

9^o Seminario
Internacional
de Investigación
sobre la calidad de la educación

DOCENTES

BOGOTÁ D.C.



GOBIERNO
DE COLOMBIA



MINEDUCACIÓN

icfes
mejor saber

The Impact of Penalties for Wrong Answers on the Gender Gap in Test Scores

David Klinowski

Santiago Centre for Experimental Social Sciences

Katherine B. Coffman

Harvard Business School

Exámenes estandarizados de admisión universitaria

- Comúnmente utilizados
 - Brasil, Chile, China, Colombia, India, Irán, Israel, Japón, Polonia, Turquía, EEUU, ...
- Altas consecuencias en educación y mercado laboral
 - Admisión universitaria
 - Ubicación a programas
 - Asignación de ayuda económica
 - Predicción de éxito

Preguntas de selección múltiple

- ¿Cuál es la respuesta correcta?
 - A.
 - B.
 - C.
 - D.
 - E.
- Usadas parcial o totalmente en muchos exámenes de admisión
- Se consideran una manera objetiva y eficiente de medir conocimiento

Penalización y posible sesgo de género

- Pregunta típica en la PSU (y SAT) antes del 2015
 - 5 opciones (una sola correcta)
 - Correcta: +1 pt Incorrecta: -0.25 pt En blanco: 0 pt
- Responder es siempre óptimo para agente neutral al riesgo
 - Totalmente al azar: $20\% \times (1 \text{ pt}) - 80\% (0.25 \text{ pt}) = 0 \text{ pt}$
 - Algo informado: $E(\text{pt}) > 0 \text{ pt}$
- Sin embargo, la práctica de dejar en blanco es común
 - Tal vez por aversión al riesgo o por falta de autoconfianza
- Matemáticamente a esta persona le va a ir peor
 - Problemático para el examen porque factores ajenos al conocimiento afectan el puntaje

Penalización y posible sesgo de género

- Mujeres tienden a ser más aversas al riesgo [Eckel y Grossman \(2008\)](#), [Croson y Gneezy \(2009\)](#), [Charness y Gneezy \(2012\)](#), [Charness, Eckel y Gneezy \(2018\)](#)
- Mujeres tienden a tener menor autoconfianza [Lichtenstein, Fischhoff, Phillips \(1982\)](#), [Beyer \(1990, 1999\)](#)
- Mujeres en el lab dejan más preguntas en blanco y obtienen peores puntajes que hombres con mismo conocimiento. Eliminar la penalización acaba con las preguntas en blanco y reduce la brecha de puntajes [Baldiga \(2014\)](#)

Examen estandarizado en el laboratorio (Baldiga, 2014)

- Prueba tipo SAT en el laboratorio
 - Con penalización por respuesta incorrecta
 - Incentivado con un pago en \$\$ proporcional al rendimiento
- Ronda 1: alumnos toman examen tipo (20 preguntas, pueden dejar en blanco)
- ¿Están las mujeres menos dispuestas a aventurarse a adivinar?
- Ronda 2: alumnos toman el mismo examen obligados a responder cada pregunta
- El laboratorio permite observar y controlar diferencias de conocimiento, además de medir tolerancia al riesgo, nivel de autoconfianza, etc.

Examen en el laboratorio (Baldiga, 2014)

- En el experimento, las mujeres dejan en promedio aprox. el doble de preguntas en blanco que los hombres
- Esto a pesar de tener el mismo nivel de conocimiento
- Las mujeres puntúan 0.08 SD peor, en este examen de solo 20 preguntas. La mayoría de esta brecha es explicada por la tendencia a dejar más preguntas en blanco
- Eliminar la penalización cierra la brecha en preguntas en blanco y reduce la brecha en puntajes
- ¿Se generaliza a exámenes de admisión universitaria propiamente?

PSU Chile

- Examen centralizado de admisión, una vez al año
- Casi 200.000 – 300.000 inscritos al año
- 6 exámenes
 - 2 obligatorios: matemática y verbal
 - 4 electivos: biología, física, química, ciencias sociales
- Cada examen entre 70-80 preguntas de selección múltiple, con 5 opciones

Cambio en la regla de puntuación

- Antes del 2015:
 - Correcta: 1 pt
 - Incorrecta: -0.25 pt
 - En blanco: 0 pt

- A partir del 2015
 - Correcta: 1 pt
 - Incorrecta: 0 pt
 - En blanco: 0 pt

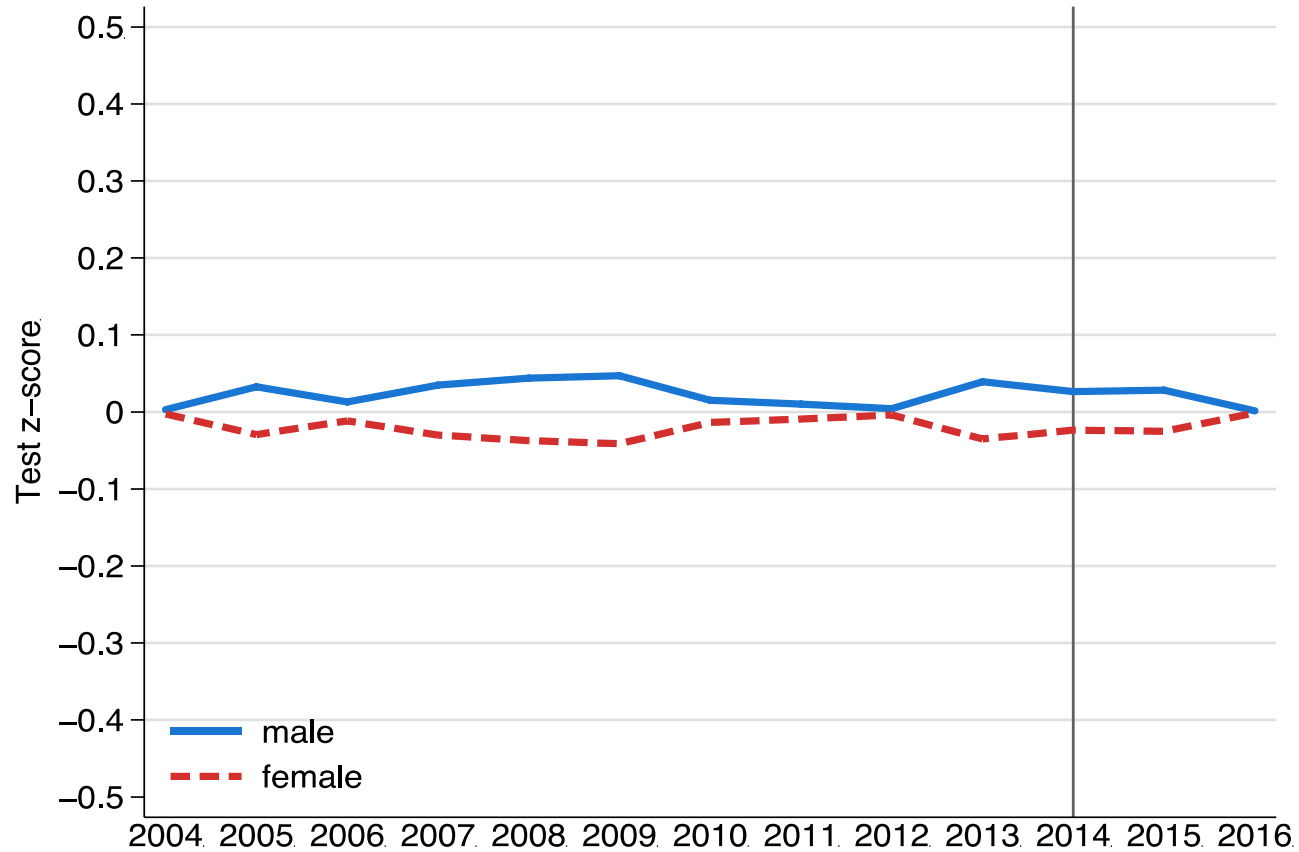
Estrategia empírica

- Comparar pre 2015 y pos 2015
 - Brecha de género en preguntas en blanco
 - Brecha de género en puntajes
- Preocupación de no capturar tendencias existentes distintas al posible impacto de la reforma
- Tomamos varias medidas descritas a lo largo de la presentación

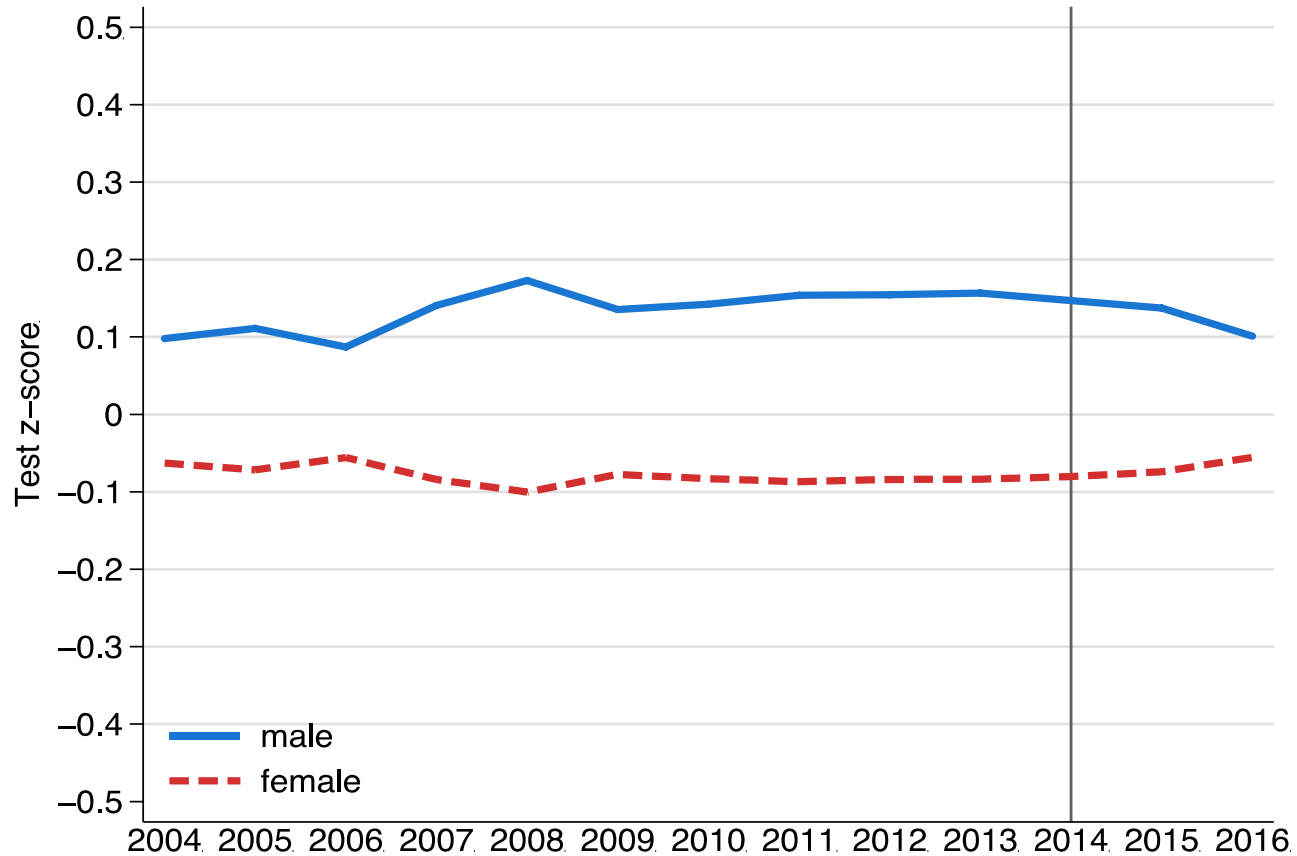
Datos

- Datos desde 2004 al 2016 para cada individuo
- Número de preguntas correctas, incorrectas y en blanco
- Datos sociodemográficos y de rendimiento escolar en educación media

