



EXPECTED AND UNEXPECTED EFFECTS OF SEXISM ON WOMEN'S MATH PERFORMANCE

Vanessa Smith-Castro¹

Eiliana Montero-Rojas

Universidad de Costa Rica

Tania Elena Moreira-Mora

Instituto Tecnológico de Costa Rica

José Andrey Zamora-Araya

Universidad Nacional de Costa Rica

ABSTRACT

Research has shown that gender differences in math performance are partially predicted by sociocultural aspects such as sexist ideologies and stereotypes. This study examined sexist ideologies as predictors of women's performance in standardized math tests, and the mediation role of math-gender stereotypes and math self-efficacy on this relationship, while controlling for abstract reasoning. Data were analyzed in samples from High School girls and university women majoring in Social Sciences, Humanities and STEM. In secondary school, the results showed the indirect, albeit expected, effect of gender stereotypes on mathematical performance through mathematical self-efficacy. The model fit was lower at a university level, and an unexpectedly positive relationship emerged between hostile sexism and mathematical performance among STEM students. The results suggest several mechanisms by which gender ideologies and stereotypes affect women's mathematical performance.

Keywords

Sexism; Stereotypes; math Self-efficacy; Standardized math Tests.

RESUMEN

La investigación ha mostrado que las diferencias de género en el desempeño matemático están parcialmente predichas por aspectos socioculturales como las ideologías sexistas y los estereotipos. Se examinaron las ideologías sexistas como predictores del desempeño en pruebas estandarizadas de matemáticas y el papel mediador de los estereotipos de género y la autoeficacia matemática en esta relación, controlando por las diferencias individuales en razonamiento abstracto. Los datos fueron analizados en muestras de estudiantes de escuelas secundarias y en estudiantes universitarias de las áreas de Ciencias Sociales, Humanidades y STEM. En secundaria, los resultados mostraron el efecto indirecto esperado de los estereotipos de género sobre el desempeño matemático a través de la autoeficacia matemática. En las universitarias el ajuste del modelo fue menor y emergió una inesperada relación positiva entre el sexismo hostil y el desempeño matemático en estudiantes de carreras STEM. Los resultados sugieren varios mecanismos mediante los cuales las ideologías y los estereotipos de género afectan el desempeño matemático de las mujeres.

Palabras Clave

Sexismos; Estereotipos; Autoeficacia Matemática; Pruebas estandarizadas de matemática

¹ Correspondence about this article should be addressed to Vanessa Smith-Castro. Email: VANESSA.SMITH@UCR.AC.CR

Decades of research have shown that, beyond individual differences in ability, socioeconomic status, academic preparation or educational opportunities and decisions, gender differences in math performance and achievement are also partially predicted by complex sexist ideologies and cultural stereotypes (Guiso, Monte, Sapienza & Zingales, 2008; Brown & Leaper, 2010).

Sexism has been traditionally defined as the endorsement of discriminatory or prejudicial beliefs and feelings based on sex, usually linked with stereotypical conceptions of the sexes and the adoption of a traditional gender-role ideology (Moya. Morales & Expósito. 2001). Currently, psychologists identify two primary types of sexist ideologies: hostile and benevolent (Glick and Fiske 1996). Hostile sexism is a derogatory view of women, based on resentment and distrust, and the perception that women are seeking control over men. Benevolent sexism is understood here as a subjectively positive view of women as 'pure creatures', who need to be protected and adored, based on the perception of women as weak, and therefore best relegated to traditional gender roles.

Literature from different theoretical frameworks within social and educational psychology suggests various mechanisms by which sexist ideologies might indirectly affect math performance and achievement among women (Dardenne, Dumont, & Bollier, 2007; Eccles, 1983; Glick and Fiske, 2011; Nosek, Smyth, Siram, Lindner, Devos, et al., 2009; Owens, 2007/2008; Spencer, Steele, & Quinn, 1999).

For instance, within the framework of the Ambivalent Sexism Theory (Glick & Fiske, 2011), research has provided compelling evidence of how sexism affects self-perceptions, through gender stereotypes. According to this theory, benevolent and hostile sexism are, in fact, ideologies regarding gender differentiation reinforcing stereotyped views of men and women. Specifically, hostile sexism is rooted in competitive gender differentiation (men are more competent than women); while benevolent sexism is rooted in complementary gender differentiation (women are pure and warm, and match "perfectly" with men, who are more capable and competent). Thus, internalization of hostile and benevolent sexism may lead people to perceive substantial differences between genders (Hyde, 2005), which might in turn affect self-perceptions and motivations.

Within the framework of the Stereotype Threat Theory (Steele & Aronson, 1995), another substantial body of evidence of the indirect effects of sexism on academic performance has been amassed. According to Steele and his colleagues (Steele & Aronson, 1995; Steele, Spencer, & Aronson, 2002), stereotypes can coerce behavior when a member of a stereotyped group is placed in a situation in which poor performance could be seen as evidence of the individual's stereotypical group deficiencies. This situational "threat in the air" disturbs individuals' performance through mechanisms such as anxiety, disengagement or evaluation apprehension, and produces the feared lackluster performance.

Furthermore, additional evidence describing the indirect effects of sexism in academic performance has been provided within the Eccles Expectancy-valued model (Eccles, 1983). This model posits that cultural stereotypes and norms determine behaviors through two core variables: success expectancies, that is, the perceived probability of success in a particular task; and subjective task value, which refers to extent to which a task provides intrinsic interest and is perceived as useful and essential by the individual (Eccles, 1983).

Finally, the Eco-cultural Mathematical Identity Model (Owens, 2007/2008) also postulates that in order to engage in mathematical thinking, it is necessary to become a self-regulated, confident learner, with a secure sense of oneself as a mathematical thinker. In this model, the social milieu, which provides beliefs, values, learning experiences, social interactions, and technological support, plays a central role in shaping mathematic identity through self-efficacy, resilience in problem-solving and sense of ownership of mathematical thinking.

Empirical data support these general principles, showing that the more strongly women endorse BS, the more likely they are to exaggerate the size of gender differences in several



domains, including math tests, academic abilities and science interests (Zell, Strickhouser, Lane, & Teeter, 2016). Data have also shown that women's performance is impaired relative to men's in the context of salient negative beliefs about women's skills in intelligence or math ability (Nguyen & Ryan, 2008; Walton & Spencer, 2009), and that even high-achieving and motivated women in math-intensive majors are affected by this stereotype threat (Good, Aronson, & Harder, 2008). Finally, evidence shows that girls whose abilities in math are repetitively questioned by their surrounding environment develop a lower math self-concept, less confidence in their math aptitude, and are less motivated than boys in the Mathematics domain (Brown & Leaper, 2010; Hyde, Fennema, Ryan, Frost, & Hopp, 1990). Moreover, lack of Mathematical confidence is one of the major factors in women's decision not to persist with Calculus (Ellis, Fosdick, & Rasmussen, 2016).

