



Revista Electrónica Educare  
ISSN: 1409-4258  
Universidad Nacional. CIDE

Montero-Rojas, Eiliana; Moreira-Mora, Tania Elena;  
Zamora-Araya, José Andrey; Smith-Castro, Vanessa  
Una nueva mirada teórica y metodológica a diferencias de género en pruebas  
de matemática: Razonamiento, actitudes psicosociales y modelos multinivel  
Revista Electrónica Educare, vol. 25, núm. 1, 2021, Enero-Abril, pp. 143-163  
Universidad Nacional. CIDE

DOI: 10.15359/ree.25-1.8

Disponible en: <http://www.redalyc.org/articulo.oa?id=194165972008>

- Cómo citar el artículo
- Número completo
- Más información del artículo
- Página de la revista en [redalyc.org](http://redalyc.org)

[Cierre de edición el 01 de Enero del 2021]

<http://doi.org/10.15359/ree.25-1.8><http://www.una.ac.cr/educare>

educare@una.ac.cr

# Una nueva mirada teórica y metodológica a diferencias de género en pruebas de matemática: Razonamiento, actitudes psicosociales y modelos multinivel

***A New Theoretical and Methodological Look at Gender Differences in Mathematics Tests: Reasoning, Psychosocial Attitudes, and Multilevel Models***

***Um novo olhar teórico e metodológico às diferenças de gênero em provas de matemática: Razoamento, atitudes psicossociais e modelos multinível***

Eiliana Montero-Rojas

Universidad de Costa Rica

San José, Costa Rica

eiliana.montero@ucr.ac.cr



<https://orcid.org/0000-0002-6770-792X>

Tania Elena Moreira-Mora

Instituto Tecnológico de Costa Rica

Cartago, Costa Rica

tmoreira@itcr.ac.cr



<https://orcid.org/0000-0002-8955-0804>

José Andrey Zamora-Araya

Universidad Nacional

ROR: <https://ror.org/01t466c14>

Heredia, Costa Rica

jzamo@una.ac.cr



<https://orcid.org/0001-6050-5850>

Vanessa Smith-Castro

Universidad de Costa Rica

San José, Costa Rica

VANESSA.SMITH@ucr.ac.cr



<https://orcid.org/0001-6348-4223>

Recibido • Received • Recebido: 03 / 05 / 2019

Corregido • Revised • Revisado: 10 / 09 / 2020

Aceptado • Accepted • Aprovado: 26 / 11 / 2020

**Resumen:** Reconociendo que el género, como constructo, es expresión de una estructura de poder, se buscó explicar diferencias observadas a favor de los varones en dos pruebas estandarizadas de contexto matemático: componente matemático de la Prueba de Admisión de la Universidad de Costa Rica y la Prueba de Matemática del Bachillerato de la Educación Formal. La muestra fue de 487 estudiantes que en 2015 cursaban el undécimo año en 10 secundarias públicas diurnas, de la gran área metropolitana de Costa Rica, 269 eran mujeres. Se estimaron modelos de regresión multinivel con el puntaje en cada

prueba como variable de respuesta. El modelo reducido solo incluyó como predictor el sexo de cada estudiante, y el modelo completo incluyó, además, un indicador de habilidades de razonamiento y los puntajes en las siguientes escalas: sexismo hostil y benevolente, actitud hacia la equidad de género en matemática y autoeficacia matemática. Los resultados apuntan a que las diferencias de género se reducen considerablemente al controlar habilidades de razonamiento. El sexismo benévolos predice el desempeño en las pruebas, aunque en menor medida. También se generó evidencia del efecto moderador del colegio para la relación entre el sexo y el puntaje en la prueba de admisión, y, especialmente, para los promedios de la prueba de bachillerato. La autoeficacia matemática también resultó de relevancia predictiva en esta última. Estos resultados evidencian la necesidad de que el personal docente promueva activamente esta dimensión de autoeficacia en sus estudiantes, especialmente en las mujeres. Las considerables diferencias en los promedios de las pruebas de Bachillerato, según colegio, aun siendo todos públicos y de la gran área metropolitana, sugieren la presencia de inequidades que deberían estudiarse con mayor profundidad, incluidos centros educativos de las diversas regiones del país, los cuales podrían presentar todavía mayores desigualdades.

**Palabras claves:** Diferencias de género; estereotipos de género; habilidades de razonamiento; modelos multinivel; pruebas de matemática; sexismo ambivalente.

**Abstract:** Acknowledging that gender, as a construct, is an expression of a power structure, observed differences favoring males were explained for two standardized tests of mathematical context: the Mathematics section of the University of Costa Rica's Admission Test and the Mathematics test from the Secondary School's Exit Exam. The sample was 487 students in the last year of high school in ten public schools from the Greater Metropolitan Area of Costa Rica; 269 were women. Multilevel regression models were estimated using the scores in these tests as dependent variables. The reduced model only included sex as a predictor, whereas the complete model included, additionally, an indicator of reasoning abilities and the following scales: hostile sexism, benevolent sexism, attitude toward gender equity in mathematics, and self-efficacy in mathematics. Results point out that gender differences are reduced by controlling for reasoning abilities. Benevolent sexism predicted performance in the tests to a lesser degree. There was also evidence of the moderating effect of the school for the relationship between sex and the score in the admission test, and, especially, for the means in the secondary school exit test. Mathematics self-efficacy also had considerable predictive power in the latter. These findings provide evidence of the need for teachers to actively promote self-efficacy in their students, especially women. The considerable differences in the average scores of the Secondary School's Exit Exam, even though they are all public and from the Greater Metropolitan Area of Costa Rica, suggests the presence of inequities that should be studied more in-depth, including schools beyond this area, which could reveal even more disadvantages.

**Keywords:** Ambivalent sexism; gender differences; gender stereotypes; Math performance; multilevel models; reasoning abilities.

**Resumo:** Dado que o gênero, como construto, é uma expressão de uma estrutura de poder, este trabalho buscou explicar diferenças observadas a favor dos homens nas duas provas padronizadas do contexto matemático: a seção matemática na prova de admissão da Universidade da Costa Rica e a prova de matemática do ensino médio da educação formal. A amostra foi de 487 estudantes da Grande Área Metropolitana da Costa Rica que cursaram o último ano do ensino médio em colégios públicos diurnos no ano de 2015. Do total, 269 são mulheres. Foram ajustados modelos de regressão multinível com a pontuação



dessas duas provas como variável resposta. O modelo reduzido somente incluiu como preditor o sexo do estudante. Já o modelo completo, inseriu ainda um indicador de habilidade de raciocínio e as seguintes escalas: sexismo hostil e benevolente, atitude para a equidade de gênero em matemática e autoeficácia matemática. Os resultados apontam que as diferenças de gênero são reduzidas ao controlar habilidades de raciocínio. O sexismo benévolos prediz a performance nas provas, embora haja pouco impacto. Também se gerou evidência do efeito moderador do colégio para a relação entre o sexo e a pontuação nas provas de admissão, especialmente para as médias das provas do ensino médio. A autoeficácia matemática também resultou ser de considerável relevância preditiva nessa última prova. Esses resultados mostram a necessidade dos professores promoverem ativamente essa dimensão da autoeficácia em seus alunos, principalmente nas mulheres. As diferenças consideráveis nas médias das provas de Bacharelado, segundo as escolas, mesmo sendo todas públicas e da grande região metropolitana, sugerem a presença de iniquidades que devem ser estudadas com maior profundidade, inclusive nos centros de ensino das diversas regiões do país, que podem apresentar desigualdades ainda maiores.

**Palabras-chave:** Diferencias de género; estereotipo de género; habilidades de raciocínio; modelos multinivel; provas de matemática; sexism ambivalente.

