

El examen al examen MIR 2015. Aproximación a la validez estructural a través de la teoría clásica de los tests

Jaime Baladrón, José Curbelo, Fernando Sánchez-Lasheras, José M. Romeo-Ladredo, Tomás Villacampa, Ana Fernández-Somoano

Introducción. En España, para ejercer como médico especialista se requiere poseer la titulación de la correspondiente especialidad médica. Los procedimientos reglados de formación médica especializada imponen la superación de la prueba MIR para acceder a aquélla. Superada la prueba, los programas de formación MIR a los que se accede en los diferentes hospitales y unidades docentes son relativamente homogéneos.

Objetivo. Efectuar una aproximación a la validez estructural del examen MIR de la última convocatoria (2015), realizado el 6 de febrero de 2016, haciendo especial énfasis en sus aspectos medibles.

Sujetos y métodos. La base de datos utilizada en el presente trabajo corresponde a las respuestas a las preguntas del examen MIR de 2015 de una muestra de 3.712 examinandos.

Resultados. El índice de dificultad medio de todas las preguntas fue de 0,6882, mientras que el índice de dificultad corregido fue de 0,5422; el de discriminación, de 0,2492, y el de correlación biserial puntual, de 0,2954. También se aplicó la fórmula número 21 de Kuder-Richardson y el alfa de Cronbach, que dieron como resultado 0,9459 y 0,9579, respectivamente. Los resultados obtenidos se compararon con los de los exámenes MIR comprendidos entre las convocatorias de 1989 y de 1993.

Conclusiones. A la vista de los resultados psicométricos, es posible afirmar que el examen MIR es un examen objetivo, de dificultad y discriminación adecuadas, así como estructuralmente válido.

Palabras clave. Datos numéricos. Estadísticas. Estándares. Estudiantes de medicina. Mediciones educativas. Prueba MIR. Psicometría.

Examination of the MIR exam. An approach to the structural validity through the classical test theory

Introduction. In Spain, to practice as a medical specialist, it is required to have the certification of the appropriate medical specialty. In order to have access to a medical specialist training programme, it is mandatory to overcome the MIR test. Passed the test, MIR training programs accessed in different hospitals and teaching units are relatively homogeneous.

Aim. To approach to the structural validity of the MIR examination of the last call (2015), held on February 6, 2016, with particular emphasis on those measurable aspects of it.

Subjects and methods. The database used in this study corresponds to the answers to the questions of the MIR exam of 2015 of a total of 3,712 examinees.

Results. The average rate of difficulty of all questions was 0.6882, while the corrected index of difficulty was 0.5422, the discrimination of 0.2492 and 0.2954 the value of the point biserial correlation. The formula number 21 of Kuder-Richardson and the Cronbach's alpha were also applied giving as results 0.9459 and 0.9579 respectively. These values were compared with those obtained for the MIR test in the range from the calls 1989 to 1993.

Conclusions. In view of the psychometric results, it can be said that the examination MIR is an objective, with an adequate level of difficulty and discrimination and also structurally valid.

Key words. Educational measurements. Medical students. MIR exam. Numerical data. Psychometrics. Standards. Statistics.

Introducción

En España, para ejercer la profesión médica como médico especialista se requiere poseer la titulación de la correspondiente especialidad médica. Los pro-

cedimientos reglados de formación médica especializada y de obtención del título de médico especialista vienen marcados, en nuestro país, por los siguientes hitos legislativos: la Ley de 20 de julio de 1955 sobre enseñanza, título y ejercicio de las espe-

Director del Curso Intensivo MIR Asturias; Oviedo, Asturias (J. Baladrón). Servicio de Medicina Interna; Hospital Universitario La Princesa; Madrid (J. Curbelo). Departamento de Construcción e Ingeniería de Fabricación; Universidad de Oviedo; Gijón, Asturias (F. Sánchez-Lasheras). Editor del blog MIRentrelazados; Zaragoza (J.M. Romeo-Ladredo). Director del Curso Atención Primaria Asturias; Clínica Oftalmológica Villacampa; Avilés, Asturias (T. Villacampa). CIBER de Epidemiología y Salud Pública, CIBERESP; Instituto de Salud Carlos III; Madrid (A. Fernández-Somoano). IUOPA-Área de Medicina Preventiva y Salud Pública; Departamento de Medicina; Universidad de Oviedo; Oviedo, Asturias, España (A. Fernández-Somoano).

Correspondencia:

Dr. Fernando Sánchez Lasheras. Departamento de Construcción e Ingeniería de Fabricación. Universidad de Oviedo. Pedro Puig Adam, s/n. Sede Departamental Oeste. Módulo 5, 1.ª planta. E-33203 Gijón (Asturias).

E-mail:

sanchezfernando@uniovi.es

Recibido:

14.06.16.

Aceptado:

16.06.16.

Conflicto de intereses:

No declarado.

Competing interests:

None declared.

© 2016 FEM

cialidades médicas, y los Reales Decretos 2015/78, de 15 de junio de 1978, por el que se regula la obtención del título de médico especialista [1], y 127/84, de 11 de enero de 1984, por el que se regula la formación médica especializada y la obtención del título de médico especialista [2].