In summary, the literature suggests that internalization of sexist ideologies and knowledge of socio-cultural stereotypes can influence women's academic outcomes by shaping their self-identities and self-confidence. The present research aims to test these ideas in different educational environments.

Current research

We propose a general model of indirect effects of sexist ideologies on women's performance in standardized math tests. We argue that hostile and benevolent ideologies about women (i.e., "women are delicate creatures that ought to be protected") increase the accessibility of specific cultural math-gender stereotypes (i.e., "Girls/women are perhaps not as good as boys/men in math"). This goes on to affect the knowledge about one's own gender identity (i.e., "I am a girl/woman") and one's self-concept and confidence (i.e., "math is not for me"), which in turn affects girls' and women's performance in standardized math tests.

In order to test this general model, it is necessary to acknowledge that performance depends not only on girls' and women's beliefs and self-confidence, but also on their gradual acquisition of specific cognitive abilities. Individual differences in abstract reasoning skills are therefore statistically controlled for, so as to test the specific role of psychosocial variables in predicting performance above and below the predictive capacity of general cognitive abilities.

Another important feature of our research is that this model was tested among women attending high schools and universities. The latter were tested separately between Social Science and Humanities students, and Science, Technology, Engineering and Mathematics (STEM) students. Data show that women experience a sense of marginalization based on the culture of STEM departments. Women in STEM majors are outnumbered by their male peers in their courses and find few female role models and professors (Hughes, 2011). The specific characteristics of the environment of many STEM majors might affect the way in which these constructs relate. Thus, we consider the separate exploration of the model in these three different sites to be of great value, but we don't advance a specific hypothesis for each environment, because of the lack of previous theoretical and empirical antecedents.

In summary, the current study aims to examine sexist ideologies as predictors of women's performance in standardized math tests, and the mediation role of math-gender stereotypes and math self-efficacy on this relationship, while controlling for abstract reasoning. As shown before, previous studies have focused on the interrelationship of those constructs separately. To our knowledge, no previous published data have been modeled considering the relationship of all these constructs in a general structural model in different educational environments. In this way, we hope to contribute to a better understanding of the implications of sexism for girls' and women's performance and achievement in the academic domain.

Methods

Participants

The model was tested on data gathered from a general survey conducted initially among 926 high school students (54% girls, $N = 501$) and 903 first-year university students (50% women, $N = 453$) from urban areas in Costa Rica. Complete data from 262 High School girls, 177 university women majoring in Social Sciences and Humanities (e.g., Social Work, Sociology, Literature), and 128 university women in STEM majors (e.g., Biology, Engineering, and Mathematics) were used for the primary analyses. Girls from the High School sample came from 10 different public institutions located in the Central Valley of San José, Costa Rica. The mean age of the High School sample was 16.65 years ($SD = .79$ years). University students attended the major State universities of the Country: the University of Costa Rica, the National University of Costa Rica and the Costa Rican Institute of Technology. The mean age of the university sample was 19.02 years ($SD = 2.59$ years).

Measures

Hostile Sexism (HS) and Benevolent Sexism (BS). Participants' sexist ideologies were measured using a Spanish version of the Ambivalent Sexism Inventory (Exposito, Moya & Glick, 1998; Fiske & North, 2014; Glick, 1996). The 22-item inventory is made up of two subscales measuring two related but relatively separate constructs (HS and BS). HS measures sexist antipathy towards women based on the perception that women seek control over men, while BS measures the vision of women as delicate creatures, confined to limited roles. Examples of HS items are "Women seek to gain power by getting control over men" and "Women exaggerate problems they have at work". Examples of BS items are "Many women have a quality of purity that few men possess" and "Women should be cherished and protected by men". Participants responded to the items by using a five-point Likert scale from 1 (totally disagree) to 5 (totally agree).

Perceptions of gender equality in math abilities (EQUALITY). Participants' beliefs about men's and women's abilities in math contexts were measured as a proxy for math-gender stereotypes, using four items from the Mathematics as a Neutral Domain Scale (Forgasz, Leder & Gardner, 1999). The four items were: "Women can do just as well as men at math", "I would trust a female just as much as I would trust a male to solve important math problems", "Being good at mathematics comes as naturally to girls as to boys" and "Boys are just as likely as girls to enjoy mathematics". Forgasz and colleagues (1999) have shown that traditional scales measuring math as a Male Domain (e.g., Fennema-Sherman Mathematics Attitude Scales, 1976) might not be suitable to measure current math-gender stereotypes given that gender differences in math performance and achievement have changed over time. Thus, instead of measuring traditional math-gender stereotypes (men are naturally better than women in math), we focused on measuring the extent to which participants believe that women and men are equally good at math. Participants responded to the items by using a five-point Likert scale from 1 (totally disagree) to 5 (totally agree).

Math self-confidence (CONFIDE). Participants' beliefs about their own mathematical abilities were measured using the Fennema-Sherman (1976) Confidence in Learning Mathematics Subscale. The Subscale is intended to measure confidence in one's ability to learn and to perform well in mathematical tasks. Examples of the items are: "I know I can do well at math;" "I am sure I could do advanced work in math;" or "I do not think I could do advanced math" (reverse scored). Participants responded to the items by using a five-point Likert scale from 1 (totally disagree) to 5 (totally agree).

Reasoning Abilities (REASON). As a control variable, participants' general reasoning skills were measured using a subset of a Test of Reasoning with Figures, developed by the Department of Specific Tests of the University of Costa Rica (Montero, Castelain, Moreira, Alfaro, Cerdas, et al., 2013). The test taps fluid cognitive functioning, which involves the active maintenance of verbal and visuospatial information in working memory, and includes skills such as problem-solving, learning, and pattern recognition (Cattell, 1963). The test comprises four subtests:



series, classification, conditions, and matrices. The subset used for the present research has 17 Six-Option multiple choice seriation tasks. Participants receive one point for each correct item. Evidence for the construct validity of the test as a measure of fluid reasoning has been provided by Cliff & Montero (2010).

Performance in Standardized math Tests (MATH). Participants' scores (percentage of correct answers) on two high-stakes standardized math tests were used to address their performance in math contexts: the mathematics subset from the Admission Test for the University of Costa Rica (MATHA) and the National Mathematics High-School Exit Test, developed by the Costa Rican Ministry of Education (MATHE). The former measures reasoning skills in mathematical contexts such as induction, deduction, categorization, analog thinking, and interpretation across 30 five-option multiple-choice items (Smith-Castro, 2014). The latter measures the ability to employ mathematical concepts, procedures, and tools across 60 four-option multiple-choice items, covering geometry, algebra, statistics, and probabilities (Mena, 2015).

