

## AUTORES

**Carlos Sunyer**  
UC3M

El autor agradece a **Lucía Cobrer** su ayuda en la elaboración del texto. También agradece los comentarios y sugerencias de **Marina Díez-Rituerto**, **Javier Gardeazabal**, **Nagore Iriberr** y **Pedro Rey-Biel**.

Todos los errores o inexactitudes del texto son responsabilidad del autor. Visualización a cargo de **Lucía Cobrer**.

Línea de investigación

Educación

Dirigida por **Lucas Gortázar**

# La brecha de género en los exámenes de acceso: el caso del MIR, y qué podemos hacer para reducirla

EsadeEcPol Insight #46 Marzo 2023

## IDEAS CLAVE

- Cuando dos personas salen de la carrera de Medicina con la misma nota, las mujeres suelen obtener mejores resultados en la prueba MIR en la parte baja del ranking, pero a medida que se sube en el ranking, las mujeres obtienen peores resultados que los hombres.
- Parte de esta brecha se debe a una mayor propensión de las mujeres a dejar preguntas en blanco, prefiriendo la no respuesta a arriesgarse a descartar algunas de las opciones en el test, con la consiguiente penalización.
- En la misma línea, cuanto más competitiva es la prueba (es decir, cuantos más candidatos hay por cada plaza disponible), peores suelen ser los resultados de las mujeres en comparación con los hombres, especialmente entre los candidatos con mejores resultados.
- Esta evidencia, que se repite en otras pruebas de acceso similares, sugiere reconsiderar el sistema de puntuación, reducir el peso de estas pruebas sobre la nota de acceso final, e incrementar la coordinación entre el número de graduados en la carrera y las plazas disponibles.

## RESUMEN EJECUTIVO

Los exámenes competitivos son objeto de intenso debate tanto público como científico, ya que distribuyen oportunidades basadas en el mérito de las respuestas, y puede penalizar a ciertos perfiles. En este insight se analiza detalladamente uno de los casos más paradigmáticos: el del MIR en España, una prueba altamente competitiva que determina el futuro laboral de más de 10,000 médicos al año, y su efecto en la igualdad de género. Para ello, se utilizaron tres estudios recientes que examinaron información individual de las pruebas del MIR durante los últimos años: Conde-Ruiz et al. (2020), Díez-Rituerto et al. (2023) y Sunyer (2023).

Estos estudios muestran que, aunque en la parte baja de la distribución de resultados del MIR las mujeres superan a los hombres con la misma nota de expediente, en la parte alta las mujeres obtienen peores resultados. En la convocatoria de 2022, las mujeres con las notas más altas obtuvieron hasta 3 puntos menos que los hombres con igual expediente. Esta diferencia equivale a entre 500 y 600 puntos en el ranking final, lo que significa que las mujeres obtuvieron especialidades menos demandadas que los hombres: esa diferencia equivale a la que hay entre Cardiología (la 3ª especialidad en agotarse) y Oftamología (la 6ª).

Los estudios permiten explicar de manera acertada este fenómeno. En particular, se ha encontrado lo siguiente:

1. Parte de esta brecha podría ser atribuible a una mayor propensión de las mujeres a dejar preguntas en blanco: aunque en teoría el sistema de puntuación está pensado para que el candidato sea indiferente entre dejar una pregunta en blanco o responderla al azar, cuando se puede descartar al menos una de las opciones disponibles en el test esta equivalencia se rompe, recompensando implícitamente a quien arriesga pudiendo descartar al menos alguna de las alternativas. Esto es más frecuente entre los hombres.
2. El contexto externo también es importante: cuanto mayor es la competitividad en la prueba (es decir, el número de candidatos en comparación con el número de plazas), peor es el desempeño de las mujeres en comparación con los hombres. Esto es especialmente cierto entre los candidatos con mejores resultados.
3. Los dos fenómenos anteriores se reproducen con claridad al analizar el caso de los repetidores (22% del total de candidatos), donde se observa que, en efecto, las mujeres son más propensas a dejar respuestas en blanco, lo que hace que incrementen menos su nota en comparación con los hombres repetidores.

Dados los resultados de estos estudios, que encuentran eco en evidencia comparable para exámenes de acceso similares, para confrontar la brecha fruto del formato sugerimos reconsiderar la penalización de errores en este tipo de pruebas para promover la igualdad de género.

- Una solución que permite seguir midiendo las diferencias entre los candidatos de más nota, y que podría extrapolarse a otros exámenes, sería un sistema de pérdida de puntos donde preguntas en blanco o incorrectas penalizan, y los candidatos empiezan con la máxima puntuación posible.

Dos medidas adicionales que podrían ayudar a este objetivo, relacionadas con el contexto en que tienen lugar esta clase de pruebas, son:

1. Reducir el peso de este tipo de exámenes en la nota final de selección.
2. Mejorar la coordinación entre el número de graduados (en Medicina, o en otras carreras) y el número de plazas creadas en el sector público.

# 1. Introducción

Cada año, millones de personas en todo el mundo se presentan a exámenes altamente competitivos para demostrar sus habilidades y conocimientos, ya sea en pruebas de admisión universitarias o en procesos de selección laboral, como las oposiciones en España. En estos, incluso pequeñas diferencias en las puntuaciones pueden tener un gran impacto en futuras oportunidades de empleo y salario (Hastings et al., 2013; Bleemer y Mehta, 2022). Dada la importancia de este tipo de exámenes para la vida de tantas personas, es crucial cuestionarse si el formato actual de estas pruebas es el más adecuado para evaluar el conocimiento deseado. Si, por la razón que sea, hay determinados formatos de examen que perjudican especialmente a un determinado perfil por motivos ajenos a las habilidades que se pretenden evaluar, los resultados podrían resultar injustos e ineficientes al no seleccionar necesariamente a los candidatos más idóneos.

Es en este contexto donde los exámenes tipo test han sido objeto de intenso debate en los últimos años. En particular, aquellos exámenes en los que se aplica un sistema de penalización a las respuestas incorrectas con el fin de evitar que los candidatos adivinen al azar. El principal argumento de los críticos con este sistema es que no todos los perfiles tienen la misma predisposición a dejar preguntas en blanco ni tampoco siguen las mismas estrategias, y ello lleva a resultados diferentes por motivos que nada tienen que ver con la habilidad del candidato. Por ser más claros, ¿qué nos dice la confianza o aversión al riesgo sobre la capacidad académica o profesional del candidato? Si dichas características resultasen relevantes, quizás fuese preferible medirlas por otras vías más válidas y fiables desde el punto de vista psicométrico, y, en el caso contrario, ¿por qué, entonces, deberían afectar a los resultados de los exámenes?

Por otro lado, esta inequidad, fruto del sistema de penalización de los exámenes tipo test, podría estar relacionada con algunas de las brechas de género observadas en el mercado laboral. Es por ello por lo que, en los últimos años, ha surgido una voluminosa literatura centrada en responder si el diseño y contexto en el que se realizan los exámenes afecta diferente según el sexo. Por lo general, la mayoría de los estudios documenta cómo las mujeres son, en promedio, más aversas al riesgo y menos confiadas que los hombres en este tipo de pruebas.<sup>1</sup> Esto, en el caso de los exámenes tipo test lleva a que las mujeres respondan menos preguntas que los hombres y, como consecuencia de ello, tengan un rendimiento inferior (Baldiga, 2014; Coffman y Klinowski, 2020).