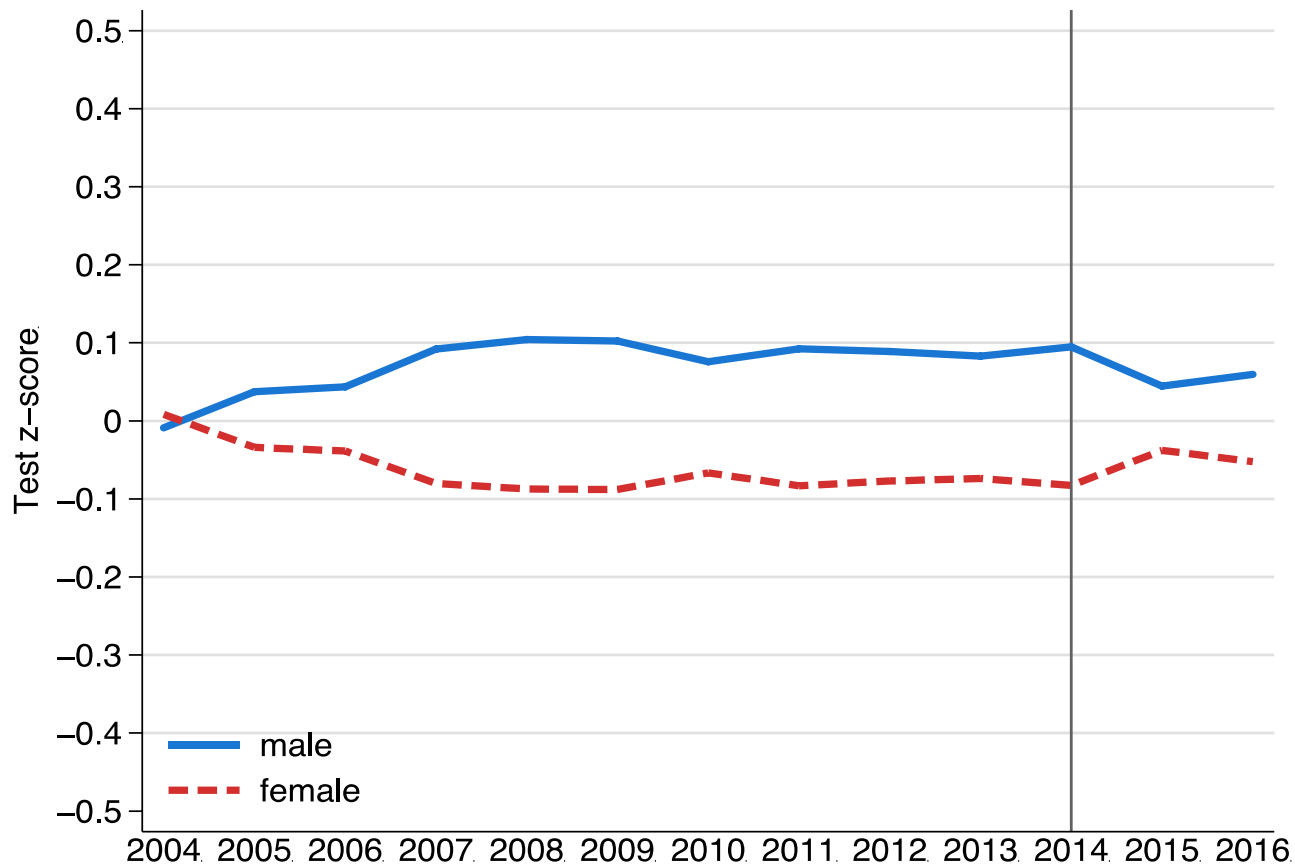
Puntaje estandarizado promedio: verbal



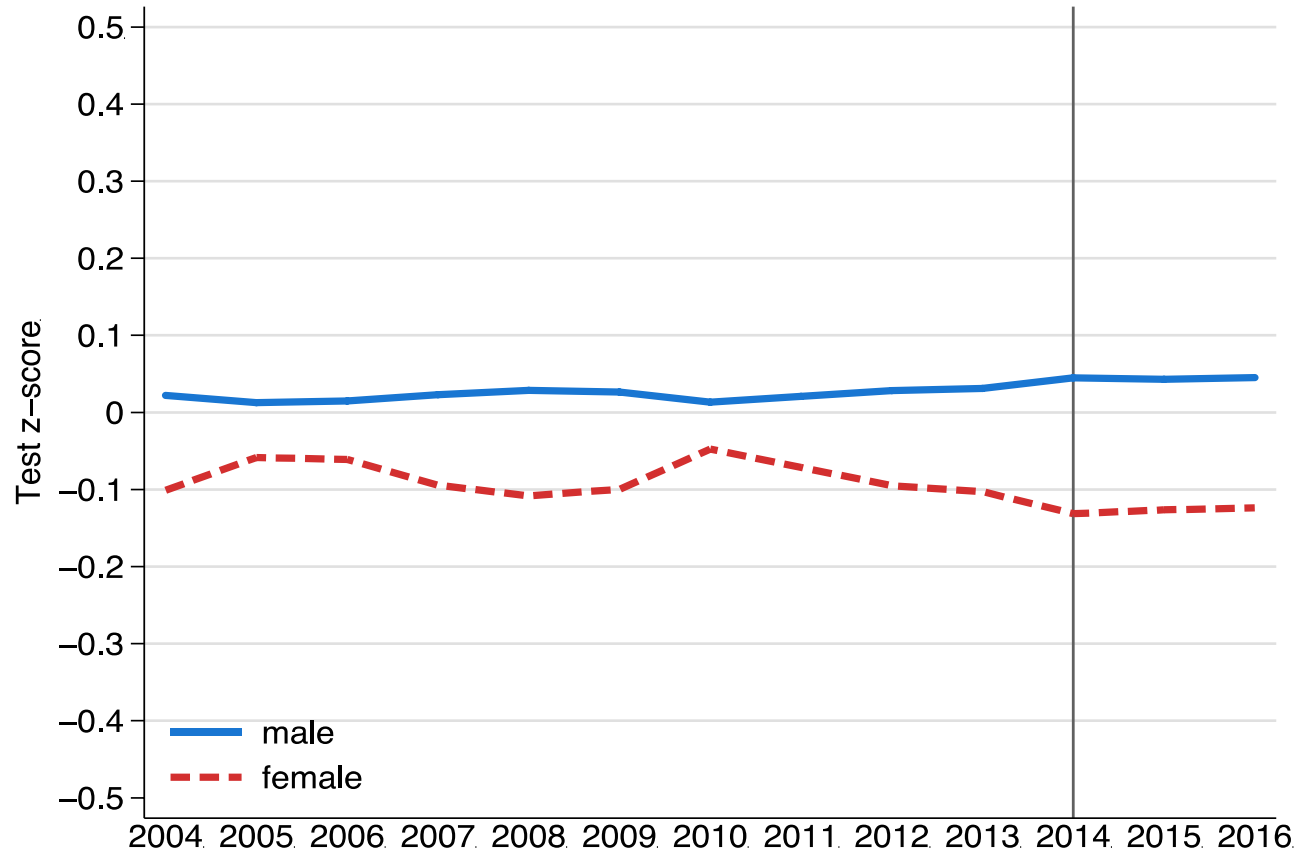
Puntaje estandarizado promedio: biología



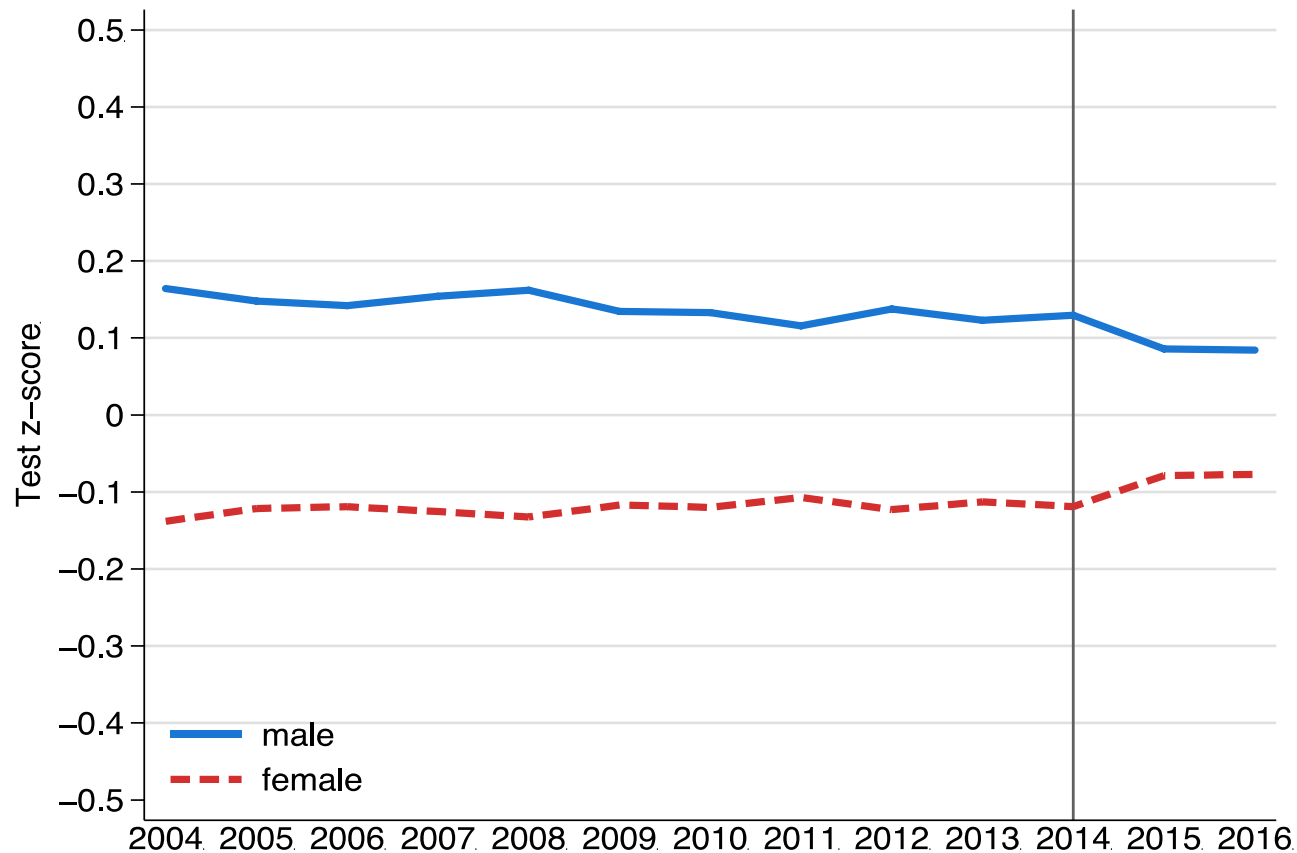
Puntaje estandarizado promedio: química



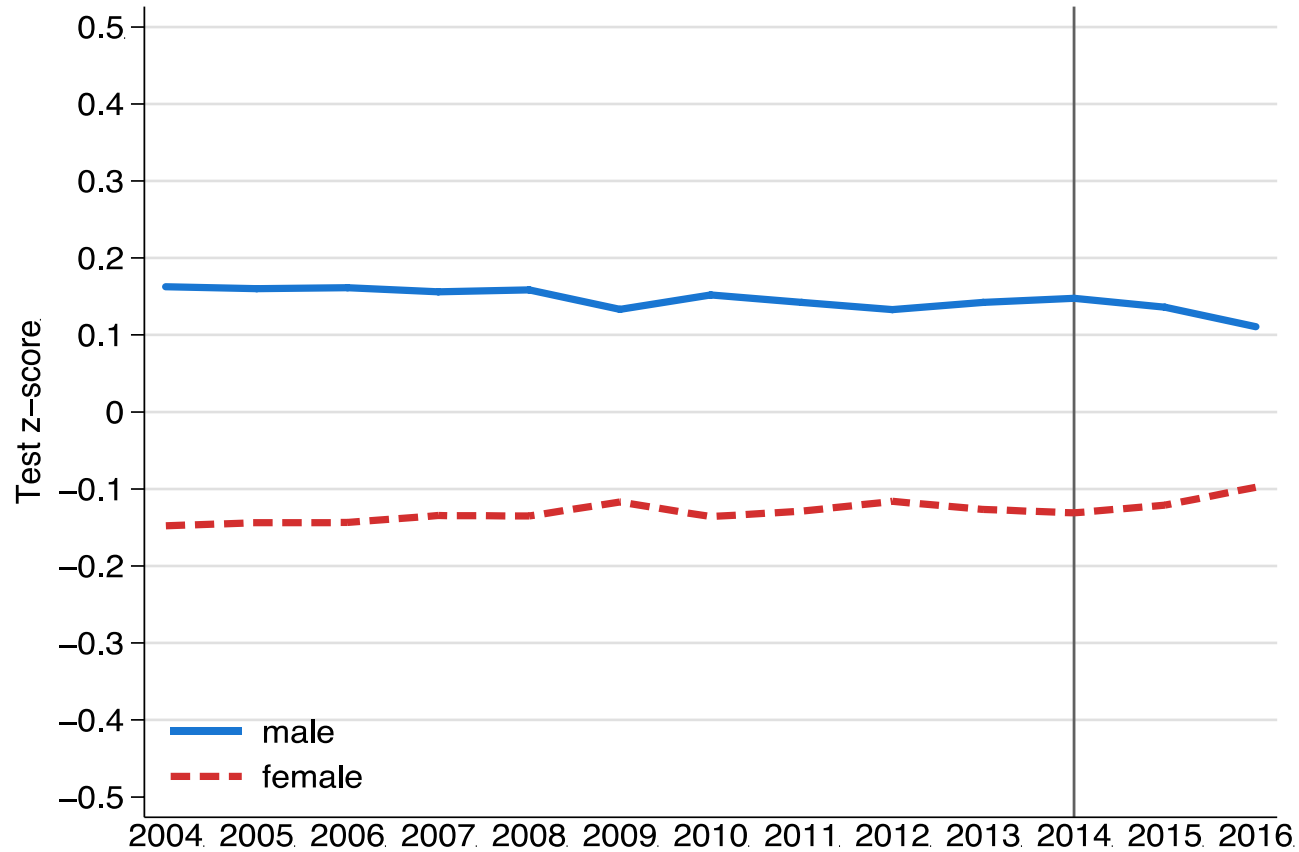
Puntaje estandarizado promedio: física



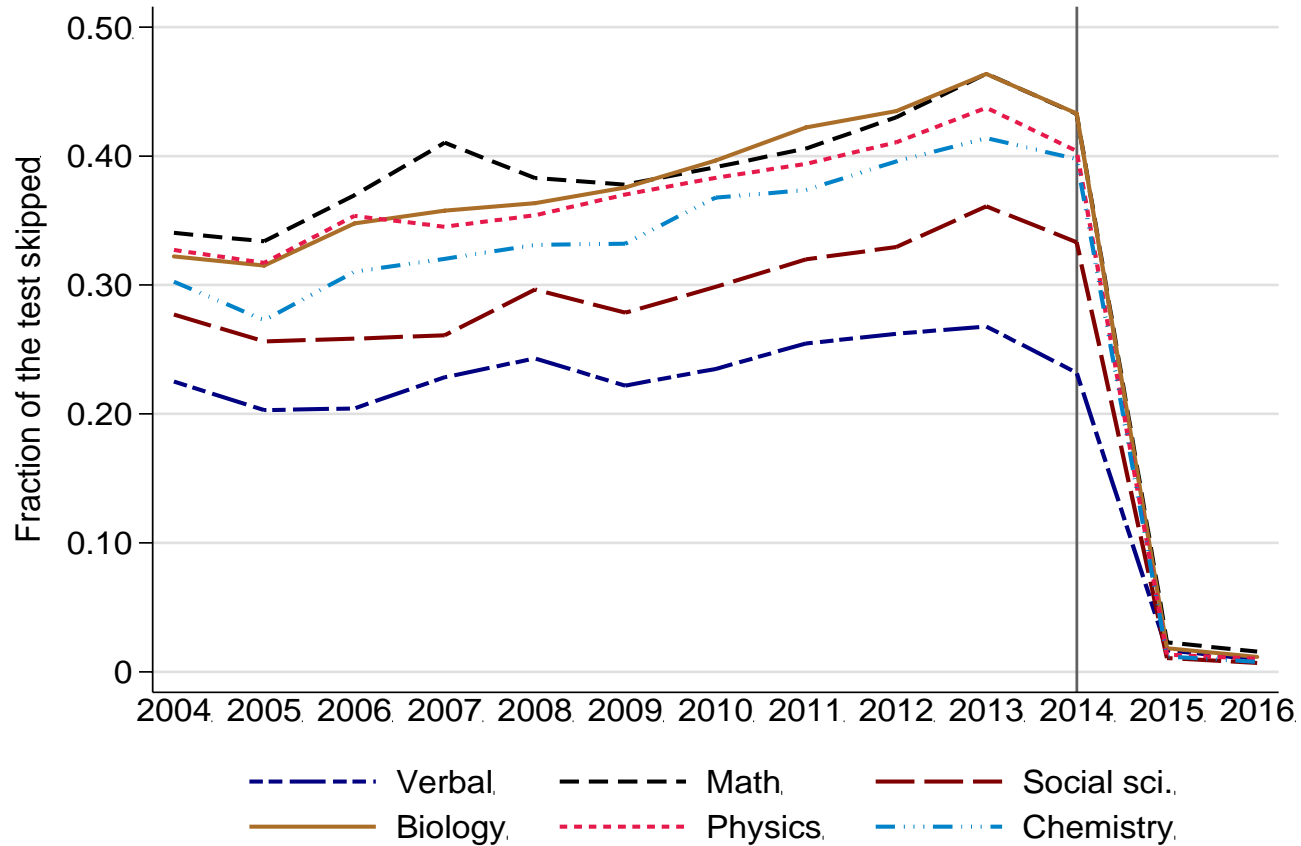
Puntaje estandarizado promedio: ciencias sociales