## Introducción

A nivel mundial, la investigación de los últimos 30 años ha evidenciado claras diferencias a favor de los varones en pruebas estandarizadas de contexto matemático ([Hedges y Nowell, 1995](#); [D. Voyer y S. D. Voyer, 2014](#)). En el caso de Costa Rica, los estudios de PISA ([Organisation for Economic Co-operation and Development \[OECD\], 2016](#)) han revelado que es uno de los países evaluados donde las diferencias resultan más grandes, con el puntaje promedio de los chicos superando al de las chicas por más de 15 puntos. Estas diferencias también han sido evidentes, históricamente, en la parte de contexto matemático de la Prueba de Admisión de la Universidad de Costa Rica (UCR) ([Montero y Villalobos, 2004](#)), y, en la Prueba de Matemática del Bachillerato de la Educación Formal ([Esquivel-Alfaro et al., 2006](#)).

Así, la meta del presente estudio es estimar en qué medida las diferencias en el desempeño de hombres y mujeres en pruebas estandarizadas de contexto matemático pueden ser atribuidas no solo a las diferencias individuales, sino también a la internalización de creencias sociales y estereotipos culturales en torno al género, y al grado de confianza en las propias destrezas matemáticas, así como a las habilidades generales de razonamiento del estudiantado. Se exploran, a la vez, los efectos moderadores de las instituciones educativas en que estudian los sujetos participantes. Se busca explicar las diferencias de género en el desempeño de las dos últimas pruebas mencionadas en el párrafo anterior, utilizando una muestra de jóvenes de 10 colegios públicos académicos diurnos de la gran área metropolitana de Costa Rica (GAM). El abordaje se realiza con una mirada nueva, considerando varias dimensiones y técnicas de control poco aplicadas en nuestro contexto. Así, el estudio pretende analizar la contribución relativa de variables psicosociales sobre el desempeño de hombres y mujeres en dos pruebas estandarizadas de contexto matemático, controlando o aislando, el efecto de habilidades básicas de razonamiento y estimando el potencial efecto moderador de los colegios.

Una de las principales novedades del estudio radica, precisamente, en la inclusión, como variable control, de las puntuaciones en una prueba de habilidades básicas de razonamiento. Esto responde a una de las críticas a los estudios sobre el efecto de las creencias y los estereotipos en el desempeño, la cual se refiere a que los estudios derivados de estas líneas de investigación (en su mayoría de carácter experimental), tienden a sobredimensionar su efecto, transmitiendo la idea de que las diferencias en el desempeño pueden desaparecer *removiendo* los estereotipos (Sackett et al., 2004). Otra novedad de esta investigación radica en el uso de modelos multinivel o modelos mixtos lineales para el análisis de los datos (Goldstein, 1987; Hox, 2010). La utilización de los modelos multinivel permite estimar los efectos específicos del sexo, las habilidades generales de razonamiento y las variables psicosociales sobre el desempeño, a la vez permite explorar el papel de los centros educativos como moderadores de la relación entre sexo y desempeño, es decir, las potenciales diferencias que produce, en las relaciones estimadas, el contexto específico de cada centro educativo.

## Referentes conceptuales

Al identificar los factores que explican las brechas de género en el desempeño académico, la investigación ha podido revelar una compleja interacción de variables evolutivas, cognitivas, ambientales, económicas y socioculturales que impactan en las motivaciones, expectativas y las decisiones vocacionales de hombres mujeres (Hyde y Mertz, 2009). En general, existe un importante consenso en reconocer que el problema es muy complejo, y que requiere de un abordaje interdisciplinario, pero también existe un importante consenso sobre el potencial impacto de las actitudes, las creencias sociales y los estereotipos de género en estas brechas (Nosek et al., 2009).

Desde la teoría del sexismo ambivalente de Glick y Fiske (2011) se postula que el sexismo es una ideología (en tanto visión de mundo y falsa conciencia), producto del sistema patriarcal, que enfatiza la diferenciación de género, la cual, a su vez, refuerza visiones estereotipadas de hombres y mujeres que pueden impactar su desempeño en contextos académicos. Se distinguen dos tipos de sexismo que afectan la forma en que hombres y mujeres son percibidos y se perciben a sí mismos o a sí mismas: el sexismo hostil, que tiene sus raíces en la supuesta diferenciación rivalizadora de género ("los hombres son más competentes que las mujeres"); y el sexismo benévolos que tiene sus raíces en la percepción estereotipada de complementariedad de género ("las mujeres son puras y cálidas, y se combinan perfectamente con los hombres, que son más capaces y competentes y que deben protegerlas"). Así, la internalización de las ideologías sexistas pueden llevar a las personas a percibir diferencias magnificadas o exageradas entre los sexos (Hyde y Mertz, 2009), lo que, a su vez, puede afectar las autopercepciones y las motivaciones.

Por su parte, la teoría de la amenaza del estereotipo de Steele y Aronson (1995), apunta a que, en los ambientes educativos, las ideologías y los estereotipos pueden afectar de manera negativa el desempeño estudiantil. En el caso particular de los estereotipos de género, la teoría



postula que las mujeres, cuando son expuestas a estereotipos negativos sobre sus habilidades y destrezas en ámbitos como la matemática, la ciencia o la inteligencia, pueden ser susceptibles a desempeñarse peor en pruebas de estas áreas (Cadinu et al., 2005).

Por otro lado, el modelo de expectativas y valor de Eccles y colaboradores (Jacobs y Eccles, 1992), postula que los estereotipos y normas culturales determinan la forma en que las personas se perciben a sí mismas y sus comportamientos a través de dos variables centrales: las expectativas de éxito, es decir, la probabilidad percibida de tener éxito en una tarea particular, y el valor subjetivo que se le atribuye a dicha tarea, que se refiere al grado en que una tarea proporciona interés intrínseco y es percibido como útil e importante por el individuo. Se postula, además, que los estereotipos culturales pueden afectar la autoeficacia de las mujeres en matemática a través de las actitudes mantenidas por importantes agentes de socialización como pares, padres y madres, y profesores o profesoras.

Los datos empíricos apoyan estos principios generales. Por ejemplo, investigaciones recientes muestran que aquellas personas que tienden a respaldar ideologías sexistas, tienden a su vez a exagerar la magnitud de las supuestas diferencias de género en varios dominios, incluidos los exámenes de matemática, las habilidades académicas y los intereses científicos (Zell et al., 2016). Los datos también han demostrado que el rendimiento de mujeres se ve disminuido en comparación con el de los hombres cuando se les expone a creencias negativas sobre las habilidades matemáticas de las mujeres (Nguyen y Ryan, 2008). Finalmente, la evidencia muestra que las niñas, cuyas habilidades en matemática son repetidamente cuestionadas, desarrollan un autoconcepto matemático más bajo, una menor confianza en su aptitud matemática y se muestran menos motivadas a estudiar matemática que los niños. Además, se ha observado que la confianza matemática es uno de los principales factores en su decisión de no persistir en materias como cálculo y, por lo tanto, en no especializarse en asignaturas con alta demanda de dominio matemático (Ellis et al., 2016).

Todas estas teorías ponen de relieve los efectos perjudiciales de la estructura social imperante, producto de un sistema patriarcal, desarrollado a través de miles de años, que ubica a las mujeres en situación de desventaja, generando impactos en todas las áreas vitales, incluyendo la educativa. Como se comentó en el párrafo anterior, un ejemplo de ello se evidencia, precisamente, en las elecciones de carrera, lo cual genera procesos de exclusión de las mujeres en diversos campos de conocimiento, tal es el caso de las carreras denominadas STEM por su acrónimo en inglés (Ciencia, Tecnología, Ingeniería y Matemática) (Ellis et al., 2016).

### **Referentes metodológicos: Los modelos multinivel**

Según diversos marcos de referencia, la conducta humana es el resultado de la influencia simultánea de factores individuales y contextuales. En estas situaciones de investigación, típicas de las ciencias sociales, es donde los modelos multinivel han hecho sus principales

contribuciones a lo largo de los últimos 30 años. Estos modelos representan una poderosa alternativa, comparados con los modelos clásicos de regresión múltiple, tradicionalmente utilizados, y, con frecuencia, de manera errónea, como herramienta para analizar estudios de factores asociados con diseños observacionales, donde variables de comportamiento o actitudes pueden ser explicadas a partir de variables individuales del sujeto, así como de variables de su entorno o ambiente (Goldstein, 1987; Hox, 2010; Raudenbush y Bryk 2002).