En 2003 se publicó la Ley de Ordenación de las Profesiones Sanitarias [3] (LOPS, Ley 44/2003, de 21 de noviembre, BOE de 23 de noviembre de 2003). En su artículo 22, la LOPS sentó las bases para un nuevo modelo de acceso a la formación médica especializada. Tal y como se expone en el mencionado artículo, el acceso a la formación sanitaria especializada se efectúa través de una convocatoria anual de carácter nacional. Las normas que regulan la convocatoria anual de dicha prueba son establecidas por el Ministerio de Sanidad, Servicios Sociales e Igualdad, previo informe del Ministerio de Educación, Cultura y Deporte y de la Comisión de Recursos Humanos del Sistema Nacional de Salud. La LOPS, en el mismo artículo, recoge que dicho examen de acceso consiste en una prueba o conjunto de pruebas que evalúa conocimientos teóricos y prácticos y las habilidades clínicas y comunicativas, así como en una valoración de los méritos académicos y, en su caso, profesionales de los aspirantes. En la LOPS se establece reglamentariamente cómo se determina el sistema de adjudicación de todas las plazas ofertadas en la convocatoria anual, que se efectúa de acuerdo con el orden decreciente de la puntuación obtenida por cada aspirante, con las peculiaridades que se establezcan respecto a las plazas de centros de titularidad privada.

El desarrollo del Título II de la LOPS se ha llevado a cabo a través de diversas normas, como el Real Decreto 1146/2006, de 6 de octubre, por el que se regula la relación laboral especial de la residencia para la formación de especialistas en ciencias de la salud [4], el Real Decreto 183/2008, de 8 de febrero, por el que se determinan y clasifican las especialidades en ciencias de la salud [5], o el Real Decreto 459/2010, de 16 de abril, por el que se regulan las condiciones para el reconocimiento de títulos extranjeros obtenidos en Estados no miembros de la Unión Europea [6]. Posteriormente, se promulgó el Real Decreto 639/2014, de 25 de julio, que derogó parte de las regulaciones anteriores y estableció una nueva redacción a algunos de los artículos de los reales decretos anteriormente citados [7].

El examen MIR se convoca anualmente desde 1978 por el Ministerio de Sanidad (hoy Ministerio de Sanidad, Servicios Sociales e Igualdad) y por el Ministerio de Educación y Ciencia, y se realiza en el mismo día y hora en toda España. En la actualidad,

la prueba selectiva consiste en un ejercicio de 225 preguntas de respuestas múltiples y 10 preguntas de reserva. Los candidatos disponen de un máximo de cinco horas para responder a las preguntas. Los contenidos versan sobre las áreas de enseñanza comprendidas en la licenciatura de medicina. La finalidad de la prueba MIR es ordenar a los médicos para la elección de especialidad. Dicha elección se realiza en función tanto del baremo del candidato como de la nota que éste obtenga en el examen MIR (peso ponderado del 90% para el examen y del 10% para el baremo académico). Dada la importancia de la prueba MIR a la hora de ordenar a los aspirantes en la elección de su especialidad, éstos vuelcan gran parte de su atención y esfuerzo en la preparación [8] de dicha prueba desde su último año de carrera.

Una vez superada la prueba, los programas de formación MIR a los que se accede en los diferentes hospitales son relativamente homogéneos, y los hospitales en los que se desarrolla la formación MIR deben someterse de forma periódica a evaluaciones para mantener su acreditación [9]. En el resto de los países del continente europeo existe una gran variabilidad en las formas de acceso a la formación médica especializada, por lo que en la actualidad no es posible hablar de un modelo común europeo de acceso a las especialidades médicas. Una de las fortalezas ya señaladas por otros autores del examen MIR es su transparencia [10], dado que todas las partes implicadas conocen cómo es y cuáles son sus normas de evaluación. Un trabajo reciente [11] señala que el principal y único factor determinante de éxito en la prueba MIR entre los egresados de las facultades de medicina de las universidades públicas españolas es la diferencia entre el número de presentados a la prueba y el de los estudiantes matriculados en primer curso seis años atrás. Esta diferencia representa, en su mayor parte, a los estudiantes que se presentan a la prueba más de un año después de haber finalizado su licenciatura/grado en medicina.

El objetivo del presente trabajo es realizar una aproximación a la validez estructural del examen MIR de la última convocatoria (2015), realizado el 6 de febrero de 2016, haciendo especial énfasis en sus aspectos medibles. La validez estructural de un examen, tal y como la define Guilbert [12], debe tener en cuenta la validez en sí misma, el grado de precisión con que el test mide lo que se pretende medir, la fiabilidad, la objetividad y la relevancia o validez del contenido. Para ello, se calcularán los índices de dificultad, dificultad corregida, discriminación, correlación biserial puntual, la fórmula número 21 de Kuder-Richardson y el coeficiente α de

Cronbach. A partir de los resultados obtenidos, se discutirá acerca de la calidad psicométrica del examen MIR.

Sujetos y métodos

La base de datos utilizada en este estudio corresponde a las respuestas a las preguntas del examen que fueron introducidas por los propios examinados del MIR de 2015 en una aplicación *ad hoc* creada por el Curso Intensivo MIR Asturias. Es decir, todos los médicos que se presentaron a la convocatoria de 2015 del examen MIR pudieron introducir sus respuestas en esta base de datos y conocer una estimación del número de orden que obtendrían en ella con los resultados de su examen y baremo académico. La citada base de datos, una vez eliminados resultados espurios y duplicados, contenía las respuestas de un total de 3.712 examinados. Dicha información anonimizada se utilizó para obtener los resultados que se presentan. A continuación se exponen las métricas fundamentales que se han empleado con el fin de analizar la validez estructural del examen MIR objeto de estudio.