En este artículo, documentamos el estado de la cuestión sobre este tema para el caso del examen MIR en España. Esta prueba, determinante para el futuro laboral de más de 10,000 médicos al año, es especialmente relevante debido a que se trata de un contexto real, a gran escala, donde la prueba es altamente competitiva, los candidatos están altamente preparados y tienen mucho en juego. Para

---

<sup>1</sup> Estudios ampliamente citados son, por ejemplo, Barber y Odean (2001), Croson y Gneezy (2009) y Charness and Gneezy (2012). No obstante, estudios más recientes, como Filippin y Crosetto (2016) y Nelson (2016), matizan y cuestionan algunos de los resultados anteriores.

ello, recurrimos a tres estudios recientes que documentan el fenómeno y se complementan entre sí: Conde-Ruiz, Ganuza y García (2020); Díez-Rituerto, Gardeazabal, Iriberni y Rey-Biel (2023); Sunyer (2023). En conjunto, los estudios muestran cómo el rendimiento de las mujeres en el MIR es inferior en comparación con el de hombres con conocimientos similares. Esta brecha de género en el rendimiento no se da en todos los candidatos y afecta, sobre todo, a los estudiantes con notas más altas. Una parte de la brecha observada es atribuible al sistema de penalización y, por lo tanto, es exclusiva de los exámenes tipo test. La otra parte de la brecha, sin embargo, tiene que ver con el contexto externo y, por tanto, podría repetirse en otros formatos de examen. En concreto, se observa que el grado de competitividad de la prueba y la presión percibida por los candidatos tiene efectos diferentes según el género.

## 2. La brecha de género en el MIR

### ¿QUÉ ES EL EXAMEN MIR?

El MIR es un examen tipo test que se exige a los graduados en Medicina que deseen acceder a una plaza de médico especialista en formación en el Sistema Nacional de Salud. Una vez realizada la prueba, los graduados son clasificados en un ranking donde el MIR tiene un peso del 90% y el expediente académico un 10%. En función de la posición alcanzada en el ranking, los candidatos eligen secuencialmente dónde quieren pasar su periodo de residencia. Dado que, en los últimos años, solo se han ofertado plazas para entre el 50-60% de los candidatos presentados y no todas las especialidades tienen la misma oferta y demanda, se trata de un examen especialmente competitivo. Desde que se consolidó este sistema de selección, en 1984, el formato del examen ha sido modificado numerosas veces.<sup>2</sup> En la actualidad, el examen consta de 200 preguntas, con un tiempo asignado de 4,5 horas. Por cada pregunta, hay cuatro opciones y solo una es correcta. Cada acierto suma tres puntos y los errores son penalizados con un punto.

### EL MIR EN LOS ÚLTIMOS AÑOS<sup>3</sup>

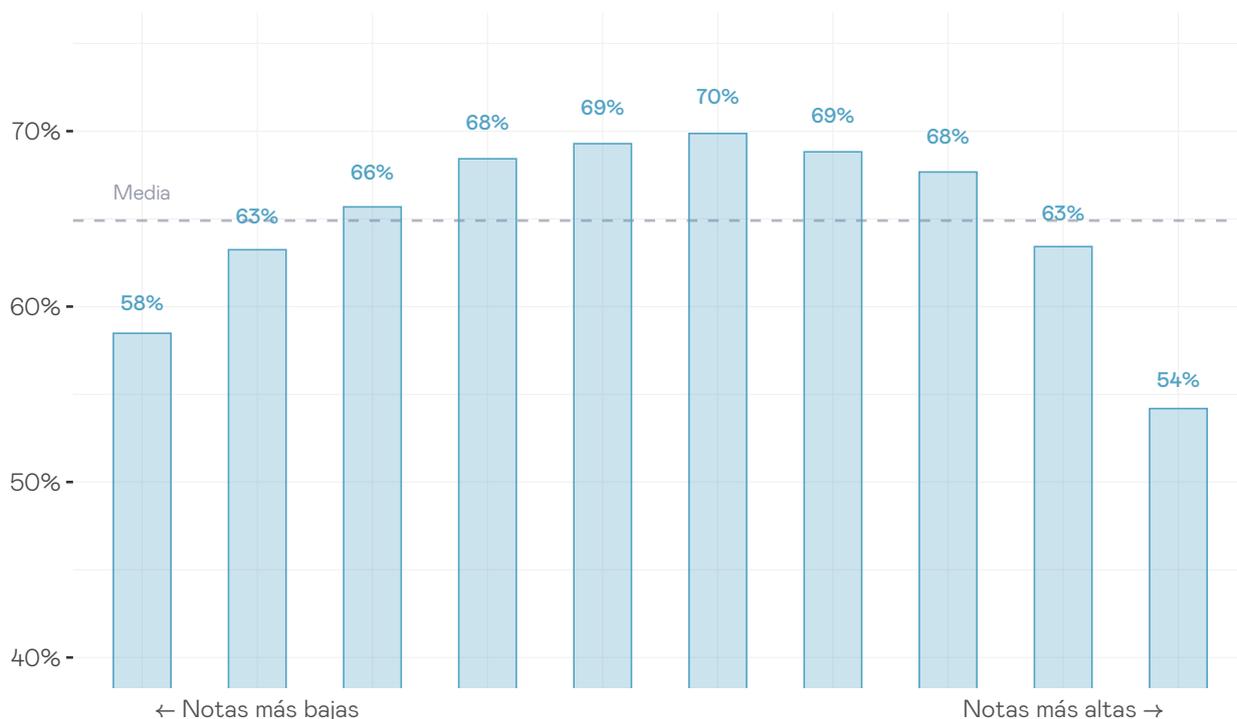
Cuando se presentan al MIR, hombres y mujeres lo hacen con un nivel de conocimiento muy parecido. Así, en el examen de 2022, mientras que las mujeres se presentaron con una media de expediente académico de 7,6, los hombres lo hicieron con un expediente académico de 7,5. Sin embargo, como se observa en la **Figura 1**, aunque las mujeres representaron el 65% de los candidatos que se presentaron al MIR durante el periodo 2018-2022, su proporción entre las notas más altas fue inferior a lo esperable. Este hecho afecta sus posibilidades de elección y requiere una explicación. Entre la batería de hipótesis propuestas más habituales en la literatura, hay cuatro que destacan: diferencias de género bajo presión, en la **disposición a adivinar respuestas, en la aversión a la pérdida y en la aversión al riesgo o la confianza.**<sup>4</sup>

2 En la sección 5.1 de Díez-Rituerto et al. (2023) se listan los principales cambios. Para un mayor detalle sobre la historia de la formación sanitaria especializada en España, véase Tutosaus et al. (2018).

3 Los resultados comentados en esta sección, aunque estén elaborados con datos de otros años, fueron observados por primera vez por Conde-Ruiz et al. (2020).

4 Véase Montolio y Taberner (2021), Coffman y Klinowski (2020), Karle et al. (2022) o Iriberry y Rey-Biel (2021).

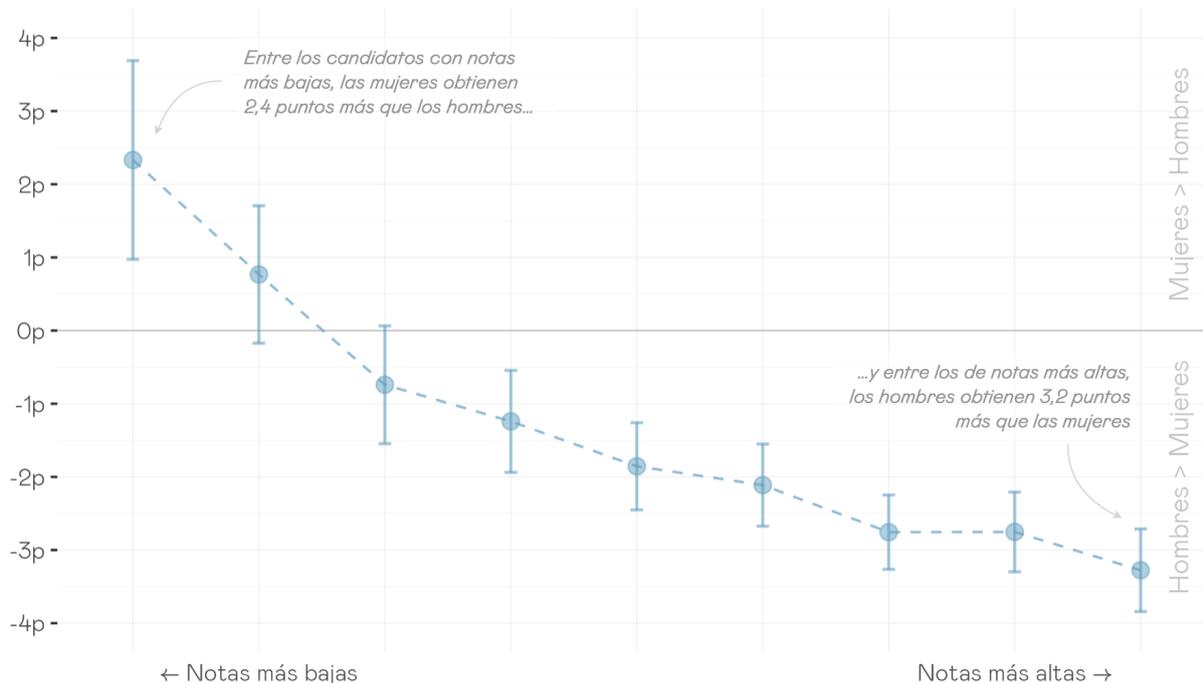
Figura 1. Proporción de mujeres según su nota en el examen MIR