Cuando se emplea un modelo de regresión multinivel (o modelos mixtos, como se les llama hoy en día en estadística), se puede, de manera simultánea, analizar los efectos o relaciones, sobre la variable dependiente, de variables predictoras en diferentes niveles de agregación, como estudiantes en nivel 1 y centros educativos en nivel 2. Además, la inclusión de los llamados *efectos aleatorios* permite valorar en qué medida las relaciones estimadas entre la variable dependiente y las variables predictoras medidas en el nivel individual pueden variar de entorno a entorno (Hox, 2010; Raudenbush y Bryk, 2002).

En la actualidad, estos modelos son utilizados en una gran cantidad de aplicaciones en las ciencias sociales, por ejemplo, los programas de pruebas internacionales PISA y TIMSS los emplean regularmente para sus estudios (OECD, 2010).

Al igual que en un modelo clásico de regresión, un modelo multinivel o mixto se concibe como una ecuación que define la variable de criterio o dependiente como una combinación lineal de un conjunto de variables independientes o predictoras. Pero, a diferencia del modelo de regresión, en el modelo multinivel estas variables predictoras se miden en diferentes niveles de agregación. Para ilustrar, tómese la ecuación básica en un modelo de 2 niveles, con una variable predictora  $X_1$ , medida en el nivel 1 (individual) y otra  $X_2$ , medida en el nivel 2 (conglomerado), es la siguiente:

Ecuación 1:

$$Y_{ij} = b_{0j} + b_{1j}X_{1ij} + b_{2j}X_{2j} + e_{ij} = b_0 + u_{0j} + (b_1 + u_{1j})X_{1ij} + b_2X_{2j} + e_{ij} \quad (1)$$

Donde  $i$  denota las unidades de primer nivel (por ejemplo, estudiantes) y  $j$  denota las unidades de segundo nivel (por ejemplo, escuelas). Los  $u_{0j}$  y  $u_{1j}$  son llamados los efectos aleatorios. En este caso indican cuánto aumenta o disminuye el coeficiente fijo ( $b_0$  o  $b_1$ ) según el conglomerado  $j$ . Ilustrativamente, en un estudio de factores asociados al rendimiento escolar,  $X_1$  podría ser el índice socioeconómico del estudiante, y  $X_2$  una escala de liderazgo académico de la persona directora. Nótese que en la ecuación anterior existen dos efectos aleatorios: uno para el intercepto y otro para el coeficiente asociado a  $X_1$ , puesto que varían según las unidades  $j$  de segundo nivel (conglomerado). Lo que esto nos dice es que las relaciones estimadas van a cambiar según el conglomerado. En el caso del intercepto y del coeficiente de  $X_1$ , su magnitud será más grande o más pequeña dependiendo de la magnitud específica que toman los efectos aleatorios,  $u_{0j}$  y  $u_{1j}$  respectivamente. Entonces, en el modelo de efectos aleatorios se deben



estimar las medias que componen la parte fija del modelo, o sea los  $b_i$ , intercepto y pendientes, las variancias de los  $b_i$  según los conglomerados  $j$  y las covariancias entre las distribuciones de estos (interceptos y pendientes). La parte fija involucra las estimaciones puntuales de los  $b_i$  y la parte aleatoria la estimación de las variancias y covariancias entre los efectos aleatorios. En el caso de la ecuación (1) la matriz  $V$  de efectos aleatorios tiene tres parámetros a estimar,  $\sigma_0^2$   $\sigma_1^2$  y  $\sigma_{10}$ . Donde  $\sigma_0^2$  es la variancia poblacional de la distribución de los interceptos a lo largo de las unidades de segundo orden (conglomerados, por ejemplo, escuelas),  $\sigma_{10}$  es la covariancia entre las distribuciones de ambos efectos aleatorios,  $b_{0j}$  y  $b_{1j}$ , y  $\sigma_1^2$  es la variancia poblacional de la distribución de las pendientes  $b_{1j}$  asociadas a la variable  $X_1$ .

## Método

### Participantes

Se trabajó con datos de 487 estudiantes que en 2015 cursaban el undécimo año en 10 secundarias públicas diurnas seleccionadas al azar del conjunto de centros educativos con esas características en la GAM de Costa Rica. El promedio de edad de la muestra fue de 16.68 años (DE = .81 años), y el porcentaje de mujeres fue de 55.2 %.

### Instrumentos

**Ideologías sexistas hostiles y benevolentes:** Para medir estos constructos se utilizó una adaptación latinoamericana de la Escala de sexismo ambivalente (ASI) ([Cárdenas et al., 2010](#)). La escala original fue desarrollada por [Glick y Fiske \(1996\)](#). La ASI es una medida directa del sexismo hostil (SH) y del sexismo benevolente (SB) que utiliza 22 ítems puntuados en una escala Likert de 5 categorías, desde *totalmente en desacuerdo* a *totalmente de acuerdo*. Once de esos ítems corresponden a cada tipo de sexismo. Ejemplos de los reactivos son *Muchas mujeres interpretan comentarios y acciones inocentes como sexistas* (SH) y *Las mujeres, en comparación con los hombres, tienden a tener una mayor sensibilidad moral* (SB). El indicador utilizado para el análisis fue el promedio obtenido por el participante en los ítems correspondientes a cada tipo de sexismo. El rango para estos indicadores varía entre 1 y 5, y los puntajes elevados implican mayor sexismo. La relación esperada a nivel empírico es inversa (a mayor sexismo, menor desempeño). Los nombres de estas variables para el análisis fueron HOSTIL y BENE.

**Equidad de género en contextos matemáticos:** Para medir la ausencia de estereotipos de género (equidad) en contextos matemáticos se utilizó una de las subescalas de la versión abreviada la *Escala de actitudes hacia las matemáticas* de Fennema y Sherman ([Doepken et al., 2003](#)). Se seleccionaron cuatro reactivos, como ejemplo se presenta el siguiente ítem: *Para resolver problemas matemáticos importantes, confiaría en una mujer tanto como confiaría en un hombre*. La escala general, junto con los términos de uso y las instrucciones de calificación, se pueden encontrar en [Doepken et al. \(2003\)](#). La medida utilizada para el análisis fue el promedio obtenido por el participante en los cuatro ítems, su rango va de 1 a 5, y puntajes altos implican



mayor percepción de equidad. La relación esperada a nivel empírico es directa (a mayor equidad, mayor desempeño) y el nombre de la variable correspondiente en el análisis es EQUIDAD.

Autoeficacia matemática: Como indicador de este constructo se utilizó la subescala de Confianza personal para la matemática, tomada, también, de la versión abreviada de la Escala de actitudes hacia la matemática de Fennema y Sherman ([Doepken et al., 2003](#)). La escala consta de 11 reactivos que se responden en una escala Likert de 5 puntos desde *totalmente en desacuerdo* a *totalmente de acuerdo*. Los siguientes son dos ejemplos de reactivos: *Estoy seguro/a de que puedo aprender matemática* y *Creo que podría resolver problemas matemáticos complejos*. Se utilizó en el análisis el promedio obtenido por el participante en la escala, valores altos implican mayor autoeficacia matemática. Desde la teoría se espera una relación directa entre este constructo y el desempeño en las pruebas de contexto matemático (a mayor autoeficacia matemática, mayor desempeño). El nombre de la variable correspondiente en el análisis es AUTOEFLCA.

Habilidades generales de razonamiento: Con el fin de controlar el efecto de las diferencias individuales en habilidades generales de razonamiento sobre el desempeño en las pruebas, se utilizó la Prueba de razonamiento con figuras (PRF), desarrollada por el Instituto de Investigaciones Psicológicas de la UCR. La PRF está fundamentada sobre el concepto de *inteligencia fluida* (Gf), definido junto con el concepto de *inteligencia cristalizada* (Gc) en la Teoría de la inversión (Investment Theory) de [Cattell \(1963\)](#). La prueba incluye cuatro componentes: series, clasificación, matrices y condiciones. Se evalúan diversas operaciones cognitivas como identificación, seriación, clasificación, y comparaciones ([Montero-Rojas et al., 2013](#)). Cada subtest cuenta con un tiempo limitado para su resolución. Su duración total no excede los 20 minutos ([Villarreal Galera, 2011](#)). Para el presente estudio se utilizaron solamente las subescalas de series y matrices, por poseer los mejores indicadores de confiabilidad psicométrica. El indicador utilizado en el análisis se llamó PRF y es el porcentaje de respuestas correctas en esta prueba. Se espera una relación directa con las pruebas de contexto matemático (a mayor puntaje en la PRF mayor puntaje en las pruebas).