Índice de dificultad

Se entiende por índice de dificultad (DF) de un ítem la proporción de sujetos que lo aciertan de entre todos los que intentan resolverlo [12,13]. Para el cálculo de dicho índice, en el presente trabajo y con el fin de poder efectuar comparaciones con trabajos previos, se utilizará la fórmula que en su momento se empleó en los análisis del examen MIR por parte de Ministerio de Sanidad, Servicios Sociales e Igualdad [12]:

$$DF = \frac{F + D}{N_1 + N_2} \quad (\text{ecuación 1}),$$

donde F es el número de respuestas correctas en el grupo fuerte (se define como grupo fuerte el formado por el 27% de los estudiantes que obtuvo la mejor puntuación en el test); D , el número de respuestas correctas en el grupo débil (se define como grupo débil el formado por el 27% de los estudiantes que obtuvo las peores puntuaciones en el test); N_1 , el número de estudiantes que respondió a la pregunta en el grupo fuerte; y N_2 , el número de estudiantes que respondió a la pregunta en el grupo débil.

Tal y como se puede deducir de la ecuación 1, el valor del índice de dificultad depende de los sujetos

que intentan contestarla. No es, por tanto, una propiedad intrínseca de cada pregunta, sino que depende de la población de individuos que intentan responderla. Así, por ejemplo, si consideramos una pregunta relativa a la especialidad de pediatría, el índice de dificultad que se obtendrá si se evalúa sobre los examinados del MIR será inferior al que resultaría si se les plantease a los médicos asistentes a un congreso de dicha especialidad. Nótese que el índice de dificultad presenta siempre valores entre 0 y 1, y que a menor dificultad de una pregunta, mayor será el valor de su índice.

Debido al menor conocimiento de los alumnos del grupo débil si se compara con el de los del grupo fuerte, es de esperar que el número de alumnos en el grupo débil que conteste a cada una de las preguntas sea menor que el que lo hace en el grupo fuerte. Con el fin de equilibrar la diferencia en el número total de respuestas entre los dos grupos, se ha propuesto [14] un factor de corrección (FC) que se suma a D , consistente en el número de estudiantes del grupo débil que respondería correctamente a cada pregunta en el caso de que la hubiera contestado, y cuya fórmula se detalla a continuación:

$$FC = (N_1 - N_2) \times \frac{D}{N_2} \quad (\text{ecuación 2}).$$

El significado de las variables es el mismo que en la ecuación 1. En el presente trabajo, se ha calculado el valor del índice de dificultad introduciendo esta corrección.

Índice de dificultad con corrección de los efectos del azar

Existe otra fórmula para el cálculo del índice de dificultad que permite corregir los efectos del azar [12]. Esta fórmula es la que se presenta en la ecuación 3. En el presente trabajo se utilizarán ambas fórmulas.

$$ID = \frac{A - \frac{E}{K-1}}{N} \quad (\text{ecuación 3}),$$

donde A es el número de sujetos que aciertan el ítem; E , el número de sujetos que fallan el ítem; K , el número de alternativas de respuesta del ítem; y N , el número total de sujetos que ha realizado el examen.

En el presente trabajo, y siguiendo una práctica generalizada, el cálculo de las fórmulas se realizará

Tabla I. Valoración de los índices de dificultad y de dificultad corregida con los efectos del azar.

Índice de dificultad		Índice de dificultad corregida con los efectos del azar	
Valores	Categorías	Valores	Categorías
> 0,80	Fácil	> 0,80	Muy fácil
> 0,60 - 0,80	Aceptable	> 0,66 - 0,80	Fácil
> 0,50 - 0,60	Óptimo	> 0,33 - 0,66	Óptimo
> 0,30 - 0,50	Aceptable	> 0 - 0,33	Difícil
0 - 0,30	Difícil	-0,33 - 0	Muy difícil

considerando que los sujetos que han dejado el ítem sin responder lo han fallado.

Los valores del índice de dificultad con la corrección del azar son menores que los de dificultad no corregida para cada una de las preguntas analizadas. La clasificación de las preguntas en función de su índice de dificultad se hará teniendo en cuenta los valores de corte que se exponen en la tabla I. Esta tabla, además de emplearse comúnmente en la teoría clásica de los test [12], también se ha utilizado en publicaciones anteriores [14,15] que abordaron el estudio de las pruebas selectivas para el acceso a plazas de formación de médicos especialistas. En la mencionada tabla se incluyen también los valores de corte que se utilizarán para el índice de dificultad corregida, propuestos por los autores del presente artículo.

Índice de discriminación

Se dice que un ítem tiene poder discriminativo si es capaz de distinguir (discriminar) entre los sujetos que puntúan alto en el test y los que puntúan bajo. El índice de discriminación se define como la correlación entre las puntuaciones de los sujetos en el ítem y sus puntuaciones en el test. En el presente trabajo emplearemos dos índices de discriminación; a continuación se presenta la ecuación del primero de ellos [14]:

$$DS = 2 \times \frac{F - D}{N_1 + N_2} \quad (\text{ecuación 4}),$$

donde F es el número de respuestas correctas en el grupo fuerte; D , el número de respuestas correctas en el grupo débil; N_1 , el número de estudiantes que

respondió a la pregunta en el grupo fuerte; y N_2 , el número de estudiantes que respondió a la pregunta en el grupo débil.

El índice de discriminación, definido tal y como se presenta en la ecuación 3, puede tomar valores entre -1 y $+1$. La clasificación de las preguntas según su índice de discriminación se llevará a cabo teniendo en cuenta los siguientes valores, empleados con anterioridad tanto en estudios de teoría clásica de los tests [13] como en publicaciones previas [14,15] que abordaron el análisis del examen MIR:

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

Al igual que en el caso del índice de dificultad, el índice de discriminación se ha calculado utilizando la corrección propuesta en la ecuación 2.