Fuente: Elaboración propia a partir de los datos de las convocatorias 2018-2022 | EsadeEcPol

A continuación, en la **Figura 2** mostramos los resultados cuando se analiza la brecha de género en las notas del examen MIR por deciles de resultados en la prueba. Mientras que, entre las notas más bajas, las mujeres obtienen algo más de 2 puntos por encima que los hombres, esa diferencia rápidamente se torna negativa, hasta alcanzar los 3 puntos por debajo entre las notas más altas. Pero conviene no olvidar que el examen MIR solo es una parte de la nota final, y que lo que más interesa a los candidatos es la posición alcanzada en el ranking, que es lo que determina en última instancia sus posibilidades de elección. ¿A cuántas posiciones equivale esa diferencia de 3 puntos en la parte alta de la distribución? La respuesta es muy relevante y nos permite comprender hasta qué punto la prueba condiciona oportunidades futuras. Para 2022, se estima que esa diferencia de 3 puntos equivale a entre 500 y 600 posiciones en el ranking, dependiendo de la especificación econométrica utilizada. Para dar una idea de la magnitud de la cifra anterior, es conveniente señalar que de las 45 especialidades de 2022, solo Medicina Familiar y Comunitaria ofreció un número de plazas superior a 500; o, si resulta más visual, es una diferencia equivalente a la distancia que hubo entre la tercera especialidad antes agotada –Cardiología– y la sexta –Oftalmología.

Figura 2. **Diferencia entre mujeres y hombres en la nota del examen MIR**  
Medida en puntos: por decil de notas



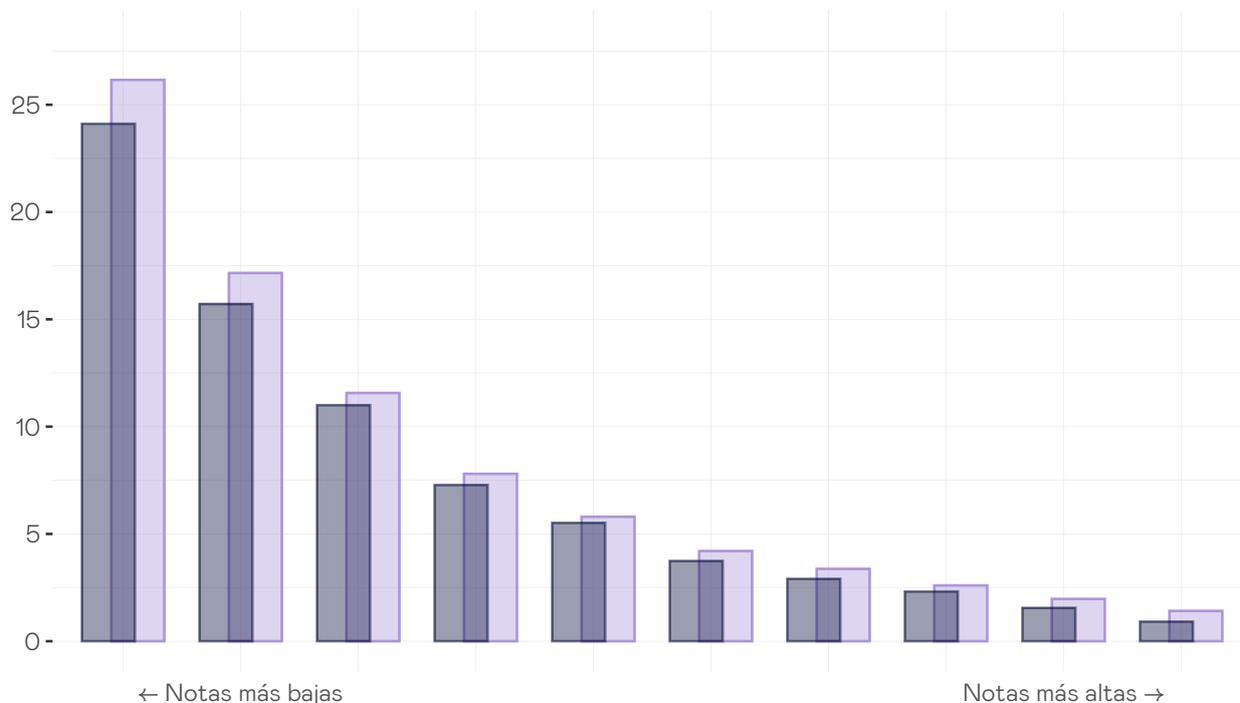
Fuente: Elaboración propia a partir de los datos de las convocatorias 2022 | EsadeEcPol

Parte de esta brecha podría ser atribuible a una mayor propensión de las mujeres a dejar preguntas en blanco. Aunque en la teoría, la penalización se diseña de tal modo que la expectativa entre adivinar al azar una pregunta y dejarla en blanco es la misma (cero), en la práctica, esa equivalencia no pocas veces se rompe, en concreto cuando se es capaz de descartar alguna de las opciones. Por ejemplo, en el MIR, dado que cada pregunta errónea quita un punto, cada acierto añade tres puntos y hay cuatro opciones, responder al azar, en promedio, da como resultado cero puntos ( $3$  puntos cuando se acierta, por su probabilidad  $(1/4)$ , menos  $1$  punto cuando se falla, por su probabilidad  $(3/4)$ ). Exactamente el mismo valor esperado que si se deja la pregunta en blanco. En cambio, si el candidato fuese capaz de descartar una de las opciones, respondiendo al azar, obtendría, de media,  $1/3$  de punto, mientras que, si dejase la pregunta en blanco, seguiría con ningún punto.<sup>5</sup> Aun así, en el caso de que no se pudiese descartar ninguna opción, las simulaciones muestran resultados consistentes con lo observado empíricamente en el MIR. En primer lugar, que el grupo menos proclive a responder (las mujeres) presenta una distribución menos dispersa; segundo, que la diferencia de notas con el grupo más proclive a responder (los hombres) es positiva en los deciles inferiores y negativa en los superiores.<sup>6</sup> La comparativa de las preguntas que quedan en blanco por sexos se muestra en la **Figura 3**. La diferencia, aunque no muy grande, es clara: las mujeres dejan más preguntas en blanco que los hombres y esa diferencia es significativa al 5% para siete de los diez deciles.

5 Cuando se descarta una pregunta, las probabilidades cambian al pasar a decidir la respuesta sobre tres, en vez de cuatro opciones. De este modo, el valor de responder al azar es  $3 \cdot 1/3 - 1 \cdot 2/3 = 1/3$

6 Las simulaciones se describen en el Apéndice A de Sunyer (2023). Estos resultados son consistentes con lo observado por Espinosa y Gardeazabal (2020).