Procedures

After the approval of the Scientific Ethics Committee of the University of Costa Rica, students were contacted in their classrooms and invited to participate voluntarily in the study. In classrooms, students completed the REASON test and the questionnaire with HS, BS, EQUALITY, and CONFIDE. As part of the informed consent, participants were asked for permission to access their MATHE and MATHA scores. Both institutions granted access to the required data.

Statistical analyses

Since the primary purpose of the study was to test a causal model with observational (cross-sectional, correlational) data, Structural Equation Modeling (SEM) was employed. The covariance matrix of the manifest variables was estimated using Lisrel 9.1 (Jöreskog & Sörbom, 1996a, 1996b). Parameter estimates were calculated using the Maximum Likelihood method. Model fit was evaluated using global, parsimony, and incremental fit indexes. Optimal models were calculated separately for High School girls, university students attending Social Sciences and Humanities majors, and university students majoring in STEM. Given the number of items involved in some scales, we used parcels as aggregate-level manifest indicators of the constructs, averaging three or four items for each parcel, depending on the number of items in each instrument (Little, Rhemtulla, Gibson & Schoemann, 2013). Specifically, for HS (three parcels), BS (three parcels), and CONFIDE (four parcels), parcels were formed by averaging randomly allocated items. Since EQUQUALITY comprised only four items, these were not combined into parcels. For the REASON construct, four parcels were created by summing items with similar difficulty. Finally, two indicators for the MATH construct (participants' scores on MATHA and MATHE) were employed. The preliminary analysis included internal consistency tests for the manifest indicators using Cronbach's Alpha coefficients and the examination of the bivariate correlations between the manifest indicators of the variables under study. Additionally, differences across sites on the average of the manifest indicators were tested using a Multivariate Analysis of Variance (MANOVA), and Post Hoc analyses with Bonferroni corrections for multiple comparisons.

Results

Descriptive statistics and correlations

Descriptive statistics and Cronbach's Alfa for all indicators across samples are shown in Table 1, whereas Table 2 presents the bivariate Pearson correlations between them. Higher scores indicate higher levels of the constructs being measured.

Overall, internal consistency analyses yielded satisfactory Cronbach’s Alpha coefficients for most average indicators across samples, ranging from .71 to .92, with the exception of MATH among High School girls, with a coefficient of .52; and REASON and EQUALITY among women in STEM majors, with coefficients of .56, and .59, respectively (see Table 1).

The MANOVA showed a general multivariate effect of sites on all indicators (Wilk’s Lambda $F_{14,1116} = 19.833, p < .001$), supporting the idea of conducting the main analysis for each sample separately. Univariate ANOVAS indicated significant differences across sites for HS, $F_{2,564} = 14.262, p < .001, \eta^2 = .048$; BS, $F_{2,564} = 26.375, p < .001, \eta^2 = .086$; EQUALITY, $F_{2,564} = 11.265, p < .001, \eta^2 = .038$; CONFIDE, $F_{2,564} = 8.507, p < .001, \eta^2 = .029$; REASON, $F_{2,564} = 46.816, p < .001, \eta^2 = .142$; and MATH, $F_{2,564} = 101.294, p < .001, \eta^2 = .264$.

Post Hoch comparisons (see Table 1) showed that High School girls endorsed sexist ideologies more than all university women together (all $ps < .05$). On the other hand, High School girls are less convinced by the idea that women are as good as men in math compared with all university women together (all $ps < .05$). Additionally, women majoring in STEM exhibited significantly more confidence in their math abilities than women in Social Sciences and Humanities majors and High School girls together (all $ps < .001$). Finally, women majoring in STEM showed significantly better performance in reasoning and math tests than women majoring in Social Sciences and Humanities, which in turn showed a better performance than High School girls (all $ps < .001$).

Correlation matrix (see Table 2) shows that math performance was positively related to general reasoning abilities and math self-confidence across samples. Among High School girls, math performance was also positively related with perceptions of gender equality in math abilities, and negatively correlated with benevolent sexist ideologies. Confidence in the own math abilities was positively related to reasoning abilities across samples, and with perceptions of gender equality in math, but the correlation was only statistically significant among High School girls. Math self-confidence related negatively with hostile sexist ideologies among women majoring STEM. The perception that women are as good as men in math was positively correlated to math self-confidence among High School girls and women majoring in STEM. These perceptions were negatively correlated to hostile and benevolent sexist ideologies among High School girls, and to benevolent sexist ideologies among women attending STEM majors. Finally, data show a significant positive correlation between both types of sexist ideologies across samples.

Table 1
Descriptive statistics and Cronbach’s Alfa for all indicators across samples

Constructs (Indicators)	Highs school Students (N = 262)					Social Sciences and Humanities (N = 177)					STEM (N = 128)				
	Min.	Max.	Mean	SD	α	Min.	Max.	Mean	SD	α	Min.	Max.	Mean	SD	α
HS (3)	1	4.83	2.74 ^a	0.71	.74	1	4.50	2.38 ^b	0.73	.81	1	4.19	2.56 ^b	0.61	.72
BS (3)	1	4.58	2.83 ^a	0.71	.71	1	5	2.37 ^b	0.76	.81	1	4.06	2.39 ^b	0.73	.77
EQUALITY (4)	1.25	5	4.31 ^a	0.85	.76	1	5	4.57 ^b	0.67	.73	2	5	4.65 ^b	0.52	.59
CONFIDE (4)	1.5	5	3.67 ^a	0.85	.91	1	5	3.63 ^a	0.92	.92	1.58	5	4.00 ^b	0.76	.92
REASON (4)	0	100	61.76 ^a	20.86	.75	5.88	100	72.58 ^b	19.43	.77	35.29	100	80.74 ^c	13.16	.56
MATH (2)	20	85	49.80 ^a	12.16	.52	30.83	100	60.25 ^b	15.26	.78	33.33	98.34	69.99 ^b	13.43	.71

Note. HS = Hostile Sexism, BS = Benevolent Sexism, EQUALITY = Perceptions of Gender equality in math contexts, CONFIDE = math self-confidence, REASON = General reasoning, MATH = Standardized math tests. High School = High school, Social Sciences, and Humanities = Social Sciences and Humanities, STEM= Science, Technology, Engineering, and Mathematics. For descriptive purposes, averages of indicators are presented here. Different subscripts across rows indicate significant mean differences at $p < .05$, using Bonferroni correction for multiple comparisons.