Índice de correlación biserial puntual

La validez discriminativa de las preguntas se estudiará también a partir del índice de correlación biserial puntual. Dicho índice es una aplicación del índice de correlación de Pearson al caso en el que una de las variables es dicotómica y la otra cuantitativa. Su fórmula se puede expresar como [13]:

$$\rho_{bp} = \frac{\mu_p - \mu_q}{\sigma_x} \times \sqrt{\frac{ID}{1 - ID}} \quad (\text{ecuación 5}),$$

donde μ_p es la puntuación media en el test de los sujetos que aciertan el ítem; μ_q la puntuación media en el test de los individuos que fallan el ítem; σ_x la desviación típica de la puntuación total del test; e ID , el índice de dificultad del ítem (proporción de sujetos que aciertan el ítem).

En función de los resultados del coeficiente de correlación biserial puntual, las preguntas se pueden clasificar en las siguientes categorías [16]:

- *Excelente*: > 0,39.
- *Buena*: > 0,30 a 0,39.
- *Regular*: > 0,20 a 0,30.
- *Pobre*: 0 a 0,20.
- *Pésima*: < 0.

En general, se puede afirmar que cuanto mayor sea el valor de este coeficiente, mayor será la relación entre obtener una puntuación alta en el test y el hecho de haber contestado correctamente a la pregunta en cuestión.

Fiabilidad

En el contexto de la teoría clásica de los test, la fiabilidad de un test se define como la consistencia con la que éste es capaz de medir una variable dada. Para el cálculo de la fiabilidad se aplicaron tanto la fórmula número 21 de Kuder-Richardson [13] como el α de Cronbach [17]. La fórmula número 21 de Kuder-Richardson se define por la siguiente ecuación:

$$KR_{21} = \frac{K}{K-1} \times \left(1 - \frac{M \times (K-M)}{K \times S^2} \right) \quad (\text{ec. 6}),$$

donde K es el número de preguntas del test; M , la media aritmética de las puntuaciones obtenidas por todos los examinandos (en este caso se utilizará el número de respuestas acertadas por cada examinando); y S , la desviación estándar de la media.

El coeficiente α propuesto por Cronbach [17] constituye otra forma de acercarse a la fiabilidad. Este coeficiente, más que la estabilidad de las medidas, refleja el grado en el que covarían los ítems que constituyen el test. Es, por tanto, un indicador de la consistencia interna del test de aplicación muy extendido en psicometría. Su fórmula viene dada por:

$$\alpha = \frac{K}{K-1} \times \left(1 - \frac{\sum_{j=1}^n \sigma_j^2}{\sigma_x^2} \right) \quad (\text{ec. 7}),$$

donde K es el número de preguntas del test; $\sum_{j=1}^n \sigma_j^2$, la suma de la varianza de los n ítems; y σ_x^2 , la varianza de las puntuaciones totales en el test.

Los resultados numéricos que se obtienen de la aplicación a un test tanto de la fórmula número 21 de Kuder-Richardson como del α de Cronbach varían entre 0 y 1. Cuanto más se acercan ambas métricas al valor de 1, más fiable se podrá considerar el test. Nótese que, mientras que las fórmulas correspondientes al índice de dificultad, el índice de dificultad corregido, el índice de dificultad con corrección de los efectos del azar, el índice de discriminación, el índice de discriminación corregido y la correlación biserial puntual se pueden aplicar a cada una de las preguntas del test, los dos índices que determinan la fiabilidad se aplican al test en su conjunto.

Resultados

El examen MIR de la convocatoria de 2015 estuvo

compuesto por un total de 235 preguntas, de las cuales las 10 últimas son de reserva y se utilizan sólo para reemplazar a aquellas de entre las 225 primeras que sean anuladas por la comisión calificadoras.

Las preguntas del examen se pueden agrupar como pertenecientes a 33 asignaturas diferentes del grado de medicina. La asignatura que mayor número de preguntas incluye en este examen es aparato digestivo (gastroenterología, hepatología y cirugía digestiva), con un total de 21 preguntas.

El reparto de preguntas por bloques de asignaturas fue el siguiente: el 50,64% de las preguntas correspondió a aparatos (las nueve especialidades médicas y sus correspondientes especialidades quirúrgicas); el 10,64%, a asignaturas básicas (excluyendo las preguntas de microbiología, que se clasificaron dentro de enfermedades infecciosas, y las de bioestadística, que se clasificaron dentro de medicina preventiva); y el 38,72%, a otro tipo de asignaturas. En la tabla II se clasifica cada asignatura dentro del bloque al que pertenece.

Se produjo un total de cuatro anulaciones, las cuales correspondieron a una pregunta de farmacología (pregunta n.º 36), una de cardiología (pregunta n.º 61), una de bioética (pregunta n.º 189) y otra de medicina preventiva (pregunta n.º 205). En los sucesivos análisis no se han tenido en cuenta las cuatro preguntas que fueron anuladas, pero sí todas las demás constitutivas del examen, incluidas las de reserva, es decir, un total de 231 preguntas.

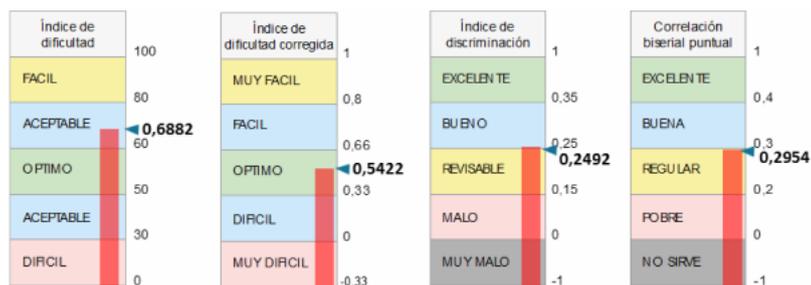
Análisis por preguntas

La clasificación de las preguntas según las categorías del índice de dificultad de la tabla I permite afirmar que, para la muestra de médicos estudiada, un 47,19% de las preguntas del examen MIR de 2015 se puede considerar como de dificultad aceptable; un 9,96%, óptima; un 6,06%, difícil; y el 36,8% restante, fácil.