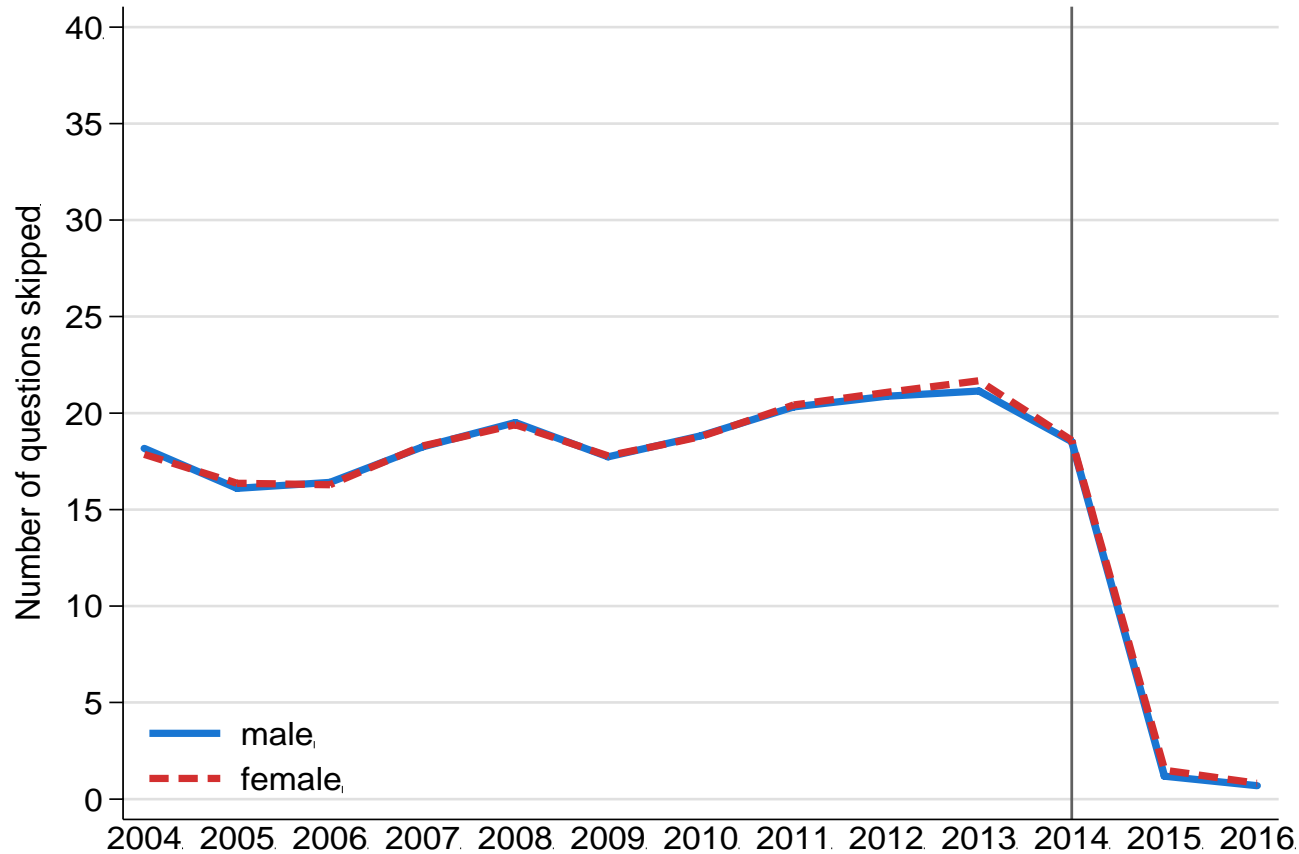
Puntaje estandarizado promedio: matemática



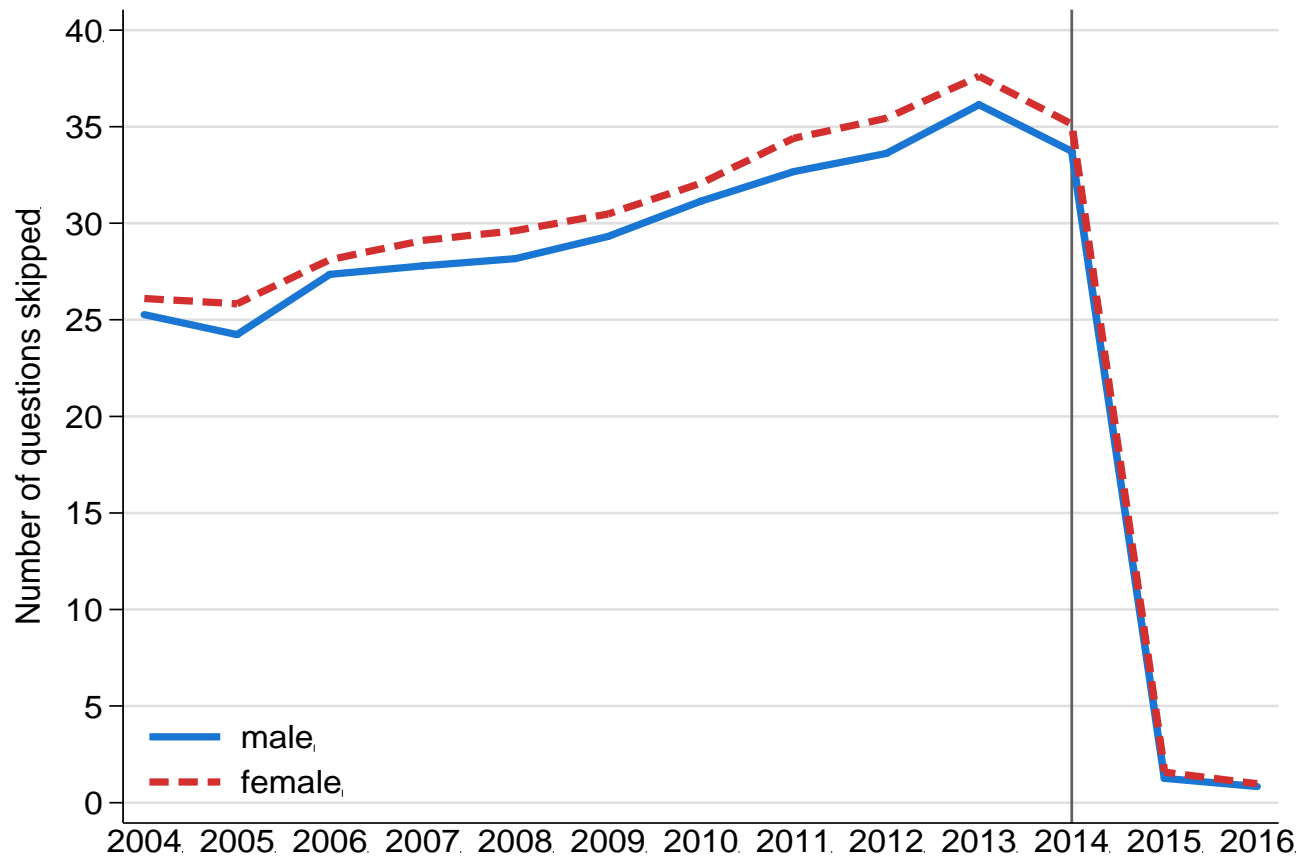
Fracción del examen en blanco



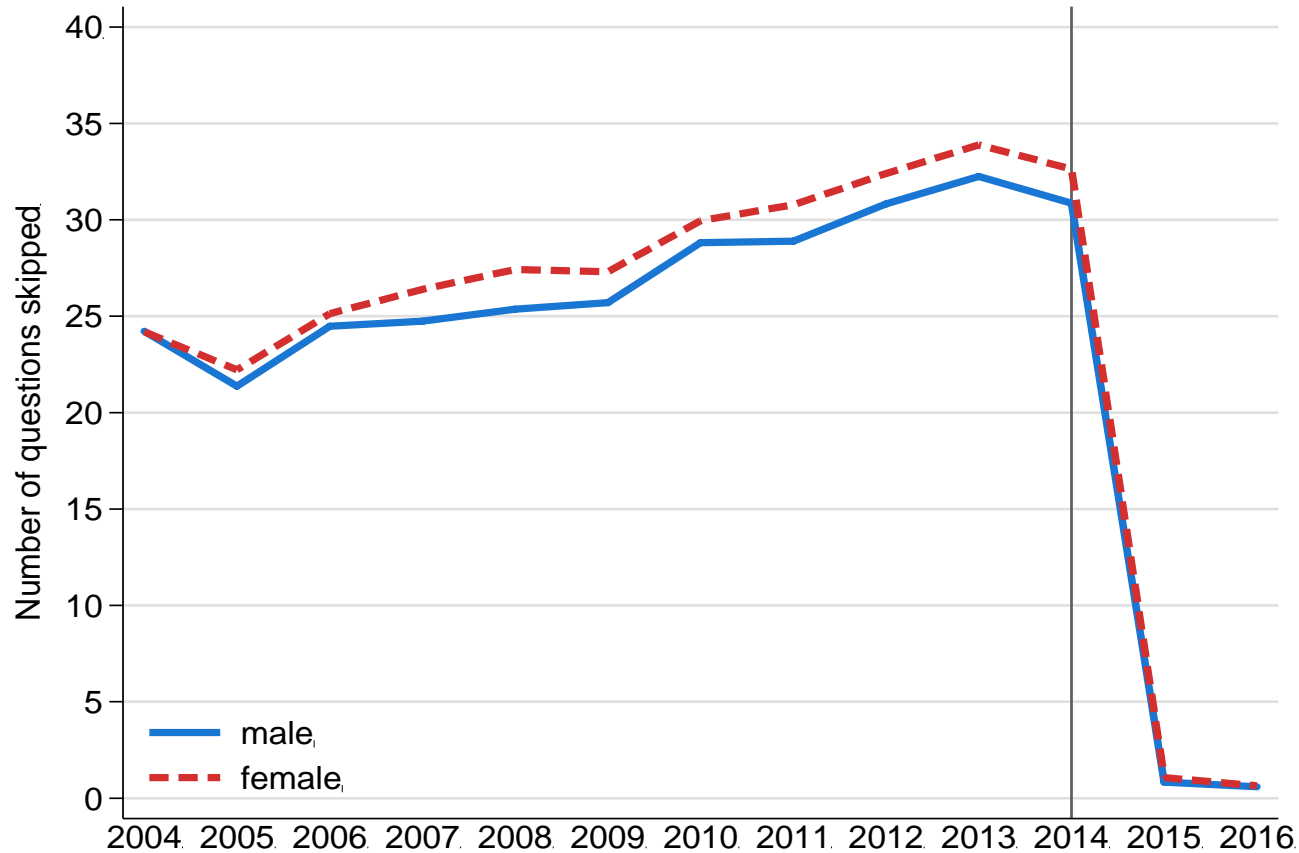
Número de preguntas en blanco promedio: verbal



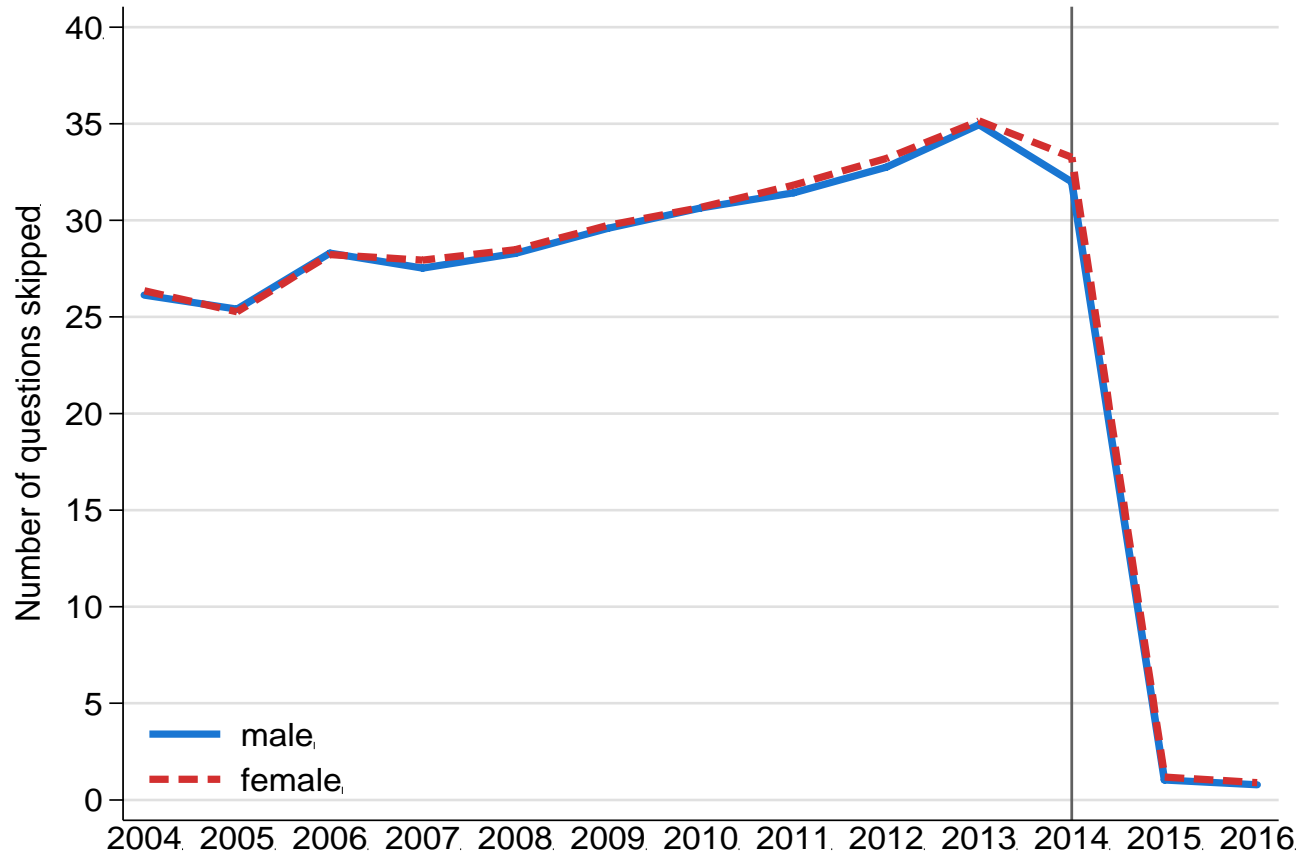
Número de preguntas en blanco promedio: biología



