

BRECHAS DE GÉNERO EN EL RENDIMIENTO ESCOLAR A LO LARGO DE LA DISTRIBUCIÓN DE PUNTAJES: EVIDENCIA PRUEBAS SABER 11**

Luz Karime Abadía¹
Gloria Bernal²

Resumen

Este artículo provee evidencia de que en Colombia las niñas tienen menos acceso que los niños a la educación superior. Usando las técnicas de regresión cuantílica (QR) y la disposición de Juhn-Murphy-Pierce, cuantificamos el efecto de la brecha en matemáticas, ciencias y puntaje global a lo largo de la distribución de puntajes de la prueba Saber 11, y sus potenciales determinantes. Encontramos una brecha significativa en matemáticas y ciencias en contra de las niñas y una brecha pequeña y significativa a favor de las niñas en lectura. Adicionalmente, para todas las áreas, la brecha ajustada se amplía en favor de los niños en la cola alta de la distribución, y el número de niñas que alcanzan puntajes para ser elegibles en la universidad es desproporcionadamente menor comparado con el número de niños. También encontramos que las características del estudiante, su familia y su escuela explican una proporción pequeña de la brecha. Finalmente, proveemos evidencia de que la brecha observada difiere significativamente por región, sugiriendo que las características específicas de cada región – por ejemplo diferencias culturales- podrían influenciar el rendimiento académico de las mujeres.

Palabras clave: brechas de género, exámenes de salida, economía de la educación

* Esta investigación recibió apoyo del Instituto Colombiano para la Evaluación de la Educación -ICFES. Las opiniones, tesis y argumentos expresados son de los autores y no representan el punto de vista del Instituto. Las autoras, profesoras-investigadoras del Departamento de Economía de la Universidad Javeriana, agradecen especialmente a Juan Manuel García Ospina por su colaboración como asistente de investigación.

¹ Profesora Asociada. Departamento de Economía (hasta Junio de 2016). Pontificia Universidad Javeriana, Bogotá-Colombia. E-mail: labadia@javeriana.edu.co

² Profesora Asistente. Departamento de Economía. Pontificia Universidad Javeriana. E-mail: gbernal@javeriana.edu.co

Abstract

This article provides evidence that in Colombia girls have less access than boys to higher education. Using quantile regression (QR) techniques and Juhn-Murphy-Pierce decomposition, we quantify the gender gap in math, science, reading and global scores along the distribution of the high school exit examinations in Colombia (Saber11 examination scores) and assess their potential determinants. We find a significant gap in math and science against girls and a small yet significant gap in reading favoring girls. In addition, for all subjects, the adjusted gap widens favoring boys at the top of the distribution, and the number of girls that achieve scores to be eligible to college is disproportionately less than boys. We also find that gender-based differences in terms of student, family, and school characteristics explain a small proportion of the gap. Finally, we provide evidence that the observed gap differs significantly by region, thus suggesting that the specific characteristics of each region—for example, cultural differences—could influence the academic performance of girls.

Key words: gender gap, exit examinations, economics of education.

1. Introducción

Para la ONU, una de las metas del Milenio planteadas en 2005 para Colombia es la de eliminar las desigualdades entre los géneros en todos los niveles de la enseñanza antes de Diciembre de 2015 (PNUD 2005). En términos de cobertura educativa en primaria y secundaria tanto para niños y niñas, el país ha cumplido la meta (UN 2015). Sin embargo, quedan algunos cuestionamientos sobre el cumplimiento de ésta en términos del rendimiento de las niñas frente a los niños en grado undécimo, y por tanto, sobre la igualdad de oportunidades que tienen las niñas para acceder a la educación superior en el país.

Las desigualdades en el rendimiento académico por género y sus consecuencias en la carrera laboral de las mujeres ha sido objeto de estudio particularmente en países desarrollados. En términos de Schrøter y Skyt (2013), si las mujeres no demuestran habilidades y conocimientos en matemáticas y lenguaje, tienen menos posibilidades de acceder a la universidad, a carreras como ingenierías y ciencias; y por consiguiente, tienen menos chances de obtener mejores oportunidades y salarios en el mercado laboral, relativo a los hombres con similares características. La menor representación de las mujeres en actividades de Ciencia, Tecnología, Ingeniería y Matemáticas (STEM) implica pérdidas económicas significativas para la sociedad (Pollitzer & Schrauder, 2015), está relacionada con segregación ocupacional y desigualdades de ganancias por género. (Loury 1997).

De hecho, trabajos como Murnane, Duhaldeborde y Tyler (2000) y Weinberger (1999) sustentan que demostrar habilidades en lectura, pero especialmente en matemáticas, tiene un impacto significativo en el ingreso futuro de los estudiantes. Un bajo rendimiento de las niñas en áreas de ciencia y matemática se materializa en escasez de profesionales en disciplinas relacionadas con ciencia, tecnología, ingeniería y matemáticas (STEM), lo que a su vez debilita la innovación y refuerza las brechas de género salariales entre hombres y mujeres (Castillo, Grazi & Tacsir 2014).

Además de las consecuencias sobre el mercado laboral, la poca representación de las mujeres en ciertas carreras, puede tener consecuencias en otras dimensiones. Puede generar déficit de innovación, creatividad y competitividad en los negocios, restringe el empoderamiento

y liderazgo de las mujeres en posiciones de poder y reduce las posibilidades de ellas de tener altos retornos (Morales & Sifontes 2014; Castillo, et al. 2014 y Schrøter et al. 2013). Además podría perpetuar ciclos negativos. Autores como Baker y Jones (1993), Güiso, Monte, Sapienza y Zingales, (2008) y Gonzales y De la Rica (2012) encuentran que las niñas se desempeñan mejor en sociedades más igualitarias en términos de género; y este resultado se explica en parte por las expectativas que tiene las mujeres en su sociedad. Es decir, si las mujeres perciben que tienen las mismas oportunidades que los hombres en la sociedad, tienen incentivos a adquirir habilidades en áreas como matemáticas y lectura, lo que a su vez les da más posibilidades de ingresar a la educación superior y de obtener empleos mejor remunerados a futuro.

En Colombia, se encuentran algunos estudios que sugieren una sociedad con altas desigualdades para las mujeres términos en ingreso y participación política. Trabajos como los de Galvis y Bonilla (2011), Hoyos, Ñopo y Peña (2010), Badel y Peña (2010), Abadía y De la Rica (2011) y Bernat (2007) indican que existe una significativa diferenciación salarial en favor de los hombres. Así mismo, la participación femenina en cargos de elección popular es todavía muy limitada en todos los niveles territoriales, incluso con la ley de cuotas políticas femeninas³; en estamentos políticos como asambleas departamentales el porcentaje de participación llega al 15% cuando la ley de cuotas establece que debe ser del mínimo 30% (PNUD, 2005). Comparado con otros países, Colombia es percibida como un país de altas desigualdades de género⁴.

A pesar de la evidente brecha de género en los ámbitos laboral y de participación política, poco se ha explorado en el país sobre las brechas que preceden estas desigualdades, como por ejemplo la brecha educativa. Este trabajo pretende aportar en este sentido. Usando la base de datos del examen SABER 11°, esta investigación busca acometer las siguientes preguntas: ¿Cómo cambia la brecha de género observada y ajustada a lo largo de la distribución de puntajes⁵? ¿Qué proporción de la brecha es explicada por diferencias en características personales, familiares y escolares de los estudiantes? ¿Qué proporción de la

³ Ley 581 del año 2000, la cual establece que las mujeres deben ocupar al menos el 30 % de los cargos directivos de los máximos niveles decisorios, en el sector público.

⁴ El Global Gender Gap Index del año 2014, calculado por el World Economic Forum, muestra que Colombia ocupa el puesto 53 entre 144 países analizados, confirmando que es un país desigual en términos de género. Este índice mide las brechas entre hombres y mujeres en cuatro categorías: Participación económica y oportunidades, educación, salud y sobrevivencia y empoderamiento político.

⁵ La brecha observada se refiere al cálculo del puntaje de las niñas menos el de los niños. En cambio, la brecha ajustada se mide también la diferencia, pero se controla por factores personales, familiares, escolares y regionales.

brecha es explicada por el retorno o impacto de las características personales, familiares y escolares? ¿En qué regiones la brecha alcanza magnitudes más altas relativas a otras regiones?

Los resultados indican que en general hay una brecha académica entre niños y niñas que se amplía a favor de ellos a lo largo de la distribución de puntajes, en el puntaje global, ciencias y matemáticas. También se encontró que hay una notable menor representación de niñas en los cuantiles altos de la distribución y que las características individuales, familiares y escolares, aunque significativas, explican una proporción pequeña de las diferencias.

Se destaca que la magnitud de la brecha a lo largo de los departamentos varía significativamente, encontrando incluso regiones donde la diferencia académica está a favor de las niñas, sugiriendo, en concordancia con Güiso et al (2008), que aspectos particulares de los Departamentos (como contexto, cultura, costumbres) pueden estar influenciando el desempeño de las niñas. Estos resultados indican que para aumentar el rendimiento de las niñas, probablemente hay que tener políticas diferenciadas por desempeño y región, pues aquellas que se hacen pensando en la niña promedio de cierta región pueden no funcionar para las niñas de alto rendimiento en otra región.

Debido a la naturaleza no experimental de los datos no es posible inferir relaciones causales de los determinantes de la brecha. Sin embargo, los hechos y las asociaciones encontradas en este trabajo proveen fuerte evidencia para asegurar que las niñas están teniendo menos oportunidades de llegar a la educación superior de calidad, y a carreras de altos ingresos, comparado con los niños. Estos hallazgos ponen en cuestionamiento el cumplimiento del compromiso del país en la ONU de eliminar las desigualdades de género, en todos los niveles de la enseñanza, antes del fin del año 2015.

El resto de esta investigación se estructura de la siguiente manera. La sección 2, contiene un resumen de la literatura relevante para el desarrollo de este trabajo. La sección 3 describe la base de datos SABER 11° suministrada por el ICFES para el estudio. La sección 4 presenta estadísticas descriptivas de las variables de interés. La sección 5 describe las metodologías usadas para la estimación de la brecha, mientras que la sección 6 presenta las regresiones y discusiones de los resultados obtenidos. En la sección 7 se encontrarán algunas

recomendaciones y sugerencias de políticas relacionadas con los hallazgos y las conclusiones de la investigación.