Table 2.
Bivariate correlations between all indicators across samples

	HS			BS			EQUALITY			CONFIDE			REASON			
	1	2	3	1	2	3	1	2	3	1	2	3	1	2	3	
BS	.38**	.42**	.36**													
EQUALITY	-.17**	-.10	-.12	-.13*	-.12	-.17*										
CONFIDE	-.06	-.11	-.27**	-.02	-.01	-.06	.19**	-.03	.29**							
REASON	.02	.04	-.03	-.16**	-.02	-.01	.28**	.12	.12	.18**	.23**	.38**				
MATH	-.09	.10	-.02	-.13*	-.03	-.03	.16*	.08	.17	.18**	.31**	.46**	.26**	.55**	.53**	

Note. 1 = High school (N = 262), 2 = Social Sciences and Humanities (N = 177), 3 = STEM (N = 128). * $p < .05$. ** $p < .01$

Testing the proposed model

Fit indices and standardized parameter estimates for the direct and indirect paths of the original model across samples can be found in Table 3; while Diagrams A, C, and E in Figure 1 depict the structural relationships between constructs for the original model across samples. The inspection of the data revealed that the proposed model exhibits a better fit among High School girls than among all university women. In the High School sample (N = 262) fit indices were: $\chi^2 = 197.259$; $df = 160$; $p = .024$; CFI = .98; NFI = .93; GFI = .93; RMSEA = .03.

Structural coefficients showed that HS ($\beta = -.14$, $B = -.04$, $SE = .03$, $t = -1.58$, $p = .115$) and BS ($\beta = -.17$, $B = -.03$, $SE = .03$, $t = -1.80$, $p = .073$) predicted EQUALITY in the expected direction (the more participants endorsed hostile and benevolent ideologies, the less they believed girls do as well as boys in math), but were *not* statistically significant. On the other hand, EQUALITY ($\beta = .17$, $B = .62$, $SE = .26$, $t = 2.35$, $p = .019$) and REASON ($\beta = .16$, $B = .73$, $SE = .33$, $t = 2.15$, $p = .032$) significantly predicted CONFIDE. The more participants believed that girls and boys are equally good in math, the more they believe in their own mathematical abilities, after controlling for their scores on abstract reasoning abilities. Finally, EQUALITY ($\beta = .19$, $B = 2.22$, $SE = 1.09$, $t = 2.03$, $p = .043$), CONFIDE ($\beta = .29$, $B = .94$, $SE = .31$, $t = 3.068$, $p = .002$) and REASON ($\beta = .42$, $B = 6.04$, $SE = 1.69$, $t = 3.58$, $p < .001$ s) significantly predicted MATH. The more participants believed in gender equality in math abilities and the more confident they felt about their own mathematic abilities, the better they performed in standardized math tests, after controlling for individual differences in reasoning abilities. The model explained 36,7 % of the variance of MATH.

To test the significance of the mediating effect of CONFIDE on the relationship between EQUALITY and MATH, we estimated a model constraining the coefficients from EQUALITY to CONFIDE to 0. The Likelihood Ratio Test yielded a Delta Chi-Squared of 5.20, with an associated p -value of .022, indicating that the indirect effect model has a better fit to the data compared to the reduced model, in which the indirect effect is equal to 0.

Among women majoring in Social Sciences and Humanities, the model showed a lesser fit to the data: $\chi^2 = 214.555$; $df = 160$; $p = .003$; CFI = .97; NFI = .93; GFI = .93; RMSEA = .04. In this sample, only three of the seven postulated structural relations were statistically significant. Specifically, REASON significantly predicted both CONFIDE ($\beta = .27$, $B = 1.27$, $SE = .42$, $t = 2.98$, $p = .003$) and MATH ($\beta = .65$, $B = 17.04$, $SE = 3.03$, $t = 5.63$, $p < .001$), and CONFIDE predicted MATH ($\beta = .20$, $B = 1.11$, $SE = .44$, $t = 2.53$, $p = .012$). The more participants believed in their own math abilities, the better they performed in standardized math tests, after controlling for their scores on reasoning abilities. HS, BS and EQUALITY did not show their expected indirect effects on math performance in this sample, therefore no mediation effects were tested. The model explained 52.4 % of the variance of MATH.

Among women in STEM, the model showed a much lesser fit to the data: $\chi^2 = 225.767$; $df = 160$; $p < .001$; CFI = .95; NFI = .85; GFI = .86; RMSEA = .06. In this sample, EQUALITY

significantly predicted CONFIDE ($\beta = .28$, $B = 1.98$, $SE = .78$, $t = 2.53$, $p = .013$), and REASON significantly predicted both CONFIDE ($\beta = .54$, $B = 2.28$, $SE = .60$, $t = 3.81$, $p < .001$) and MATH ($\beta = .78$, $B = 18.61$, $SE = 5.44$, $t = 3.42$, $p < .001$). The more participants believed in gender equality regarding math abilities, the more confident they felt about their own mathematic abilities, and the more confident they felt about their math capacities the better they performed in standardized math tests, after controlling for individual differences in reasoning abilities. Again, the expected indirect effects of sexist ideologies on mathematic performance via perceptions of gender equality in math contexts and math self-confidence were not supported by the data in this site, which made mediation tests unnecessary. The model explained 71.8% of the variance of MATH.

In summary, data show that the proposed model was only partially supported across samples. Particularly in university contexts, our model did not describe the indirect effects of sexist ideologies and gender stereotypes on math performance well.

Testing an alternative model

Given the inconclusive evidence of the expected indirect effects of sexist ideologies on math performance, we tested an alternative model, including additional direct paths from HS and BS to MATH. Table 3 presents the standardized parameter estimates for the measurement model for the alternative model across samples. As in the case of the proposed model, indicators loaded significantly and strongly on its corresponding latent variable, except for EQUALITY2 in the sample of women attending STEM majors. Table 5 shows fit indices and standardized parameter estimates for the direct and indirect paths of the alternative model across samples; while Diagrams B, D and F in Fig.1 depict the structural relationships between constructs for the alternative model across samples.

Among High School girls ($N = 262$), fit indices for the alternative model were: $\chi^2 = 191.611$; $df = 158$; $p = .035$; $CFI = .99$; $NFI = .93$; $GFI = .94$; $RMSEA = .03$. Among women majoring in Social Sciences and Humanities, fit indices were: $\chi^2 = 211.208$; $df = 158$; $p = .003$; $CFI = .97$; $NFI = .90$; $GFI = .90$; $RMSEA = .04$. Finally, among women majoring in STEM, fit indices were: $\chi^2 = 220.767$; $df = 158$; $p < .001$; $CFI = .95$; $NFI = .85$; $GFI = .86$; $RMSEA = .06$. Overall, Likelihood Ratio Tests comparing the fit of the alternative model against the original model were not statistically significant, indicating that none of the alternative models were statistically superior to the original models across samples (all $\Delta\chi^2 < 5.65$, all $dfs = 2$, all $ps > .05$).