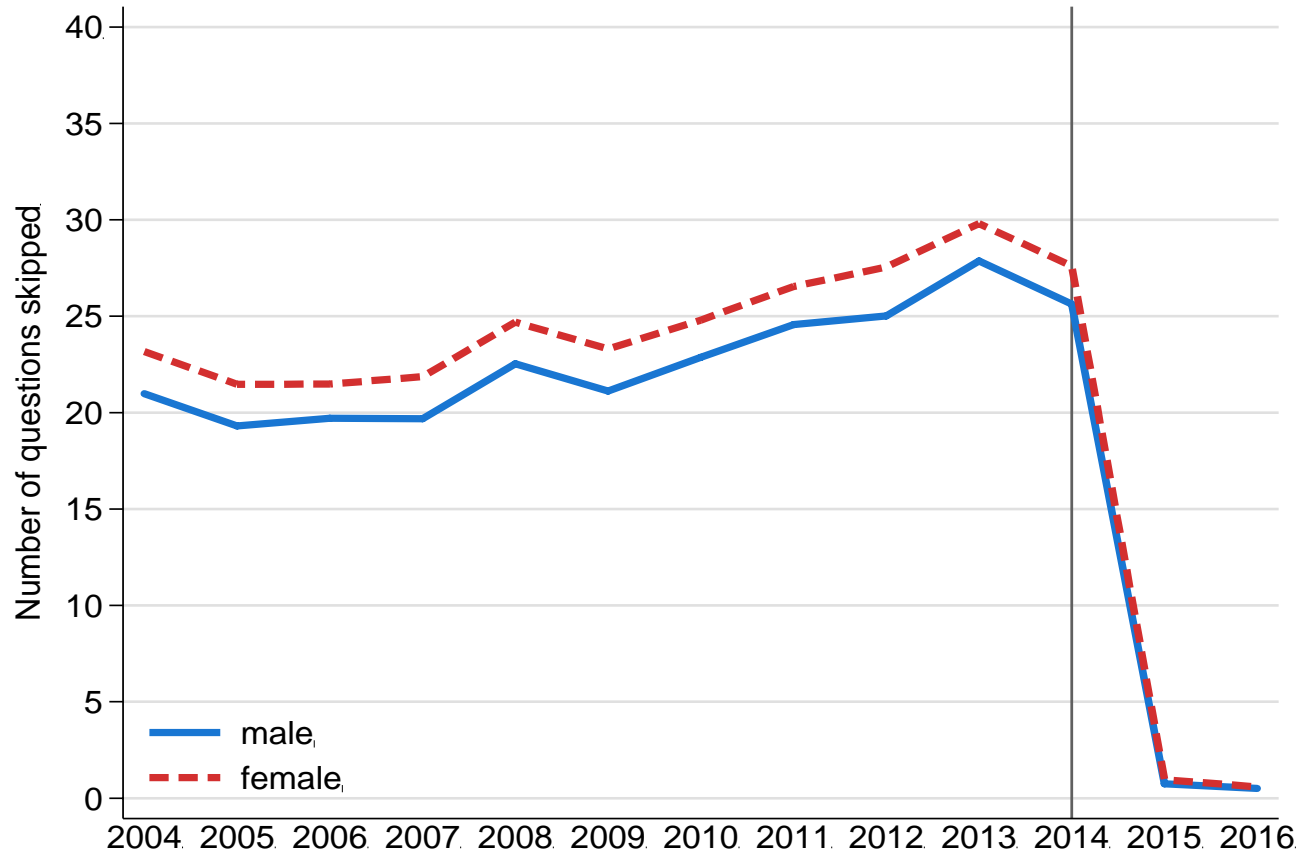
Número de preguntas en blanco promedio: química



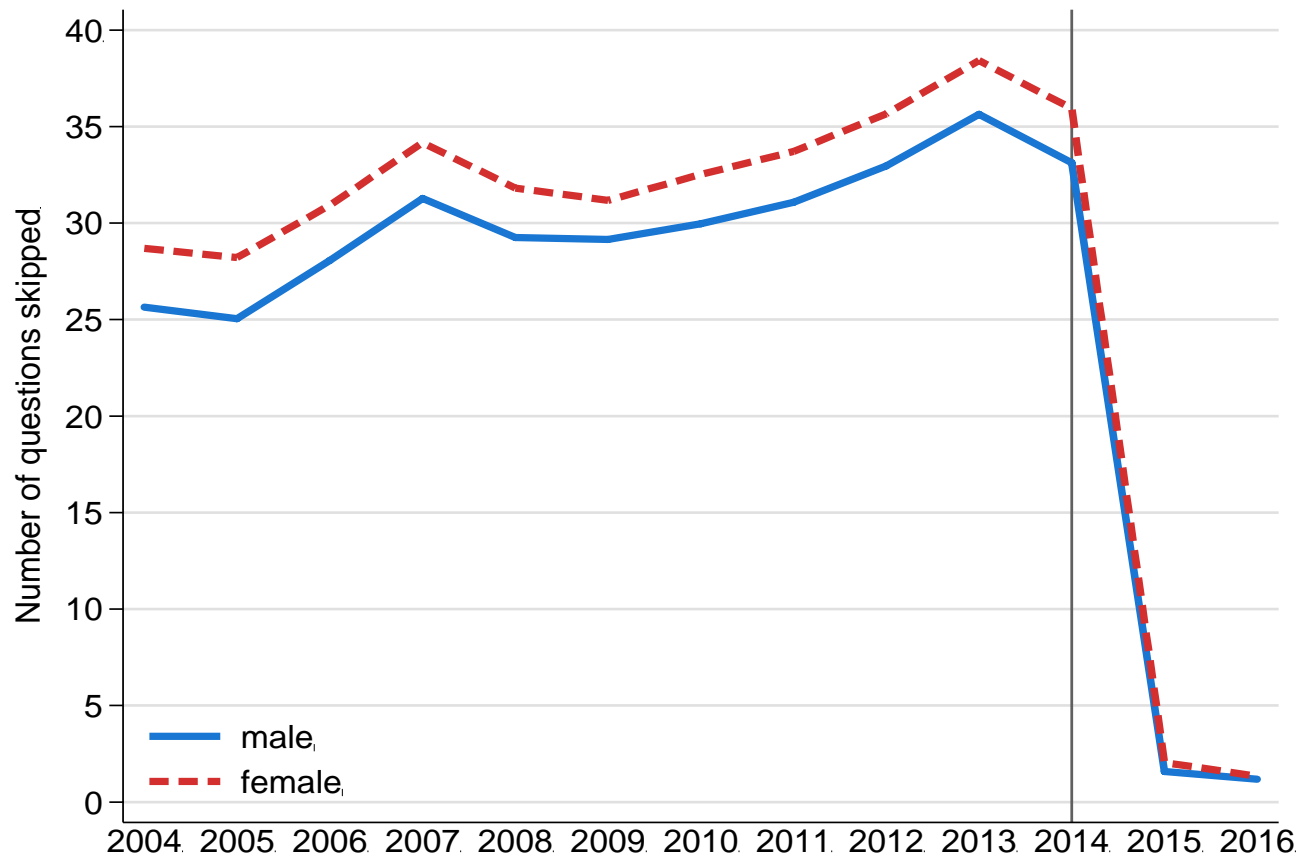
Número de preguntas en blanco promedio: física



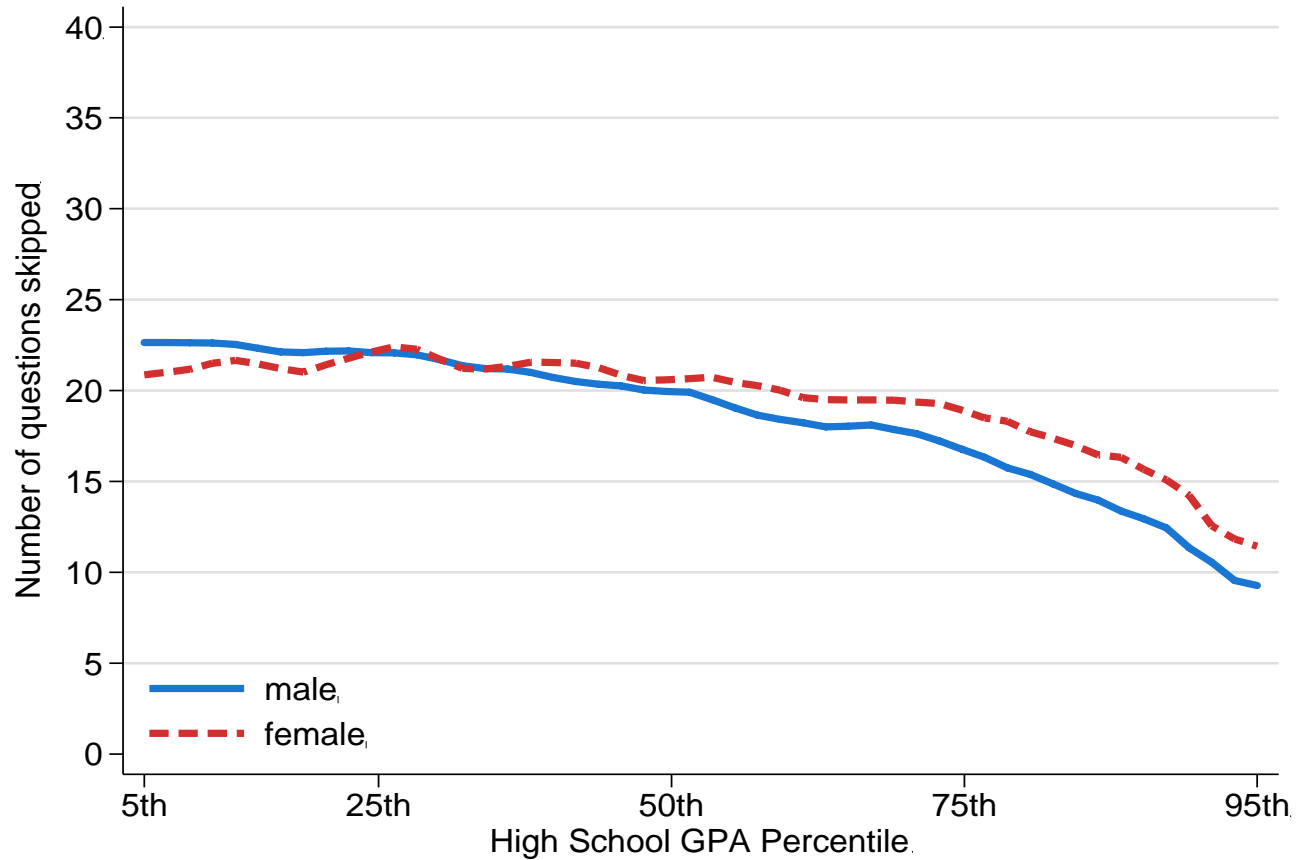
Número de preguntas en blanco promedio: c. sociales



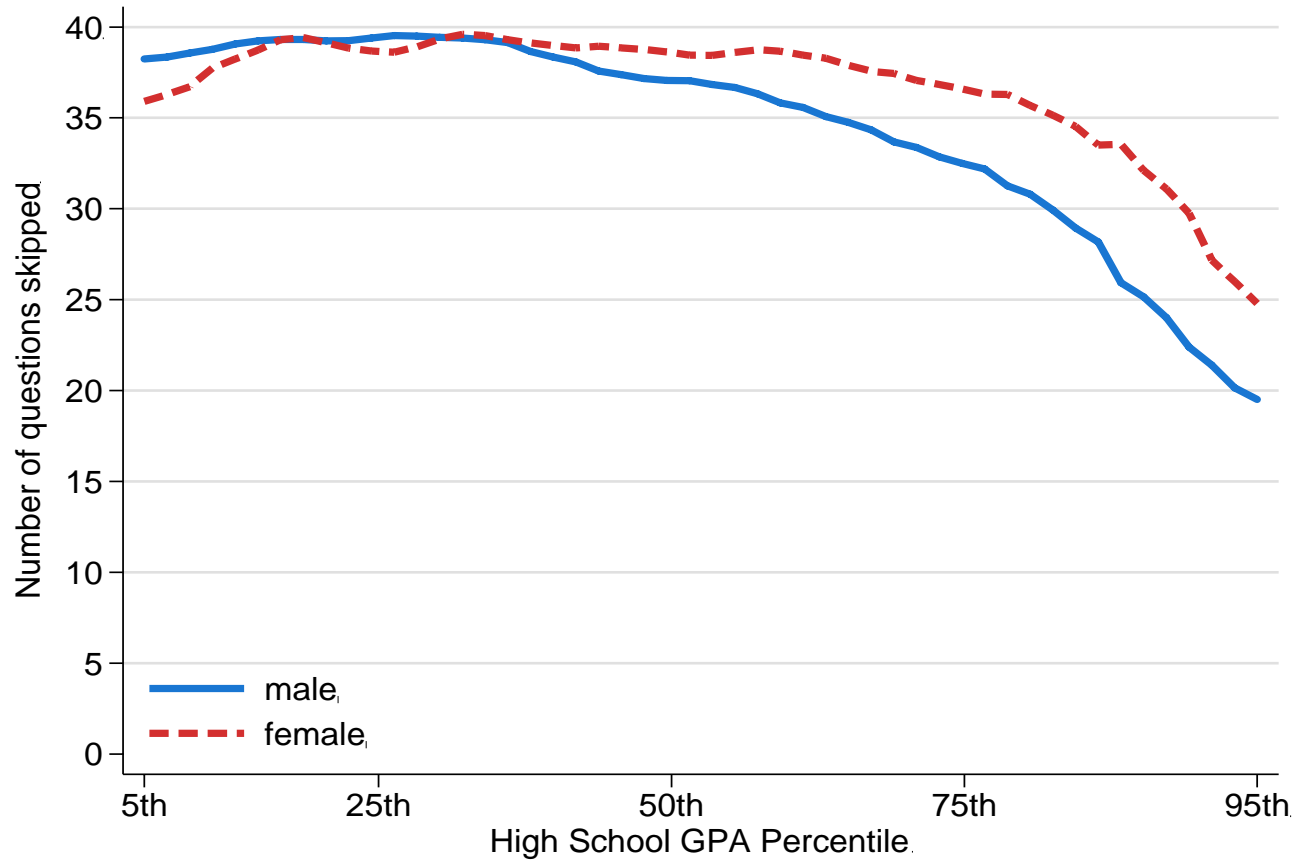
Número de preguntas en blanco promedio: matemática