Los resultados obtenidos con el índice de dificultad con corrección de los efectos del azar clasifican al 31,6% de las preguntas en la categoría de dificultad óptima; el 24,24%, en la de fácil; el 18,61%, en la de difícil; el 20,35%, en la de muy fácil; y el 5,19%, en la de muy difícil. En el caso del examen MIR de la convocatoria MIR de 2015, un total de 12 preguntas presentó un índice de dificultad con corrección de los efectos del azar con valores negativos. Esto no resulta nuevo, como puso en evidencia el estudio de los exámenes de las convocatorias de 2005 y de 2006 [17].

Si se efectúa un análisis pregunta a pregunta relativo a su índice de discriminación, existe en todo

Figura. Índice de dificultad, índice de dificultad con corrección de los efectos del azar, índice de discriminación e índice de correlación biserial puntual.



el examen un total de 71 preguntas (30,74%) de discriminación excelente y otras 43 (18,61%) con discriminación buena, mientras que 53 (22,94%) corresponden a la categoría de discriminación revisable, 57 (24,68%), a la de mala discriminación, y siete (3,03%), a la de muy mala.

Igualmente, el análisis pregunta a pregunta de sus índices de correlación biserial puntual permite afirmar que más de la mitad de las preguntas del examen (51,95%) corresponde a las categorías de discriminación excelente o buena; el 25,54% se clasifica como regular; el 21,65%, como pobre; y el 0,87%, como pésimo (solo dos preguntas).

Análisis por asignaturas

Al realizar el cálculo del índice de dificultad para cada una de las asignaturas (Tabla II), se observa que la mayoría de éstas (24 de un total de 33; 72,7%) presenta un índice de dificultad aceptable, mientras que cuatro son fáciles (anestesiología, genética, cirugía plástica y habilidades comunicativas) y cinco tienen una dificultad óptima (cuidados paliativos, anatomía patológica, cirugía maxilofacial, farmacología y geriatría).

En relación con el índice de dificultad corregido, 26 de las 33 asignaturas se clasifican como de dificultad óptima (78,8%), mientras que tres son muy fáciles (anestesiología, genética y cirugía plástica) y una es fácil (habilidades comunicativas). Las tres asignaturas restantes se clasifican como difíciles (cuidados paliativos, geriatría y fisiología).

En lo relativo al índice de discriminación, 13 de las 33 asignaturas han resultado clasificadas como con buena discriminación, mientras que tres son excelentes (otorrinolaringología, urgencias y dermatología). Otras 13 se enmarcan dentro de la categoría de revisables (fisiología, habilidades comunicati-

vas, gestión clínica, cirugía maxilofacial, farmacología, enfermedades infecciosas, digestivo, genética, neurología, oftalmología, anatomía, oncología y cuidados paliativos). Otras cuatro asignaturas se considera que tienen un índice de discriminación malo (traumatología, cirugía plástica, anestesiología y geriatría).

En relación con el índice de correlación biserial puntual, 12 de las 33 asignaturas que componen el examen se pueden clasificar como buenas, mientras que, a los mismos efectos, tres resultan excelentes (genética, otorrinolaringología y dermatología), 15 se clasifican como regulares, y otras tres, como pobres (cuidados paliativos, geriatría y fisiología).

Análisis por bloques de asignaturas

A continuación se describirán los resultados obtenidos al agrupar las preguntas individuales en los tres bloques de asignaturas previamente descritos: básicas, aparatos y otras especialidades. Tanto en lo relativo al índice de dificultad medio como a la media de la dificultad corregida por los efectos del azar (Tabla III), las preguntas correspondientes a los tres bloques (básicas, aparatos y otros) presentan valores medios de índice de dificultad muy similares. En el caso del índice de dificultad, dichos valores medios están comprendidos entre los 0,6574 del bloque de asignaturas básicas (más difíciles) y los 0,7032 del bloque de otras especialidades (más fáciles). El valor mayor de desviación estándar en los índices de dificultad corresponde a las preguntas del bloque de preguntas correspondientes a asignaturas básicas con un valor de 0,2439, aunque no existen grandes diferencias entre los tres bloques considerados. En el caso del índice de dificultad con corrección de los efectos del azar, el valor menor corresponde también al bloque de asignaturas básicas (0,4921, más difíciles), y el mayor, al de otras especialidades (0,5516, más fáciles). También, en este caso, la mayor variabilidad del índice se produce en el bloque de asignaturas básicas.

Los resultados obtenidos para cada bloque de asignaturas son muy similares en lo relativo al índice de discriminación y al de correlación biserial puntual (Tabla IV), y existen muy ligeras variaciones entre los tres bloques de asignaturas considerados. Las diferencias existentes en los valores de desviación estándar entre bloques resultan también mínimos. Así, por ejemplo, en lo relativo a la correlación biserial puntual, los valores de las preguntas de los tres bloques de asignaturas se encuentran en el entorno de 0,3 y, por tanto, en el límite entre las categorías buena y regular.

Tabla II. Índices de dificultad, de dificultad con corrección de los efectos del azar, de discriminación y de correlación biserial puntual de las preguntas del examen MIR de 2015 agrupadas por asignaturas.