2. Revisión de la literatura

La literatura sobre la existencia de brechas de género en el desempeño escolar, sus causas y su relación con el bienestar de las sociedades es diversa y extensa. Se resaltan, no obstante, dos corrientes de pensamiento que explican por qué en promedio las niñas obtienen menores puntajes que los niños en matemáticas y mayores en lectura. Por una parte, se plantea la corriente de pensamiento que sugiere que la diferencia en el rendimiento entre géneros está dada por aspectos biológicos. Trabajos como los de Lawton y Hatcher (2005), Temple y Cornish (1993), Gallagher, Levin y Cahalan, (2002), Halpern (2004), Kucian, Loenneker, Dietrich, Martin y Von Aster (2005) y Gallagher y Kaufman (2005) sugieren que las diferencias se explican porque hombres y mujeres difieren en sus habilidades visuales-espaciales, patrones cognitivos y funcionamiento del cerebro. No obstante, estas teorías no dan total cuenta de las diferencias, y resultan insuficientes para explicar hechos como la ausencia de brecha de género en algunos países o en algunas culturas.

Por otra parte, la corriente de pensamiento desarrollada principalmente por la sociología y la economía indica que las brechas de género se deben a diferencias en características sociales o culturales de los estudiantes. Autores como Baker et al (1993), Riegle-Crumb (2005), Güiso et al (2008), Hyde y Mertz (2009), Fryer y Levitt (2010), y Gonzáles et al (2012) encuentran que las niñas se desempeñan mejor en sociedades más igualitarias en términos de oportunidades para la mujer. En esta corriente toma fuerza el argumento que la baja representación de las mujeres, en comparación con los hombres, en algunas carreras (como ingenierías, matemáticas y ciencias, y en programas de doctorado), y en cargos de poder público y político y la desigualdad de oportunidades en general, hacen que las niñas vean menos utilidad que los niños en rendir en el estudio y las desincentiva a ser competitivas en matemáticas. Esto sugiere que factores asociados al contexto, y no solo aspectos biológicos, contribuyen de manera amplia y significativa a la explicación de las diferencias de género.

En esta línea, pero para 19 países en África se destaca el trabajo de Dickerson, McIntosh y Valente (2015), los cuales estiman la brecha en puntajes de matemáticas en pruebas estandarizadas para estudiantes de 13 años. Encuentra que aunque hay diferencias

significativas a favor de los niños en matemáticas, y que estas varían significativamente a través de las regiones. Muestran que, por ejemplo, en regiones donde hay mayor proporción de mujeres con baja educación, mayores tasas de fertilidad y mayor participación de población musulmana la brecha observada promedio es más alta, y que la brecha ajustada disminuye significativamente al controlar por cada una de estas características. Concluyen además que las diferencias regionales son más predictivas de la brecha de género que características como la educación de los padres o variables del colegio.

En términos de estimación de la brecha a lo largo de la distribución, se han realizado algunos estudios principalmente para el área de matemáticas. Los resultados son mixtos para la cola baja de la distribución, pero consistentes para la cola alta: la brecha observada en matemáticas es más alta en el extremo superior de la distribución que en la media (Guimarães & Sampaio, 2008 y Hedges & Nowell, 1995). Además, Penner y Paret (2008), Xie y Shauman (2003) y Ellison y Swanson (2009), al menos para Estados Unidos, encuentra que hay una notable menor representación de niñas en percentiles altos de la distribución; sugiriendo que las niñas con excepcional rendimiento están concentradas en colegios de élite y que las niñas con habilidades para alcanzar altos rendimientos en matemáticas no lo están haciendo.

Para el caso colombiano, se encuentran dos estudios que se refieren a la brecha de género en rendimiento académico. ICFES (2013) hace un análisis descriptivo de la brecha promedio observada de género en las pruebas nacionales (Saber 3°, 5°, 9° y 11°) e internacionales (TIMSS y PISA); e identifica algunos factores motivacionales, cognitivos, actitudinales y escolares que podrían estar asociados al rendimiento. En todas las pruebas, para los diferentes grados, es consistente que las niñas muestran desventajas en matemáticas respecto a los niños y ventajas en lectura; aunque para secundaria se hacen mayores las diferencias.

Por otra parte, Abadía (2014) calcula la brecha de género observada y ajustada por características personales, familiares y escolares, haciendo uso de los resultados de matemáticas y lenguaje de las pruebas PISA 2012 para Colombia. Destaca que la brecha observada está a favor de los niños en matemáticas y a favor de las niñas en lenguaje. Sin embargo, una vez se controla por características personales, familiares y escolares, la brecha de matemáticas se acentúa mientras que la de lenguaje se reduce. Estos trabajos estiman la

brecha en la media de la distribución, pero no a lo largo de esta, como es la intención de la presente investigación.

Este trabajo contribuye en varios aspectos al caso Colombiano. Es el primero que mide la brecha de género en el rendimiento escolar usando el examen de estado (Saber 11) en Colombia. Esta prueba es de carácter censal lo cual permite reducir posibles sesgos de estimación de los parámetros. Segundo, es el primer estudio en medir y analizar la brecha de ciencias. Tercero, mide la brecha de género escolar no solamente en la media sino a lo largo de la distribución de puntajes. Cuarto, se usa la técnica de Juhn, Murphy y Pierce para descomponer la brecha en diferentes puntos de la distribución de puntajes y determinar en qué medida las características individuales, familiares y escolares explican las diferencias de género. Por último, es el primer estudio que cuantifica diferencias de género en el rendimiento escolar en los diferentes Departamentos de Colombia.

3. Datos

Se usa información de la prueba SABER 11° administrada por el ICFES para el año 2014, calendario A. Se escogió 2014 por ser la base de datos más actualizada disponible al momento de realizar la investigación. Esta prueba está diseñada para medir las competencias de los estudiantes en grado undécimo y es usada como mecanismo de selección de los estudiantes para el ingreso a la educación superior, y para la obtención de becas estudiantiles⁶. Además de los puntajes, la base contiene información acerca de las características socioeconómicas de cada estudiante, así como de algunas características de sus familias, escolares y regionales.

La prueba SABER 11° tiene varias ventajas sobre otras pruebas como PISA y TIMMS, que también miden el desempeño académico de los estudiantes. A saber, la prueba SABER 11°: 1) es de carácter censal, lo que permite reducir los errores de estimación de los parámetros poblacionales; 2) tiene información regional, la cual puede ayudar a explicar las diferencias de género observadas; 3) puede ser un mejor instrumento de medición de las competencias de los estudiantes, porque estos pueden estar motivados a hacer su mejor esfuerzo, dada la

⁶ Por ejemplo, becas como Andrés Bello y Ecopetrol, Ser Pilo Paga, y algunas otras dadas por las Universidades son otorgadas a los estudiantes que obtengan determinados puntajes en la pruebas Saber 11°.

importancia de la prueba para el ingreso a la educación superior. Sin embargo, la desventaja evidente de la base es que se dificulta hacer comparaciones con otros países por ser una prueba nacional.

La prueba SABER 11° aplicada en el segundo semestre de 2014⁷ se compone de 5 áreas: matemáticas, lectura crítica, sociales y competencias ciudadanas, ciencias naturales e inglés. Además, tiene un puntaje global que es la suma ponderada de los puntajes de las pruebas individuales⁸. Los puntajes de las 5 áreas van en una escala de 0 a 100, con media 50 y desviación estándar 10. El puntaje global va en una escala de 0 a 500, con media 250 y desviación estándar 50. La distribución de estos puntajes es producto de una transformación hecha por el ICFES a los puntajes rasos, siguiendo la metodología de respuesta al ítem⁹.

Para la estimación de la brecha, esta investigación usa cuatro puntajes como variables dependientes: puntaje global, matemáticas, ciencias y lectura crítica. Las tres últimas se usan por ser las áreas usualmente analizadas en la literatura para la estimación de brechas académicas de género, y por ser éstas transversales a la mayoría de las disciplinas; y puntaje global, por ser este el criterio usado por universidades y otras organizaciones para el acceso a los programas y el otorgamiento de becas¹⁰.

La base SABER 11° del segundo semestre del año 2014 originalmente cuenta con 541.874 observaciones. Sin embargo, para tener una muestra más homogénea, se excluyeron estudiantes menores de 14 años y mayores de 21 y personas que hayan presentado un examen con características particulares (usualmente por cuestiones de discapacidad). Adicionalmente, se excluyeron individuos que no reportaron información en las variables de interés. El número de observaciones definitivo es de 504.085 estudiantes (93% de las observaciones originales).

⁷ Cabe anotar que en la muestra no se incluye a los estudiantes de calendario B que representan al 5% del total de estudiantes bachilleres del país, y cuyas condiciones económicas son en promedio más favorables que las de los estudiantes en calendario A.

⁸ El puntaje global se calcula como el promedio ponderado de las 5 áreas del examen, donde inglés tiene una menor ponderación.

⁹ A diferencia de la teoría Clásica, la teoría de Respuesta al Ítem (TRI) permite obtener mediciones invariantes respecto a los instrumentos utilizados y sujetos implicados (Muñiz 1997). Es decir, con la transformación a la TRI (mediante una función es probabilística, no lineal y abstracta), se espera que los resultados de los test no dependerán de la muestra sino que los puntajes pueden ser comparables entre cohortes.

¹⁰ Por ejemplo, el programa “ser pilo paga”, que beca a 10.000 estudiantes en 2014, asignó que los estudiantes con un puntaje global inferior a 310, y que además cumplieran con otras características eran beneficiarios de una beca completa de matrícula y sostenimiento por toda su estancia en la universidad.

4. Estadísticas descriptivas

La Tabla 1 muestra las estadísticas descriptivas de todas las variables usadas en la estimación de la brecha, en conjunto y separadas por género. Al principio de la tabla se muestran los puntajes: global, matemáticas, ciencias y lectura. La columna 1 muestra que el puntaje global presenta media de 252 y desviación estándar de 42; mientras que el promedio en matemáticas, ciencias y lectura es aproximadamente de 50.6 y desviación estándar de 10. Las medias de los puntajes están ligeramente por encima de los parámetros de la normalización hecha por el ICFES (250 puntos para puntaje global y 50 puntos para matemáticas, ciencias y lectura), debido a que las personas que quedaron fuera de la muestra presentan puntajes ligeramente más bajos que el resto.

De acuerdo con la columna 1 de la Tabla 1, en general, los estudiantes colombianos en grado undécimo calendario A se caracterizan por tener condiciones socioeconómicas desfavorables y por estar concentrados en algunos departamentos del país¹¹. En promedio, los estudiantes son de bajos recursos económicos (41% y 35% pertenecen a estratos 1 y 2, respectivamente); tienen padres con bajo nivel educativo (53% de los padres y 50% de las madres tienen como máximo nivel de educación básica primaria); estudian en colegios públicos (75%), y atienden a la jornada de la mañana (53%) o la completa (20%). Por otra parte, la mayoría de estudiantes vive en zonas urbanas (80%) y su concentración se da principalmente en los departamentos de Cundinamarca (24%), Antioquia (13%), Valle del Cauca (7.5%) y Atlántico (5.3%).