Desempeño en pruebas estandarizadas de contexto matemático: Con este fin se utilizaron dos indicadores: el porcentaje de respuestas correctas en la sección de razonamiento en contextos matemáticos de la Prueba de aptitud académica (PAA) de la Universidad de Costa Rica y el porcentaje de respuestas correctas obtenido en la Prueba de matemática del Bachillerato de la Educación Formal del Ministerio de Educación de Costa Rica. Los nombres de las variables correspondientes usadas en el análisis fueron MATE\_UCR y MATE\_BACH, respectivamente. La PAA aplicada en esa oportunidad pretendía medir habilidades de razonamiento en contextos verbales y matemáticos, habilidades que se supone son necesarias para el aprendizaje universitario. Las dos escalas se elaboraban por separado. Los ítems eran de escogencia única y con una duración máxima de tres horas para resolver 85 ítems, 50 de ellos de contexto verbal y 35 de contexto matemático. De ellos, se calificaron 30 ítems de contexto matemático y 40 de contexto verbal ([Smith-Castro, 2014](#)). Por su parte, la Prueba de matemática del Bachillerato de la Educación



Formal evalúa la competencia matemática entendida como la habilidad para usar la Matemática en diversas situaciones donde se hace necesario explicar y predecir. Los contenidos abarcan las áreas de geometría, álgebra, estadística y probabilidades (Mena Castillo, 2015).

## Procedimiento

Después de obtener los respectivos permisos institucionales, se estableció contacto con estudiantes en sus clases y se les invitó a participar voluntariamente en el estudio. Los grupos de estudiantes, de manera autoadministrada, completaron, en el primer semestre del 2015, el cuestionario que contenía las medidas de actitudes y creencias (sexismos, estereotipos de género en matemática, autoeficacia matemática). Asimismo, completaron la medida de razonamiento general (PRF). Como parte del consentimiento informado, al estudiantado se le solicitó permiso para acceder a sus calificaciones en la parte matemática de la Prueba de Aptitud Académica Universidad de Costa Rica, y a sus calificaciones en la Prueba de Matemática del Bachillerato de la Educación Formal. Estas dos pruebas se aplicaron en ese mismo año, entre setiembre y noviembre.

## Análisis de datos

Para la estimación de los modelos multinivel se utilizó el paquete estadístico Stata 14, con el método de máxima verosimilitud. De acuerdo con los referentes conceptuales del estudio, se ajustaron inicialmente dos modelos para estimar la relación entre el sexo del estudiante y su desempeño en las dos medidas criterio (PAA-Matemática y Bachillerato-Matemática): el modelo reducido de efectos aleatorios consideró solamente la corrección por la estructura de conglomerados y el coeficiente aleatorio correspondiente al sexo como variable predictora. El modelo completo de efectos aleatorios, por su parte, incluyó, además de lo anterior, el puntaje en la Prueba de razonamiento con figuras, las puntuaciones en las dos subescalas de sexismos, las puntuaciones en la escala de estereotipos de género en contextos matemáticos y las puntuaciones en la escala de autoeficacia matemática como variables predictoras para el desempeño en cada una de las dos pruebas.

## Resultados

### Estadísticas descriptivas

En la [Tabla 1](#) se presentan las estadísticas descriptivas de las variables en estudio. Las [Figuras 1, 2 y 3](#) muestran los resultados promedio de las dos variables dependientes y de la variable de control, puntaje en la prueba de razonamiento PRF, por sexo, según colegio.



**Tabla 1:** Estadísticas descriptivas y correlaciones entre las variables (n = 487)

Variables	Total M (SD)	Correlaciones						
		1	2	3	4	5	6	7
MATE_UCR (1)	39.94 (17.59)	-						
MATE_BACH (2)	64.09 (13.89)	.346**	-					
PRF (3)	66.24 (20.35)	.246**	.403**	-				
BENE (4)	3.03 (0.72)	-.084	-.138**	-.059	-			
HOSTIL (5)	2.87 (0.75)	.033	-.039	.101*	.412**	-		
EQUIDAD (6)	4.37 (0.77)	.067	.150**	.181**	-.066	-.087	-	
AUTOEFICA (7)	3.87 (0.84)	.149**	.367**	.249**	.041	.097*	.172**	-
Hombres								
MATE_UCR (1)	42.68 (18.64)	-						
MATE_BACH (2)	66.50 (14.38)	.304**	-					
PRF (3)	71.45 (18.44)	.181**	.488**	-				
BENE (4)	3.22 (0.70)	-.104	-.162*	-.102	-			
HOSTIL (5)	3.20 (0.67)	-.046	-.083	.020	.266**	-		
EQUIDAD (6)	4.39 (0.72)	-.084	.061	.021	-.049	-.083	-	
AUTOEFICA (7)	4.09 (0.77)	.076	.418**	.258**	.004	.079	.153*	-
Mujeres								
MATE_UCR (1)	37.58 (16.31)	-						
MATE_BACH (2)	62.02 (13.14)	.358**	-					
PRF (3)	61.76 (20.86)	.257**	.296**	-				
BENE (4)	2.87 (0.70)	-.145*	-.206**	-.140*	-			
HOSTIL (5)	2.58 (0.68)	-.016	-.151*	-.011	.426**	-		
EQUIDAD (6)	4.35 (0.82)	.193**	.221**	.286**	-.095	-.126*	-	
AUTOEFICA (7)	3.68 (0.85)	.160**	.281**	.166**	-.035	-.065	.182**	-

\*\* La correlación es estadísticamente significativa en el nivel 0.01 (bilateral).

\* La correlación es estadísticamente significativa en el nivel 0.05 (bilateral).

(1) MATE\_UCR: Porcentaje de correctas en la parte matemática de la PAA de la UCR

(2) MATE\_BACH: Porcentaje de correctas en la Prueba de Matemática de Bachillerato

(3) PRF: Porcentaje de correctas en la Prueba de razonamiento con figuras

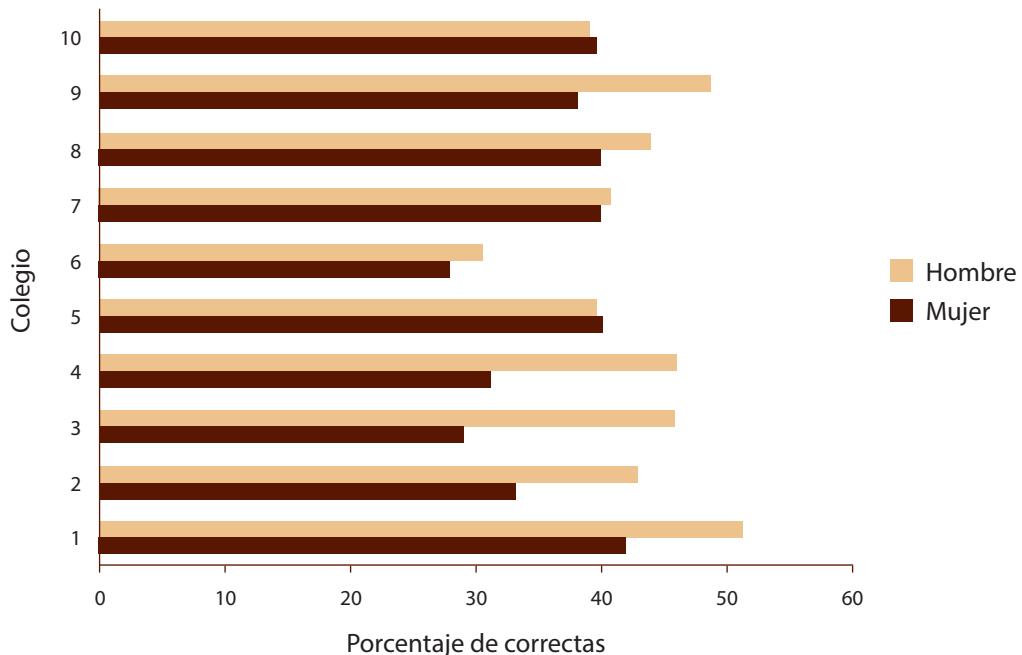
(4) BENE: Puntaje promedio del participante en la escala de Sexismo Benevolente