Across samples, all original structural coefficients in the alternative model remained similar in magnitude. None of the additional paths were statistically significant, with the exception of an unexpected significant positive effect of hostile sexism on math performance among women attending STEM majors ($\beta = .32$, $B = 2.08$, $SE = .96$, $t = 2.18$, $p = .02$). In this sample, endorsement of hostile sexism was significantly associated with better performance in standardized math tests, after controlling for individual differences in reasoning. This model explained 78.4% of the variance of MATH.

In summary, testing an alternative model adding direct effects of sexism on performance showed no better fit to the data than the original model of indirect effects. Therefore, the more parsimonious model of indirect effects seems to better reflect the complex relationship of the variables. However, the surprising direct and positive relationship between Hostile Sexism with math performance among women majoring in STEM should not be neglected.



Table 3.
Parameters estimates and fit indexes for the PROPOSED structural model across samples

Predictors	MATH								
	High School (N = 262)			Social Sciences and Humanities (N = 177)			STEM (N = 128)		
	Direct	Indirect	Total	Direct	Indirect	Total	Direct	Indirect	Total
HS		-0.035	-0.035		-0.001	-0.001		-0.037	-0.037
BS		-0.041	-0.041		-0.005	-0.005		-0.028	-0.028
EQUALITY	0.193	0.050	0.243	0.048	-0.015	0.033	0.171	0.019	0.190
CONFIDEN	0.292		0.292	0.195		0.195	0.067		0.067
REASON	0.419	0.048	0.467	0.645	0.053	0.698	0.783	0.036	0.819
Fit indexes									
χ^2 (160)		197.259			214.555			225.767	
p		0.0240			0.0026			0.0005	
RMSEA		0.0298			0.0439			0.0567	
RMSEA P CLOSE		0.996			0.740			0.256	
NFI		0.926			0.898			0.848	
CFI		0.985			0.972			0.949	
GFI		0.933			0.897			0.860	

Note. HS = Hostile Sexism, BS = Benevolent Sexism, EQUALITY = Perceptions of Gender equality in math contexts, CONFIDE = math self-confidence, REASON = General reasoning, MATH = Standardized math tests. High School = High school, Social Sciences, and Humanities = Social Sciences and Humanities, STEM= Science, Technology, Engineering, and Mathematics.

Table 4.
Parameters estimates and fit indexes for the ALTERNATIVE structural model across samples

Predictors	MATH								
	High School (N = 262)			Social Sciences and Humanities (N = 177)			STEM (N = 128)		
	Direct	Indirect	Total	Direct	Indirect	Total	Direct	Indirect	Total
HS	-0.025	-0.027	-0.046	0.175	-0.002	0.174	0.323	-0.049	0.274
BS	-0.200	-0.030	-0.230	-0.085	-0.005	-0.091	-0.115	-0.036	-0.151
EQUALITY	0.136	0.051	0.191	0.053	-0.017	0.036	0.220	0.026	0.246
CONFIDEN	0.296	-	0.296	0.212	-	0.212	0.097	-	0.097
REASON	0.372	0.048	0.418	0.625	0.057	0.683	0.825	0.056	0.881
Fit indexes									
χ^2 (158)		191.611			211.208			220.572	
p		0.0353			0.0030			0.0007	
RMSEA		0.0285			0.0436			0.0556	
RMSEA P CLOSE		0.997			0.748			0.290	
NFI		0.928			0.900			0.851	
CFI		0.986			0.972			0.952	
GFI		0.935			0.897			0.863	

Note. HS = Hostile Sexism, BS = Benevolent Sexism, EQUALITY = Perceptions of Gender equality in math contexts, CONFIDE = math self-confidence, REASON = General reasoning, MATH = Standardized math tests. High School = High school, Social Sciences, and Humanities = Social Sciences and Humanities, STEM= Science, Technology, Engineering, and Mathematics.