Adicionalmente, alrededor de la mitad de las madres de estos estudiantes se dedican al hogar (52%), mientras que casi todos sus padres tienen una ocupación fuera de casa (96%). La edad promedio de los estudiantes de la muestra es 16.8 años y solo el 2% reporta haber repetido un año escolar. Además, se observa una proporción más grande de niñas (54.5%) que de niños; que en números absolutos son 274.685 niñas y 229.400 niños. Esta desproporción es importante porque más adelante se analizará no solo la brecha de género a lo largo de la distribución de puntajes, sino además la presencia del número de niñas en las colas altas de dicha distribución.

Las columnas 3 y 5 de la Tabla 1 muestran las estadísticas de las variables por género. Los tres puntajes reportan diferencias entre niños y niñas (columna 7). En matemáticas,

¹¹ Esta muestra incluye un total de 32 departamentos más la capital del país, Bogotá.

ciencias y puntaje global la brecha promedio observada está a favor de los niños por 2.2, 2.9 y 5.8 puntos respectivamente (equivalente a 0.22, 0.29 y 0.12 desviaciones estándar). En lectura, la brecha está a favor de las niñas, pero esta solo es de 0.64 puntos (0.064 desviaciones estándar). Todas las diferencias son estadísticamente significativas (columna 4).

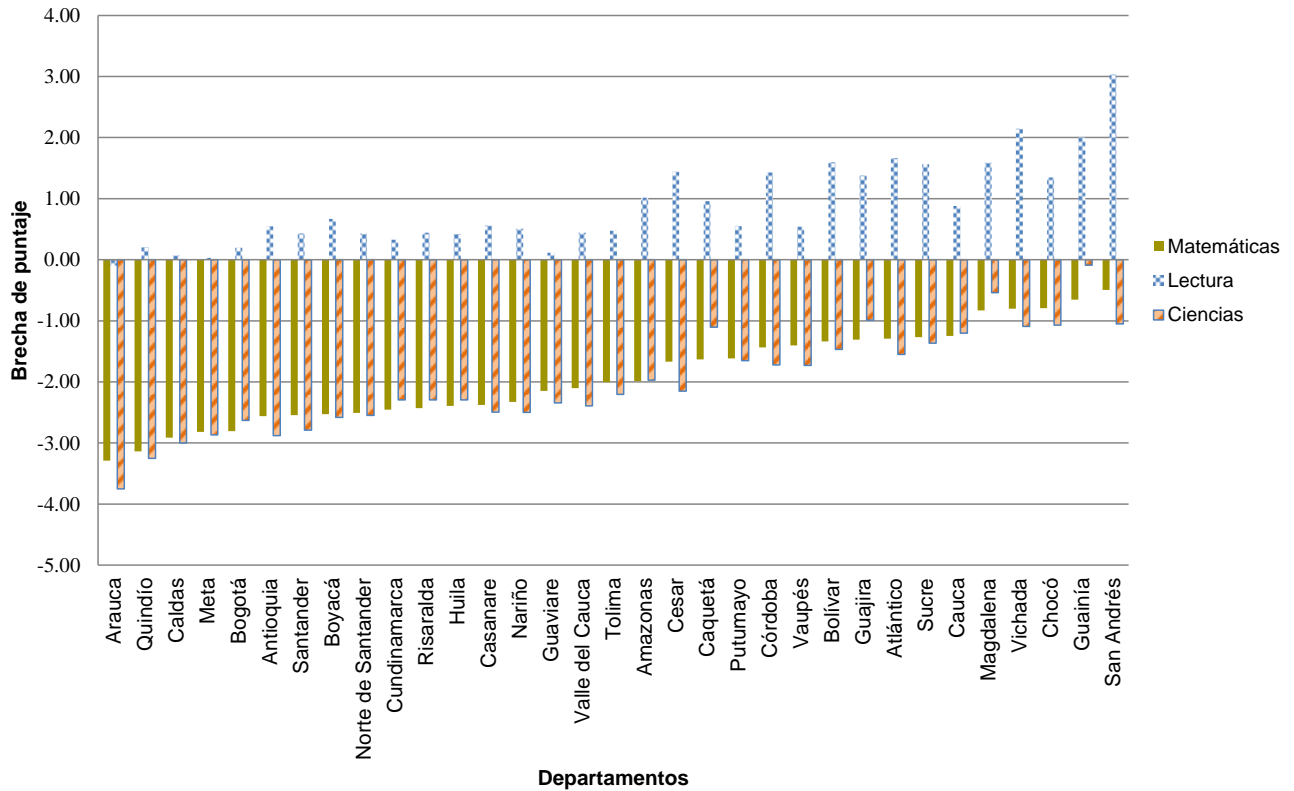
Las diferencias de género de las variables familiares, escolares y regionales en la mayoría de los casos son sutiles aunque significativas. Sin embargo, en general se resalta que las niñas están expuestas a características ligeramente menos favorables para un mejor desempeño académico comparado con los niños.

En particular, una proporción más alta de niñas (3% más) reporta estar en el estrato socioeconómico más bajo (1) comparado con los niños, mientras que para el resto de estratos los niños superan ligeramente a las niñas en proporción. Además, en promedio, un porcentaje mayor de niñas tiene padres y madres sin educación (24.4% y 19%), comparado con los niños (21.9% y 16.8%); y un porcentaje menor de niñas tiene padres y madres con educación superior (9.4% y 9.5%) comparado con los padres y madres de los niños (10.6% y 10.8%). Asimismo, las niñas tienen más probabilidades de estar inscritas en colegios públicos, cuyos desempeños son en promedio más bajos que los de los colegios privados. En cuanto a la distribución regional, las proporciones de niños y niñas son relativamente similares al promedio.

Por otro lado, la brecha observada entre regiones es bastante heterogénea. La figura 1 muestra la brecha observada en matemáticas, ciencias y lectura por departamentos. En general, la brecha de género está a favor de los niños en matemáticas y ciencias, mientras que está ligeramente a favor de las niñas en lectura. Sin embargo, la magnitud de ésta varía por regiones. En Arauca por ejemplo la brecha en matemáticas y ciencias es de -3.2 y -3.8 puntos respectivamente mientras que en San Andrés es de -0.5 y -1.1. Así mismo, la brecha en lectura tiende a cero en Arauca pero es de 3 en San Andrés¹². Además, la correlación entre las brechas observadas de matemáticas y lectura por regiones es de aproximadamente 0.88. Es decir, que regiones con mayor brecha en matemáticas en contra de las niñas tienen menor brecha en lectura a favor de las niñas. La correlación entre ciencias y lectura es similar a lo encontrado entre matemáticas y lectura. Estos resultados consistentes con la literatura (i.e. Güiso et al. 2008).

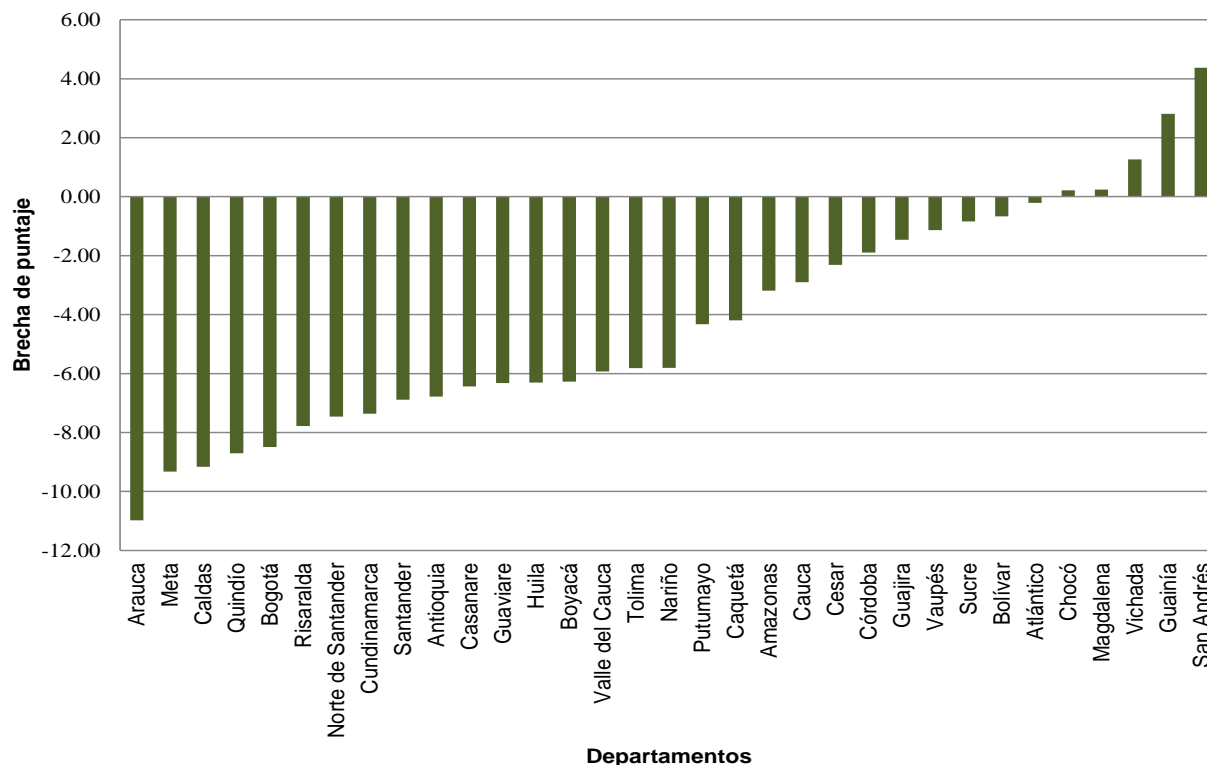
¹² Se realizaron pruebas estadísticas donde se determinó que estas brechas son significativamente diferente de cero.

Figura 1. Brecha de género en matemáticas, lectura y ciencias por Departamentos



La figura 2 muestra la brecha en puntaje global para los 32 departamentos y la capital del país. A saber, 22 departamentos muestran una brecha negativa y estadísticamente significativa. Arauca, Quindío, Caldas, Meta y Bogotá son las regiones donde se observan mayores brechas en contra de las niñas en matemáticas y ciencias; y menores o nulas diferencias a favor de las niñas en lectura. La heterogeneidad de brechas entre regiones, sin que la distribución de niños y niñas sea diferente dentro de ellas, da lugar a pensar que en efecto aspectos sociales, culturales y/o regionales pueden relacionarse con el desempeño de las niñas.