(5) HOSTIL: Puntaje promedio del participante en la escala de sexismo hostil

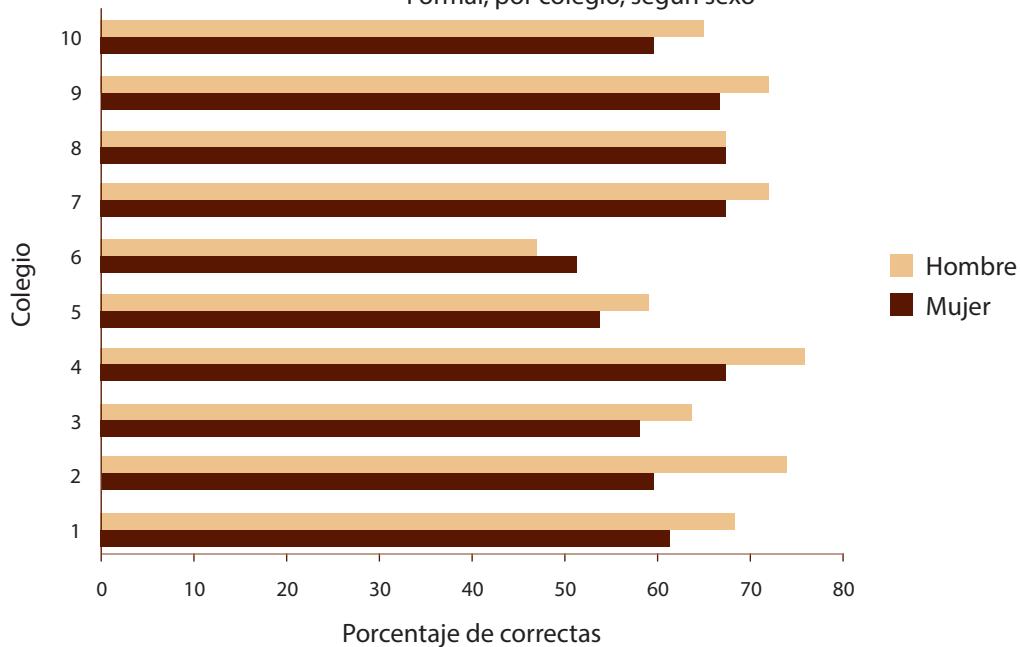
(6) EQUIDAD: Puntaje promedio del participante en la escala de equidad matemática

(7) AUTOEFICA: Puntaje promedio del participante en la escala de autoeficacia matemática

**Nota:** Elaboración propia

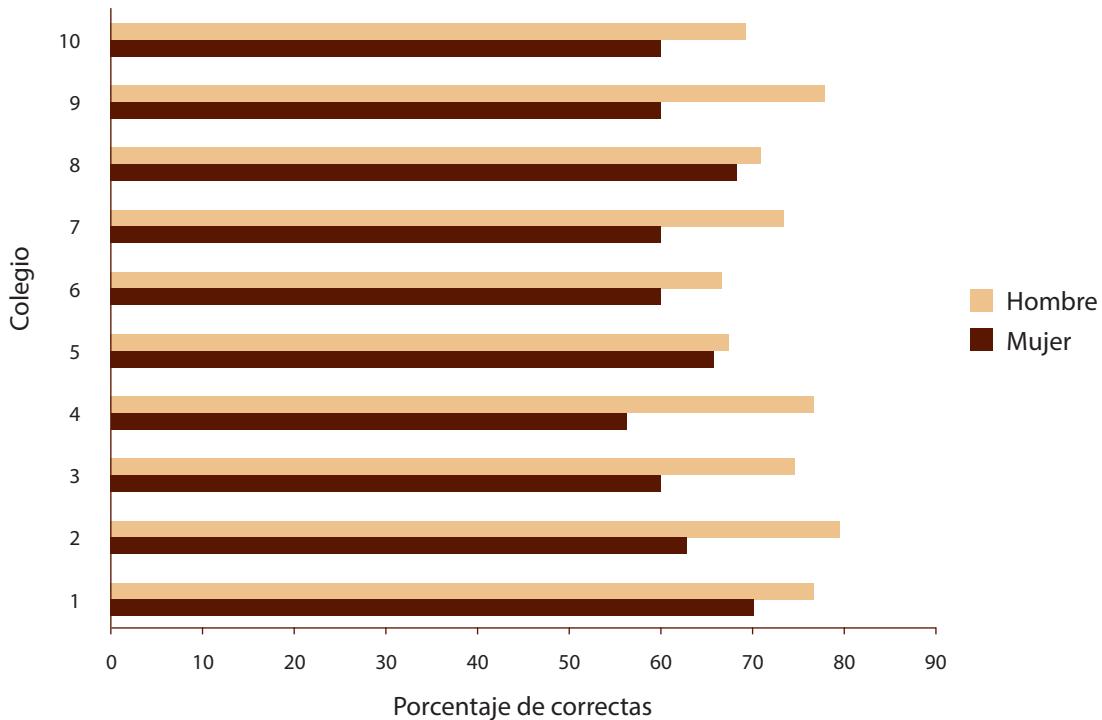
**Figura 1:** Puntaje promedio de respuestas correctas en la parte matemática de la PAA de la UCR, por colegio, según sexo.

**Nota:** Elaboración propia.

**Figura 2:** Puntaje promedio de respuestas correctas en la Prueba de Matemática del Bachillerato de la Educación Formal, por colegio, según sexo

**Nota:** Elaboración propia.

**Figura 3:** Puntaje promedio de respuestas correctas en la PRF, por colegio, según sexo



**Nota:** Elaboración propia.

Como se observa en la [Tabla 1](#), y tal como era de esperar, los promedios de hombres y mujeres difieren de manera relevante en las dos variables dependientes, encontrándose diferencias estadísticamente significativas en la prueba T de Student de comparación de promedios. Igualmente, se encontraron diferencias de importancia a favor de los varones en los puntajes de la Prueba de razonamiento con figuras, las medidas de sexismo y la autoeficacia. La única variable donde las diferencias no fueron relevantes en la comparación por sexo fue el puntaje en la medida de equidad.

### Modelos multinivel

La [Tabla 2](#) presenta los resultados de los modelos multinivel. Tal y como se indicó antes, el llamado modelo reducido incluye únicamente el sexo como variable independiente; mientras que el modelo completo incluye, además, el puntaje en la Prueba de razonamiento con figuras, los puntajes en las dos escalas de sexismo, la equidad y la autoeficacia matemática.