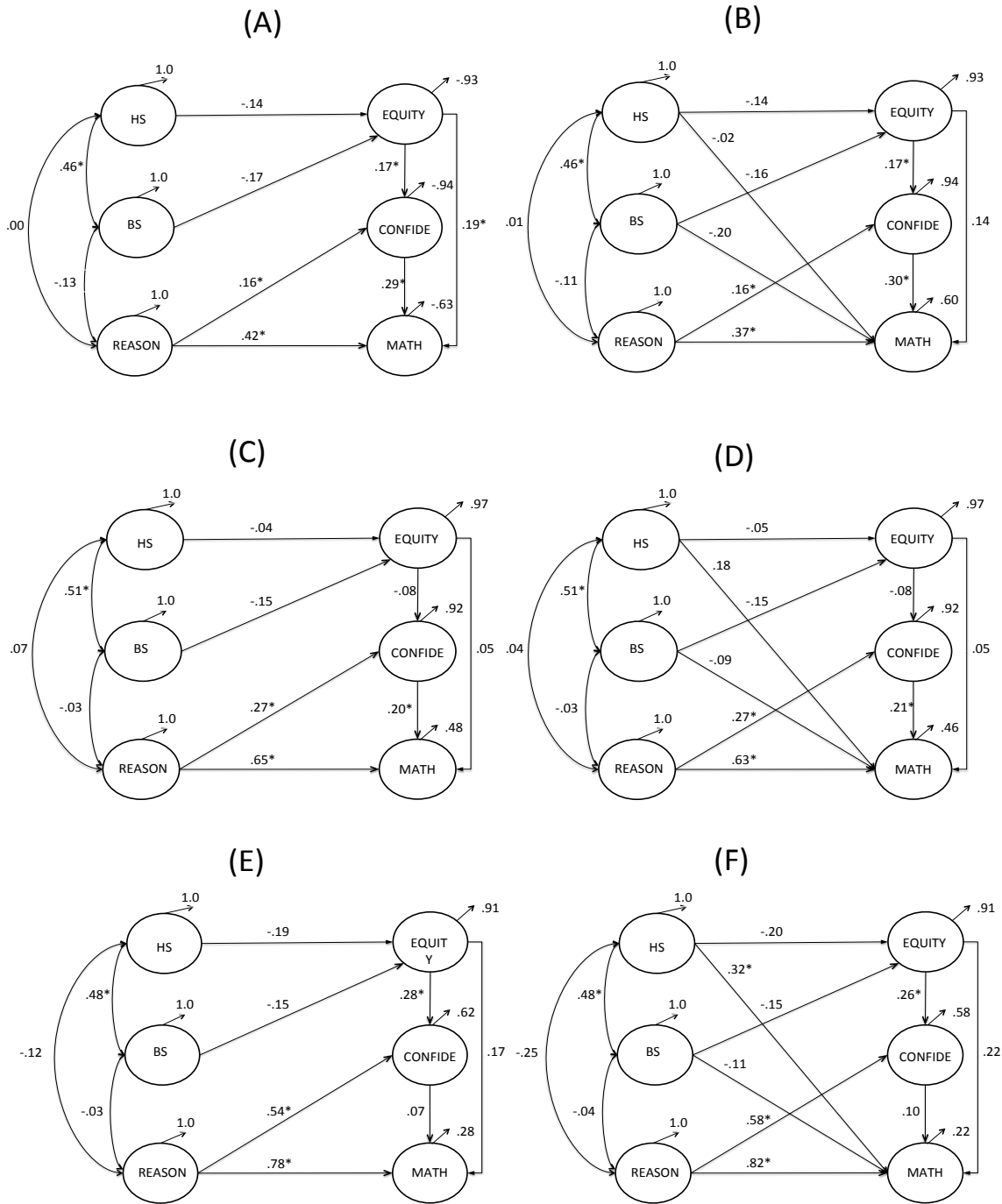


Figure 1. Standardized parameter estimates for the original and the alternative models across samples. Diagrams A, C, and E depict parameter estimates for the original model among High School girls, university women in Social Sciences and Humanities majors, and University women in STEM majors, respectively. Diagrams B, D, and F depict parameter estimates for the alternative models, in these environments, respectively.
 * = $p < .05$



Discussion

The current study aimed to examine sexist ideologies as predictors of women's performance in standardized math tests, and the mediation role of math-gender stereotypes and math self-efficacy on this relationship, while controlling for abstract reasoning.

Our data partially support this complex chain of effects among High School girls, but also highlighted substantial differences between High School environments and university contexts, and revealed some unexpected effects of sexist ideologies on performance among university women majoring in STEM. In light of this, we discuss our results taking into account the specificity of each educational environment.

High-school environments

Among high school girls, data showed that the more they endorsed sexist ideologies, the less convinced they were that girls are as good as boys in math. On the other hand, data showed that the more they believed in gender equality regarding mathematical abilities, and the more confident they felt about their mathematical capacities, the better they performed in standardized math tests. These findings are notable in several ways.

First, after statistically controlling for individual differences in their cognitive abilities, the most important predictor of girls' mathematical performance was their confidence in their mathematical abilities. This result highlights the need for a more systematic approach to the role of affective variables in girls' academic performance. It is possible that self-efficacy affects girls' academic effort and persistence in their preparation for exams, and that this academic engagement is responsible for the subsequent performances. It is also possible that the lack of confidence increases girls' levels of anxiety while performing tests, leading to impairments in their performance. These hypotheses are consistent with previous research showing that mathematical confidence is one of the major predictors of girls' math performance, and one of the major factors in women's decision not to persist with math-intensive courses (Ellis, Fosdick & Rasmussen, 2016). More research is needed to understand the ways in which self-confidence affects performance. Overall, however, our data support the assumption that in order to promote girls' mathematical thinking, they require encouragement in the development of a secure sense of self as mathematical thinkers (Owens, 2007/2008).

Second, our findings show that even after controlling for their actual and perceived math abilities, stereotypical beliefs about gender equality in math abilities remained a significant predictor of performance. Our data are consistent with research based on Eccles' model (Eccles, 1983), showing that the cultural transmission of gender-role stereotypes influences individuals' goals and general self-schemata, which in turn influence specific behaviors and performances. These findings are also consistent with research on Stereotype and Social Identity Threat (e.g., Steele and Aronson 1995; Steele et al. 2002), suggesting that the mere salience of stereotypes can directly impair performance in tests, above and beyond actual and perceived cognitive abilities (Steele, 1997).

Although our study does not directly address issues of career choices, our results provide relevant information for teachers, school authorities, and policymakers in the context of the worldwide debate around the gender gap in STEM fields. Several authors have suggested that stereotypes such as "men are naturally more brilliant and interested in math and science than girls" influence the educational aspirations and achievements of boys

and girls, as well as the occupational choices of men and women (Bian, Leslie & Cimpian, 2017; Hill, Corbett & St. Rose, 2010). Previous research has shown that women who endorse such stereotypes also report less interest in math and science, and are less likely to pursue a math or science degree (Schmader, Johns & Barquissau, 2004). Experimental data have also shown that reminding women of the “math-male” stereotype, or just unobtrusively emphasizing their gender, is enough to diminish their performance on a subsequent math or engineering test (Nosek, Smyth, Sriram, Lindner, Devos, et al, 2009; Hill, Corbett & St. Rose, 2010). Our data add more evidence along these lines, showing that stereotypes affect, both directly and indirectly, girls’ performance in two high stakes standardized math tests that are crucial in shaping their future: the University of Costa Rica admission test and the National Secondary Exit Test.

Third, our results suggest that sexist ideologies have a marginal distal influence on performance, primarily through gender stereotypes. Although not statistically significant, the sign and magnitude of the relationship between Hostile and Benevolent Sexism and perceptions of gender equality in math capacities should not be neglected, highlighting the need to continue examining the role of sexism in stereotypes in the academic domain.

The low magnitude of the observed effects and the lack of their significance are likely due to the fact that the content of HS and BS focus on competitive and complementary gender differentiation along with the interpersonal domain, rather than direct comparisons of men and women capacities in the academic domain. The putative effects of sexism on stereotypes were therefore found here to be only indirect and marginal. More research is needed in order to capture those aspects of sexism that directly influence gender stereotypes and performance in academic domains. For instance, in a series of experiments, Dardenne, Dumont & Cattell (2007) found that benevolent sexism per se (rather than paternalism) impaired women’s performance (Experiment 3), but also that the impaired performance was fully mediated by mental intrusions about the participants’ sense of competence (Experiment 4). These results, together with our data, highlight multiple paths in which sexism negatively affects performance in academic domains which deserve more research in the future.

University environments

Among university students, the expected indirect effect of sexist ideologies and stereotypes on performance was, overall, not fully supported. Rather than gender ideologies or cultural beliefs about gender differentiation, the most important predictors of their performance were their sense of math self-confidence and their reasoning abilities.

Perhaps these findings relate to the typical self-selection processes involved in the admission to any university. Previous research by Correll (2001) has shown that beliefs about gender differences in mathematics impact individuals’ assessments of their mathematical competence, which in turn leads to gender differences in decisions to persist on a path toward a STEM career. The author posits that shared cultural beliefs attached to various tasks affect not only how individuals are conducted into particular academic activities by others, but also how individuals “self-select” into academically relevant activities, which might contribute to the large gap between the number of male and female students who choose STEM majors.