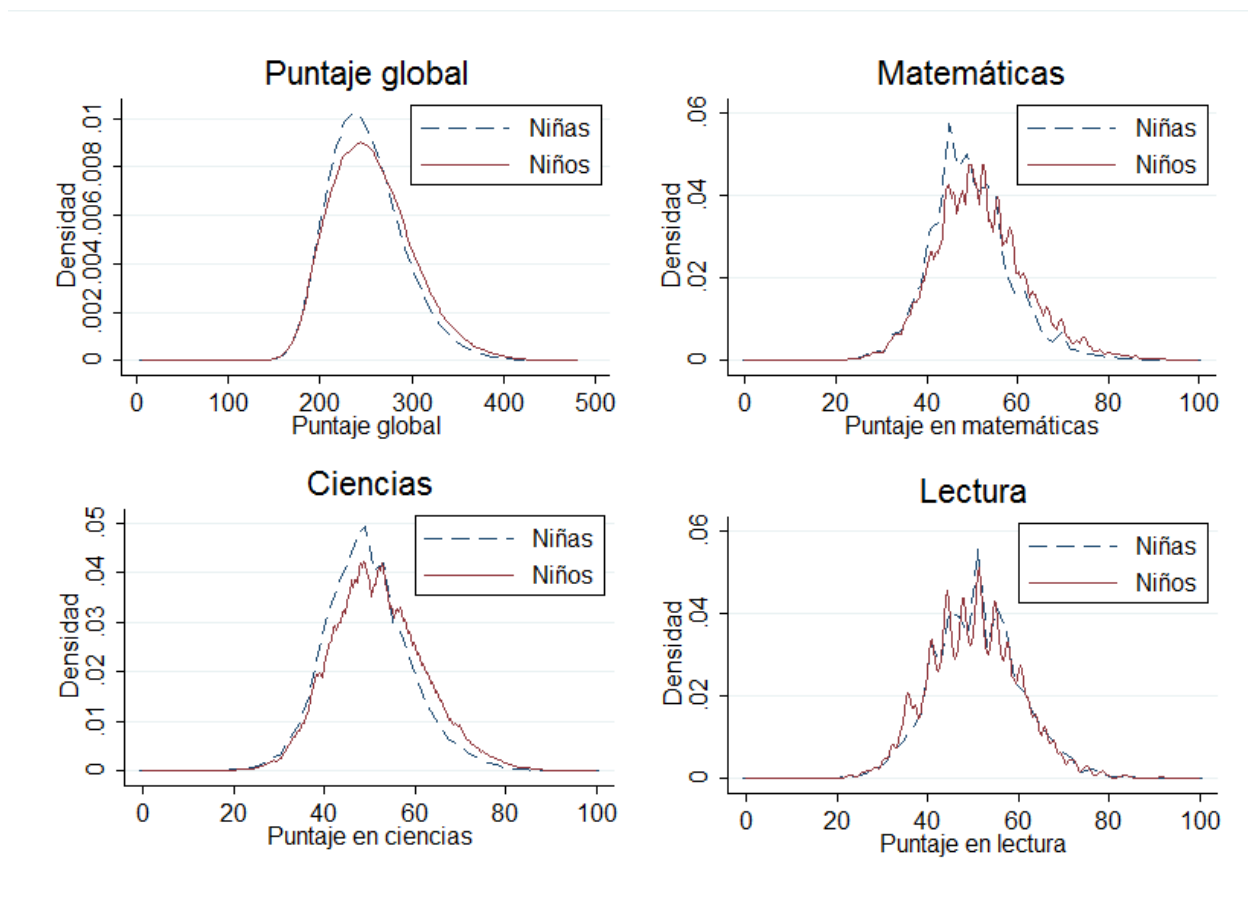
Figura 2. Brecha de género en puntaje global por Departamentos



La figura 3 muestra los kernels de las distribuciones de puntaje global, matemáticas, ciencias y lectura, respectivamente, de los niños y las niñas. Se observa que en la cola baja de las tres distribuciones hay coincidencias. Sin embargo, para puntaje global, matemáticas y ciencias, la distribución de puntajes de las niñas está a la izquierda de la de los niños para la parte media y alta de la distribución, insinuando que la brecha de puntajes por género se amplía en la medida en que aumentan estos puntajes. La brecha para lectura es menos concluyente a partir del kernel. La diferencia de la brecha a lo largo del kernel, resalta la importancia de estimar la brecha no solo en la media, sino también a lo largo de la distribución¹³.

¹³ La prueba Kolmogorov-Smirnov permite rechazar la hipótesis nula de igualdad entre las distribuciones de niñas y niños de los cuatro puntajes analizados.

Figura 3. Distribuciones Kernel de los puntajes global, matemáticas, ciencias y lectura



Adicionalmente, la tasa de niñas/niños a lo largo de la distribución es desigual. La tabla 2 reporta las proporciones de niños a niñas a lo largo de los percentiles de la distribución. La proporción de niñas/niños en toda la muestra es de 1.2. Si los niños y niñas estuvieran igualmente distribuidos a lo largo de los puntajes, se esperaría que esta proporción se replicara en todos los cuantiles de la distribución. Sin embargo, las niñas están sub-representadas en la cola alta de la distribución en puntaje global, matemáticas y ciencias (en el cuantil 95 la proporción es 0.75, 0.60 y 0.63 respectivamente). En otras palabras, pese a que hay más niñas en el sistema educativo que niños, hay un menor número de niñas en los cuantiles altos de la distribución, sugiriendo que un menor número de niñas comparado con el número de niños, tiene oportunidades de acceder a educación superior.

5. Metodología

En esta sección se describen las metodologías econométricas usadas para calcular la brecha de género ajustada en el rendimiento escolar (la brecha entre niñas y niños con similares características observables), así como para determinar qué proporción de la brecha observada se debe a que niños y niñas tienen características promedio diferentes y qué proporción al retorno de dichas características.

Para calcular la brecha de género ajustada en el rendimiento escolar, se usan técnicas econométricas, tales como Mínimos Cuadrados Ordinarios (MCO) y Regresión Cuantílica (RQ). Estas técnicas, permiten calcular la brecha ajustada en promedio y a través de diferentes cuantiles de la distribución de puntajes, respectivamente. Las dos metodologías se respaldan sobre las bases de la función de producción de educación propuesta por Hanushek (1979), donde los inputs son los atributos del estudiante, las características de la familia, y los recursos escolares, mientras el output es el rendimiento del estudiante; en este caso, el puntaje global, de matemáticas, ciencias o lectura del estudiante en las pruebas SABER 11.

La ecuación 1 representa la función a estimar por MCO.

$$Y_i = \beta_0 + \beta_1 G + \beta_2 \bar{P}_i + \beta_3 \bar{F}_i + \beta_4 \bar{S}_i + \varepsilon_i \quad (1)$$

Donde Y_i es el puntaje del estudiante (global, matemáticas, ciencias o lectura); \bar{P}_i , \bar{F}_i , \bar{S}_i son vectores de características personales, familiares y escolares del estudiante i , respectivamente. G es una dummy igual a 1 si el estudiante i es una niña y cero 0 si es niño y ε_i corresponde a un término de error que recoge la perturbación aleatoria. El parámetro de interés para esta investigación es entonces β_1 . Este parámetro indica la brecha en puntaje entre niños y niñas, con similares características observables. Si β_1 es negativo, significa que hay una brecha en contra de las niñas.

Aunque la regresión por MCO es útil como referente de la magnitud de la brecha entre niños y niñas, otros métodos estimativos resultan más apropiados para hacer un análisis más comprehensivo de la brecha. Ese es el caso de la regresión cuantílica (RQ), desarrollada por Koenker y Basset (1978). Este método permite estimar el impacto de un regresor (en este caso

género) sobre una variable de interés en diferentes puntos de la distribución condicional. A diferencia de MCO, el estimador de la regresión cuantílica es más robusto en presencia de datos que se alejan del promedio, usualmente presentes en grandes bases de datos como esta. El modelo a estimar puede representarse como:

$$y_i = \beta_{Gq}G_i + \beta_{Pq}\bar{P}_i + \beta_{Fq}\bar{F}_i + \beta_{Sq}\bar{S}_i + \mu_{qi}, \text{ y}$$

$$Q_q(y_i|G_i, \bar{P}_i, \bar{F}_i, \bar{S}_i) = \beta_{Gq}G_i + \beta_{Pq}\bar{P}_i + \beta_{Fq}\bar{F}_i + \beta_{Sq}\bar{S}_i \quad (2)$$

La variación con respecto a la notación de MCO está en la adición de la función Q_q a la estimación. Donde q es un número entre cero y uno, $Q_q(y_i|G_i, \bar{P}_i, \bar{F}_i, \bar{S}_i)$, es el cuantil del puntaje escolar condicional al conjunto de regresores de las características personales, familiares y escolares del estudiante i . El estimador cuantílico de q , $\hat{\beta}_q$, minimiza sobre β_q la siguiente función objetivo:

$$Q(\beta_q) = \sum_{i:y_i \geq x'_i\beta} q|y_i - P_i\beta_{Pq} + F_i\beta_{Fq} + S_i\beta_{Sq}| + \sum_{i:y_i < x'_i\beta} (1-q)|y_i - P_i\beta_{Pq} + F_i\beta_{Fq} + S_i\beta_{Sq}|$$

Así por ejemplo, si $q = 0.90$ la función da mayor importancia a la predicción de observaciones $y \geq x'\beta$ que a las observaciones con $y < x'\beta$.

Adicionalmente, se usa la metodología de Juhn, Murphy y Pierce – JMP (1993) para descomponer la brecha observada en la media y a lo largo de la distribución de puntajes. Esta técnica permite descomponer la brecha de género observada en el rendimiento escolar, en diferencias en la distribución de características observables (efecto cantidades), en diferencias en la distribución de los retornos de dichas características (efecto precio) y en cambios en la distribución de los residuos (efecto cantidades y precios de las habilidades no observables).

La distribución de puntajes globales, matemáticas, ciencias o lectura puede ser expresada como:

$$Y_i^G = \hat{\beta}^G X_i^G + F^{G-1}(\theta_i^G | X_i^G) \quad (4)$$

Donde $G = m, f$ indica el género, niños y niñas respectivamente; X_i un vector de características personales, familiares y escolares; y el último término de la ecuación (4), $\varepsilon_i^G = F^{G-1}(\theta_i^G | X_i^G)$ es el inverso de la función acumulada de residuos, la cual se compone de dos

partes: el cuantil de un individuo i en la distribución de residuos de θ_i^G y la función de distribución de residuos, F_i .

La diferencia entre el puntaje global, de matemáticas, ciencias o lectura de niños y niñas ($Y_m - Y_f$), en determinado cuantil, puede ser descompuesto de la siguiente forma:

$$\hat{\beta}^f X_i^m + F^{f-1}(\theta_i^m | X_i^m) - \hat{\beta}^f X_i^f + F^{f-1}(\theta_i^f | X_i^f) \quad (5)$$

$$\hat{\beta}^m X_i^f + F^{m-1}(\theta_i^f | X_i^f) - \hat{\beta}^f X_i^f + F^{m-1}(\theta_i^f | X_i^f) \quad (6)$$

$$\hat{\beta}^f X_i^f + F^{f-1}(\theta_i^f | X_i^f) - \hat{\beta}^f X_i^f + F^{m-1}(\theta_i^f | X_i^f) \quad (7)$$

Donde la ecuación (5), (6) y (7) representan respectivamente, las diferencias en la distribución de características observables (efecto cantidades), las diferencias en la distribución de los retornos de dichas características (efecto precio) y los cambios en la distribución de los residuos (efecto residuos).