Figura 3. **Media de preguntas sin responder en el examen MIR**  
Diferencias entre mujeres y hombres: por decil de notas



Fuente: Elaboración propia a partir de los datos de las convocatorias de 2018-2022 | EsadeEcPol

## UNA MIRADA CON MÁS PERSPECTIVA

Por el momento, lo anterior sugiere que la mayor propensión de las mujeres a dejar respuestas en blanco podría estar detrás de las brechas de género observadas en el MIR. Sin embargo, esta explicación es parcial por al menos tres motivos. El primero es que, históricamente, no siempre las mujeres han respondido menos preguntas que los hombres en el MIR. El segundo es que la explicación no aborda la causa subyacente de esta diferencia. Y tercero, la explicación tampoco es suficiente para explicar toda la brecha, especialmente entre las notas más altas, donde, de manera generalizada, los candidatos apenas dejan preguntas sin responder (**Figura 3**).

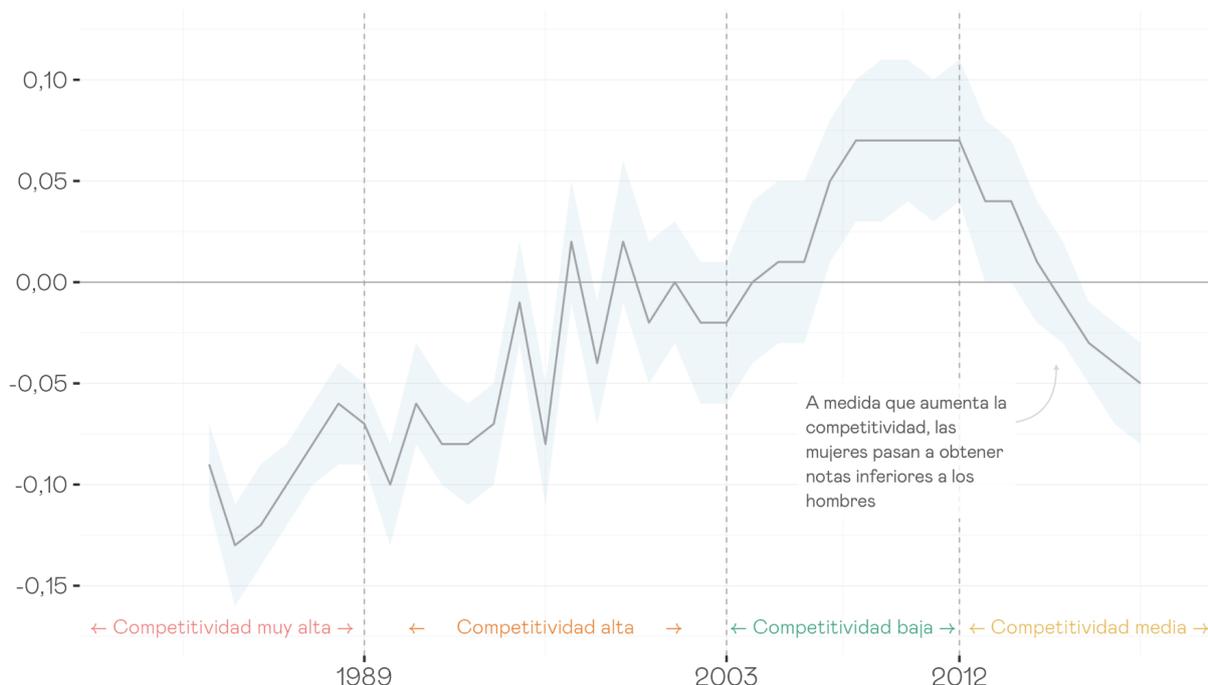
Es en este punto, donde el trabajo de Díez-Rituerto et al. (2023) puede arrojar algo de luz. Con datos del MIR para más de 30 años, desde 1983 a 2019, encuentran que cuanto más alta es el grado de competitividad de la prueba, menor es el rendimiento relativo de las mujeres respecto a los hombres, especialmente entre las notas más altas. Este efecto de la competitividad sobre la brecha de género aumenta la propensión en las mujeres a dejar respuestas en blanco, pero también a cometer más errores. Es decir, la competitividad afecta al rendimiento general y, por tanto, no tiene por qué ser algo exclusivo de la estructura de los exámenes tipo test (donde el comportamiento estratégico con las respuestas en blanco cobra un papel esencial). De hecho, otros estudios realizados en contextos diferentes documentan efectos similares (Gneezy et al., 2003; Niederle y Vesterlund, 2007).

En este trabajo, la competitividad se mide como el cociente entre el número de plazas ofertadas y el número de candidatos. Cuanto mayor sea este cociente, menos competitivo será el proceso. De este modo, los autores distinguen cuatro periodos –mostrados como líneas verticales en la **Figura 4**– según su nivel de competitividad. El primero, entre 1983 y 1988, donde esta fue muy alta, con una probabilidad inferior al 20% (esto es, 20 plazas por cada 100 candidatos); un segundo periodo (1989-2002), donde la competitividad se moderó, situándose entre el 20% y el 40%; un periodo de competitividad bajo (2003-2011), donde la probabilidad fue superior al 60%; por último, un repunte de la competitividad de 2012 a 2019, donde ésta pasó a situarse entre el 40% y el 60%.

La **Figura 4** muestra cómo en periodos donde la competitividad ha sido más alta, mayor ha sido la brecha de género, y viceversa. Este patrón se hace aún más evidente cuando se controla por la competitividad (**Figura 5**).<sup>7</sup> Si bien en la década de los 80, la brecha de género era negativa y muy acentuada, con los años se fue revirtiendo hasta alcanzar valores positivos con el cambio de siglo. En la actualidad está cercana a cero, aunque las mujeres siguen obteniendo un rendimiento inferior entre las notas más altas, tal y como se muestra en la **Figura 2**.

Figura 4. **Diferencia entre mujeres y hombres en la nota del examen MIR**

Según competitividad = número de plazas por candidato; en desviaciones estándar



Fuente: Díez-Rituerto et al. (2023) | EsadeEcPol

7 Para más detalle sobre las especificaciones econométricas, véase el Apéndice A.1 de Díez-Rituerto. (2023).

Figura 5. **Diferencia entre mujeres y hombres en la nota del examen MIR**

Controlando por competitividad = número de plazas por candidato; en desviaciones estándar

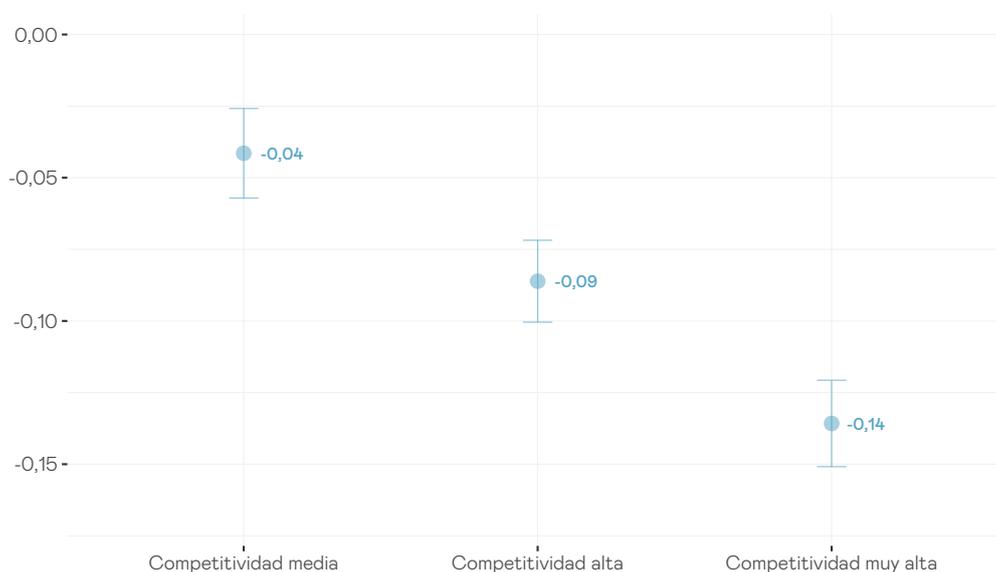


Fuente: Díez-Rituerto et al. (2023) | EsadeEcPol

Cuando distinguimos los periodos según su nivel de competitividad, se obtienen los resultados mostrados en la **Figura 6**. El mensaje es inequívoco: a mayor competitividad, menor es el rendimiento medio relativo de las mujeres en comparación con los hombres. Tal y como se muestra en la Tabla 3 de Díez-Rituerto et al., (2023), estos cambios en el rendimiento asociados a la competitividad se dan tanto en el número de preguntas en blanco como en las preguntas acertadas.