**Tabla 2:** Resultados de los modelos multivariados para predecir el porcentaje de correctas en la sección de razonamiento en contextos matemáticos de la Prueba de Aptitud Académica de la Universidad de Costa Rica y el porcentaje de correctas en la Prueba de Matemática del Bachillerato de la Educación Formal del Ministerio de Educación de Costa Rica

Predictores	MATE_UCR			MATE_UCR			MATE_BACH			MATE_BACH			
	Modelo reducido			Modelo completo			Modelo reducido			Modelo completo			
	B	Beta	Valor p	B	Beta	Valor p	B	Beta	Valor p	B	Beta	Valor p	
SEXO (1)	5.351	0.165	0.003	3.587	0.115	0.042	4.805	0.173	0.000	1.962	0.070	0.094	
PRF	-	-	-	0.167	0.191	0.000	-	-	-	0.210	0.307	0.000	
BENE	-	-	-	-2.513	-0.099	0.033	-	-	-	-1.524	-0.079	0.058	
HOSTIL	-	-	-	0.151	0.003	0.901	-	-	-	-0.919	-0.049	0.256	
EQUIDAD	-	-	-	0.221	0.009	0.829	-	-	-	0.688	0.038	0.326	
AUTOEFICA	-	-	-	1.657	0.077	0.087	-	-	-	4.415	0.266	0.000	
Constante	37.357	0.000	27.016	0.000	61.429	0.000	0.000	35.972	0.000	0.000	0.000	0.000	
R <sup>2</sup>	2.21%		8.41%			2.26%			2.26%			26.10%	
Efectos aleatorios	Estimado	Estandarizado	Estimado	Estandarizado	Estimado	Estandarizado	Estimado	Estandarizado	Estimado	Estandarizado	Estimado	Estandarizado	
Constante	1.888	0.103	0.000	0.000	5.560	0.400	5.005	0.360					
Sexo	2.490	0.113	0.000	0.084	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	

(1) SEXO: 1 si el estudiante es hombre, 0 si es mujer.

**Nota:** Elaboración propia.



En el caso de la parte matemática de la Prueba de Admisión de la UCR, para el modelo de efectos fijos, el valor del coeficiente para la variable sexo fue de 5.35. Esta cifra indica que, en promedio y ajustando por la estructura de conglomerados de la muestra, los hombres presentaron 5.35 puntos más en la nota porcentual de la parte matemática de la prueba de admisión, comparados con las mujeres. Esta diferencia posee importancia práctica, pues cinco puntos porcentuales en este componente del promedio de admisión pueden representar el ingreso o no a la UCR, o a una carrera específica. Ahora bien, cuando se obtiene este mismo coeficiente con el modelo completo, que incluye la medida de habilidades de razonamiento, los dos indicadores de sexismo, la equidad y la autoeficacia matemática, su valor baja a 3.59, lo cual representa alrededor de un 33 % de disminución con respecto a la diferencia entre hombres y mujeres estimada con el modelo reducido. Para la Prueba de Matemática de Bachillerato, las disminuciones son todavía más dramáticas, puesto que la diferencia entre hombres y mujeres, controlando solamente por la estructura de conglomerados es de 4.80, mientras que, en el modelo completo, que incluye, además del sexo, las habilidades de razonamiento, los dos indicadores de sexismo, la equidad y la autoeficacia matemática, es de 1.96, lo que representa una disminución de 59 % con respecto al resultado del modelo reducido. Aun en el modelo completo, con los predictores adicionales, el sexo sigue siendo una variable relevante para explicar las diferencias en desempeño en el caso de la prueba de la UCR, dado que su coeficiente de regresión estandarizado, beta, es 0.11, estadísticamente significativo, y superior a 0.10, valor de umbral para considerarlo de importancia práctica. Sin embargo, en el caso de la prueba de Matemática de Bachillerato, el coeficiente estandarizado para sexo disminuye a 0.07 en el modelo completo, y no resulta estadísticamente significativo.

Por otro lado, la variable que más fuerza posee para la predicción de ambas variables dependientes es la medida de habilidades de razonamiento. Para la parte matemática de la prueba de admisión de la UCR el valor de beta es 0.19, mientras que para la prueba de Matemática de Bachillerato es de 0.31, donde los dos coeficientes son estadísticamente significativos. En el caso de la prueba de admisión de la UCR, la autoeficacia matemática no resulta significativa, pero sí el sexismo benevolente con un coeficiente de -0.10. Por el contrario, en la prueba de Matemática de Bachillerato resulta de mayor importancia predictiva la autoeficacia matemática, con un coeficiente beta bastante alto de 0.27.

Finalmente, a partir de estos resultados, no hay evidencia que indique que los desempeños promedio de los estudiantes en la parte Matemática de la prueba de la UCR varían sistemáticamente de colegio a colegio, una vez controladas las variables predictoras del modelo, pues el valor estimado, con variables estandarizadas, de la variabilidad entre colegios para el intercepto es 0.00. Sin embargo, el modelo sí arroja un nivel de variabilidad, de colegio a colegio, en la relación estimada entre el sexo y el desempeño en la prueba, pues la variabilidad estimada para esta relación en variables estandarizadas es de 0.08. En el caso de la prueba de Matemática de Bachillerato, hay evidencia para afirmar que el desempeño promedio varía



grandemente de colegio a colegio, aun si se controlan las variables predictoras en el modelo, dado que la variabilidad estimada del intercepto para variables estandarizadas es 0.36; por el contrario, la relación entre el sexo y el desempeño en la prueba parece ser la misma en todos los colegios, una vez controladas las variables predictoras.

## Discusión

En función del objetivo sustantivo de la investigación, relativo al estudio de las diferencias de género en pruebas estandarizadas de contexto matemático, es claro que la inclusión y control de variables de razonamiento y de actitud ayudaron a explicar las diferencias observadas entre hombres y mujeres, y consecuentemente, a disminuir la magnitud de las diferencias observadas según sexo en los modelos estimados. Entonces, si no se controla este tipo de variables por medio de modelos multivariados, su efecto *confusor* se vería expresado en las diferencias brutas entre hombres y mujeres. Precisamente este es uno de los aportes del presente enfoque de análisis.

En esta investigación, la diferencia inicial entre hombres y mujeres en la parte matemática de la prueba de admisión de la UCR disminuye de 5.35 a 3.59 puntos porcentuales cuando se incluyen las variables de control en el modelo, y de 4.80 a 1.96 cuando la diferencia entre los sexos se calcula para los puntajes en la prueba de Matemática de Bachillerato. Es en esta última donde la reducción porcentual es más dramática (alrededor de 33 % para la prueba de admisión de la UCR y de 59 % para la de Bachillerato).

Por otro lado, la capacidad predictiva de las variables para la parte matemática de la prueba de la UCR resultó ser bastante más baja que para la prueba de Matemática de Bachillerato (cerca del 8 % y del 26 % respectivamente). Aunque se trata exactamente de la misma muestra de estudiantes, la disminución de la diferencia entre hombres y mujeres es bastante menor para la prueba de la UCR, y en este momento solo se puede especular acerca de las posibles razones. Una podría ser la diferencia en los constructos que intentan medir estas dos pruebas, a pesar de que sus ítems se presentan en contextos matemáticos.

A nivel predictivo claramente las habilidades de razonamiento, representadas en este estudio por el puntaje en la PRF (Prueba de razonamiento con figuras) resultan el mejor predictor en los modelos multinivel, para ambas pruebas, con coeficientes estandarizados de 0.193 y 0.307 respectivamente. Si esta variable no se hubiera incluido como control, su relación estaría confundida en la estimación de las diferencias entre hombres y mujeres, brindando estimaciones sesgadas y mayores. Además, la muestra de varones aventaja de manera relevante a las mujeres en este indicador, donde el promedio de aquellos es 71.45 y de estas 61.76. Consecuentemente, una buena parte de las diferencias entre los puntajes brutos de hombres y mujeres en las dos pruebas se debe a las diferencias en el indicador de razonamiento. Una explicación posible para estas diferencias en razonamiento es el conocido efecto de autoselección (*self-selection*)

(Heckman, 1979). En este caso particular, la muestra de varones que llega a último año de secundaria pública está más autoselecciónada, comparada con la muestra de las mujeres. Para ofrecer evidencia a favor de esta hipótesis se calculó el índice de masculinidad para los estudiantes que realizaron la Prueba de Aptitud Académica de la Universidad de Costa Rica en el 2015 y se comparó con el índice de masculinidad correspondiente a la población de hombres y mujeres entre 17 y 18 años, estimada, en ese mismo año por el INEC (Instituto Nacional de Estadística y Censos de Costa Rica). Los resultados de esta comparación mostraron que la razón de masculinidad a nivel poblacional es de alrededor 106; mientras que ese mismo indicador entre quienes realizaron la PAA de la UCR en el 2015 era apenas de 81, evidenciando una tendencia diferencial a favor de las mujeres. También debe apuntarse que la UCR tiene una oferta académica diversa, no concentrada en carreras mayoritariamente femeninas o masculinas.