Our data show that, overall, women in university contexts reported significantly fewer sexist beliefs and cultural gender stereotypes than High School girls, suggesting that this self-selection hypothesis might not only apply to STEM majors but also the general



pursuit of high-level academic paths. Following the “leaking pipeline” metaphor, we believe that on the path towards university, girls with less progressive gender attitudes might have leaked out more than those with more progressive attitudes, leaving us with a self-selected population of university women with homogenous beliefs and attitudes about gender roles. This, of course, can only be adequately tested with longitudinal designs and indicates the need for more longitudinal studies in this field. These results also point to the need for the inclusion of different populations in the empirical studies, so as to acknowledge the moderating role of different educational and socio-demographic contexts in the relationships under consideration and to avoid sweeping generalizations.

Perhaps the most surprising finding in university contexts is the positive direct coefficient between hostile sexist ideologies and math performance among women majoring in STEM, that is, those who exhibit higher levels of math performance also exhibit more hostile sexism. Could this be a reflection of the boys’ club mentality that prevails in STEM, male-dominated environments? Previous research has documented that the competitive climate of many STEM majors, combined with the masculine language and culture that predominates in those environments affect women's ability to fit within these majors (Hughes, 2011), and that some ‘stayers’ in STEM majors are able to persist because they take on more “acceptable” gender roles and values (Ong, 2005). This complex process might account for the negative attitudes toward the specific “types” of women pictured in the HS measure. However, our data are not conclusive in this regard and should be taken with caution, given the potential suppression effect suggested by the differences between the zero-order correlations and the beta weights in this sample. Nevertheless, this unexpected result indeed suggests interesting research paths to tackle these issues and to continue studying women in STEM environments.

As a final methodological note, it is worth pointing out that the reasoning abilities test used as a control variable for math performance proved to accomplish a very good job in all the models, presenting standardized path coefficients higher than .37 in all cases. It shows the importance of employing controls such as this for models with endogenous variables that involve intellectual skills. Since attitudes and other psychological traits might correlate with reasoning abilities in observational studies, our recommendation is to always consider the use of this kind of control variable when working with observational data. In this way, the possible confounding effects generated by the association between psychological traits and basic reasoning abilities are neutralized.

Limitations

Our observational data inhibit us from establishing reliable causal inferences, especially in university samples, in which performance data were retrospectively collected. However, it is important to point out that, in High School samples, we measured the endorsement of sexist beliefs & stereotypes and math self-efficacy months before girls took their standardized tests, ruling out the suggestion that their attitudes were influenced by their actual performance in the standardized tests. Nevertheless, longitudinal data are still necessary in order to test the causal relationship between the variables correctly. Likewise, there might be some concern regarding the low reliability of some of the measures. However, by forming latent factors out of sub-sets of the measures, it was possible to define the constructs in terms of the shared commonalities among the parcels

and, therefore, to control for the measurement error. In other words, measurement issues cannot fully account for the pattern of associations observed here. However, attempts to improve our measures should be a constant goal in our research.

Finally, one might have questions as to the elevated percent of missing data, especially among High School students. In this regard, it is important to notice that in order to take the Exit Exams and the Admission Tests, High School students are required to complete the eleventh grade. In this scenario, a portion of missing data is due to the fact that some students either dropped out of school before taking the standardized tests, or could not successfully complete their eleventh grade. It should be noted that in 2014, the dropout rate in Costa Rican high schools was around 9%, and that the failure rate for the tenth and eleventh grades was around 20% (Programa Estado de la Nación, 2015). In such circumstances, missing data rates in studies like ours are expected to be relatively high. Future studies should employ several methods to minimize attrition, including a) school/community engagement, b) contact and communication strategies, c) tracing, d) flexibility of data collection, and e) incentives.

Conclusion

Our results provide noteworthy evidence of how sexist ideologies and gender stereotypes influence girls' and women's academic self-efficacy, and how these self-appraisals directly influence their performance in standardized math tests. The effect of ideologies and stereotypes on math self-efficacy and math performance was greater among high school girls. In university environments, on the other hand, math self-efficacy showed a substantive effect on performance, after controlling for individual differences in abstract reasoning. The unexpected positive relationship between hostile sexism and math performance among women majoring in STEM might reflect adaptation mechanisms in a male-dominated learning environment that should be further studied. The use of a reasoning test to neutralize the possible confounding effects of basic intellectual ability on the relations between math performance and the socio-affective traits (sexism, gender equality in math contexts, and math self-confidence) proved to be a fortunate decision, since basic reasoning abilities explain an important part of math performance variance.



References

- Beyer, S. & Bowden, E. (1997). Gender Differences in Self-Perceptions: Convergent Evidence from Three Measures of Accuracy and Bias. *Personality and Social Psychology Bulletin*, 23, 157-172.
- Bian, L., Leslie, S. J., & Cimpian, A. (2017). Gender stereotypes about intellectual ability emerge early and influence children's interests. *Science*, 355(6323), 389–391.
- Brown, C. S., & Leaper, C. (2010). Latina and European American girls' experiences with academic sexism and their self-concepts in mathematics and science during adolescence. *Sex Roles: A Journal of Research*, 63, 860-870.
- Cattell, R.B. (1963). Theory of fluid and crystallized intelligence: A critical experiment *Journal of Educational Psychology*, 54, 1-22.
- Cliff, A. y Montero, E. (2010). El Balance entre Excelencia y Equidad en Pruebas de Admisión: Contribuciones de Experiencias en Sudáfrica y Costa Rica [The balance between excellence and equity in admission tests: Contributions from South Africa and Costa Rica]. *Revista Iberoamericana de Evaluación Educativa*, 3 (2), 7-28.
- Correll, S.J., (2001). Gender and the career choice process: the role of biased self-assessments. *American Journal of Sociology* 106 (6), 1691-1730.
- Dardenne, B., Dumont, M., & Bollier, T. (2007). Insidious dangers of benevolent sexism: Consequences for women's performance. *Journal of Personality and Social Psychology*, 93, 764–779.
- Eccles, J. (1983). Expectancies, values, and academic behaviors. In J. T. Spence (Ed.), *Achievement and achievement motives: Psychological and sociological approaches* (pp. 75–146). San Francisco: Freeman.
- Ellis J, Fosdick BK, 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. doi:10.1371/journal.pone.0157447
- Expósito, F., Moya, M., & Glick, P. (1998). Sexismo ambivalente: medición y correlatos [Ambivalent sexism: measurement and correlates]. *Revista de Psicología Social*, 13, 159-169.
- Fennema, E., & Sherman, J. (1976). Fennema-Sherman Mathematics Attitudes Scales: Instruments Designed to Measure Attitudes toward the Learning of Mathematics by Females and Males. *Journal for Research in Mathematics Education*, 7(5), 324-326. doi:10.2307/748467
- Fiske. S.T.. & North. M.S. (2014). Measures of stereotyping and prejudice: Barometers of bias. In G. Boyle & D. Saklofske (Eds.). *Measures of Personality & Social Psychological Constructs* (pp. 684-718). Elsevier/Academic Press.
- Forgasz, H.J., Leder, G.C. & Gardner, P.L. (1999). The Fennema-Sherman Mathematics as a Male Domain Scale Reexamined. *Journal for Research in Mathematics Education*, 30 (3), 342-348.
- Glick. P. & Fiske. S.T. (1996). The Ambivalent Sexism Inventory: Differentiating hostile and benevolent sexism. *Journal of Personality and Social Psychology*. 70. 491-512.
- Glick, P. & Fiske, S. T. (2011). Ambivalent sexism revisited. *Psychology of Women Quarterly*, 35(3), 530–535. doi:10.1177/0361684311414832