Mediante el uso de estas metodologías, se espera capturar la magnitud de la brecha de género y la relación de esta con diferentes atributos del estudiante. Sin embargo, es importante resaltar que esta medición puede distar de una estimación causal, porque puede haber variables omitidas que sesguen de alguna manera los coeficientes. Este sesgo se podría *mejorar* sin embargo, con la inclusión de una base de datos longitudinal, con la cual no se dispone para el momento de esta investigación.

6. Resultados

6.1 Brecha de Género Ajustada

Los resultados de la estimación de la brecha de género ajustada para puntaje global, matemáticas, ciencias y lectura son presentados en la tabla 3. La primera columna muestra la brecha estimada en la media (estimación MCO), mientras que el resto de columnas muestra la estimación de la brecha en los cuantiles 20, 40, 60, 80 y 95 (regresión cuantílica).

La variable dependiente es el puntaje (global, matemáticas, ciencias o lectura crítica) y la variable de interés es el coeficiente de la *dummy* de género. Adicionalmente, tanto en la estimación de MCO como en la de regresión cuantílica se incluyen como controles: i. variables

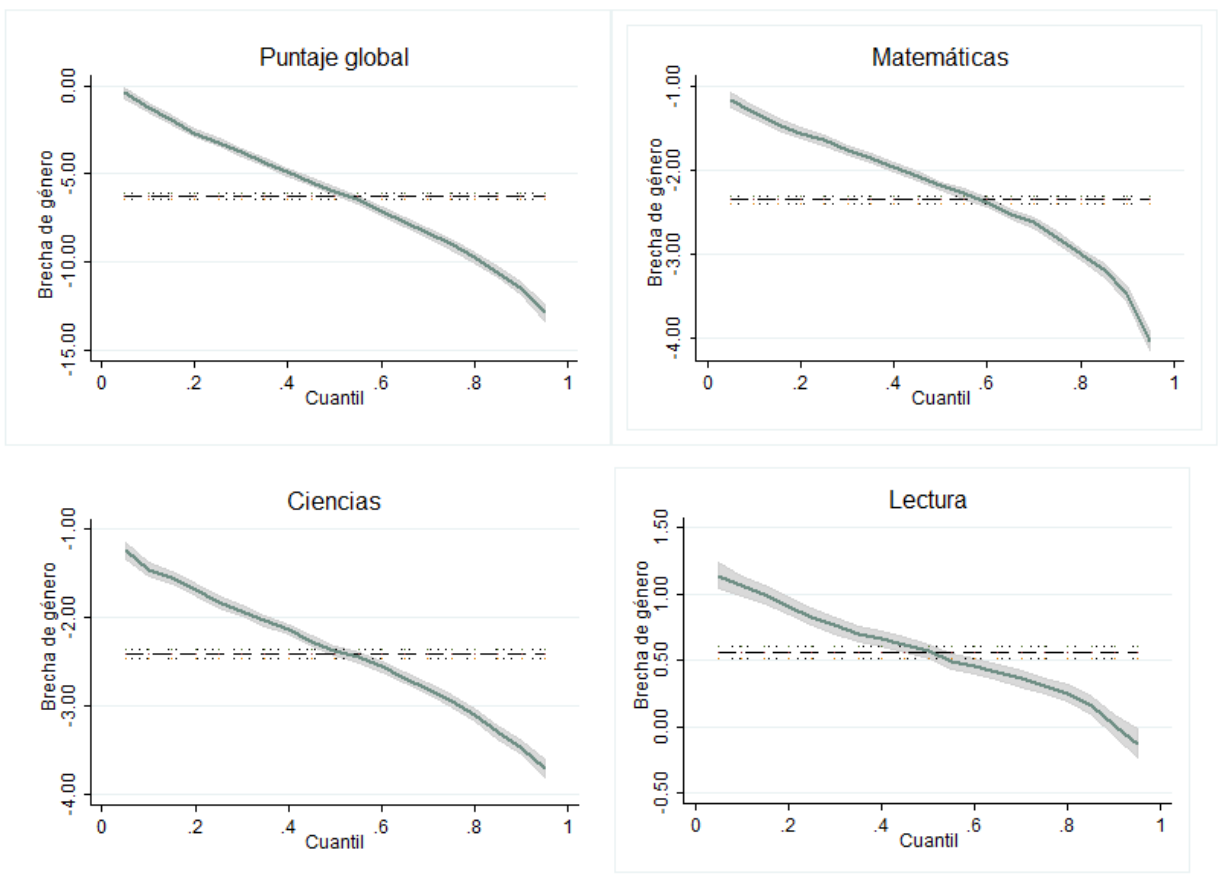
personales tales como género, edad, edad al cuadrado y una *dummy* que indica si el estudiante repitió el examen; ii. variables familiares tales como nivel educativo y ocupación de los padres, estrato socioeconómico, departamento donde vive el estudiante; y iii. variables escolares como jornada, un indicador de la zona donde se ubica el colegio y si el colegio es público o privado.

Los resultados de las estimaciones revelan que el desempeño de las niñas es inferior al de los niños con similares características observables, en el puntaje global, matemáticas y ciencias. Así mismo, se encuentra que a las niñas les va mejor en lectura, aunque las diferencias son muy pequeñas.

La figura 4 ilustra el comportamiento de la brecha ajustada a lo largo de la distribución de puntajes¹⁴.

¹⁴ En cada una de las estimaciones de regresión cuantílica (puntaje global, matemáticas, ciencias y lectura), la prueba de Wald permite rechazar la hipótesis nula, por tanto, los coeficientes estimados de la *dummy* de género son estadísticamente diferentes entre sí a un nivel de significancia del 99%.

Figura 4. Brecha de género a lo largo de las distribuciones de puntajes



El primer panel de la Tabla 3, muestra la brecha estimada para el puntaje global de la prueba. La brecha es negativa y estadísticamente significativa en la media y en los cinco cuantiles analizados. Es decir, a las niñas les va peor que a los niños con similares características a lo largo de la distribución. En la media, la magnitud de la brecha alcanza -6.25 puntos (0.15 desviaciones estándar) y se observa que ésta es ligeramente mayor a la brecha observada promedio (-5.8 puntos o 0.12 desviaciones estándar).

Así mismo, se observa que la brecha es monótonicamente creciente, acentuándose la diferencia en contra de las niñas en la cola alta de la distribución. Es así como en el 20% de los puntajes más bajos, la brecha tiene una magnitud de -2.7 puntos (0.05 desviaciones estándar), mientras que la diferencia de género llega a ser de casi -13 puntos (0.26 desviaciones estándar) comparando niños y niñas dentro del 5% de los puntajes más altos.

Los resultados de las estimaciones en el desempeño escolar de las áreas de matemáticas y ciencias son similares a los del puntaje global, aunque por supuesto de menor magnitud. En el caso de matemáticas, la *dummy* de género es negativa y estadísticamente significativa en la media y a lo largo de la distribución. En promedio, la diferencia es de aproximadamente -2.4 puntos (0.24 desviaciones estándar) en contra de las niñas. Sin embargo, la brecha se acentúa a medida que aumentan los puntajes; en el cuantil 20, la diferencia de género en el puntaje de matemáticas es de -1.6 puntos (0.16 desviaciones estándar) y se va incrementando hasta alcanzar una diferencia de -4 puntos (0.4 desviaciones estándar) en el cuantil 95 (Figura 4).

En el área de ciencias, la brecha ajustada es en promedio de -2.4 puntos (0.24 desviaciones estándar) en contra de las niñas y ésta se acentúa en los puntajes altos. En el primer cuantil analizado ésta es de -1.7 puntos (0.17 desviaciones estándar) y alcanza una magnitud de -3.7 (0.37 desviaciones estándar) en el último cuantil.

Finalmente, el último panel de la Tabla 3 presenta la estimación de la brecha de género en el puntaje de lectura. Los resultados confirman que las niñas tienen ventaja en lectura, aunque su magnitud es pequeña. La *dummy* de género es positiva y estadísticamente significativa en la media y a lo largo de la distribución de puntajes, excepto en el cuantil 95. En promedio, las niñas obtienen cerca de 0.6 puntos (0.06 desviaciones estándar) más que los niños. En el cuantil 20 la diferencia es de 0.9 puntos (0.09 desviaciones estándar), disminuye monótonicamente y se vuelve negativa al comparar los estudiantes que obtuvieron el 5% de los puntajes más altos (-0.127 puntos y 0.013 desviaciones estándar).

En resumen, al controlar por características personales, familiares y escolares, las niñas presentan mayores puntajes en lectura mientras que los niños en matemáticas, ciencias y en el puntaje total. Tanto en matemáticas y ciencias como en el puntaje global, la diferencia en el rendimiento escolar es más pronunciada en la cola alta de la distribución. Sin embargo, es importante resaltar que la brecha a favor de las niñas en lectura es muy baja e inclusive negativa en la cola superior de la distribución¹⁵.

Los resultados obtenidos en la media, son coherentes con lo encontrado en estudios previos en otros países alrededor del mundo en pruebas internacionales estandarizadas de

¹⁵ Como una prueba de robustez de los resultados, se realizó un análisis similar usando información de la prueba Saber 11° de las cohortes de los años 2009 y 2012. A pesar de que en años anteriores al 2014 la prueba Saber 11° tenían una estructura diferente, se encuentra que las características de los estudiantes, así como, las magnitudes y dirección de las brechas de género estimadas son muy similares a las del 2014 en desviaciones estándar.

matemáticas y lectura, tales como las pruebas PISA y TIMS, los cuales muestran que después de controlar por características observables la brecha en contra de las niñas persiste en matemáticas y ciencias (Güiso et. al 2008; Gonzales et. al 2012, Marks 2008 y Dickerson, McIntosh y Valente, 2015).

6.2 Descomposición de Juhn, Murphy y Pierce

En la sección anterior se confirmó la existencia de brechas de género en el puntaje global, matemáticas, ciencias y lectura, aun controlando diferentes atributos observados de los estudiantes. Lo anterior permite concluir que existen variables no tenidas en cuenta en las estimaciones y posiblemente no observables que pueden explicar las diferencias en el desempeño escolar de niños y niñas. Por tanto, en esta sección se muestran los resultados de la descomposición de Juhn, Murphy y Pierce, con el fin de cuantificar la importancia de factores observables y no observables para explicar las brechas de género observadas.

Como se mencionó anteriormente, la técnica de JMP permite determinar qué proporción de la brecha observada se atribuye a diferencias en los atributos de los estudiantes, a diferencias en los coeficientes estimados y en los residuos. La tabla 4 muestra los resultados de esta técnica.