Otra de las similitudes entre los modelos para ambas pruebas es que, entre las medidas relacionadas con la dimensión de género, solamente el sexismo benevolente brinda evidencia de una relación de cierta importancia, resulta inversa en ambos casos y consistente con la teoría y la investigación previa (Nosek et al., 2009). El respaldo a la idea de que las mujeres son más vulnerables y dependientes se asocia con menores rendimientos en las pruebas estandarizadas de contexto matemático, lo que muestra la importancia de intervenir sobre las creencias estereotipadas acerca del género que pueden afectar el desempeño de los estudiantes.

Por otro lado, los modelos difieren en términos de la relevancia de la autoeficacia matemática como variable de control, su peso es bastante mayor para la prueba de Matemática de Bachillerato, con un coeficiente beta igual a 0.27, mientras que para la parte matemática de la Prueba de Admisión es solamente de 0.08, relativamente bajo. Ambos resultados van en la dirección esperada: a mayor autoeficacia matemática, mejor desempeño en ambas pruebas. Este hallazgo concuerda con la investigación previa que indica que la autoeficacia se perfila como uno de los principales predictores del desempeño en pruebas estandarizadas. Es más, dada la diferencia de promedios, a favor de los varones, en esta escala, es claro que parte de las diferencias entre sexos para los desempeños en las pruebas, especialmente la de Matemática de Bachillerato, es debida a los menores niveles de autoeficacia matemática de las mujeres, resultado que concuerda con la teoría, al representar una internalización de los estereotipos y reflejar la asimetría en las relaciones de poder entre hombres y mujeres. Se hace crucial, por tanto, intervenir formalmente desde el proceso educativo, para promover mayores niveles de autoeficacia matemática en las estudiantes.

Es precisamente en hallazgos empíricos como estos donde se manifiestan, en los procesos de desarrollo personal y social de la vida cotidiana, las dimensiones de subjetividad e intersubjetividad, inherentes a la diferenciación de género, y consistentes con nuestro enfoque teórico. Así, en los programas de intervención que se diseñen alimentados por estos resultados, es necesario hacer explícitas estas dimensiones.



Finalmente, otros resultados de interés se refieren a la estimación de los efectos aleatorios, tanto del intercepto como del sexo en ambas variables de criterio. Se debe recordar que un efecto aleatorio de importancia implica que la relación que estamos estimando, en este caso entre el sexo y el desempeño en la prueba, variará de conglomerado a conglomerado, es decir, de colegio a colegio. Esto significa que el ambiente de la institución educativa produce un efecto moderador en la relación, haciéndola mayor o menor, tal y como se ha observado en otros contextos (Cassady y Johnson, 2002; Hernández Uralde et al., 2006).

En el presente estudio, el efecto aleatorio para el intercepto del modelo completo en la prueba de bachillerato en Matemática es de 5 puntos porcentuales, lo que indica que dos personas, con las mismas habilidades, actitudes psicosociales y autoeficacia, pero en diferentes colegios, pueden presentar puntajes que varían hasta en 10 puntos porcentuales (5 puntos hacia arriba o hacia abajo del valor observado). En otras palabras, si el puntaje verdadero de una estudiante es de 70 puntos, su puntaje observado podría estar entre 65 y 75, dependiendo del colegio en que se encuentre cursando sus estudios. Este hallazgo resulta particularmente preocupante en virtud de que evidencia inequidades en el sistema educativo, a pesar de que todos los colegios participantes en él pertenecen a la modalidad de académicos diurnos públicos, y todos son de la GAM. Por el contrario, para la parte matemática de la PAA, el coeficiente aleatorio del sexo es 0.00, indica que el colegio no parece tener un efecto importante en la relación entre el sexo del estudiante y su rendimiento en la prueba. Esto es, de colegio a colegio la relación estimada se mantiene.

## Conclusiones

Este estudio ha generado hallazgos para alimentar investigaciones explicativas que permitan valorar las causas de las diferencias entre hombres y mujeres en pruebas estandarizadas de contexto matemático. Asimismo, se evidencia la necesidad de generar programas de intervención dirigidos a disminuir estas diferencias.

En primer lugar, los resultados mostrados para la prueba de Matemática de Bachillerato, siendo la autoeficacia matemática el segundo predictor más importante, después de las habilidades de razonamiento, enfatizan la importancia de que el personal docente promueva activamente esta dimensión de autoeficacia en sus estudiantes, especialmente en las mujeres, considerando que los menores niveles en autoeficacia matemática de ellas explican parcialmente las diferencias observadas en las pruebas estandarizadas de contexto matemático. Lamentablemente, diversas investigaciones han mostrado que, por el contrario, muchas de las actitudes y comportamientos de docentes contribuyen más bien a perpetuar las percepciones estereotipadas tradicionales acerca de las mujeres y la matemática, generando un menor sentido de autoconfianza matemática para ellas (Gamboa-Araya, 2012). Este comportamiento no es casual, sino una expresión de la internalización de los estereotipos, producto de la

estructura patriarcal de nuestra sociedad. Se hace necesario, entonces, intervenir de manera formal y continua desde el sistema educativo, para retar esta estructura y generar espacios de aprendizaje con mayor equidad.

En segundo lugar, dadas las características del sistema educativo costarricense, se evidencia la necesidad de que en las investigaciones se controle el efecto de las habilidades generales de razonamiento del estudiantado, para evitar la magnificación de las diferencias entre los sexos en los puntajes de las pruebas. Igualmente, aunque en menor medida, resulta pertinente controlar por factores psicosociales, tales como la ideología de sexismo benevolente.

En tercer lugar, según se muestra en la estimación de los efectos aleatorios para el intercepto en el modelo multinivel, en la prueba de Matemática de Bachillerato, los promedios estudiantiles varían considerablemente, según colegio, aun si se controlan todos los predictores mencionados. Esta situación debe ser investigada con más profundidad, pues podría reflejar inequidades de relevancia entre los contextos de los centros educativos públicos de la GAM. Aún más, es un hallazgo que lleva a reflexionar sobre qué sucedería con estas diferencias, si el estudio se extendiera a colegios públicos de todo el país, incluyendo zonas urbana y rural, por ejemplo.

En cuarto lugar, a nivel metodológico, se ha evidenciado la importancia de utilizar modelos estadísticos como la regresión multinivel, que permitan estimar el efecto moderador del centro del educativo sobre los desempeños promedio en las pruebas, así como sobre la relación entre el sexo y el desempeño en las pruebas. El ser capaz de identificar, por medio de estos modelos, las variables propias de los individuos y del ambiente del colegio que generan variaciones en las relaciones, permite orientar, de manera más precisa, los programas de intervención educativa para reducir las brechas observadas.

Finalmente, a pesar de que la inclusión de las habilidades de razonamiento y las actitudes psicosociales, y la autoeficacia matemática permitieron explicar, en buena parte, las diferencias observadas entre hombres y mujeres en los puntajes de las dos pruebas, existen aún discrepancias no explicadas por los modelos estimados en este estudio, por lo que se hace necesario continuar con la investigación de otros posibles factores.

## Agradecimiento

Agradecemos al Consejo Nacional de Rectores (CONARE) por el financiamiento para esta investigación, bajo el proyecto 723-B3-307 *Nuevas formas de medir viejas ideologías: El caso del sexismoy sus implicaciones en el ámbito académico*, así como al Ministerio de Educación Pública de Costa Rica por suministrarnos los datos de la Prueba de Bachillerato de Matemática para la muestra de estudiantes participantes. Finalmente, nuestro profundo agradecimiento a los diez colegios públicos de secundaria que tuvieron a bien abrirnos sus puertas para recolectar la información de sus estudiantes.