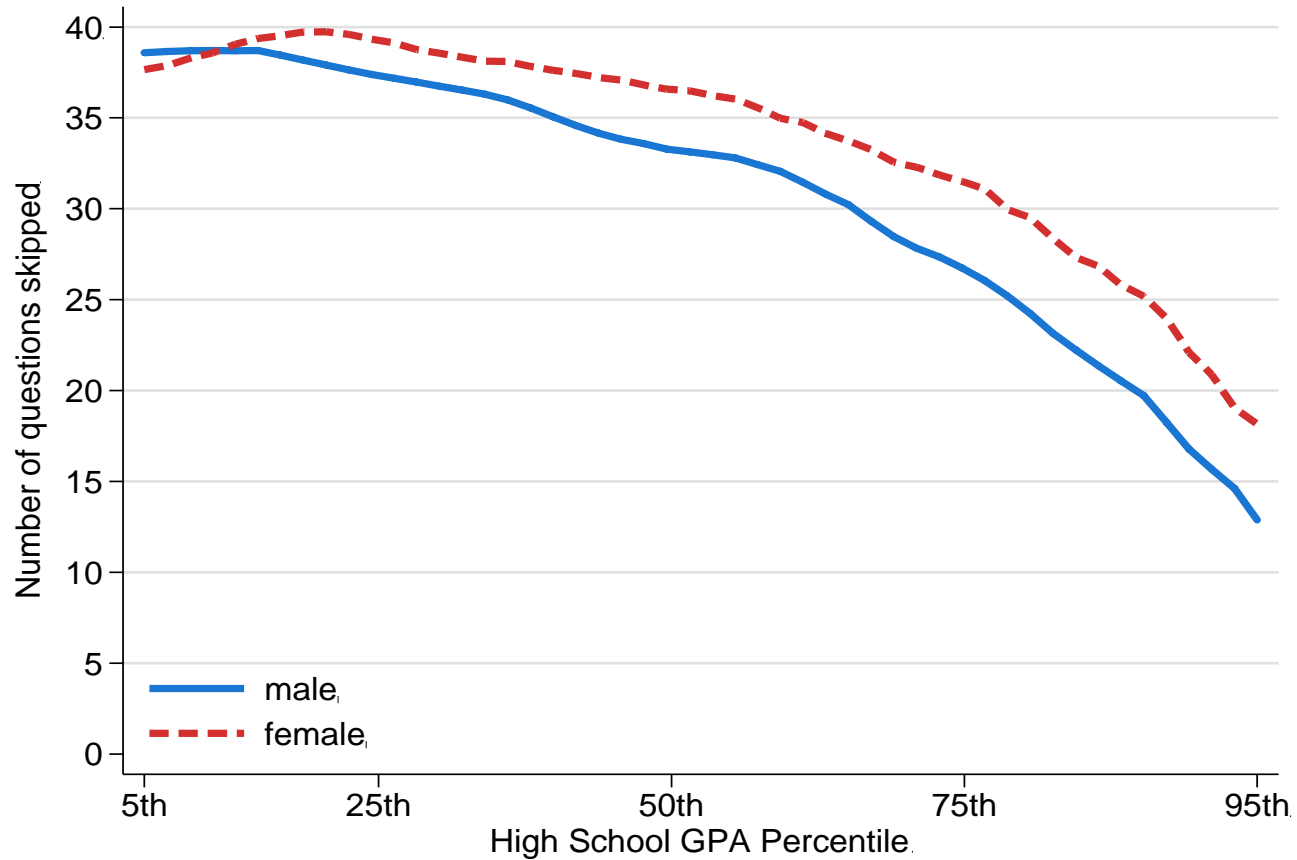
Heterogeneidad por habilidad: verbal



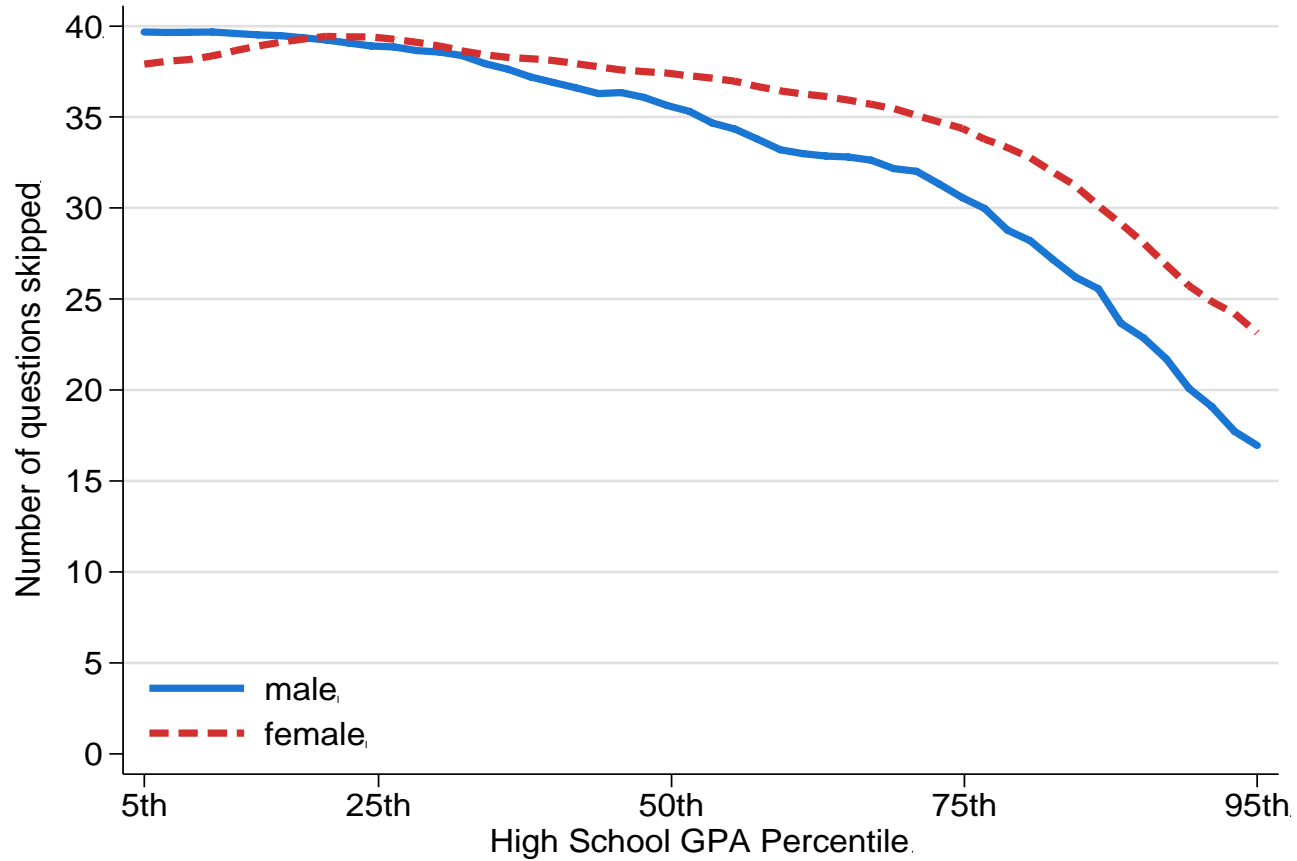
Heterogeneidad por habilidad: biología



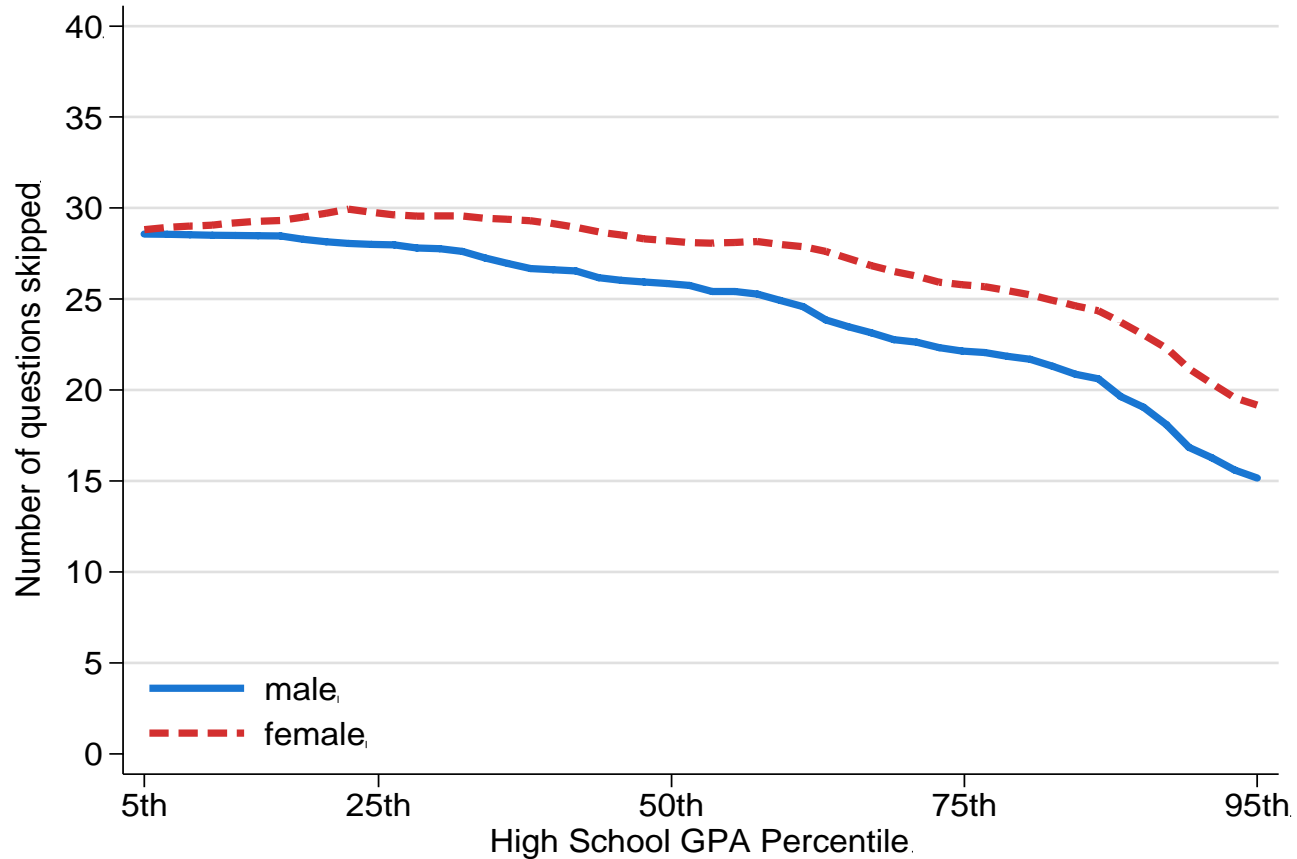
Heterogeneidad por habilidad: química



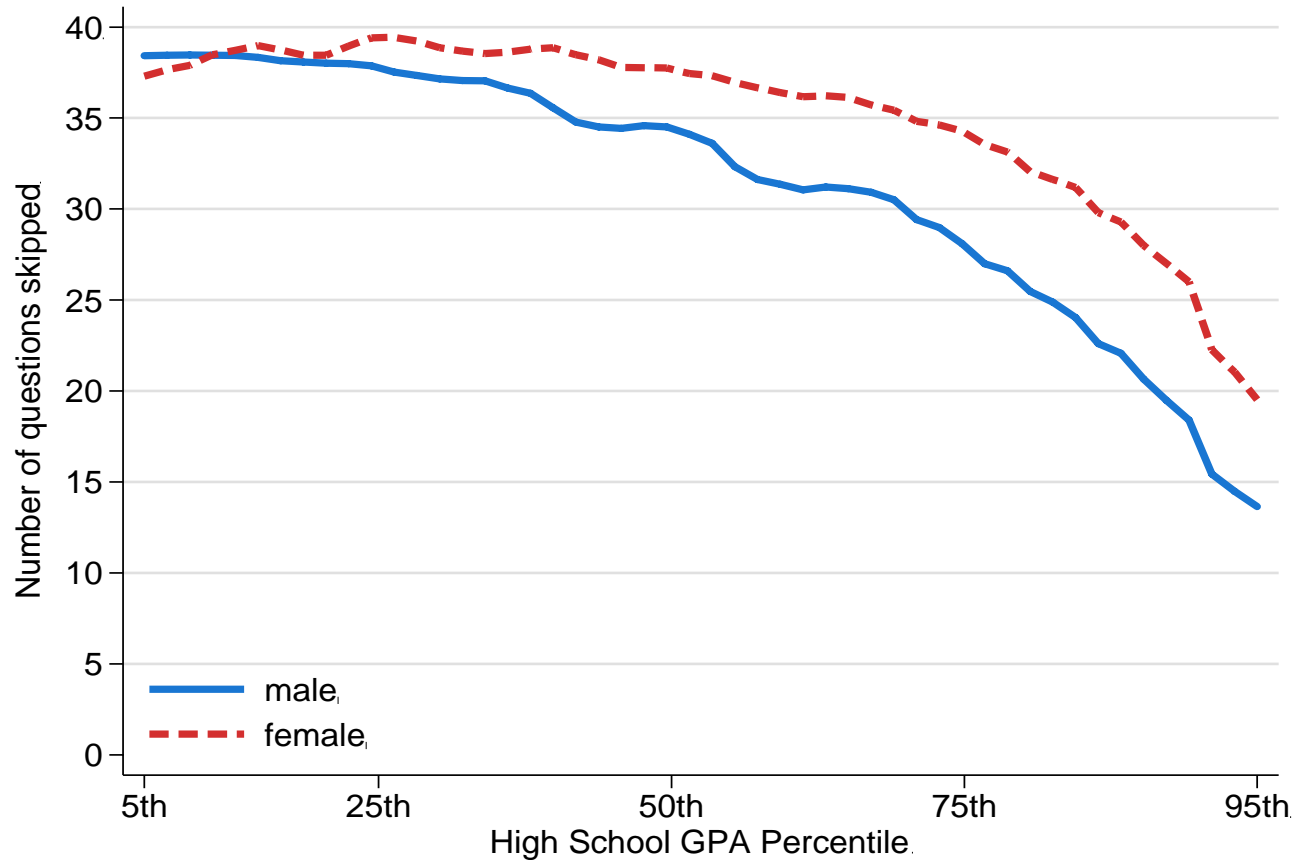
Heterogeneidad por habilidad: física



Heterogeneidad por habilidad: ciencias sociales



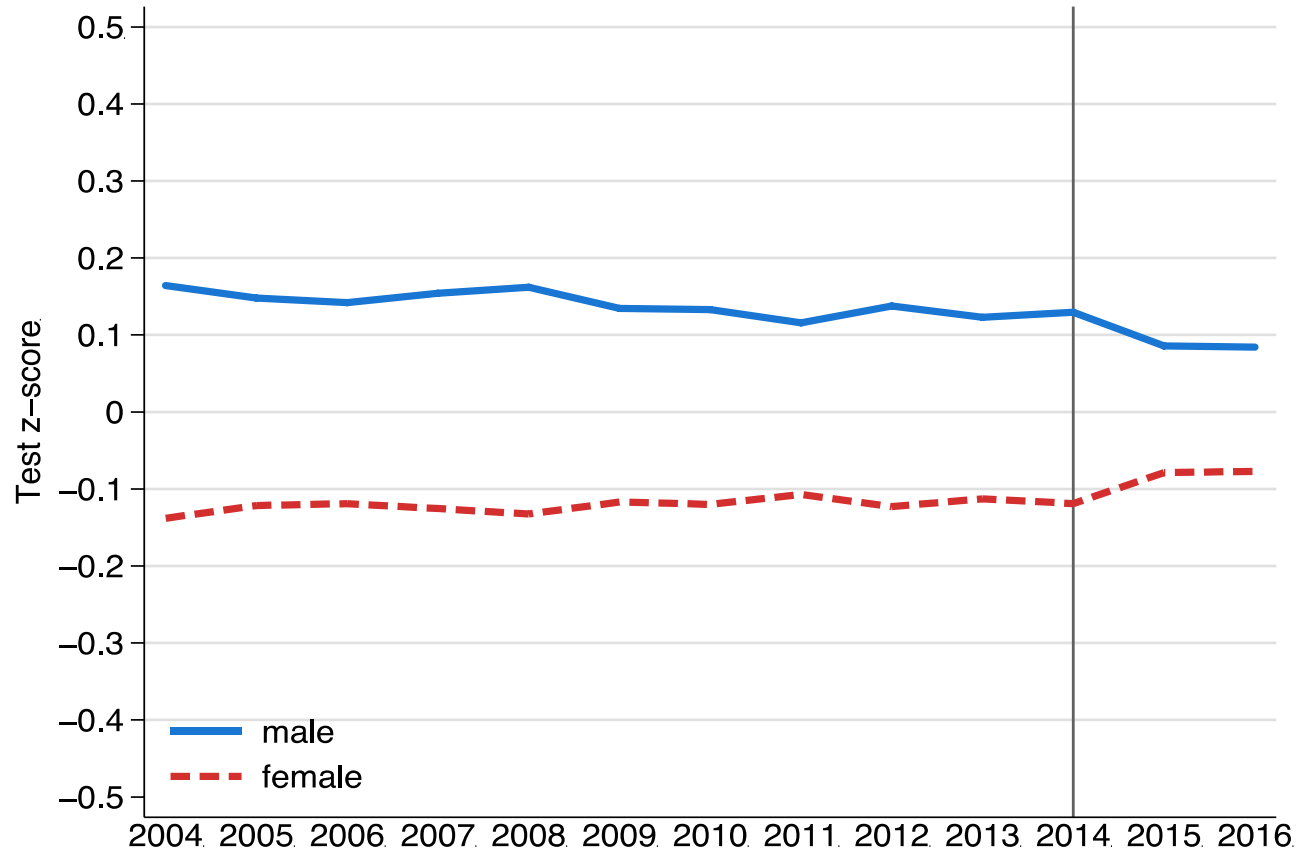
Heterogeneidad por habilidad: matemáticas



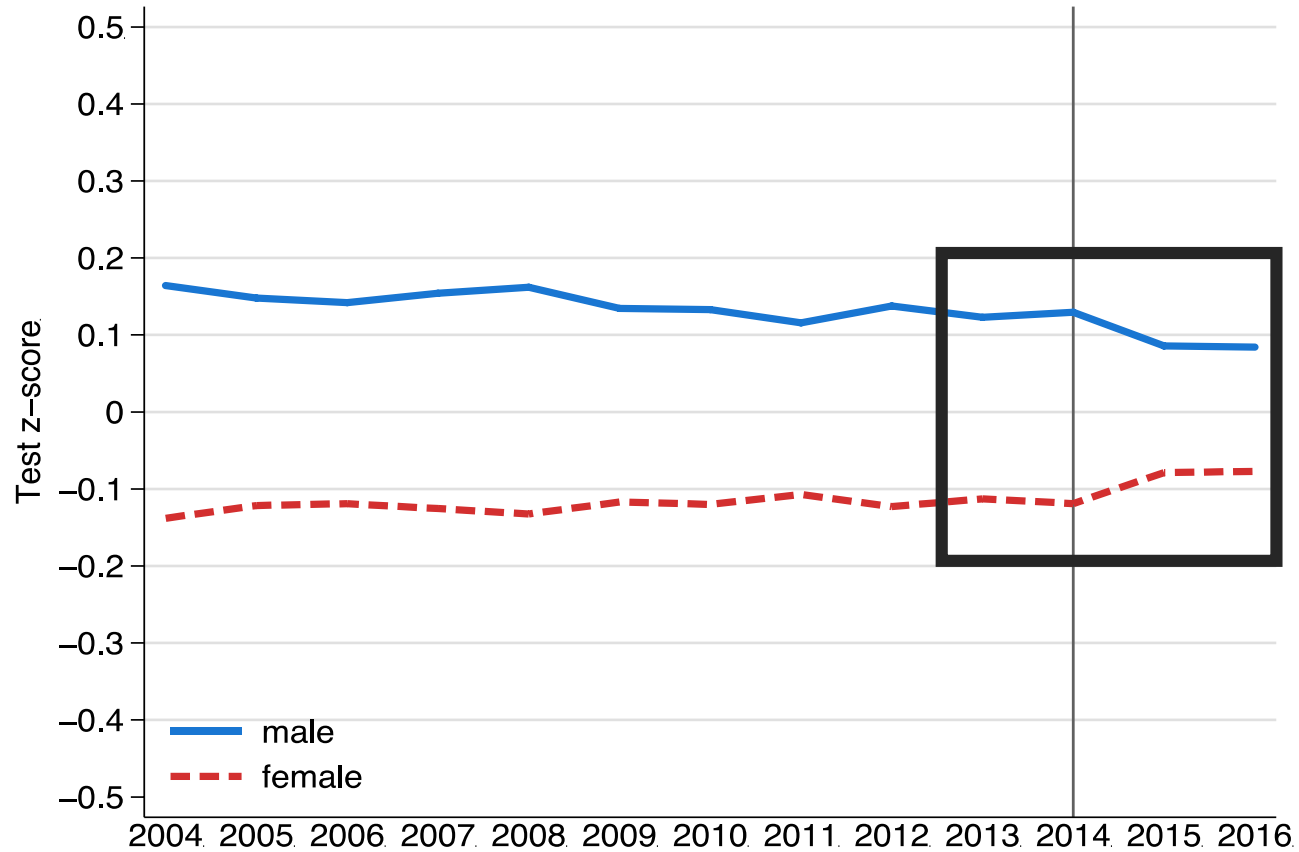
