

Contenidos y calidad del examen MIR

Jordi PALÉS ARGULLÓS

El objetivo de esta presentación es abordar de forma resumida la pertinencia de los contenidos del examen MIR y su calidad técnica.

Objetivo del examen MIR

Para iniciar este análisis, la primera pregunta que se plantea es si los contenidos del examen MIR son pertinentes. La respuesta dependerá de cuál consideremos que es el objetivo o finalidad del examen MIR. Se pueden plantear varias posibilidades en función de lo que consideremos que es este examen.

¿Podríamos preguntarnos si el examen MIR es un instrumento para evaluar los conocimientos que se deberían haber adquirido al finalizar el Grado de Medicina, o un instrumento para medir el grado de preparación que el aspirante tiene como médico generalista que una vez asegurado este nivel iniciaría su formación especializada, o bien un instrumento para medir el grado de preparación del aspirante en las diferentes especialidades médicas? ¿O sería finalmente una 'prueba de fuego' que ha de pasar cualquier médico para ejercer como tal en el futuro? Los contenidos podrían variar en función del objetivo.

Inicialmente, el examen MIR se creó para 'medir el nivel de conocimientos adquiridos por el candidato durante la licenciatura' y, de hecho, evalúa conocimientos propios del médico generalista, pero también conocimientos propios de las diferentes especialidades. Sin embargo, no evalúa competencias importantes como las habilidades de comunicación u otras que por ley deberían evaluarse [1].

A pesar de todo lo que pueda ser, no debe olvidarse que, ante todo, el examen MIR es una prueba que sirve para clasificar a los aspirantes, sin otorgar niveles de suficiencia a aquellos que la superen. Y ello sin tener en cuenta que el nivel mínimo exigido para considerar superada la prueba es muy bajo (el 30% de la media de las diez mejores puntuaciones).

Tampoco se puede obviar que el examen MIR, y este sería un efecto indeseado, mide el grado de entrenamiento de los aspirantes para responder exámenes tipo test.

Departamento de Biomedicina.
Facultad de Medicina y Ciencias de la Salud. Universitat de Barcelona.
Fundación Educación Médica.

© 2020 FEM

Pertinencia de los contenidos del examen MIR

Una vez analizado el objetivo del examen MIR, se pueden analizar sus contenidos y la pertinencia de estos. En las últimas ediciones, el examen MIR consta de preguntas con imágenes (15% en la edición de 2018), preguntas de conocimientos no contextualizadas (34,8% en la edición de 2018) y preguntas contextualizadas, con viñetas clínicas (50,2% en la edición de 2018).

Respecto al contenido de las preguntas, se exploran contenidos comunes a todas las especialidades, algunos básicos que serían pertinentes y algunos accesorios que no lo serían tanto, y por otra parte, se exploran contenidos propios de una especialidad, algunos de ellos pertinentes –es decir, se considera que cualquier médico, al margen de su especialidad, debería tenerlos– y otros accesorios o no pertinentes –es decir, tan específicos de una especialidad que, en todo caso, deberían tener solo aquellos que ejercerían aquella especialidad–.

Este análisis indica que es necesario replantearse seriamente los contenidos del examen MIR en función del objetivo principal del examen, que debería también quedar bien explicitado.

Calidad de las preguntas. Calidad del examen

Otro aspecto a considerar, además de la pertinencia de las preguntas en lo que se refiere a los conocimientos que evalúan, es la calidad de dichas preguntas desde un punto de vista técnico.

Los exámenes tipo test de elección múltiple se han convertido con el tiempo en uno de los instru-

Tabla. Comparativa de defectos técnicos en los exámenes MIR de las convocatorias de los años 2010 y 2018.

ERROR TÉCNICO	FREC.	FREC.
	2010	2018
Enunciados negativos (Deben evitarse, los NO, excepto, indicar la opción falsa o incorrecta, etc)	35	64
Distractores con distinta longitud o grado de concreción o de complejidad. (La respuesta más larga o más completa suele ser la correcta)	10	2
Las opciones no pueden leerse de forma continuada al enunciado. Redacción incorrecta del enunciado y distractores. Todos los términos que puedan estar en el enunciado deben eliminarse de las opciones	10	14
Un distractor es completamente diferente de los otros (le resta valor como opción)	8	4
El enunciado contiene juicios de valor u opiniones personales (¿Cuál cree, considera?, le parece, etc.)	7	9
Se repiten términos en varios de los distractores y no en otros (hace perder valor a estos últimos)	7	2
Distractores mal formulados (sin sentido con respecto al enunciado o la pregunta, o sin correspondencia lógica)	6	
Aparición de adverbios como "puede ser", "podría", "es probable" "siempre" raramente (se deben evitar)	4	12
El enunciado da pistas (p.e. el enunciado propone contestar cual o cuales (da pistas a favor de las opciones que consideren varios aspectos)	3	
Incorrecta colocación de las cifras	2	
Opciones demasiado obvias (les hace perder valor)	2	5
En el enunciado no se dan datos que permitan considerar todos y cada uno de los distractores, por lo que pierden valor como posibles respuestas correctas	2	1
Las opciones son todas demasiado largas	1	9
Las viñetas son muy largas	0	23
Opciones que engloban a otras	1	
Distractores incompletos. (p.e. si uno indica para una prevalencia una ubicación geográfica, los otros también deben hacerlo, ¿Dónde?)	1	
Distractores negativos con enunciados negativos	1	1
	100	132