## Referencias

- Cadinu, M., Maass, A., Rosabianca, A. y Kiesner, J. (2005). Why do women underperform under stereotype threat? Evidence for the role of negative thinking. *Psychological science*, 16(7), 572-578. <https://doi.org/10.1111/j.0956-7976.2005.01577.x>
- Cárdenes, M., Lay, S.-L., González, C., Calderón, C. y Alegría, I. (2010). Inventario de sexismo ambivalente: Adaptación, validación y relación con variables Psicosociales. *Revista Salud & Sociedad*, 1(2), 125-135. <https://doi.org/10.22199/S07187475.2010.0002.00006>
- Cassady, J. C. y Johnson, R. E. (2002). Cognitive test anxiety and academic performance. *Contemporary Educational Psychology*, 27(2), 270-295. <https://doi.org/10.1006/ceps.2001.1094>
- Cattell, R. B. (1963). Theory of fluid and crystallized intelligence: A critical experiment. *Journal of Educational Psychology*, 54(1), 1-22. <https://doi.org/10.1037/h0046743>
- Doeppen, D., Lawsky, E. y Padwa, L. (2003). *Modified fennema-sherman attitude scales*. <teacherleaders.files.wordpress.com/2013/07/modified-fennema-math-attitude.doc>
- Ellis J., Fosdick B. K. y Rasmussen C. (2016). Women 1.5 times more likely to leave STEM pipeline after calculus compared to men: Lack of mathematical confidence a potential culprit. *PLoS ONE*, 11(7), e0157447. <https://doi.org/10.1371/journal.pone.0157447>
- Esquivel-Alfaro, J. M., Montero-Rojas, E., Sosa-Jara, D., Hernández-Rodríguez, A. C., Corella-Espinoza, M. I. y Fallas-Monge, J. (2006). Evaluación externa de las pruebas nacionales de bachillerato de la educación media [Informe final]. Universidad de Costa Rica.
- Gamboa-Araya, R. (2012). ¿Equidad de género en la enseñanza de las matemáticas? *Revista Electrónica Educare*, 16(1), 63-78. <https://doi.org/10.15359/ree.16-1.6>
- Glick, P. y Fiske, S. T. (1996). The ambivalent sexism inventory: Differentiating hostile and benevolent sexism. *Journal of Personality and Social Psychology*, 70(3), 491-512. <https://doi.org/10.1037/0022-3514.70.3.491>
- Glick, P. y Fiske, S. T. (2011). Ambivalent sexism revisited. *Psychology of Women Quarterly*, 35(3), 530-535. <https://doi.org/10.1177/0361684311414832>
- Goldstein, H. (1987). *Multilevel models in educational and social research*. Griffin.
- Heckman, J. J. (1979). Sample selection bias as a specification error. *Econometrica*, 47(1), 153-161. <https://doi.org/10.2307/1912352>
- Hedges, L. V. y Nowell, A. (1995). Sex differences in mental test scores, variability, and numbers of high-scoring individuals. *Science*, 269(5220), 41-45. <https://doi.org/10.1126/science.7604277>



- Hernández Uralde, J., Márquez Jiménez, A. y Palomar Lever, J. (2006). Factores asociados con el desempeño académico en el EXANI-I. Zona Metropolitana de la Ciudad de México 1996-2000. *Revista Mexicana de Investigación Educativa*, 11(29), 547-581. [http://www.scielo.org.mx/scielo.php?script=sci\\_arttext&pid=S1405-66662006000200547&lng=es&nrm=iso](http://www.scielo.org.mx/scielo.php?script=sci_arttext&pid=S1405-66662006000200547&lng=es&nrm=iso)
- Hox, J. J. (2010). *Multilevel analysis: Techniques and applications* (2.<sup>a</sup> ed.). Routledge. <https://doi.org/10.4324/9780203852279>
- Hyde, J. S. y Mertz, J. E. (2009). Gender, culture, and mathematics performance. *Proceedings of the National Academy of Sciences of the United States of America*, 106(22), 8801-8807. <https://doi.org/10.1073/pnas.0901265106>
- Jacobs, J. E. y Eccles, J. S. (1992). The impact of mothers' gender-role stereotypic beliefs on mothers' and children's ability perceptions. *Journal of Personality and Social Psychology*, 63(6), 932-944. <https://doi.org/10.1037/0022-3514.63.6.932>
- Mena Castillo, P. J. (2015). Desarrollo en la prueba nacional de bachillerato de matemática: Una necesidad. *Cuadernos de Investigación y Formación en Educación Matemática*, 10(13), 53-66. <https://revistas.ucr.ac.cr/index.php/cifem/article/view/19144/19200>
- Montero-Rojas, E., Castelain, T., Moreira, T. E., Alfaro-Rojas, L., Cerdas-Núñez, D., García-Segura, A., Roldán-Villalobos, M. G. (2013). Evidencias iniciales de validez de criterio de los resultados de una prueba de razonamiento con figuras para la selección de estudiantes indígenas para la Universidad de Costa Rica y el Instituto Tecnológico de Costa Rica. *Revista Educación*, 37(2), 103-117. <https://doi.org/10.15517/revedu.v37i2.12928>
- Montero, E. y Villalobos, J. (2004). *Estudio comparativo del promedio de admisión a la Universidad de Costa Rica y sus componentes: Aplicación del año 2001*. Programa Permanente de la Prueba de Aptitud Académica, Instituto de Investigaciones Psicológicas, Universidad de Costa Rica.
- Nguyen, H.-H. D. y Ryan, A. M. (2008). Does stereotype threat affect test performance of minorities and women? A meta-analysis of experimental evidence. *Journal of Applied Psychology*, 93(6), 1314-1334. <https://doi.org/10.1037/a0012702>
- Nosek, B. A., Smyth, F. L., Sriram, N., Lindner, N. M., Devos, T., Ayala, A., Bar-Anan, Y., Bergh, R., Cai, H., Gonçalves, K., Kesebir, S., Maliszewski, N., Neto, F., Olli, E., Park, J., Schnabel, K., Shiomura, K., Tudor Tulbure, B., Wiers, R. W., Samogyi, M., ... Greenwald, A. G. (2009). National differences in gender-science stereotypes predict national sex differences in science and math achievement. *Proceedings of the National Academy of Sciences*, 106(26), 10593-10597. <https://doi.org/10.1073/pnas.0809921106>



OECD. (2010). *PISA 2009 results: What makes a school successful? Resources, Policies and Practices* (Vol. 4). Autor. <http://dx.doi.org/10.1787/9789264091559-en>

OECD. (2016). *PISA 2015 results excellence and equity in education* (Vol. 1). Autor. <https://doi.org/10.1787/9789264266490-en>

Raudenbush, S. W. y Bryk, A. S. (2002). *Hierarchical linear models: Applications and data analysis method* (Advanced Quantitative Techniques in the Social Sciences series 1, 2.<sup>a</sup> ed.). Sage Publications.

Sackett, P. R., Hardison, C. M. y Cullen, M. J. (2004). On interpreting stereotype threat as accounting for African American-White differences on cognitive tests. *American Psychologist*, 59(1), 7-13. <https://doi.org/10.1037/0003-066X.59.1.7>

Smith-Castro, V. (comp.) (2014). *Compendio de instrumentos de medición IIP-2014*. Serie: Cuadernos Metodológicos. Instituto del IIP, Universidad de Costa Rica.

Steele, C. M. y Aronson, J. (1995). Stereotype threat and the intellectual test performance of African Americans. *Journal of Personality and Social Psychology*, 69(5), 797- 811. <https://doi.org/10.1037/0022-3514.69.5.797>

Villarreal Galera, M. P. (2011). Construcción de un modelo psicométrico cognitivo para una prueba de inteligencia fluida [Trabajo final de investigación]. Universidad de Costa Rica.

Voyer, D. y Voyer, S. D. (2014). Gender differences in scholastic achievement: A meta-analysis. *Psychological bulletin*, 140(4), 1174-1204. <https://doi.org/10.1037/a0036620>

Zell, E., Strickhouser, J. E., Lane, T. N. y Teeter, S. R. (2016). Mars, Venus, or Earth? Sexism and the exaggeration of psychological gender differences. *Sex Roles*, 75, 287-300. <https://doi.org/10.1007/s11199-016-0622-1>