- Good, C., Aronson, J., & Harder, J. A. (2008). Problems in the pipeline: Stereotype threat and women's achievement in high-level math courses. *Journal of Applied Developmental Psychology, 29*, 17-28.
- Guiso L, Monte F, Sapienza P, & Zingales L. (2008). Culture, gender and math. *Science, 320*, 1164–1165.
- Hill, C., Corbett, C., & St Rose, A. (2010). *Why so few? Women in science, technology, engineering, and mathematics*. American Association of University Women. 1111 Sixteenth Street NW, Washington, DC 20036.
- Hyde, J. S. (2005). The gender similarities hypothesis. *American Psychologist, 60*, 581–592.
- Hyde J.S, Fennema E., Ryan, M., Frost, L.A., Hopp, C. (1990). Gender comparisons of mathematics attitudes and affect: A meta-analysis. *Psychol Women Quart. 1990, 14*, 299–324.
- Hughes, R.M. (2011). Are the predictors of women's persistence in STEM painting the full picture? A series of comparative case studies. *Int. J. of Gender, Science and Technology, 3* (3), 547-570.
- Jacobs, J. E., & 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*, 932–944.
- Jöreskog, K. G., & Sörbom, D. (1996a). PRELIS: User's reference guide. Chicago, IL: Scientific Software International.
- Jöreskog, K. G., & Sörbom, D. (1996b). LISREL 8: User's reference guide. Chicago, IL: Scientific Software International. Chicago, IL: Scientific Software International Inc.
- Little, T. D., Rhemtulla, M., Gibson, K., & Schoemann, A. M. (2013). Why the items versus parcels controversy needn't be one. *Psychological Methods, 18*, 285-300.
- Mena, P. (2015). Desarrollo en la prueba nacional de bachillerato de Matemática: una necesidad [Development of the National High-schools Exit Test in Mathematics]. *Cuadernos de Investigación y Formación en Educación Matemática, 10* (13), 53-66.
- Montero, E.; Castelain, T.; Moreira, T.; Alfaro, L.; Cerdas, D.; García A.; Roldán, 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 [Initial criterion validity of the Test of Reasoning with Figures for the selection of Indigenous People at the University of Costa Rica and the Costa Rica Technological Institute]. *Revista Educación 37*(2), 103-117.
- Moya. M., Morales. E., & Expósito. F. (2001). Nuevas formas. viejos intereses: Neosexismo en varones españoles [New forms. old interests: Neosexism among Spanish men]. *Psicothema. 13* (4). 643-649.
- Nguyen, H.-H. D., & 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*, 1314-1334.
- Nosek, B. A., Smyth, F. L., Sriram, N., Lindner, N. M., Devos, T., Ayala, A., Bar-Anan, Y., Bergh, B., Cai, H., Gonsalkorale, K., Kesebir, S., Maliszewski, N., Neto, F., Olli, E., Park, J., Schnabel, K., Shiomura, K., Tulbure, B. T., Wiers, R., Somogyi, M., Akrami, N., Ekehammar, B., Vianello, M., Banaji, M. R., & Greenwald, A. G. (2009). Nationgender-science in gender-science stereotypes predict national sex



- differences in science and math achievement. *Proceedings of the National Academy of Sciences*, 106, 10593-10597.
- Ong, M., (2005). Body Projects of Young Women of Color in Physics: Intersections of Gender, Race, and Science. *Social Problems*, 52, 593-617.
- Owens, K. (2007/2008). Identity as a mathematical thinker. *Mathematics Teacher Education and Development*, 9, 36-50.
- Programa Estado de la Nación. (2015). *Quinto Informe Estado de la Educación en Costa Rica. [Fifth informof the state of the Education in Costa Rica]*. San José: Programa Estado de la Nación.
- Schmader T, Johns M, Barquissau M. (2004). The costs of accepting gender differences: The role of stereotype endorsement in women's experience in the math domain. *Sex Roles*, 50, 835–850.
- Shih, M., Pittinsky, T. L., & Ambady, N. (1999). Stereotype susceptibility: Identity salience and shifts in quantitative performance. *Psychological Science*, 10, 81–84.
- Smith-Castro, V. (comp.) (2014). *Compendio de Instrumentos de Medición IIP-2014 [IIP Measurement Instruments Compendium 2014]*. Serie Cuadernos Metodológicos. San José, CR: Instituto de Investigaciones Psicológicas.
- Spencer, S. J., Steele, C. M., & Quinn, D. M. (1999). Stereotype threat and women's math performance. *Journal of Experimental Social Psychology*, 35, 4–28.
doi:10.1006/jesp.1998.1373.
- Steele, C. M. (1997). A threat in the air: How stereotypes shape intellectual identity and performance. *American Psychologist*, 52, 613-629.
- Steele, C. M., & Aronson, J. (1995). Stereotype threat and the intellectual test performance of African Americans. *Journal of Personality and Social Psychology*, 69, 797–811.
doi:10.1037/0022-3514.69.5.797
- Steele, C. M., Spencer, S. J., & Aronson, J. (2002). Contending with group image: The psychology of stereotype and social identity threat. In M. Zanna (Ed.), *Advances in experimental social psychology* (Vol. 34, pp. 379–440). New York, NY: Academic Press.
- Walton G. M., & Spencer S. J. (2009). Latent ability: Grades and test scores systematically underestimate the intellectual ability of negatively stereotyped students. *Psychological Science*, 20, 1132–1139 10.1111/j.1467-9280.2009.02417.x
- Zell, E., Strickhouser, J. E., Lane, T. N., & Teeter, S. R. (2016). Mars, Venus, or Earth? Sexism and the exaggeration of psychological gender differences. *Sex Roles*, 75, 287-300.

Received: 12/31/2017
Accepted: 05/30/2019