Resumen hasta ahora

- Existía una brecha de género en preguntas en blanco
 - Mayor para estudiantes más habilidosos
- La reforma eliminó la práctica de dejar preguntas en blanco y cerró esa brecha de género
- ¿Ayudó la reforma a mejorar los puntajes relativos de las mujeres?

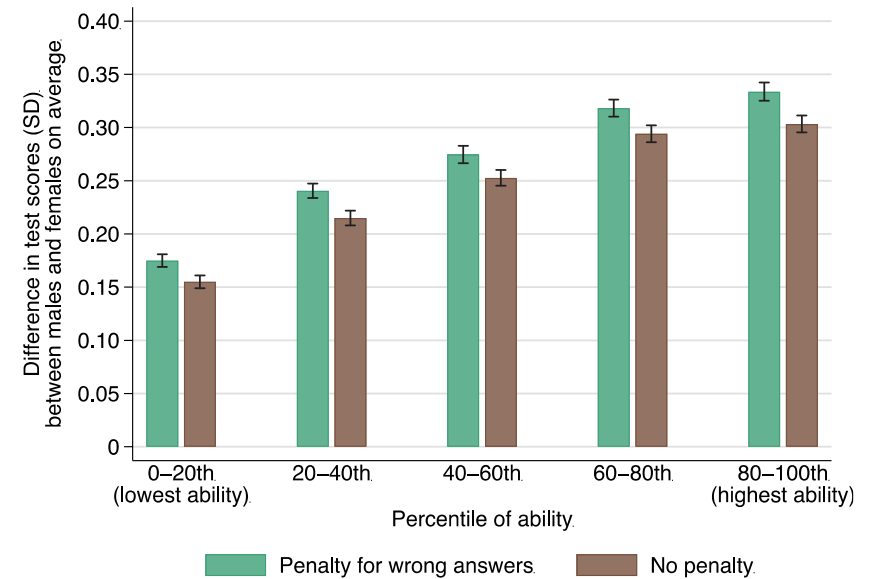
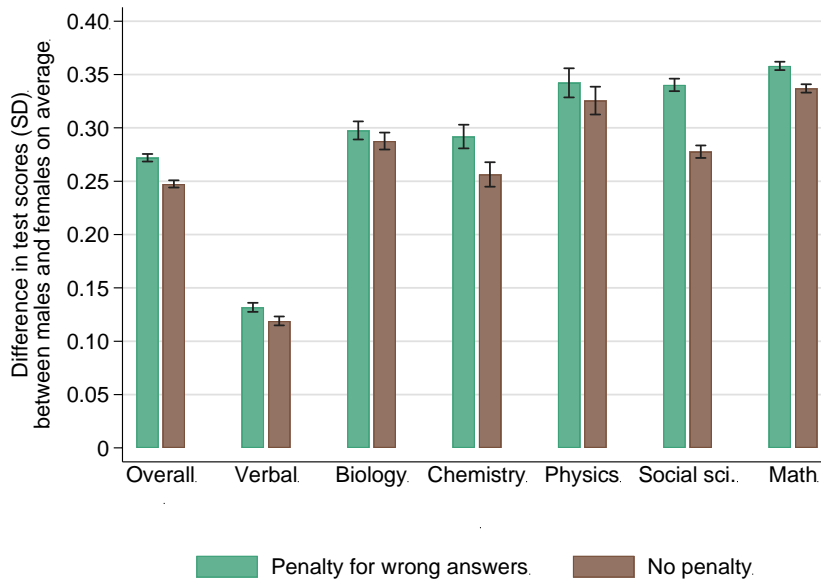
Estrategia: ventana angosta



Estrategia: ventana angosta



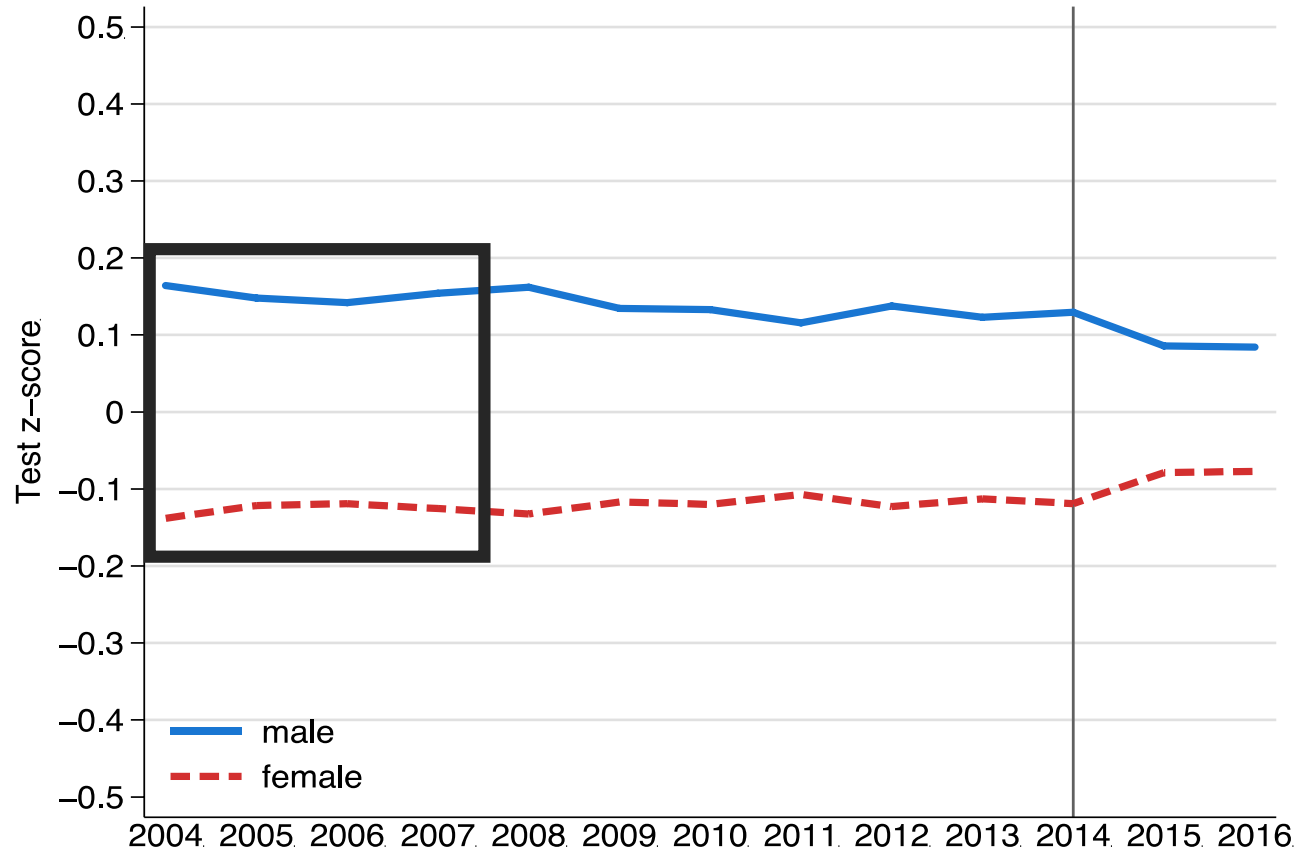
Brecha de género en puntajes estandarizados