Figura 6. **Diferencia entre mujeres y hombres en la nota del examen MIR**

Respecto a un escenario de baja competitividad = más plazas por candidato; en desviaciones estándar

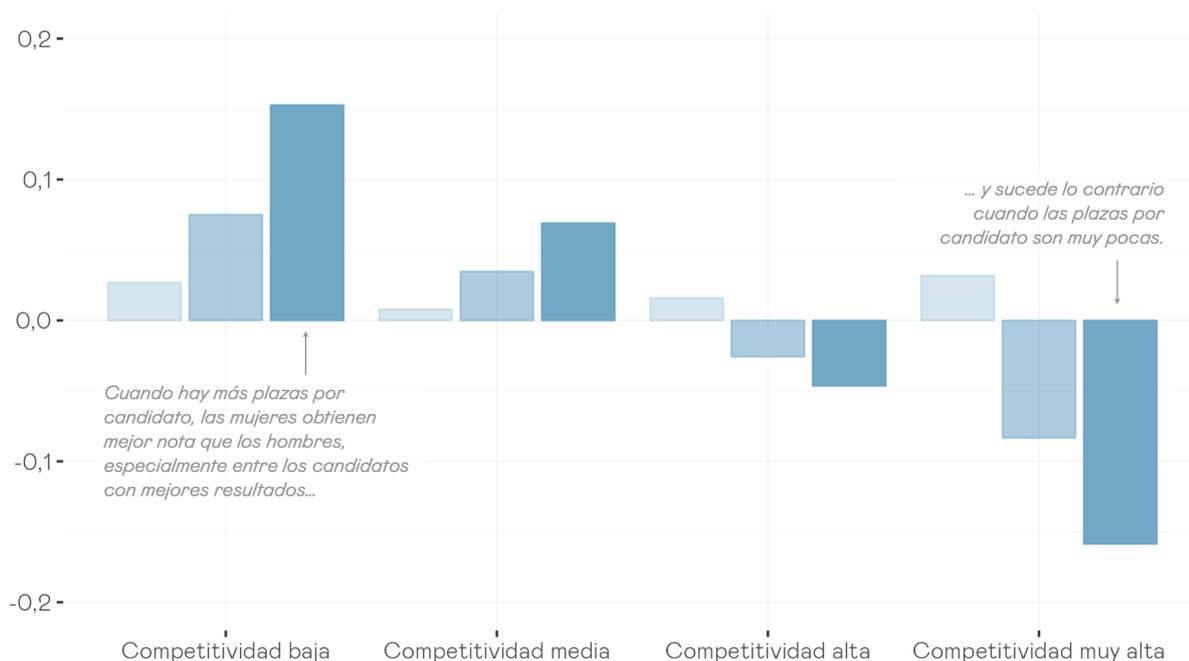


Fuente: Díez-Rituerto et al. (2023) | EsadeEcPol

Por último, es importante examinar si el efecto de la competitividad es heterogéneo según el nivel de conocimiento del candidato, medido este a través del expediente académico del grado en Medicina.<sup>8</sup> Para ello, Díez-Rituerto et al. (2023) evalúan los efectos de la competitividad sobre la brecha de género para tres perfiles de expedientes académicos: el 10% inferior (percentil 10), la mediana (percentil 50) y el 10% superior (percentil 90). Tal y como se muestra en la **Figura 7**, cuanto más alto es el nivel académico del candidato, más sensible se es a cambios en la competitividad.<sup>9</sup> Las implicaciones de estos resultados son importantes puesto que la competitividad es determinante tanto para el tamaño como para el signo de la brecha de género. De este modo, en contextos de competitividad baja, las mujeres con mejores notas obtienen un rendimiento superior al de los hombres con mejores notas, y viceversa, cuando la competitividad es alta. Este resultado está en línea con las observaciones realizadas anteriormente y con resultados obtenidos de contextos diferentes al MIR (Iriberry y Rey-Biel, 2019; Conde-Ruiz et al., 2020).

Figura 7.

**Diferencia entre mujeres y hombres en la nota del examen MIR en función del nivel de competitividad**  
Notas más bajas (percentil 10), intermedias (mediana), más altas (percentil 90); en desviaciones estándar



Fuente: Elaboración propia a partir de Díez-Rituerto et al. (2023) | EsadeEcPol

8 En la última convocatoria de la que se tienen datos (2022), la correlación entre la nota obtenida en el examen y el expediente académico fue superior a 0.60.

9 Esta gráfica es una reelaboración a partir de la Figura 5 de Díez-Rituerto et al. (2023). Nótese que, con el fin de facilitar el mensaje, no se ha tenido en cuenta los intervalos de confianza originales.

## ¿QUÉ OCURRE ENTRE LOS REPETIDORES?

Recapitulando lo comentado hasta ahora, se pueden destacar los siguientes mensajes. Las mujeres, en promedio, responden menos preguntas que los hombres, y esto afecta negativamente a sus resultados. Esta diferencia en la propensión a responder explica parcialmente la brecha de género. La otra parte de la explicación parece estar asociada al grado de competitividad percibido por los candidatos: a mayor competitividad, más negativamente afecta con el rendimiento de las mujeres y esto afecta especialmente a las candidatas con mayor rendimiento. Por un lado, aumenta la propensión a dejar respuestas en blanco, y, por otro, induce a un mayor número de errores (cuando la competitividad es alta).

Para completar el argumento, puede resultar interesante inspeccionar qué ocurre entre los candidatos repetidores de la prueba del MIR. Bajo el supuesto de que a los candidatos no les interesa repetir el examen (la nota obtenida, a diferencia de otras oposiciones, no se guarda, por lo que pierden más tiempo y esto les genera más incertidumbre), es razonable pensar que aquellos que ya han repetido, afrontan la prueba percibiendo más presión que el resto de los candidatos.<sup>10</sup> De ser cierta esta hipótesis, y dado que la literatura relacionada sugiere un comportamiento diferente de hombres y mujeres bajo presión,<sup>11</sup> es posible que, en las notas obtenidas por los repetidores, se produjesen cambios en la brecha de género. Esto es precisamente lo que trata de contrastar Sunyer (2023). Para este ejercicio, el MIR es un examen idóneo debido a su alto número de repetidores.<sup>12</sup> De la base de datos construida por el autor a partir de los resultados publicados entre 2018 y 2022, un 22% de los candidatos repitió el examen al menos una vez, y hasta 231 repitieron el examen los 5 años. Por lo general, los repetidores se caracterizan por haber obtenido unos resultados inferiores a las del resto de candidatos. Pero esto no es siempre cierto. Por ejemplo, en 2021, de los 1.626 repetidores identificados, 351 obtuvieron una nota superior a la de la media de ese año.