La brecha observada promedio en el puntaje global es aproximadamente de -5.8 puntos (0.12 desviaciones estándar) en el puntaje global. El componente “Efecto Cantidad” es negativo aunque pequeño, es decir, si las niñas tuvieran las características individuales, familiares y escolares de los niños la brecha sería ligeramente mayor, excepto en los puntajes más altos (cuantiles 90 y 95), donde los niños tienen ventaja. Sin embargo, la mayor proporción de la brecha es explicada por los retornos asociados a dichas características (Efecto Precio), el cual favorece a los niños y se incrementa con el puntaje. Adicionalmente, las diferencias en los residuos (el componente de las variables no observadas) cobran importancia para explicar brecha en los puntajes altos.

Los resultados de la descomposición en las áreas de matemáticas y ciencias son similares a los del puntaje global. La brecha observada está a favor de los niños en promedio y en todos cuantiles, excepto en el cuantil 5 para matemáticas. De nuevo, la mayor parte de la brecha es explicada porque los niños presentan mayor impacto de las características observables y éstos se incrementan en la cola derecha de la distribución (este componente

explica un 63% y 60% de la brecha en matemáticas y ciencias respectivamente, el cuantil 95). Así mismo, diferencias en las características no observadas, están positivamente correlacionados con la brecha observada en los puntajes altos. Adicionalmente, aunque las niñas presentan características que las favorecen, la magnitud de este componente es pequeño y no compensa el Efecto Precio.

El último panel de la Tabla 5, presenta la descomposición del área de lectura. Los resultados son diferentes a los observados en matemáticas, ciencias y puntaje global. Como se había mencionado anteriormente, la brecha en lectura está a favor de las niñas aunque su magnitud es pequeña (es en promedio 0.675 puntos, 0.068 desviaciones estándar) y esta ventaja solo ocurre en los puntajes más bajos.

Los anteriores resultados revelan que aspectos como la educación y ocupación de los padres, el estrato, el tipo de colegio, el lugar de residencia, entre otras características observables explican una porción muy pequeña de la brecha, y que existen otro tipo de variables no incluidas en el modelo, como por ejemplo, diferencias de género en motivación, aspectos culturales y del ambiente escolar en el que se desenvuelven niñas y niños en Colombia, así como, diferencias en el rol que ocupan hombres y mujeres en la sociedad colombiana, que podrían explicar estas diferencias¹⁶. Desafortunadamente la base de datos de Saber 11° no cuenta con información que permita dar cuenta de estos factores motivacionales y del ambiente escolar y no hay en Colombia buenas fuentes de información regionales sobre el desempeño de la mujer relativo al hombre en la sociedad Colombiana.

7. Discusión y posibles implicaciones de política

En Colombia existen diferencias importantes de género en el desempeño escolar. Los niños tienen ventaja resolviendo pruebas estandarizadas de matemáticas y ciencias mientras que las niñas la tienen en lectura. Usando información de la prueba de Estado SABER 11°, los resultados de este trabajo revelan que una vez se controla por características personales, familiares y escolares observables, persisten brechas en los puntajes obtenidos por niños y niñas. Se encuentra una brecha negativa y estadísticamente significativa (incluso mayor a 0.3 desviaciones estándar en el percentil 95) que favorece a los niños en puntaje global,

¹⁶ Según el Global Gender Gap Report 2014 (Reporte de brecha de Género 2014), Colombia ocupa el puesto 53 entre 144 países. Mostrando que hay importantes rezagos en término de igualdad de género en el país.

matemáticas y ciencias, la cual se acentúa entre estudiantes con puntajes altos. En el área de lectura, la brecha es inversa. Las niñas obtienen puntajes ligeramente mayores a los niños, pero esta diferencia es cada vez menor a lo largo de la distribución, al punto de volverse a favor de ellos en el último cuantil. La ampliación de las brechas a lo largo de las distribuciones, y el hecho de que hay una menor representación de niñas en puntajes sobresalientes, agudiza el resultado de que las niñas presentan un desempeño desfavorable frente a los niños en el país.

Al realizar un análisis de desempeño por género entre regiones, se encuentra que las brechas difieren significativamente a través de los Departamentos de Colombia. En puntaje global la diferencia en el desempeño escolar está en contra de las niñas en todas las regiones analizadas (donde Arauca, Quindío, Caldas, Meta y Bogotá son las 5 regiones con las brechas más altas). Así mismo, en regiones donde las niñas gozan de mayores ventajas en lectura con respecto a los niños, ellas poseen menores desventajas (e incluso brechas estadísticamente no diferentes de cero en algunos Departamentos) en matemáticas y ciencias. Esta evidencia va en favor de los argumentos que superponen determinantes sociales, culturales y regionales sobre biológicos para explicar las diferencias en rendimiento por género (Güiso et. al 2008).

Se encuentra además, por medio de la descomposición de Juhn, Murphy y Pierce, que una pequeña proporción de la brecha es explicada por diferencias en características personales, familiares y escolares observables; lo cual hace descartar estos factores como ampliamente determinantes de las diferencias. En cambio, una alta proporción de la brecha es atribuida a factores no observados, los cuales cobran mayor importancia en la cola alta de la distribución. Por lo tanto, se infiere que factores no incluidos en las estimaciones, y posiblemente no observables (como variables de motivación, expectativas, ambiente escolar y culturales determinantes del rol de la mujer), posiblemente contribuyen a explicar parte importante de la brecha. Futuras investigaciones podrían explorar esta inferencia.

En cuanto a la implicación de los hallazgos, aunque no es posible inferir relaciones causales de los determinantes de la brecha debido a la naturaleza no-experimental de los datos, los hechos y las asociaciones encontradas en este trabajo proveen evidencia para asegurar que existe una brecha académica significativa en contra de las niñas. De acuerdo con la literatura, este es un problema que tiene consecuencias indeseables para la sociedad (OECD 2011). Un menor rendimiento escolar de las niñas en asignaturas como ciencia y matemáticas, por razones ajenas a su talento o preferencias, implica que las niñas tienen menos

oportunidades de acceso a carreras de Ciencias, Tecnología, Ingeniería y Matemáticas (STEM). Este hecho alimenta la sub-representación de mujeres en estas áreas lo cual representa significativas pérdidas económicas para las sociedades modernas, pues restringe la innovación, creatividad y competitividad en los negocios; coarta el empoderamiento y liderazgo de mujeres en cargos de poder; y reduce la posibilidad de que ellas tengan ingresos futuros altos, propios de las profesiones STEM (Morales & Sifontes 2014, Schrøter & Skyt 2013). Por tanto, este trabajo soporta la necesidad de identificar políticas efectivas que promuevan el desempeño de las niñas en estas áreas, lo que probablemente redundará en: un aumento del número de científicas e ingenieras; una mayor investigación e innovación y un mayor crecimiento económico (OECD 2011).

Algunas acciones de política pueden mitigar estas diferencias y llevar a las niñas a que se desempeñen de maneras más acorde a su potencial. Más concretamente, campañas de sensibilización que aminoren los estereotipos de género en hogares y escuelas (tales como: “las matemáticas son cosa de niños” o “los niños dedíquense a las ingenierías y las niñas a las humanidades o a la cocina”); y la promoción de charlas que involucren ejemplos a seguir de mujeres que ocupan importantes posiciones en la sociedad y en campos como las STEM. Combatir los estereotipos prejuiciosos ayuda a las niñas a motivarse y a esforzarse más en obtener buenos resultados académicos. En algunos países como USA y Canadá, leyes y políticas para combatir el sexismo de los salones de clase han sido analizadas y aplicadas durante años (Ramsey 1998 y Coulter 1996). En Colombia sin embargo, este tema es en el mejor de los casos un tímido debate.

Independientemente de las estrategias que se estudien y acojan, esta investigación sugiere que las medidas a diseñar deben ajustarse al contexto en el que las niñas viven, y apuntar más allá del clásico refuerzo en materias en las que ellas son deficientes. Dadas las grandes diferencias de brechas encontradas a lo largo de la distribución y las regiones, se advierte que las intervenciones para nivelar a las niñas deben ser diferenciadas por las características socio-culturales de las regiones donde las niñas crecen, así como por el grado de desempeño que presentan las estudiantes (alto, medio o bajo). En otras palabras, no es de esperarse que una medida que se hace pensando en la niña promedio de cierta región funcione para las niñas de alto rendimiento en otra región. Así mismo, dada la alta correlación entre brechas de matemáticas, ciencias y lectura, el refuerzo puntual en un tema o un área despierta

menos importancia, pues una efectiva intervención (de carácter motivacional o incluso de mejora puntual en matemáticas), probablemente también jalone el rendimiento de ellas en las otras áreas que no sean tratadas directamente.

Finalmente, en apoyo a los objetivos del milenio de la ONU, este trabajo sugiere el fortalecimiento y/o construcción de escenarios donde las niñas puedan desarrollar mejor su potencial académico. Esto muy probablemente contribuya a una sociedad más igualitaria en género, donde la mujer tenga mejores posibilidades de ingresar a la universidad, de estudiar carreras acordes as su potencial, y de alcanzar posiciones laborales que hagan justicia de sus capacidades.

8. Referencias

Abadía, L. (2014). Performance and Gender Score Gaps of Colombia in PISA Test. Working paper.

Abadía Alvarado L.K and De la Rica S. (2011) Changes in the Gender Wage Gap and the Role of Education and other Job Characteristics Colombia: 1994- 2010. *Vniversitas Económicas*, Vol. 11 No. 8. Departamento de Economía, Pontificia Universidad Javeriana.

Badel, A & Peña, X. (2010). Descomponiendo La Brecha Salarial De Genero Con Ajuste De Sesgo De Selección: El Caso Colombiano. *Revista de Análisis Económico*, Vol. 25, N° 2, pp. 169-191.

Baker, David P., and Deborah Perkins Jones. (1993) Creating Gender Equality: Cross-national Gender Stratification and Mathematical Performance. *Sociology of Education* 66.2 :91–103.

Bernat L.F. (2007) ¿Quiénes son las Mujeres Discriminadas?: enfoque distributivo de las diferencias salariales por género. Borradores de Economía y Finanzas, Universidad ICESI, Cali-Colombia

Castillo, R; Grazzi, M & Tacsir, E. (2014). Women in Sience and Technology. Interamerican Development Bank.

Coulter, R. P. (1996). Gender equity and schooling: Linking research and policy. *Canadian Journal of Education/Revue canadienne de l'education*, 433-452.

Dickerson, A., McIntosh, S. and Valente, C. (2015). Do the maths: An analysis of the gender gap in mathematics in Africa. *Economics of Education Review*. Vol. 46, 1-22.

Ellison G. and Swanson A. (2010). The Gender Gap in Secondary School Mathematics at High Achievement Levels: Evidence from the American Mathematics Competitions. *Journal of Economic Perspectives*, Vol. 24, No. 2, 109–128.

Fryer Jr, R. G., & Levitt, S. D. (2004). Understanding the black-white test score gap in the first two years of school. *Review of Economics and Statistics*, 86(2), 447-464.