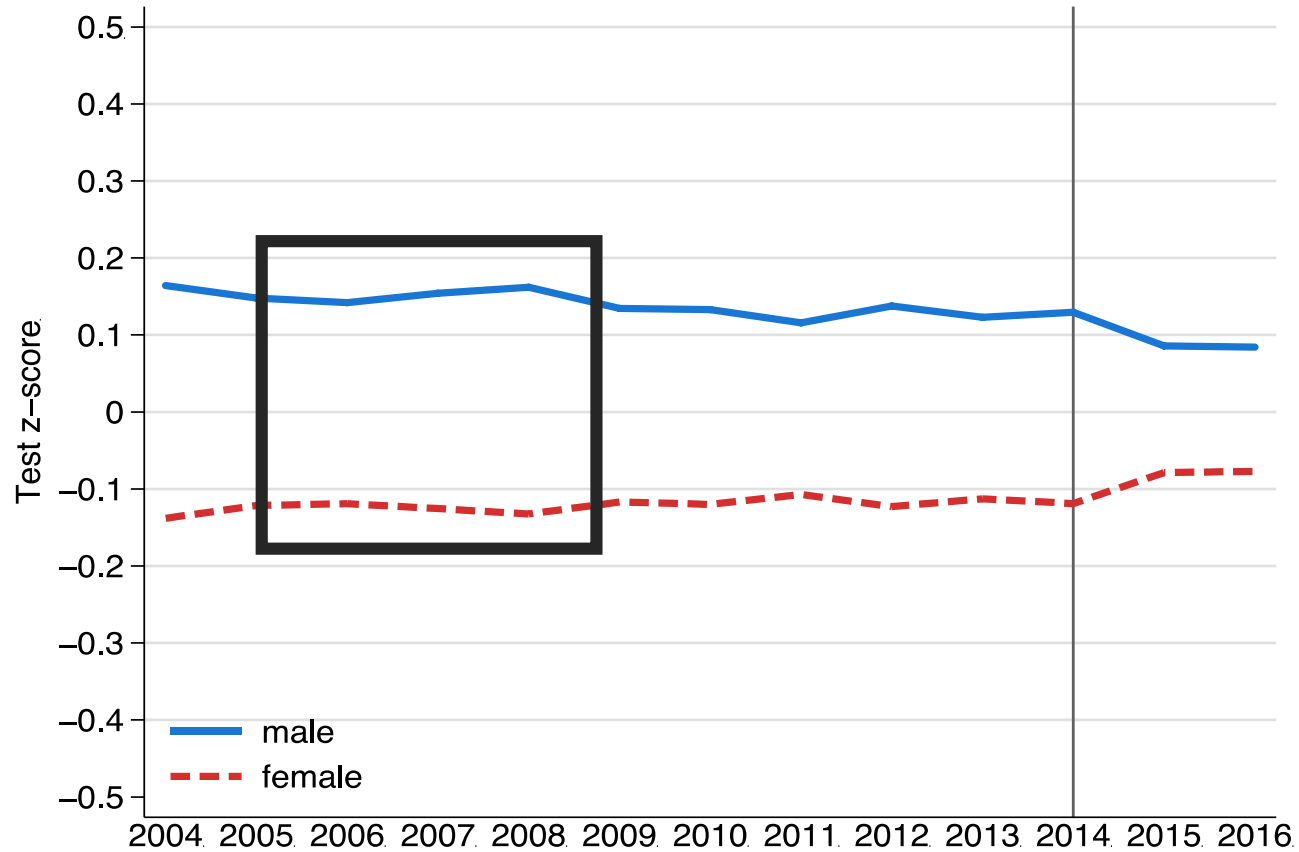
Estrategia: análisis placebo

- Replicar el análisis para distintas ventanas, suponiendo que la reforma ocurrió en un determinado año
- Comparar la magnitud de los efectos de estas reformas placebo con la magnitud del efecto de la verdadera reforma