Entre una convocatoria y la siguiente, los repetidores incrementaron, en promedio, su nota en 5 puntos, lo que, por ejemplo, para la convocatoria de 2021 equivale a un avance en torno a 1.000 puestos en el ranking final. ¿Hubo alguna diferencia de género en la variación de notas entre convocatorias para los repetidores? En una primera aproximación, la **Figura 8** da respuesta a la anterior pregunta. Si bien no hay diferencias significativas entre los primeros deciles, a partir del decil 5 la brecha se vuelve negativa para las mujeres hasta alcanzar un diferencial de 1 punto. Es decir, entre los repetidores que más incrementaron su nota, las mujeres lo hicieron en menor medida que los hombres. Todo esto una vez se ha controlado por numerosas variables.<sup>13</sup>

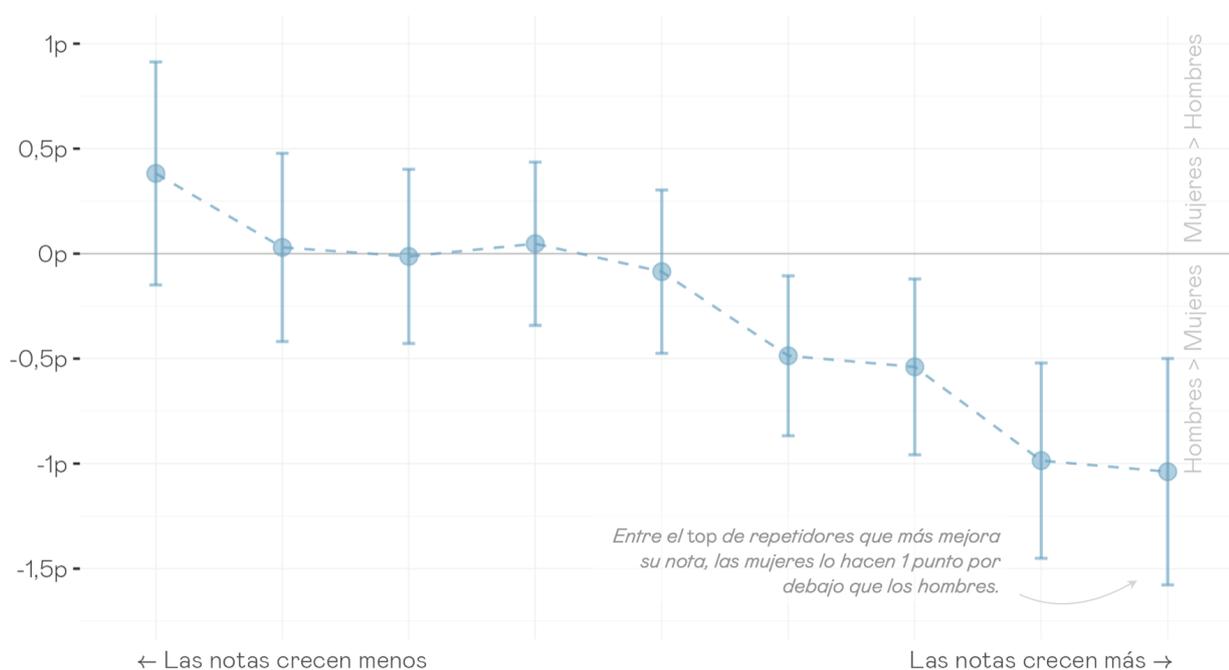
10 De hecho, para diferentes especificaciones econométricas, se observa que los candidatos que repiten habiendo obtenido plaza, tienden a cometer menos errores, pero dejar más preguntas en blanco.

11 Por ejemplo, Montolio y Taberner (2021).

12 En los exámenes de acceso a la universidad de China y Estados Unidos se registran tasas de repetición en torno al 28% y 54%, respectivamente (Goodman et al., 2020; Kang et al., 2021).

13 Nacionalidad, expediente académico, año de la convocatoria, número de veces que se ha repetido y nota que se obtuvo en la anterior convocatoria.

Figura 8. Diferencia entre mujeres y hombres en la mejora de notas cuando repiten el examen MIR  
Según los puntos de mejora respecto el examen anterior; por deciles

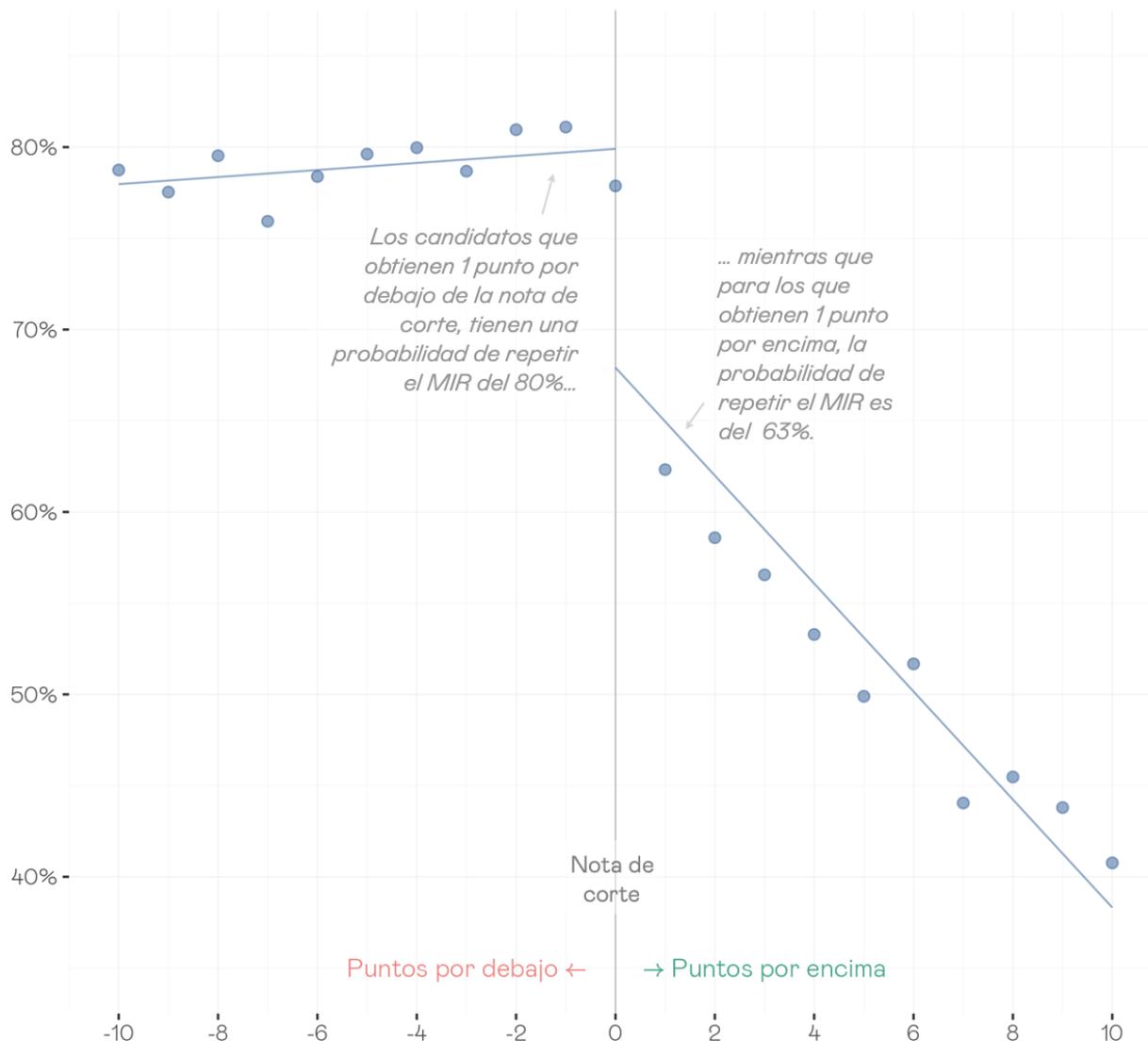


Fuente: Elaboración propia a partir de los datos de las convocatorias 2018-2022 | EsadeEcPol

Sin embargo, repetir el MIR es una decisión voluntaria y endógena, en la que, potencialmente, influyen variables no observadas, como, por ejemplo, el nivel socioeconómico de los candidatos. No tener esto en cuenta, puede sesgar los resultados. Para mitigar este problema, se hace uso de la nota de corte, es decir, la nota del último candidato que accede a una plaza, como potencial fuente de variación exógena. En concreto, para el periodo de estudio, esta nota de corte estuvo determinada por el 35% de la media de las 10 mejores puntuaciones de la convocatoria. La idea para estimar el efecto de repetir es intuitiva. Situarse a un lado u otro de la nota de corte afecta a la probabilidad de repetir el examen. El principal supuesto es que este cambio en la probabilidad venga dado de forma exógena, lo cual es razonable, puesto que, en el momento de la realización del examen, los candidatos desconocen cuál será la nota de corte. Bajo esta condición, es posible comparar a los repetidores a un lado y otro de la nota de corte y asumir que toda diferencia observada es producto del acto de repetir. Es lo que en econometría se conoce como Diseño de Regresión Discontinua.<sup>14</sup> En la **Figura 9** se muestra cómo, en efecto, la probabilidad de repetir da un salto, estimado en un 17%, a un lado y otro de la nota de corte.