Gallagher and Kaufman (2005). *Gender Differences in Mathematics: An Integrative Psychological Approach*. Cambridge University Press.

Gallagher, A., Levin, J., Cahalan, C. (2002) *Cognitive Patterns of Gender Differences on Mathematics Admissions Test*. ETS Research Report.

Galvis, L & Bonilla, L. (2011). *Las Desigualdades En La Distribución Del Nivel Educativo De Los Docentes En Colombia*. Documentos De Trabajo Sobre Economía Regional 008918, Banco De La República - Economía Regional.

Gonzales A. and De la Rica S. (2012). *Gender Gaps in PISA Test Scores: The Impact of Social Norms and the Mother's Transmission of Role Attitudes*. IZA DP No. 6338.

Guimaraes J. and Sampaio B. (2008) *Mind the Gap: Evidences from Gender Differences in Scores in Brazil*.

Güiso, L., Monte, F., Sapienza, P. and Zingales, L. (2008). *Culture, Gender and Math*. *Science*, 320(5880), pp. 1164-1165.

Halpern D. (2004). *A Cognitive-Process Taxonomy for Sex Differences in Cognitive Abilities*. *American Psychological Society*, Vol. 13, No. 4.

Hanushek, E. A.. (1979). *Conceptual and Empirical Issues in the Estimation of Educational Production Functions*. *The Journal of Human Resources*, 14(3), 351–388.

Hedges L. and Nowell A. (1995). *Sex differences in mental test scores, variability, and numbers of high-scoring individuals*. *Science* 7 July 1995: Vol. 269 no. 5220 pp. 41-45.

Hoyos, Alejandro and Ñopo, Hugo and Peña, Ximena, *The Persistent Gender Earnings Gap in Colombia, 1994-2006* (May 2010). IDB Working Paper No. 54.

Hyde J. and Mertz J. (2009). *Gender, culture, and mathematics performance*. Department of Psychology, University of Wisconsin.

ICFES (2013). *Análisis de las Diferencias de Género en el Desempeño de Estudiantes Colombianos en Matemáticas y Lenguaje*.

Juhn C., Murphy K. and Pierce B. (1993), *Wage Inequality and the Rise in Returns to Skills*, *the Journal of political Economy*, Vol. 101, NO. 3, 410-442.

Koenker R. and Basset G. (1978), Regression Quantiles, Vol. 46, No. 1, pp 33-50.

Kucian K., Loenneker T., Dietrich T., Martin E. and Von Aster M. (2005). Gender differences in brain activation patterns during mental rotation and number related cognitive tasks. *Psychology Science*. 47(1): pp. 112-131.

Lawton C. and Hatcher D. (2005), Gender Differences in Integration of Images in Visuospatial Memory. *Sex Roles*, Vol. 53, Issue 9-10, pp 717-725.

Loury, L. (1997). The Gender Earnings Gap among College-Educated Workers. *Industrial & Labor Relations Review* July 1997 50: 580-593

Marks G. (2008). Accounting for the gender gaps in student performance in reading and mathematics: Evidence from 31 countries. *Oxford Review of Education* 34 (1), 89-109

Morales R. y Sifontes D. (2014) "Desigualdad de Género en Ciencia y Tecnología: un estudio para América Latina. *Observatorio Laboral Revista Venezolana*. Vol. 7, No. 13.

Murnane, Richard J., John B. Willett, Yves Duhaldeborde, and John H. Tyler. 2000. "How Important Are the Cognitive Skills of Teenagers in Predicting Subsequent Earnings?" *Journal of Policy Analysis and Management*, 19(4): 547-68.

Murnane Richard J., John B. Willett, and Niederle and Vesterlun (2014). Do Women Shy Away from Competition? Do Men Compete too Much?. *The Quarterly Journal of Economics*.

Niederle, M. and Yestrumskas, A. (2008). Gender Differences in Seeking Challenges. The Role of Institutions. National Bureau of Economic Research. Working Paper No.13922.

Organization for Economic Cooperation and Development (OECD), 2011, Report on the gender initiative: Gender Equality in Education, Employment and Entrepreneurism (OECD publishing).

Penner A. and Paret M. (2008). Gender differences in mathematics achievement: Exploring the early grades and the extremes. *Social Science Research* 37, 239–253.

PNUD. (2005). Hacia una Colombia equitativa e incluyente. Informe de Colombia. Objetivos de desarrollo del milenio 2005. DNP.

Pollitzer, E. and M. Schraudner, 2015, "Integrating Gender Dynamics into Innovation Ecosystems, *Sociology and Anthropology*, Vol 3 No 11.

Ramsey, C. B. (1998). Subtracting sexism from the classroom: Law and policy in the debate over all-female math and science classes in public schools. *Tex. J. Women & L.*, 8, 1.

Riegle-Crumb, C. (2005). The Cross-national Context of the Gender Gap in Math and Science. In L. Hedges & B. Schneider (Eds.), *The social organization of schooling* pp. 227–243.

Schrøter J. and Skyt H. (2013). *Math and Gender: Is Math a Route to a High-Powered Career?*. IZA DP No. 7164.

Temple and Cornish (1993) Recognition Memory for Words and Faces in Schoolchildren: A Female Advantage for Words. *British Journal of Developmental Psychology*, 11: 421–426. doi: 10.1111/j.2044-835X.1993.tb00613.x.

Weinberger, Catherine J. (1999). Mathematical College Majors and the Gender Gap in Wages. *Industrial Relations*, 38 (July):407-413

Xie, Yu, and Kimberlee A. Shauman. 2003. *Women in Science: Career Processes and Outcomes*. Boston, MA. Harvard University Press.

Anexos

Tabla 1. Estadísticas descriptivas por género

Variables	Total	DE	Niños	DE	Niñas	DE	Diferencia
	Media o proporción	(2)	Media o proporción	(4)	Media o proporción	(6)	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Personales							
Género (1=niña ; niño=0)	0.545	0.498					
Puntaje Global (de 0 a 500)	252.6	42.43	255.8	44,51	250	40,43	5,8000 ***
Matemáticas (de 0 a 100)	50.59	9.831	51,8	10,33	49,5	9,26	2,2400 ***
Ciencias (de 0 a 100)	50.57	9.913	51.82	10.29	49.53	9.462	2,2900 ***
Lectura (de 0 a 100)	50.57	9.804	50,2	9,99	50,8	9,63	-0,6400 ***
Edad	16.85	1.103	16,9	1,12	16,7	1,07	0,1900 ***
Repetición	0.0196	0.139	0,023	0,15	0,017	0,13	0,0061 ***
Familiares							
Edu. Padre Ninguna	0.232	0.422	0,219	0,41	0,244	0,43	-0,0250 ***
Edu. Padre Primaria	0.304	0.460	0,304	0,46	0,305	0,46	-0,0010
Edu. Padre Secundaria	0.268	0.443	0,276	0,45	0,261	0,44	0,0150 ***
Edu. Padre Técnica	0.0560	0.230	0,055	0,23	0,057	0,23	-0,0014 ***
Edu. Padre Superior	0.0992	0.299	0,106	0,31	0,094	0,29	0,0123 ***
Edu. Padre no sabe	0.0403	0.197	0.0401	0.196	0.0405	0.197	#####
Edu. Madre Ninguna	0.180	0.384	0,168	0,37	0,190	0,39	-0,0220 ***
Edu. Madre Primaria	0.322	0.467	0,315	0,46	0,328	0,47	-0,0130 ***
Edu. Madre Secundaria	0.309	0.462	0,322	0,47	0,297	0,46	0,0250 ***
Edu. Madre Técnica	0.0736	0.261	0,070	0,26	0,077	0,27	-0,0069 ***
Edu. Madre Superior	0.101	0.301	0,108	0,31	0,095	0,29	0,0129 ***
Edu. Madre no sabe	0.0142	0.118	0.0164	0.127	0.0124	0.111	#####
Padre empresario	0.0195	0.138	0,022	0,15	0,017	0,13	0,0053 ***
Padre empresario pequeño	0.0218	0.146	0,023	0,15	0,021	0,14	0,0028 ***
Padre empleado gerente	0.0191	0.137	0,021	0,14	0,018	0,13	0,0027 ***
Padre empleado directivo	0.0103	0.101	0,011	0,11	0,010	0,10	0,0015 ***
Padre empleado técnico profesional	0.0773	0.267	0,081	0,27	0,074	0,26	0,0065 ***
Padre empleado auxiliar administrativo	0.0204	0.141	0,020	0,14	0,020	0,14	0,0000
Padre obrero/operario	0.230	0.421	0,230	0,42	0,230	0,42	0,0000
Padre independiente	0.0267	0.161	0,028	0,16	0,026	0,16	0,0016 ***
Padre trabajador por cuenta propia	0.374	0.484	0,372	0,48	0,376	0,48	-0,0040 ***
Padre se dedica al hogar	0.00955	0.0972	0,009	0,10	0,010	0,10	-0,0001
Padre pensionado	0.0307	0.172	0,032	0,18	0,030	0,17	0,0026 ***
Padre se dedica a otro oficio	0.161	0.367	0,151	0,36	0,169	0,38	-0,0180 ***
Madre empresaria	0.00882	0.0935	0,010	0,10	0,008	0,09	0,0016 ***
Madre empresaria pequeña	0.0197	0.139	0,021	0,14	0,019	0,14	0,0016 ***
Madre empleado gerente	0.00980	0.0985	0,011	0,10	0,009	0,09	0,0017 ***
Madre empleado directivo	0.0122	0.110	0,013	0,11	0,012	0,11	0,0013 ***
Madre empleado técnica profesional	0.0625	0.242	0,065	0,25	0,061	0,24	0,0037
Madre empleado auxiliar administrativo	0.0485	0.215	0,048	0,22	0,049	0,22	-0,0002 *
Madre obrera/operario	0.0735	0.261	0,074	0,26	0,073	0,26	0,0012
Madre independiente	0.0248	0.155	0,025	0,16	0,025	0,16	0,0001
Madre trabajadora por cuenta propia	0.119	0.324	0,120	0,33	0,119	0,32	0,0010
Madre se dedica al hogar	0.524	0.499	0,523	0,50	0,525	0,50	-0,0020 **
Madre pensionada	0.00918	0.0954	0,010	0,10	0,009	0,09	0,0007 ***
Madre se dedica a otro oficio	0.0876	0.283	0,082	0,28	0,092	0,29	-0,0097 ***
Zona urbana (1= cierto, 0=zona rural)	0.806	0.395	0,806	0,40	0,806	0,40	0,0000 **

Tabla 1. Estadísticas descriptivas por género (continuación)