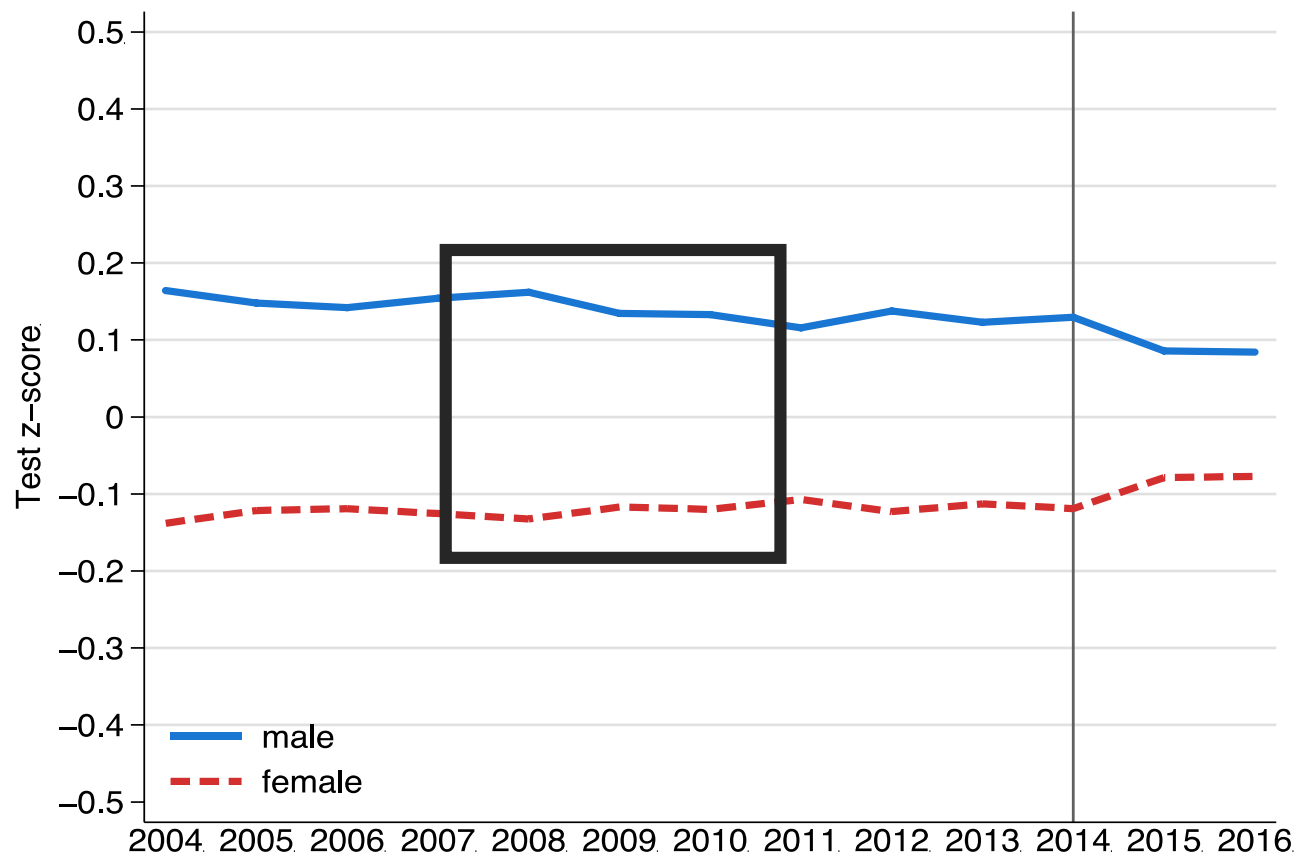
Estrategia: análisis placebo



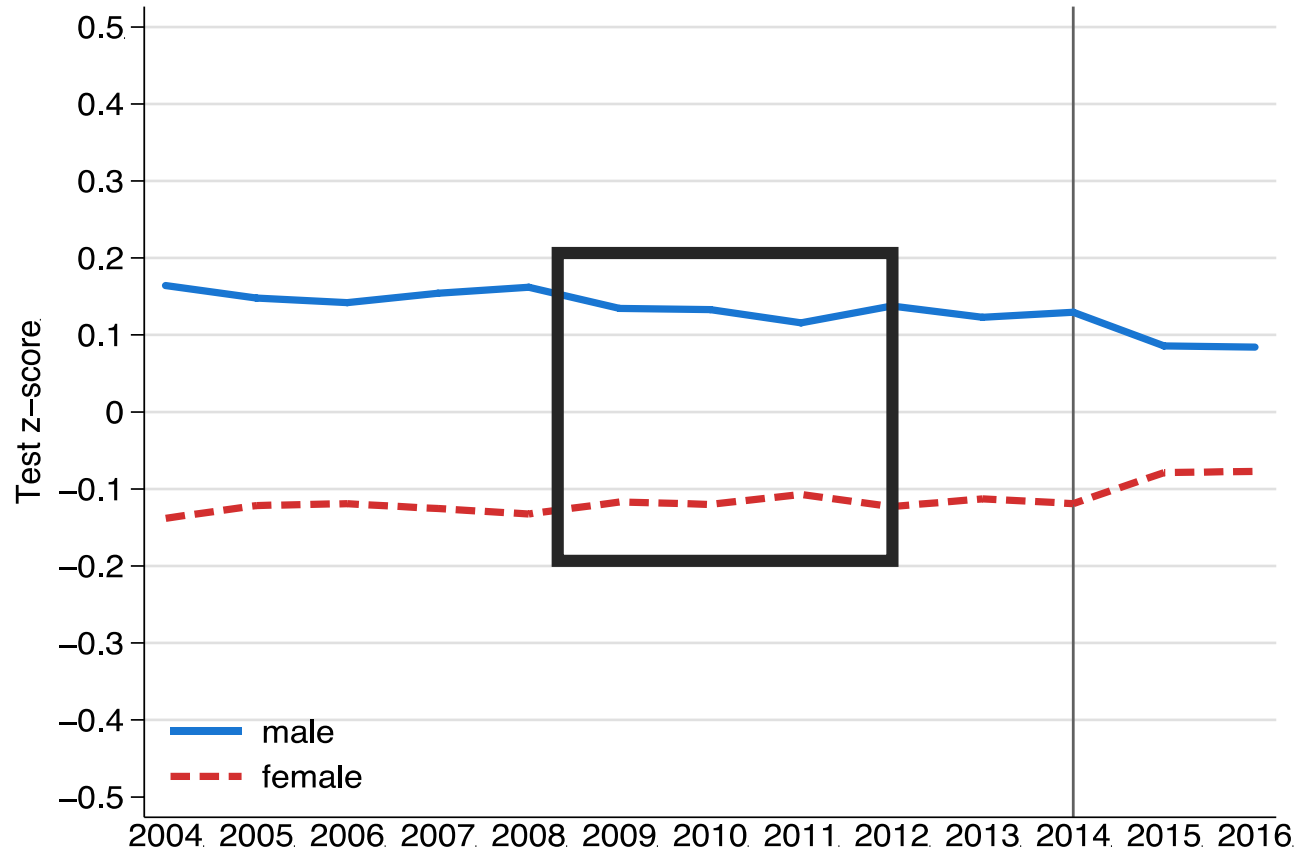
Estrategia: análisis placebo



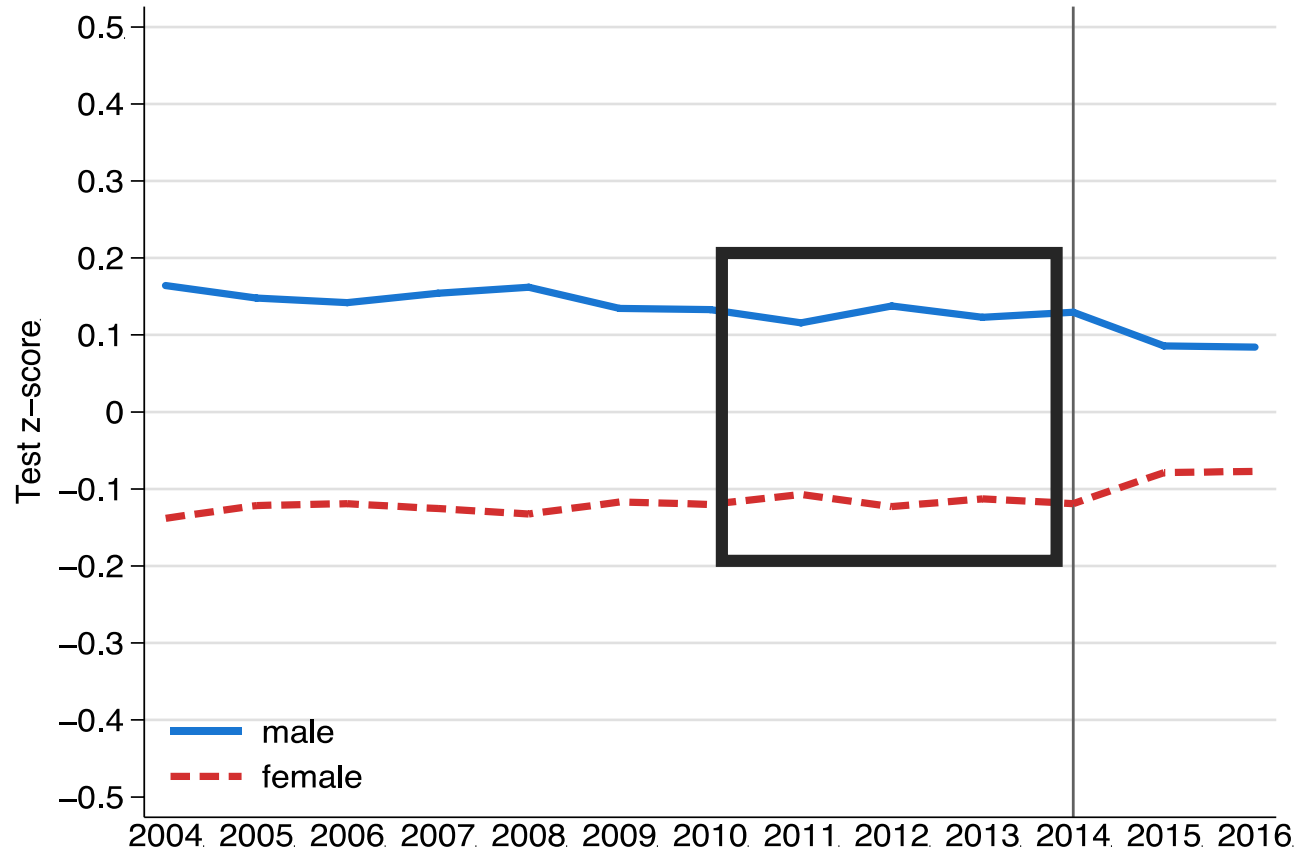
Estrategia: análisis placebo



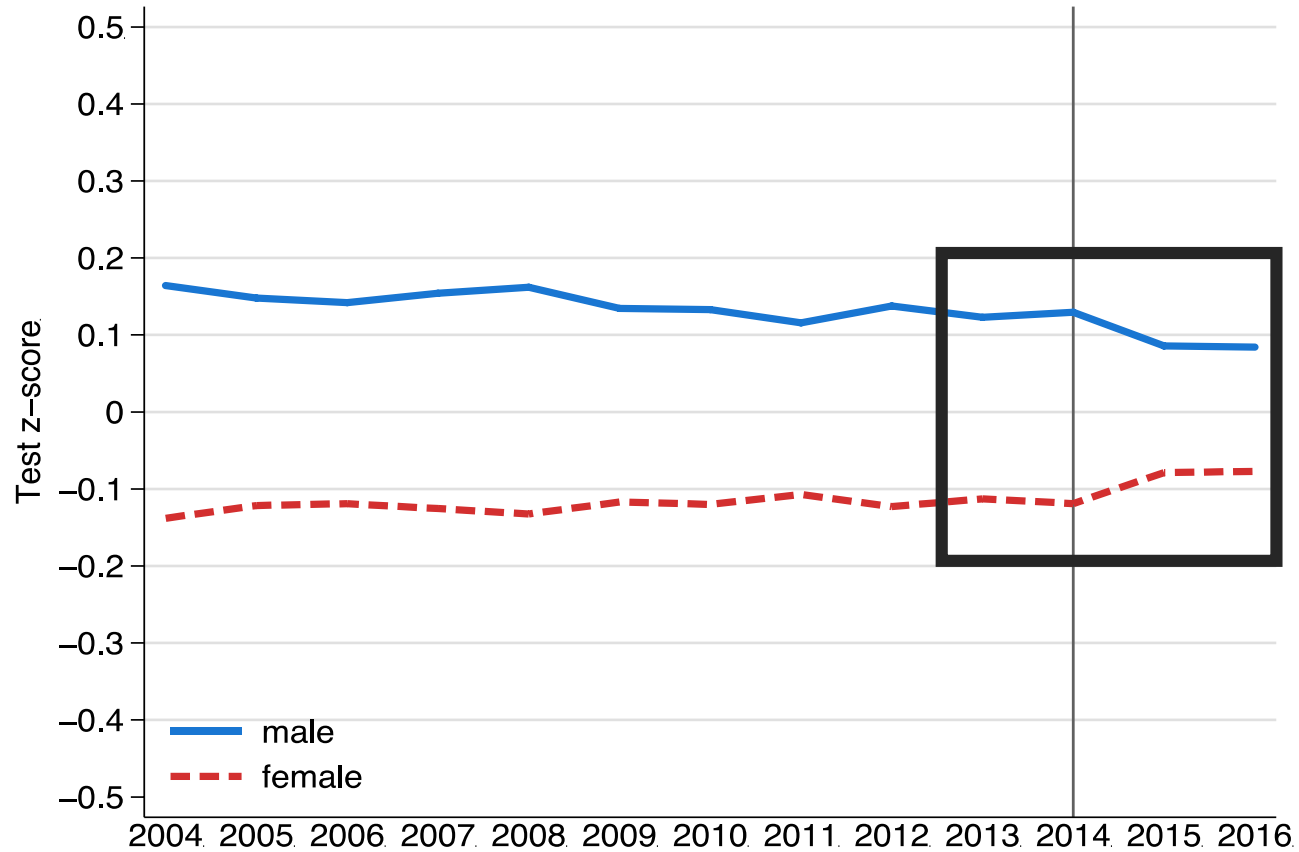
Estrategia: análisis placebo



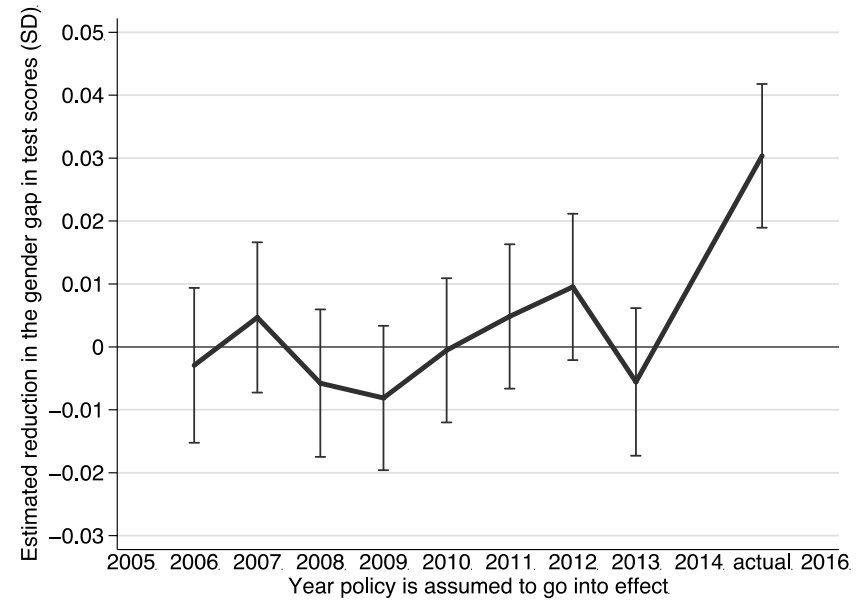
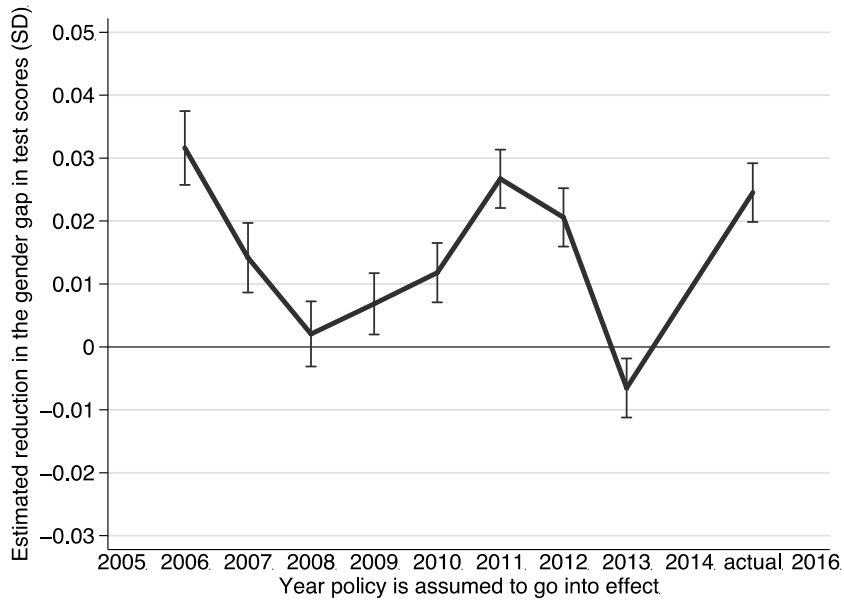
Estrategia: análisis placebo



Estrategia: análisis placebo



Estrategia: análisis placebo

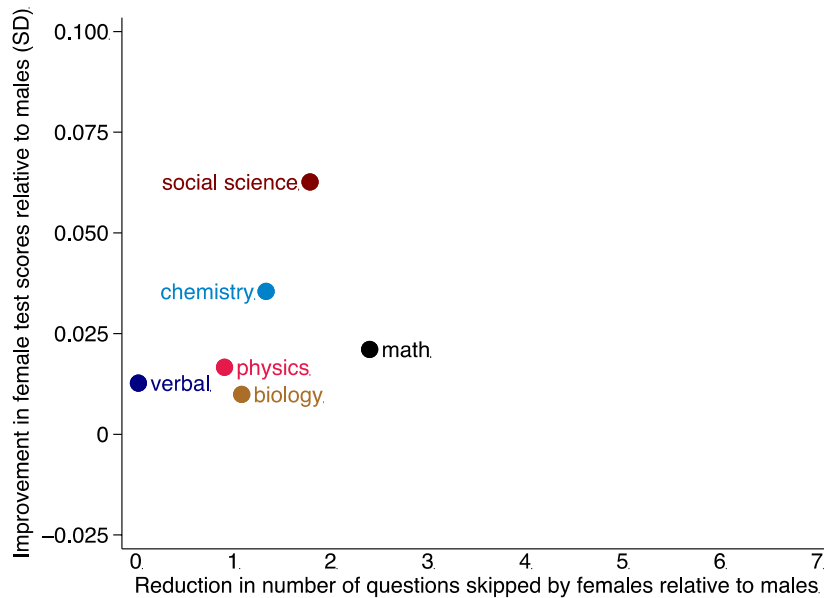


Estrategia: relación entre preguntas en blanco y puntajes

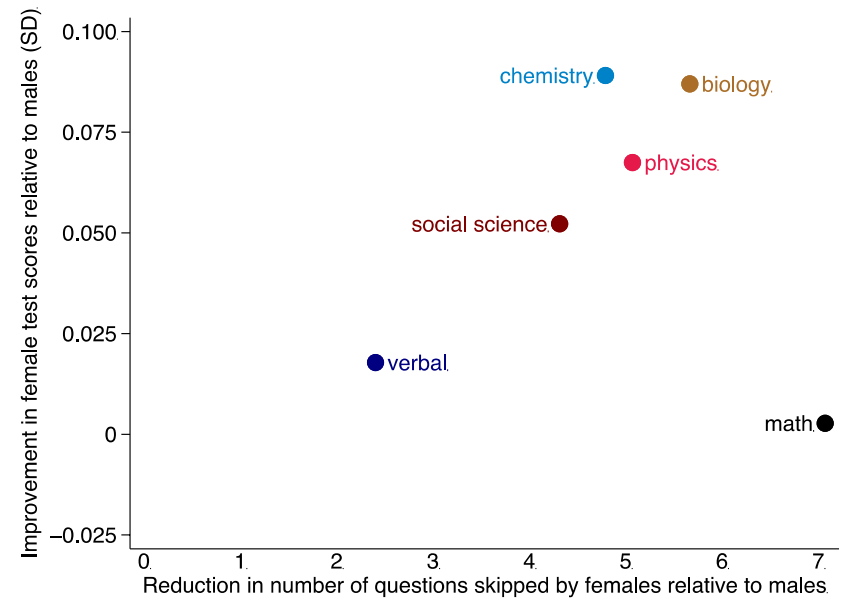
- Si la reforma benefició a las mujeres porque redujo la brecha de preguntas en blanco, se esperaría un mayor beneficio en las áreas donde hubo un mayor cierre en la brecha de preguntas en blanco

Estrategia: relación entre preguntas en blanco y puntajes

Todos (corr=0.46, 0.77)



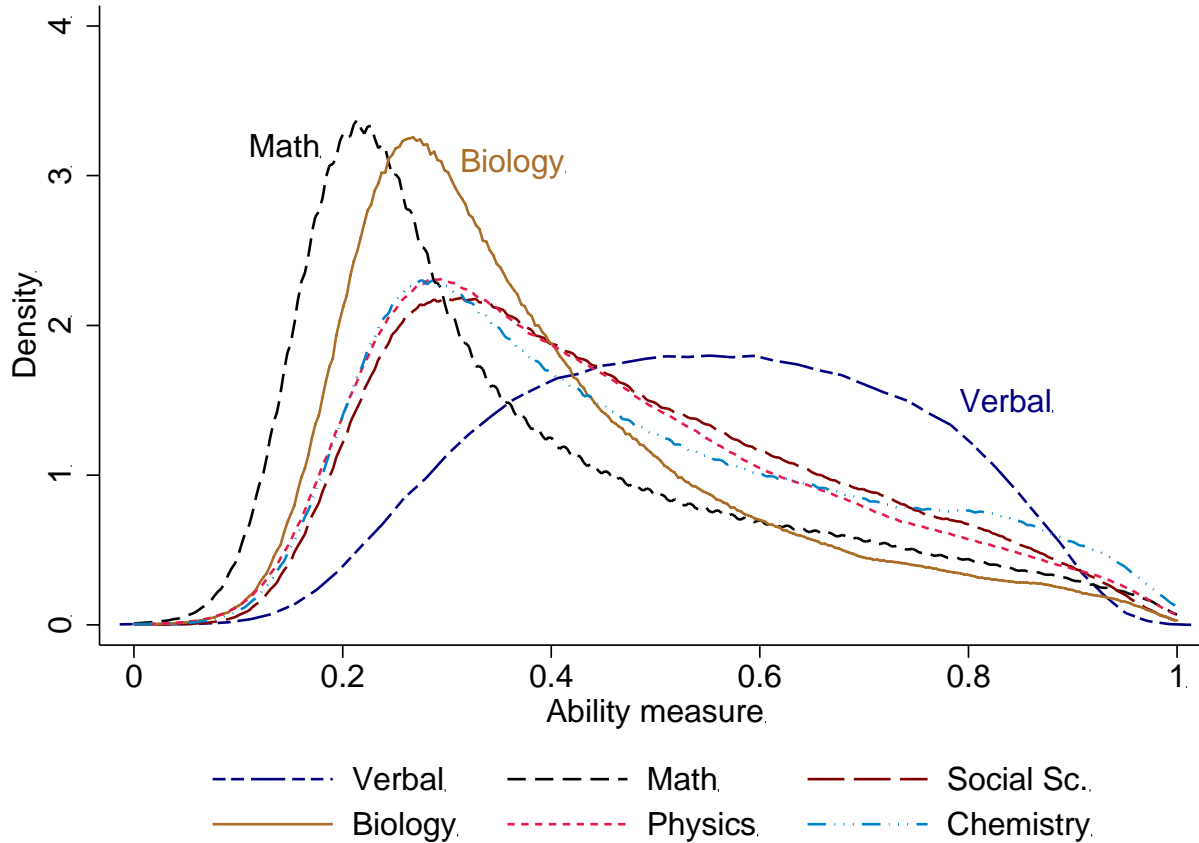
Más habilidosos (corr=0.01, 0.93)



¿Por qué matemáticas es la excepción?

- Una posibilidad es que matemáticas admite en menor nivel adivinar de manera correcta, al ser más difícil descartar opciones incorrectas cuando no se conoce la respuesta
- Matemáticas se vuelve menos predictivo del puntaje promedio escolar luego de la reforma
- Matemáticas es el área más difícil relativo a la habilidad de la población que lo toma

Distribución de puntajes en 2015



Conclusión

- Estudiamos una reforma en un contexto de gran escala y altas consecuencias
- Antes de la reforma, documentamos un significativa brecha de género en preguntas dejadas en blanco y en puntajes en contra de las mujeres
- Eliminar la penalización por respuestas incorrectas elimina prácticamente la práctica de dejar preguntas en blanco y la brecha de género en esta práctica
- Estimamos que la reforma cierra la brecha en puntajes en 0.03 SD, o un 9 por ciento