14 Para más detalles sobre la metodología de este caso concreto, consúltese la sección 4.1 de Sunyer (2023).

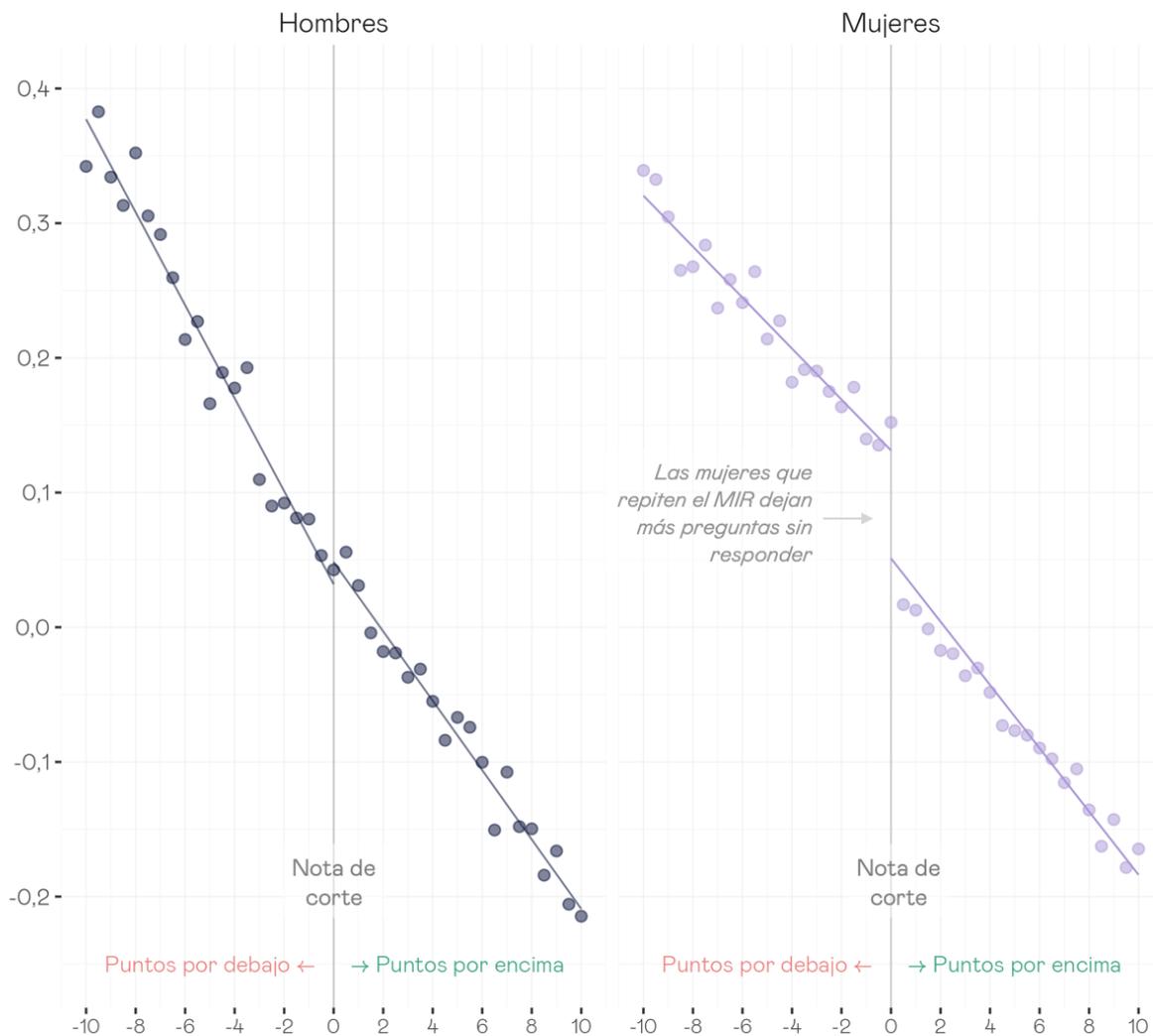
Figura 9. Probabilidad de repetir del examen MIR  
Según la distancia respecto a la nota de corte



Fuente: Sunyer (2023) | EsadeEcPol

Una vez se compara a los repetidores y sus contrafactuales en torno a la vecindad de la nota de corte, se obtiene lo siguiente. Mientras que ambos sexos reducen en proporciones parecidas el número de respuestas incorrectas, las mujeres repetidoras pasan a dejar más respuestas sin responder, tal y como se puede observar en la **Figura 10**. Esta diferencia en el número de respuestas en blanco tiene como consecuencia que las mujeres repetidoras incrementen su nota menos que los hombres repetidores. Este resultado refuerza la hipótesis de partida de que las mujeres son, en promedio, más propensas a dejar respuestas en blanco que los hombres, y que parte de esa propensión es atribuible a diferencias en el comportamiento bajo presión.

Figura 10. Efecto de repetir el MIR sobre el número de preguntas que se dejan en blanco  
Según la distancia respecto a la nota de corte; medido en desviaciones estándar



Fuente: Sunyer (2023) | EsadeEcPol

### 3. ¿Cómo promover una mayor igualdad de género a través del MIR?

En los últimos años, se ha producido un intenso debate acerca de si los exámenes tipo test con penalización a las respuestas incorrectas perjudican a ciertos perfiles de persona por razones ajenas a las evaluadas. Este debate ha alcanzado la arena pública, dando lugar a importantes cambios en el formato de exámenes tan relevantes como la prueba de acceso a la universidad en Estados Unidos. En este *insight*, analizamos la brecha de género para el caso del MIR.

En primer lugar, mostramos cómo la brecha de género es negativa para las mujeres y se produce, principalmente, en las candidatas con un rendimiento superior. La principal implicación de esta diferencia de resultados es un menor número de mujeres entre los hospitales y especialidades más demandados por razones potencialmente ajenas a sus conocimientos como médicos. Esta brecha de género se produce por al menos dos vías. La primera, está relacionada con el formato del examen, y consiste en una mayor propensión de las mujeres a dejar preguntas en blanco. La segunda, tiene que ver más con el contexto en el que se da la prueba, y consiste en diferencias de comportamiento según el grado de competitividad de la prueba o de la presión percibida. En concreto, mostramos cómo el grado de competitividad de la prueba afecta al rendimiento diferente según el sexo, perjudicando especialmente a las mujeres con mejor expediente. Esta diferencia en el comportamiento según el grado de competitividad afecta tanto a la propensión a dejar respuestas en blanco como a la tasa de aciertos. Finalmente, obtenemos que la presión percibida es otro factor relevante a partir de un análisis sobre cómo repetir afecta al comportamiento según el sexo. Se muestra que, aunque tanto hombres como mujeres reducen en igual medida sus errores al repetir, las mujeres tienden a responder menos preguntas, lo cual provoca que su nota aumente menos.

En conjunto, los resultados sugieren que, tanto el diseño del MIR como el contexto en el que este se presenta, afectan diferente a hombres y mujeres. Por tanto, se tratan de factores relevantes de cara a la elaboración de políticas públicas enfocadas a promover la igualdad de género.