	Asignatura	Número de preguntas	Índice de dificultad	Índice de dificultad con corrección de los efectos del azar	Índice de discriminación	Correlación biserial puntual
Aparatos	Digestivo	21	0,6737 (0,1967)	0,5440 (0,2847)	0,2273 (0,1255)	0,2657 (0,1115)
	Enfermedades infecciosas	17	0,6914 (0,2065)	0,5615 (0,2814)	0,2296 (0,1314)	0,2801 (0,1350)
	Neumología	14	0,7451 (0,1712)	0,6359 (0,2498)	0,2673 (0,1474)	0,3305 (0,1446)
	Cardiología	13	0,6408 (0,2038)	0,4875 (0,2757)	0,2589 (0,1200)	0,2595 (0,0922)
	Nefrología	12	0,6485 (0,2161)	0,4692 (0,2762)	0,2605 (0,1828)	0,2870 (0,1591)
	Neurología	11	0,7350 (0,2104)	0,6113 (0,2717)	0,2109 (0,1594)	0,2719 (0,1320)
	Hematología	10	0,6931 (0,2636)	0,5512 (0,3413)	0,3107 (0,2137)	0,3493 (0,1763)
	Reumatología	10	0,6702 (0,1801)	0,5199 (0,2683)	0,2714 (0,1405)	0,3090 (0,1221)
	Endocrinología	10	0,6224 (0,1574)	0,4762 (0,2122)	0,2601 (0,1789)	0,2621 (0,1397)
Básicas	Anatomía patológica	5	0,5893 (0,3007)	0,4334 (0,4079)	0,2935 (0,2224)	0,3165 (0,1965)
	Inmunología	4	0,6396 (0,2172)	0,4810 (0,3406)	0,3211 (0,0835)	0,3351 (0,0858)
	Anatomía	4	0,7064 (0,1838)	0,5279 (0,2685)	0,1931 (0,047)	0,2241 (0,0298)
	Farmacología	4	0,5993 (0,2637)	0,3528 (0,3223)	0,2360 (0,1331)	0,2170 (0,1121)
	Genética	2	0,9613 (0,0103)	0,8680 (0,0040)	0,2111 (0,0433)	0,4855 (0,0751)
Otras	Preventiva	15	0,7986 (0,1406)	0,6447 (0,2358)	0,2668 (0,1228)	0,3601 (0,0945)
	Pediatría	13	0,6856 (0,1767)	0,5276 (0,2379)	0,3010 (0,1756)	0,3163 (0,1377)
	Ginecología y obstetricia	11	0,7135 (0,2013)	0,6073 (0,2766)	0,2790 (0,1493)	0,3255 (0,1046)
	Psiquiatría	8	0,7411 (0,2330)	0,6296 (0,3164)	0,2760 (0,1361)	0,3643 (0,0592)
	Traumatología	7	0,6406 (0,3839)	0,4575 (0,4704)	0,1305 (0,1046)	0,2423 (0,1472)
	Habilidades comunicativas	6	0,8617 (0,0702)	0,7828 (0,0822)	0,1565 (0,0794)	0,2704 (0,0474)
	Fisiología	5	0,4311 (0,2936)	0,2152 (0,3962)	0,1979 (0,2012)	0,1907 (0,1579)
	Oftalmología	4	0,7533 (0,2105)	0,5706 (0,3101)	0,2044 (0,1302)	0,3086 (0,0823)
	Gestión clínica	4	0,6524 (0,2851)	0,4963 (0,3974)	0,2372 (0,0324)	0,3039 (0,1107)
	Otorrinolaringología	3	0,7781 (0,1506)	0,5970 (0,2141)	0,4170 (0,2068)	0,4142 (0,0681)
	Dermatología	3	0,7525 (0,1102)	0,6273 (0,1503)	0,4007 (0,1355)	0,3984 (0,067)
	Oncología	3	0,6891 (0,3131)	0,5246 (0,4239)	0,1894 (0,0465)	0,2934 (0,1200)
	Paliativos	3	0,5065 (0,0499)	0,3163 (0,0965)	0,1719 (0,057)	0,1609 (0,0432)
Geriatría	3	0,5419 (0,2138)	0,3229 (0,2663)	0,0814 (0,1160)	0,1230 (0,0907)	
Cirugía maxilofacial	2	0,5880 (0,1283)	0,3553 (0,1318)	0,2371 (0,0914)	0,2437 (0,0853)	
Urgencias	1	0,4970	0,3307	0,5174	0,3856	
Cirugía plástica	1	0,9301	0,8598	0,0992	0,2959	
Anestesiología	1	0,9506	0,8831	0,0925	0,2768	
Cirugía vascular	1	0,6427	0,4171	0,2627	0,2567	
Total		231	0,6882 (0,2115)	0,5422 (0,2861)	0,2492 (0,1472)	0,2954 (0,1268)

Tabla III. Índice de dificultad e índice de dificultad con corrección de los efectos del azar de las preguntas del examen MIR de la convocatoria de 2015 para cada bloque de preguntas. Valores medios, desviaciones estándar y clasificación en categorías.

	Índice de dificultad	Índice de dificultad (categorías)				Índice de dificultad con corrección de los efectos del azar	Índice de dificultad con corrección de los efectos del azar (categorías)				
		Fácil	Aceptable	Óptima	Difícil		Muy fácil	Fácil	Óptima	Difícil	Muy difícil
Básicas	0,6574 (0,2439)	7	5	5	2	0,4921 (0,3270)	5	3	5	4	2
Aparatos	0,6812 (0,1982)	41	61	10	6	0,5427 (0,2715)	16	38	37	22	5
Otras	0,7032 (0,2217)	37	43	8	6	0,5516 (0,2973)	26	15	31	17	5
Total	0,6882 (0,2115)	85	109	23	14	0,5422 (0,2861)	47	56	73	43	12

Tabla IV. Índices de discriminación y de correlación biserial puntual de las preguntas del examen MIR de la convocatoria de 2015 agrupadas por bloques de asignaturas. Valores medios, desviaciones estándares y clasificación en categorías.