mentos más utilizados para la evaluación en diferentes disciplinas y muy especialmente en educación médica. Sin embargo, con su uso a lo largo de mucho tiempo, hemos aprendido ciertos aspectos acerca de sus ventajas e inconvenientes.

Por una parte, los exámenes tipo test de elección múltiple tienen poca semejanza con lo que es el trabajo de un profesional médico y poseen un papel limitado en la evaluación del conocimiento factual y en la capacidad de aplicación. También se sabe que su fiabilidad es alta siempre que estén bien diseñados, y que dicha fiabilidad tiene mucha relación con asegurar una buena calidad en su elaboración. Finalmente, sabemos que los formatos basados en escenarios cortos y con cuatro opciones son mejores que los basados en escenarios extremadamente largos y con cinco opciones. En los últimos tiempos, alguno de estos aspectos, como el número de opciones, se han corregido y han pasado de cinco a cuatro.

Respecto a la calidad técnica de las preguntas, en los últimos años se han publicado diferentes trabajos que han analizado dicha calidad en las últimas ediciones del examen MIR. Estos trabajos señalan algunos puntos a tener en cuenta. Así, por ejemplo, se ha detectado que siguen existiendo en los exámenes preguntas con baja calidad psicométrica que no son anuladas [2]. También se ha detectado que la dificultad y la discriminación de las preguntas ha

ido disminuyendo en el período 2009-2017, lo cual obliga a mejorar la calidad psicométrica [3].

Finalmente, otro trabajo ha puesto de manifiesto que existe un porcentaje elevado de preguntas con al menos un defecto técnico (68-72%, convocatorias de 2009 a 2013), y que se sigue manteniendo la presencia de opciones no plausibles, o defectos técnicos relativos a los ítems negativos [4].

De un estudio realizado por el autor (no publicado) sobre los principales defectos técnicos que deben evitarse en la confección de preguntas tipo test según el National Board of Medical Examinees, se deduce que en dos convocatorias separadas entre sí por ocho años (2010-2018) persisten, a pesar del tiempo transcurrido entre ambas, importantes defectos. Estos errores técnicos se aprecian en la tabla. Entre los más frecuentes cabe destacar la presencia de enunciados negativos. También destaca un gran número de preguntas con viñetas excesivamente largas. En total, en el examen de la convocatoria de 2010, se identificaron 100 defectos técnicos en todo el examen, y en la convocatoria de 2018, 132 defectos. Todo ello indica que la calidad técnica de las preguntas no mejora con las convocatorias y que debería hacerse un esfuerzo para mejorarlal porque, en caso contrario, el examen pierde fiabilidad.

Esto pone en cuestión el proceso de confección del examen, que sin duda requeriría una sustancial mejora con la participación de expertos en la elaboración de este tipo de tests.

Conclusiones

En resumen, de todo lo expuesto se pueden derivar algunas las siguientes recomendaciones para la mejora del examen tipo test desde el punto de vista técnico. Se debería:

- Definir claramente el objetivo principal del examen, si existe (además de ordenar los candidatos), y confeccionarlo en consecuencia.
- Ajustar la pertinencia de los contenidos en función de los objetivos.
- Confeccionar exámenes que evalúen mejor la contextualización de los conocimientos y la competencia clínica.
- Arbitrar mecanismos para evaluar las habilidades (comunicativas) y los procedimientos.
- Mejorar el proceso de elaboración de las preguntas, introduciendo un procedimiento de validación de su calidad.
- Evitar los defectos técnicos en muchas preguntas.
- Hacer participar en la confección, además de expertos en contenidos, a expertos en psicometría.

- Tender hacia exámenes con otros tipos de preguntas y diferentes formatos (*online* e informatizados), que permitan evaluar aspectos que actualmente no se evalúan.

Referencias

1. Ley 44/2003 de Ordenación de las Profesiones Sanitarias. Boletín Oficial del Estado, n.º 280, de 22 de noviembre de 2003.
2. Sánchez-Lasheras F, Baladrón J, 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.
3. 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.
4. 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: 47.