Respecto al diseño de los exámenes tipo test y su impacto sobre la propensión a dejar preguntas en blanco, proponemos dos alternativas. O bien modificar la prueba mediante la eliminación de la penalización de los errores, o bien modificar el sistema de puntuación.

En el caso de la eliminación de la penalización a los errores, existe evidencia del efecto que esta política tuvo sobre la brecha de género en Chile y su equivalente a la EvAU (Coffman y Klinowski, 2020). Esta medida redujo en un 70% la diferencia en el número de preguntas en blanco y, como resultado, la brecha de género en el resultado final del examen se redujo un 13%, sobre todo entre los estudiantes con notas más altas. Así mismo, Karle et al. (2022), con un experimento a escala más

reducida, también muestran resultados positivos para esta medida. Por lo general, son abundantes los estudios que abogan por eliminar las diferencias en la puntuación para las preguntas en blanco y las respuestas erróneas, también para el caso particular de las recompensas a las preguntas en blanco (Iriberry y Rey-Biel, 2021). El principal inconveniente de esta política es que la penalización de errores, aunque afecte heterogéneamente a la población de candidatos, es un mecanismo efectivo para mejorar la precisión en las pruebas como el MIR. Eliminarla, por tanto, implica perjudicar a los candidatos más capacitados, al resultar más complicado diferenciar entre grados de conocimiento (Funk y Perrone, 2017; Akyol et al., 2022).

La otra alternativa, recientemente planteada, tiene que ver con cómo se contextualiza el puntaje, lo que en la literatura se conoce como *framing*. Balart et al. (2022) proponen un marco de pérdida donde los candidatos empiezan con la máxima puntuación posible y tanto preguntas en blanco como respuestas incorrectas penalizan. Aunque matemáticamente este sistema sea equivalente a la penalización de errores, efectuando este experimento, los autores encuentran que las preguntas en blanco se reducen entre un 18-20%.

Sin embargo, tal y como sugieren los resultados mostrados en este *insight*, el diseño del examen no es el único factor determinante de la brecha de género entre notas. Otras características, ajenas al diseño del examen, como el grado de competitividad de la prueba o la presión percibida por los candidatos, también afectan diferente según el género.

En este sentido, una alternativa destinada a incrementar la igualdad entre hombres y mujeres podría consistir en reducir el peso del MIR en la nota final. Esa es al menos la perspectiva adoptada por los estudios que han analizado los efectos de la presión percibida por los candidatos, usando el peso del examen en la nota final como *proxy* (Azmat et al., 2016; Montolio y Taberner, 2021). Un potencial problema de esta medida es que podría inducir selectivamente a algunas universidades a inflar el expediente académico de sus alumnos, derivando así en otro tipo de desigualdad.

Por último, para reducir el sesgo originado por la competitividad, también sería deseable incrementar la coordinación entre el número de graduados en Medicina y el número de médicos que el Sistema Nacional de Salud es capaz de absorber. Esta propuesta, realizada con anterioridad por numerosos organismos,<sup>15</sup> requiere de una planificación centralizada, adaptada a las necesidades cambiantes del sistema, tal y como se hace en otros países con sistemas similares, véase Francia. Con esta medida, previsiblemente, se mejoraría el ajuste en el cociente entre el número de plazas ofertadas y el número de candidatos, lo que, a su vez, reduciría el grado de competitividad de la prueba.

Sea cual fuere la medida adoptada, sería conveniente implementarla gradualmente y facilitando las herramientas disponibles para su evaluación, especialmente apostando por la facilidad en el acceso a los datos necesarios.

---

15 Véase: <http://lgestioimporta.cat/arees/brechas-de-genero-en-el-acceso-a-plazas-mir/?lang=es>

# Referencias

- Azmat, G., Calsamiglia, C., & Iriberry, N. (2016). Gender differences in response to big stakes. *Journal of the European Economic Association*, 14(6), 1372-1400.
- Balart, P., Ezquerro, L., & Hernandez-Arenaz, I. (2022). Framing effects on risk-taking behavior: evidence from a field experiment in multiple-choice tests. *Experimental Economics*, 25(4), 1268-1297.
- Baldiga, K. (2014). Gender differences in willingness to guess. *Management Science*, 60(2), 434-448.
- Barber, B. M., & Odean, T. (2001). Boys will be boys: Gender, overconfidence, and common stock investment. *The quarterly journal of economics*, 116(1), 261-292.
- Bleemer, Z., & Mehta, A. (2022). Will studying economics make you rich? A regression discontinuity analysis of the returns to college major. *American Economic Journal: Applied Economics*, 14(2), 1-22.
- Charness, G., & Gneezy, U. (2012). Strong evidence for gender differences in risk taking. *Journal of Economic Behavior & Organization*, 83(1), 50-58.
- Coffman, K. B., & Klinowski, D. (2020). The impact of penalties for wrong answers on the gender gap in test scores. *Proceedings of the National Academy of Sciences*, 117(16), 8794-8803.
- Conde-Ruiz, J. I., Ganuza, J. J., and García, M. (2020). Gender gap and multiple choice exams in public selection processes. *Hacienda Publica Espanola*, 235:11-28.
- Croson, R., & Gneezy, U. (2009). Gender differences in preferences. *Journal of Economic literature*, 47(2), 448-474.
- Díez-Rituerto, M., Gardeazabal, J., Iriberry, N., and Rey-Biel, P. (2023). Gender gaps in access to medical intern positions: The role of competition. Working Paper.
- Espinosa, M. P., & Gardeazabal, J. (2020). The gender-bias effect of test scoring and framing: A concern for personnel selection and college admission. *The BE Journal of Economic Analysis & Policy*, 20(3).
- Funk, P., & Perrone, H. (2016). Gender differences in academic performance: The role of negative marking in multiple-choice exams. Working Paper.
- Gneezy, U., Niederle, M., & Rustichini, A. (2003). Performance in competitive environments: Gender differences. *The quarterly journal of economics*, 118(3), 1049-1074.
- Goodman, J., Gurantz, O., & Smith, J. (2020). Take two! SAT retaking and college enrollment gaps. *American Economic Journal: Economic Policy*, 12(2), 115-158.
- Hastings, J. S., Neilson, C. A., & Zimmerman, S. D. (2013). Are some degrees worth more than others? Evidence from college admission cutoffs in Chile (No. w19241). National Bureau of Economic Research.

- Iriberry, N., & Rey-Biel, P. (2019). Competitive pressure widens the gender gap in performance: Evidence from a two-stage competition in mathematics. *The Economic Journal*, 129(620), 1863-1893.
- Iriberry, N., & Rey-Biel, P. (2021). Brave boys and play-it-safe girls: Gender differences in willingness to guess in a large scale natural field experiment. *European Economic Review*, 131, 103603.
- Kang, L., Lei, Z., Song, Y., & Zhang, P. (2021). Gender Differences in Reactions to Failure in High-Stakes Competition: Evidence from the National College Entrance Exam Retakes. *SSRN Electronic Journal*.
- Karle, H., Engelmann, D., & Peitz, M. (2022). Student performance and loss aversion. *The Scandinavian Journal of Economics*, 124(2), 420-456.
- Nelson, J. A. (2016). Not-so-strong evidence for gender differences in risk taking. *Feminist Economics*, 22(2), 114-142.
- Niederle, M., & Vesterlund, L. (2007). Do women shy away from competition? Do men compete too much?. *The quarterly journal of economics*, 122(3), 1067-1101.
- Montolio, D., & Taberner, P. A. (2021). Gender differences under test pressure and their impact on academic performance: a quasi-experimental design. *Journal of Economic Behavior & Organization*, 191, 1065-1090.
- Sunyer, C. (2023). Does Retaking Increase the Gender Gap? The Effects of Retaking on Performance and Behavior. Working Paper.
- Tutosaus Gómez, J. D., J. Morán-Barrios, and F. Pérez Iglesias (2018). Historia de la formación sanitaria especializada en España y sus claves docentes. *Educación Médica* 19(4), 229–234.