	Índice de discriminación	Índice de discriminación (categorías)					Índice de correlación biserial puntual	Índice de correlación biserial puntual (categorías)				
		Excelente	Buena	Revisable	Mala	Muy mala		Excelente	Buena	Regular	Pobre	Pésima
Básicas	0,2574 (0,1345)	5	6	5	2	1	0,2978 (0,1391)	5	5	6	2	1
Aparatos	0,2513 (0,1499)	39	23	26	25	5	0,288 (0,1325)	26	28	36	27	1
Otras	0,245 (0,1476)	27	14	22	30	1	0,3043 (0,1173)	24	32	17	21	0
Total	0,2492 (0,1472)	71	43	53	57	7	0,2954 (0,1268)	55	65	59	50	2

Análisis del examen en su conjunto

La figura muestra los valores medios del índice de dificultad, el índice de dificultad con corrección de los efectos del azar, el índice de discriminación y el índice de correlación biserial puntual media de todas las preguntas del examen MIR de 2015. El valor medio del índice de dificultad es de 0,6882 y se encuentra dentro de la categoría de aceptable, mientras que el índice de dificultad corregido con los efectos del azar, con un valor de 0,5422, puede considerarse como óptimo. En lo referente al índice de discriminación, el valor medio obtenido es de 0,2492. Este valor se encuentra en la frontera entre las categorías de bueno y revisable. Algo similar ocurre con el índice de correlación biserial puntual, que, con 0,2954, se encuentra en el límite de las categorías de bueno y regular. El valor obtenido para el índice de dificultad medio de todas las preguntas se comparó a través del test *t* con los valores medios calculados para los exámenes MIR de las convocatorias comprendidas entre 1988-1989 y 1992-1993 [14]. El resultado del test permite afirmar la

existencia de diferencias estadísticamente significativas ($p < 0,001$; $t = -54,45$) entre los valores medios del índice de dificultad de las convocatorias MIR mencionadas cuando se comparan con el valor medio de la dificultad del examen MIR de la convocatoria de 2015. Así, el índice de dificultad de este último es superior al de los exámenes MIR con los que se compara. Nótese que un valor superior del índice de dificultad significa que el examen es más fácil. El mismo tipo de análisis, efectuado para el índice de discriminación, arroja también un resultado estadísticamente significativo ($p = 0,042$; $t = 2,94$). En este caso, el índice de discriminación del examen MIR de 2015 resulta ligeramente inferior al de las convocatorias con las que se compara (todas las comprendidas entre la de 1988-1989 y la de 1992-1993). El promedio del índice de discriminación correspondiente a las pruebas MIR de las convocatorias mencionadas ha sido publicado por el Ministerio de Sanidad [3] y se encuentra comprendido entre 0,25 y 0,29 (bueno), y, por tanto, es muy similar al obtenido en el presente estudio para el examen MIR de 2015.

A continuación, se aplicaron tanto la fórmula 21 de Kuder-Richardson como el α de Cronbach al examen MIR de la convocatoria de 2015 usando la muestra de respuestas disponibles, y se obtuvo como resultados 0,9459 y 0,9579, respectivamente.

El valor obtenido del coeficiente de fiabilidad según la fórmula 21 de Kuder-Richardson es similar al obtenido en un trabajo precedente [14], que obtuvo un valor de 0,97 para las pruebas MIR de las convocatorias 1988-1989, 1989-1990, 1990-1991, 1991-1992 y 1992-1993.

Discusión

El Ministerio de Sanidad publicó estudios sobre la validez estructural de los exámenes MIR de las convocatorias de 1988 a 1992 [14,15]. También cedió los datos necesarios para el análisis de los exámenes de las diferentes profesiones sanitarias de las convocatorias de 2005 y 2006, realizado por Bonillo [18]. Ambos estudios psicométricos se realizaron basándose en el análisis de las respuestas de todos los presentados al examen.

El Ministerio de Sanidad modificó, a partir de la convocatoria 2009, el número de preguntas del examen, que pasó de estar formado por 250 más 10 preguntas de reserva, a las 225 más 10 preguntas de reserva actuales. A partir de esa convocatoria se incluyeron preguntas vinculadas a imágenes. En la convocatoria 2015, el Ministerio de Sanidad modificó el formato de las preguntas, que pasó de cinco opciones de respuesta a cuatro. Como no se ha publicado hasta ahora ningún análisis psicométrico sobre el nuevo formato de examen, decidimos acometerlo, comparando los datos con los publicados de convocatorias anteriores.

En relación con las limitaciones que presenta este estudio, a diferencia de los publicados sobre el formato anterior de examen, no analiza la información correspondiente a todos los examinandos, sino a una muestra de 3.712, que supone alrededor de un tercio del total de 11.227 médicos presentados a la prueba el 6 de febrero de 2016. En la base de datos están incluidas las respuestas anonimizadas tanto de alumnos como de no alumnos del centro de preparación del examen MIR que facilitó los datos. La muestra estudiada tiene un sesgo debido a la inclusión en ella de una mayor proporción de médicos con puntuaciones altas de examen y una menor proporción de médicos con puntuaciones bajas que en la población de todos los presentados al MIR. Este sesgo se expresa en la mediana de preguntas netas de los médicos de la muestra (128,67 pregun-

tas netas), más alta que la de todos los médicos presentados al examen MIR de 2015 (115,67 preguntas netas). Las preguntas netas se calculan restando a las preguntas válidas de cada aspirante un tercio de sus preguntas erróneas. Desde el punto de vista de los autores, este hecho también resulta determinante a la hora de interpretar la comparación del índice de dificultad del examen MIR de 2015 con los de las convocatorias 1988-1992, dado que la puntuación obtenida en el examen MIR por el subconjunto de médicos que han participado en el estudio es superior a la de la población general y, por tanto, para ellos, el examen ha resultado de menor dificultad de la que realmente tiene para el conjunto de la población. Sería conveniente comparar en el futuro estos resultados con el análisis de los datos de todos los presentados al examen, datos que sólo posee el Ministerio de Sanidad, Servicios Sociales e Igualdad.

