preguntas de ésta que aparecen en la prueba MIR y que, de forma ideal, debería ser similar a los créditos cursados durante el grado de Medicina. Así, la previsibilidad del peso en el examen de las distintas asignaturas facilitaría a los opositores la preparación de éste, al poder dedicar más tiempo e intensidad de estudio a las asignaturas que el Ministerio considere más importantes a la hora de intentar medir el conocimiento de un médico general en la prueba de acceso a las especialidades médicas. Por tanto, el Ministerio influye, a través del diseño de la prueba, en el tipo de conocimiento que desea primar en los nuevos médicos en el momento en que éstos acceden a los diferentes programas de residencia en los hospitales españoles.

Tabla VI. Resultados obtenidos de la aplicación del test de Mann-Whitney a la comparación de los valores promedio que toman los índices de dificultad, dificultad con corrección de los efectos del azar, discriminación, correlación biserial puntual, dificultad según la teoría de respuesta al ítem (TRI) y discriminación según la TRI en las pruebas MIR de 2020 y 2021 en los tres bloques de preguntas.

	2020		2021	
	H	Valor de <i>p</i>	H	Valor de <i>p</i>
Índice de dificultad	2,05	0,359	0,45	0,798
Índice de dificultad con corrección de los efectos del azar	2,46	0,292	0,6	0,742
Índice de discriminación	4,49	0,106	14,41	0,01
Correlación biserial puntual	4,35	0,113	11,56	0,003
Dificultad según la TRI	1,81	0,405	1,8	0,406
Discriminación según la TRI	3,08	0,214	9,37	0,009

Bibliografía

1. Real Decreto 589/2022, de 19 de julio, por el que se regulan la formación transversal de las especialidades en Ciencias de la Salud. BOE núm. 173, de 20 de julio de 2022, p. 102998-3025.
2. Ley 44/2003, de 21 de noviembre, de ordenación de las profesiones sanitarias. BOE núm. 280, de 22 de noviembre de 2003, p. 41442-58.
3. Sitges-Serra A. El examen MIR, a examen. FEM 2020; 23 (Supl 1): S23-5.
4. Palés-Argullós J. Contenidos y calidad del examen MIR. FEM 2020; 23 (Supl 1): S11-3.
5. Lobato RD, Lagares A, Villena V, Alen JF, Jiménez-Roldan L, Munárriz PM, et al. El método de selección de los residentes en España. Análisis del examen MIR y propuesta de una nueva metodología. Neurocirugía 2015; 26: 53-63.
6. Ibáñez L. EUNACOM: un esfuerzo señero de las facultades de Medicina chilenas. Rev Chil Ped 2013; 84: 365-6.
7. Rivière E, Quinton A, Neau D, Constans J, Vignes JR, Dehail P. Analyse docimologique des premières épreuves classantes nationales informatisées (ECNi) de 2016: propositions d'amélioration. Revue de Médecine Interne 2019; 40: 47-51.
8. Bonillo A. Pruebas de acceso a la formación sanitaria especializada para médicos y otros profesionales sanitarios en España: examinando el examen y los examinados. Gac San 2012; 26: 231-5.
9. Rodríguez-Díez MC, Alegre M, Díez N, Arbea L, Ferrer M. Technical flaws in multiple-choice questions in the access exam to medical specialties ('examen MIR') in Spain (2009-2013). BMC Med Educ 2016; 16: 1-8.
10. 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 test. FEM 2016; 19: 217-26.
11. 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.
12. 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 de 2009 a 2017. FEM 2018; 21: 181-93.

13. Baladrón J, Sánchez-Lasheras F, Romeo-Ladrero JM, Villacampa T, Curbelo J, Jiménez-Fonseca P, et al. The MIR 2018 exam: psychometric study and comparison with the previous nine years. *Medicina (Kaunas)* 2019; 55: 751.
14. 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.
15. Muñiz J. Teoría clásica de los test. Madrid: Pirámide; 2002.
16. Guilbert JJ. Educational handbook for health personnel (offset publication, n.º 35). 1 ed. Geneva: WHO; 1977.
17. Pruebas selectivas para el acceso a plazas de formación de médicos especialistas (1982-1992). Madrid: Ministerio de Sanidad y Consumo; 1993.
18. Pruebas selectivas para el acceso a plazas de formación de médicos especialistas. Validez estructural, diseño y capacidades exploradas (1988-1992). Madrid: Ministerio de Sanidad y Consumo; 1993.
19. Lord FM. Applications of item response theory to practical testing problems. Hillside, NJ: Erlbaum; 1980.