<i>Variables</i>	Total		Niños		Niñas		Diferencia
	Media o proporción	DE	Media o proporción	DE	Media o proporción	DE	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Estrato 1	0.414	0.493	0,390	0,48	0,420	0,49	-0,0300 ***
Estrato 2	0.351	0.477	0,350	0,47	0,340	0,47	0,0100 ***
Estrato 3	0.182	0.386	0,190	0,39	0,170	0,38	0,0200 ***
Estrato 4	0.0374	0.190	0,030	0,19	0,030	0,18	0,0000 ***
Estrato 5	0.0109	0.104	0,011	0,10	0,010	0,99	0,0010 ***
Estrato 6	0.00424	0.0650	0,005	0,06	0,003	0,06	0,0015 ***
Escolares							
Colegio público (1= cierto, 0=no oficial)	0.754	0.431	0,737	0,44	0,769	0,42	-0,0320 ***
Jornada Mañana	0.536	0.499	0,205	0,40	0,203	0,40	0,0020 ***
Jornada Tarde	0.168	0.373	0,532	0,50	0,540	0,50	-0,0080 ***
Jornada Noche	0.0499	0.218	0,055	0,23	0,046	0,21	0,0092 ***
Jornada Completa	0.204	0.403	0,044	0,20	0,042	0,20	0,0017 **
Jornada Sabática	0.0428	0.202	0,165	0,37	0,170	0,38	-0,0050 ***
Regionales							
Amazonas	0.00122	0.0349	0,001	0,04	0,001	0,03	0,0004 ***
Antioquia	0.132	0.338	0,127	0,33	0,136	0,34	-0,0090 ***
Arauca	0.00484	0.0694	0,005	0,07	0,005	0,07	-0,0003 *
Atlántico	0.0532	0.224	0,055	0,23	0,052	0,22	0,0037 ***
Bogotá	0.175	0.380	0,180	0,38	0,171	0,38	0,0090 ***
Bolívar	0.0448	0.207	0,045	0,21	0,045	0,21	0,0002
Boyacá	0.0326	0.178	0,034	0,18	0,032	0,18	0,0018 ***
Caldas	0.0213	0.144	0,021	0,15	0,021	0,14	0,0001
Caquetá	0.00697	0.0832	0,007	0,08	0,007	0,08	-0,0002
Casanare	0.0103	0.101	0,010	0,10	0,010	0,10	-0,0003
Cauca	0.0248	0.155	0,024	0,15	0,025	0,16	-0,0007 *
Cesar	0.0227	0.149	0,023	0,15	0,023	0,15	0,0001
Chocó	0.00678	0.0821	0,007	0,08	0,007	0,08	-0,0002
Córdoba	0.0348	0.183	0,036	0,19	0,034	0,18	0,0014 ***
Cundinamarca	0.0640	0.245	0,065	0,25	0,064	0,24	0,0011 *
Guainía	0.000367	0.0192	0,000	0,02	0,000	0,02	-0,0001
Guaviare	0.00166	0.0408	0,002	0,04	0,002	0,04	0,0000
Huila	0.0258	0.158	0,026	0,16	0,026	0,16	-0,0003
Guajira	0.0122	0.110	0,011	0,11	0,013	0,11	-0,0017 ***
Magdalena	0.0276	0.164	0,028	0,17	0,027	0,16	0,0011 ***
Meta	0.0208	0.143	0,021	0,14	0,021	0,14	0,0003
Nariño	0.0305	0.172	0,030	0,17	0,031	0,17	-0,0011 ***
Norte de Santander	0.0285	0.167	0,028	0,17	0,029	0,17	-0,0004
Putumayo	0.00699	0.0833	0,007	0,08	0,007	0,08	-0,0003 *
Quindío	0.0129	0.113	0,013	0,11	0,013	0,11	-0,0003
Risaralda	0.0193	0.138	0,019	0,14	0,019	0,14	-0,0001
San Andrés	0.00121	0.0347	0,001	0,03	0,001	0,04	0,0000
Santander	0.0488	0.215	0,049	0,22	0,049	0,22	0,0004
Sucre	0.0203	0.141	0,021	0,14	0,020	0,14	0,0007 *
Tolima	0.0317	0.175	0,032	0,18	0,032	0,18	0,0002
Valle del Cauca	0.0751	0.264	0,072	0,26	0,078	0,27	-0,0059 ***

Tabla 1. Estadísticas descriptivas por género (continuación)

<i>Variables</i>	Total		Niños		Niñas		Diferencia
	Media o proporción	DE	Media o proporción	DE	Media o proporción	DE	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Vaupés	0.000482	0.0220	0.001	0.02	0.000	0.02	0.0002 ***
Vichada	0.000794	0.0282	0.001	0.03	0.001	0.03	0.0002 ***
Observaciones	504,085		229,400		274,685		

DE= desviación estándar.

***, ** y * representan significancia estadística al 99, 95 y 90% respectivamente.

Tabla 2. Proporción de niñas y niños por quintiles de las distribuciones

Cuantil	Distribuciones			
	Puntaje global	Matemáticas	Ciencias	Lectura
1er quintil	1,276	1,497	1,522	1,021
2do quintil	1,395	1,434	1,413	1,219
3er quintil	1,312	1,286	1,278	1,261
4to quintil	1,163	1,077	1,074	1,268
5to quintil	0,908	0,795	0,793	1,246
Cuantil 95	0,749	0,600	0,629	1,212

La proporción es el cociente del número de niñas entre el número de niños en el quintil/cuantil

Tabla 3. Estimación MCO y Regresión Cuantílica: Puntaje global, Matemáticas, Ciencias y Lectura

<i>Variables</i>	(1)	(2)	(3)	(4)	(5)	(6)
	Puntaje global					
	MCO	Cuantil 20	Cuantil 40	Cuantil 60	Cuantil 80	Cuantil 95
Genero (1=niña ; niño=0)	-6.250*** (0.100)	-2.686*** (0.123)	-4.904*** (0.122)	-7.145*** (0.128)	-9.719*** (0.151)	-12.800*** (0.228)
Constante	445.994*** (8.999)	393.792*** (10.518)	453.447*** (10.684)	497.673*** (11.377)	548.394*** (13.523)	621.616*** (20.975)
Observaciones	504,085	504,085	504,085	504,085	504,085	504,085
R-cuadrado/R-Cuadrado ajustado	0.332	0.150	0.174	0.193	0.213	0.239
	Matemáticas					
	MCO	Cuantil 20	Cuantil 40	Cuantil 60	Cuantil 80	Cuantil 95
Genero (1=niña ; niño=0)	-2.356*** (0.024)	-1.562*** (0.031)	-1.972*** (0.029)	-2.389*** (0.029)	-3.010*** (0.035)	-4.013*** (0.058)
Constante	97.806*** (2.208)	73.930*** (2.880)	91.036*** (2.613)	106.105*** (2.573)	126.387*** (3.111)	166.033*** (4.536)
Observaciones	504,085	504,085	504,085	504,085	504,085	504,085
R-cuadrado/R-Cuadrado ajustado	0.256	0.102	0.119	0.137	0.165	0.211
	Ciencias					
	MCO	Cuantil 20	Cuantil 40	Cuantil 60	Cuantil 80	Cuantil 95
Genero (1=niña ; niño=0)	-2.413*** (0.024)	-1.686*** (0.034)	-2.138*** (0.029)	-2.555*** (0.030)	-3.105*** (0.035)	-3.707*** (0.051)
Constante	106.789*** (2.206)	90.177*** (2.971)	103.604*** (2.605)	120.145*** (2.704)	134.848*** (3.069)	146.837*** (4.542)
Observaciones	504,085	504,085	504,085	504,085	504,085	504,085
R-cuadrado/R-Cuadrado ajustado	0.266	0.122	0.138	0.156	0.178	0.217
	Lectura					
	MCO	Cuantil 20	Cuantil 40	Cuantil 60	Cuantil 80	Cuantil 95
Genero (1=niña ; niño=0)	0.565*** (0.024)	0.907*** (0.032)	0.652*** (0.030)	0.454*** (0.029)	0.263*** (0.035)	-0.127** (0.054)
Constante	72.421*** (2.208)	63.309*** (2.768)	71.102*** (2.661)	75.468*** (2.709)	78.907*** (3.221)	99.318*** (4.893)
Observaciones	504,085	504,085	504,085	504,085	504,085	504,085
R-cuadrado/R-Cuadrado ajustado	0.265	0.136	0.146	0.152	0.152	0.153

Errores estándar robustos en paréntesis.

***, ** y * representan significancia estadística al 99, 95 y 90% respectivamente.

En todos los casos se incluyen controles de características individuales, familiares y escolares

Adicionalmente, se incluye treinta y dos dummies para departamentos y una para Bogotá D.C. están incluidas

Tabla 4. Descomposición de Juhn-Murphy-Pierce

Puntaje global				
	DT	Q	P	U
Media	5.829	-0.418	6.246	0.001
Cuantil 5	-0.384	-0.724	5.004	-4.665
Cuantil 10	0.000	-1.209	5.353	-4.143
Cuantil 25	1.923	-1.445	5.985	-2.616
Cuantil 50	5.769	-0.508	6.513	-0.236
Cuantil 75	8.846	-0.158	6.817	2.187
Cuantil 90	11.538	0.538	6.644	4.354
Cuantil 95	13.076	0.950	6.494	5.631
Matemáticas				
Media	2.214	-0.581	2.214	0.581
Cuantil 5	0.000	-0.67	1.918	-1.247
Cuantil 10	2.000	0.193	2.275	-0.468
Cuantil 25	1.000	-0.289	2.166	-0.877
Cuantil 50	2.000	-0.287	2.353	-0.065
Cuantil 75	3.000	0.001	2.492	0.506
Cuantil 90	4.000	-0.002	2.58	1.421
Cuantil 95	4.000	-0.204	2.533	1.67
Ciencias				
Media	2.257	-0.528	2.257	0.528
Cuantil 5	1.000	-0.32	2.328	-1.008
Cuantil 10	1.000	-0.443	2.350	-0.908
Cuantil 25	2.000	0.093	2.416	-0.509
Cuantil 50	2.000	-0.334	2.431	-0.097
Cuantil 75	3.000	-0.404	2.527	0.876
Cuantil 90	3.000	-0.445	2.411	1.034
Cuantil 95	4.000	0.259	2.382	1.358
Lectura				
Media	-0.675	-0.673	-0.675	0.673
Cuantil 5	-1.000	0.550	-0.677	-0.874
Cuantil 10	0.000	0.044	0.077	-0.121
Cuantil 25	-1.000	-0.411	-0.437	-0.152
Cuantil 50	-1.000	-0.188	-0.710	-0.102
Cuantil 75	0.000	0.197	-0.578	0.381
Cuantil 90	0.000	0.188	-0.667	0.479
Cuantil 95	0.000	0.243	-0.857	0.615

DT: Diferencias en el puntaje (niños-niñas)

Q: Contribución de las diferencias en características observables

P: Contribución de las diferencias en los retornos

U: Diferencias en cantidades y precios no observables

Grupo de referencia: Niños