Otra de las posibles limitaciones del presente trabajo es que sólo realiza un análisis psicométrico del examen MIR de la convocatoria de 2015 desde el punto de vista cuantitativo, sin tener en cuenta otras consideraciones cualitativas acerca de las preguntas. Así, en un trabajo reciente que analiza los errores en la redacción de las preguntas de los exámenes MIR del período comprendido entre los años 2009 y 2013, se observó la existencia de un gran número de preguntas con defectos técnicos de redacción [19]. A pesar de esto, los autores calificaron dichos defectos como menores y afirmaron que el número de defectos técnicos se había ido reduciendo con los años. Según dichos autores, la existencia de estos defectos favorece a los examinandos que conocen las técnicas generales de respuesta y búsqueda de soluciones correctas en preguntas de opción múltiple, más allá de su conocimiento real. A pesar de que la existencia de dicho tipo de preguntas puede afectar a la validez del test, según los autores mencionados, los resultados obtenidos hacen que la validez del test se pueda considerar alta. Estos hallazgos son coherentes con los encontrados en nuestro estudio, que permiten afirmar que el poder de discriminación del examen MIR de la convocatoria de 2015 es solo ligerísimamente inferior al de los exámenes MIR de las convocatorias 1989 a 1993, publicados anteriormente por el Ministerio de Sanidad [15].

Como hemos señalado anteriormente, sería recomendable realizar este 'examen del examen' utilizando las respuestas de todos los examinados presentados y no sólo las de una muestra no aleatoria de ellos. La realización de una evaluación psicométrica previa a la anulación de preguntas por parte

de la comisión calificadora de la prueba permitiría, a nuestro juicio, mejorar el proceso de anulaciones de preguntas que realiza la comisión, excluyendo de la plantilla definitiva tanto las preguntas incorrectas o con varias respuestas válidas, como las pocas preguntas con muy mala capacidad discriminativa, que no servirían para cumplir el objetivo que persigue la prueba MIR de ordenar a los aspirantes según sus distintos niveles de conocimiento médico.

Bibliografía

1. Real Decreto 2015/1978, de 15 de julio, por el que se regula la obtención de títulos de especialidades médicas. BOE núm. 206, de 29 de agosto de 1978. p. 20172-4.
2. Real Decreto 127/1984, de 11 de enero, por el que se regula la formación médica especializada y la obtención del título de Médico Especialista. BOE núm. 26, de 31 de enero de 1984. p. 2524-8.
3. 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.
4. Real Decreto 1146/2006, de 6 de octubre, por el que se regula la relación laboral especial de la residencia para la formación de especialistas en Ciencias de la Salud. BOE núm. 240, de 7 de octubre de 2006. p. 34864-70.
5. Real Decreto 183/2008, de 8 de febrero, por el que se determinan y clasifican las especialidades en Ciencias de la Salud. BOE núm. 45, de 21 de febrero de 2008. p. 10020-35.
6. Real Decreto 459/2010, de 16 de abril, por el que se regulan las condiciones para el reconocimiento de títulos extranjeros, obtenidos en Estados no miembros de la Unión Europea. BOE núm. 107, de 3 de mayo de 2010. p. 39032-56.
7. Real Decreto 639/2014, de 25 de julio, por el que se regula la troncalidad, la reespecialización troncal y las áreas de capacitación específica, se establecen las normas aplicables a las pruebas anuales de acceso a plazas de formación y otros aspectos del sistema de formación sanitaria especializada en Ciencias de la Salud, y se crean y modifican determinados títulos de especialista. BOE núm. 190, de 6 de agosto de 2014. p. 63130-67.
8. Torres M, Cardellach F, Bundó M, Capdevila JA, en representación de la Comisión de Medicina y Especialidades Relacionadas del Consell Català d'Especialitats en Ciències de la Salut. Sistema formativo MIR: propuesta de cambios para la adecuación a las necesidades del modelo sanitario. *Med Clin (Barc)* 2008; 131: 777-82.
9. Llupià A, Costas L, Grau J, Trilla A. Conocimientos, capacidades e intereses de los médicos internos y residentes (MIR) al incorporarse al hospital. *Med Clin (Barc)* 2009; 133: 107-11.
10. Aranda-Sánchez M. Examen al examen MIR. *Med Clin (Barc)* 2016; 146: 443-5.
11. Sentí M, Pérez J, Baños JE. Factores predictores de resultados en la prueba MIR en las universidades públicas. Análisis de la cohorte 2008-2014. *FEM* 2016; 19: 155-60.
12. Guilbert JJ. Educational handbook for health personnel (offset publication, n.º 35). 1 ed. Geneva: WHO; 1977.
13. Muñiz J. Teoría clásica de los test. 2 ed. Madrid: Pirámide; 2002.
14. Pruebas selectivas para el acceso a plazas de formación de médicos especialistas (1982-1992). Madrid: Ministerio de Sanidad y Consumo; 1993.
15. 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.
16. Ebel RL, Frisbie DA. Essentials of education measurement. 5 ed. Englewood Cliffs, NJ: Prentice Hall; 1990.
17. Cronbach LJ. Coefficient alpha and the internal structure of tests. *Psychometrika* 1951; 16: 297-334.
18. 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 Sanit* 2012; 26: 231-5.
19. 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). *Med Educ* 2016; 16: 47.