Intelligence 67 (2018) 76–83

Contents lists available at ScienceDirect

Intelligence

journal homepage: www.elsevier.com/locate/intell

Sex differences in ability tilt in the right tail of cognitive abilities: A 35-year examination

Jonathan Wai[a,*], Jaret Hodges[b], Matthew C. Makel[b]

[a] Autism & Developmental Medicine Institute, Geisinger Health System, Department of Psychology, Case Western Reserve University, USA
[b] Talent Identification Program, Duke University, USA

ARTICLE INFO

Keywords:
Sex differences
Ability tilt
Cognitive abilities
STEM
Historical examination

ABSTRACT

Sex differences in cognitive ability level and cognitive ability pattern or tilt (e.g., math > verbal) have been linked to educational and occupational outcomes in STEM and other fields. The present study examines cognitive ability tilt across the last 35 years in 2,053,265 academically talented students in the U.S. (SAT, ACT, EXPLORE) and 7119 students in India (ASSET) who were in the top 5% of cognitive ability, populations that largely feed high level STEM and other occupations. Across all measures and samples, sex differences in ability tilt were uncovered, favoring males for math > verbal and favoring females for verbal > math. As ability tilt increased, sex differences in ability tilt appeared to increase. Additionally, sex differences in tilt increased as ability selectivity increased. Broadly, sex differences in ability tilt remained fairly stable over time, were consistent across most measures, and replicated across the U.S. and India. Such trends should be carefully monitored given their potential to impact future workforce trends.

1. Introduction

The underrepresentation of women in high level science, technology, engineering, and mathematics (STEM) careers is widely researched and discussed. Given the importance of ensuring the full development of female talent for STEM fields (National Academy of Sciences, 2010), understanding the origins of and solutions to such underrepresentation remains an important area of inquiry. Although recent research suggests that female representation has been improving on many indicators (e.g., Ceci, Ginther, Kahn, & Williams, 2014; Miller & Wai, 2015), women still hold only about 7–16% of tenured faculty positions and < 30% of doctorates and bachelor's degrees in math-intensive fields (Ceci et al., 2014). Many interlocking factors have been proposed to explain this differential, including interests, encouragement, and bias (Ceci & Williams, 2010; Halpern et al., 2007; Moss-Racusin, Dovidio, Brescoll, Graham, & Handelsman, 2012).

1.1. Ability differences in the extreme right tail of the distribution

Another factor that has received substantial attention that may contribute to explaining female underrepresentation in STEM fields are differences in representation in the extreme right tail or top 5% to 0.01% of the distribution of math ability (Benbow & Stanley, 1980, 1983; Wai, Cacchio, Putallaz, & Makel, 2010), which may be linked to greater male variability in various aspects, such as personality (Borkenau, McCrae, & Terracciano, 2013), brain structure (Ritchie et al., 2017), and physical parameters (Lehre, Lehre, Laake, & Danbolt, 2008). Representation differences at these select ability levels may matter because even within the top 1% of math ability, higher scores at age 13 are related to significantly higher STEM educational and occupational outcomes decades later, including earning a STEM PhD, STEM publication, STEM patent, STEM university tenure, and having a job in a STEM field (e.g., Park, Lubinski, & Benbow, 2007; Wai, Lubinski, & Benbow, 2005). Although studies suggest that at least on some math measures females have improved their representation among top scorers in recent years (Makel, Wai, Peairs, & Putallaz, 2016), males continue to have higher representation in the right tail of math measures broadly and such a difference has been apparent for at least the last 35 years.

1.2. Ability pattern or "tilt" differences in the extreme right tail of the distribution

However, math abilities in isolation, especially relative to factors such as interests (e.g., Su, Rounds, & Armstrong, 2009), are likely a lesser factor explaining female STEM underrepresentation (e.g., Ceci et al., 2014; Miller & Wai, 2015). In addition to ability level, another factor that remains understudied is ability pattern or "tilt" in the

* Corresponding author.
E-mail address: waijon@hotmail.com (J. Wai).

https://doi.org/10.1016/j.intell.2018.02.003
Received 15 May 2017; Received in revised form 4 January 2018; Accepted 14 February 2018

extreme right tail of cognitive abilities. Ability tilt can refer broadly to the pattern and structure of multiple abilities within an individual or group. For the purposes of this study, we examine two abilities, math and verbal (e.g., math > verbal, or verbal > math). Ability tilt on the SAT and ACT college entrance exams predict college majors and jobs in STEM and other fields (Coyle, Purcell, Snyder, & Richmond, 2014; Coyle, Snyder, & Richmond, 2015) among general population samples. These findings have been proposed to support investment theories, the idea that investment in one area such as math relates positively to complimentary math and STEM outcomes, but negatively to non-complimentary verbal or humanities outcomes (Coyle, 2018).

1.3. Ability tilt predicts real world outcomes decades later

Additionally, because intra-individual discrepancies in ability scores appear larger for gifted students in the right tail of cognitive abilities in comparison to general population counterparts (e.g., Lohman, Gambrell, & Lakin, 2008), male-female tilt differences could have more salience for the academic, occupational, and creative pursuits for high ability populations. For students within the top 1% of ability, students who scored higher on math relative to verbal ability at age 13 (on the SAT) tended toward STEM occupations decades later, whereas students who scored higher on verbal relative to math ability at age 13 tended toward humanities occupations (Lubinski, Webb, Morelock, & Benbow, 2001; Park et al., 2007). Such trends have also been found in even more select samples of the top 0.01% (1 in 10,000 for their age group), where the pattern of ability, not just the magnitude of ability is associated with subsequent educational, occupational, and creative accomplishments (Kell, Lubinski, & Benbow, 2013; Makel, Kell, Lubinski, Putallaz, & Benbow, 2016). Moreover, individuals who score well in both math and verbal domains have been found to be less likely to pursue careers in STEM fields than individuals who only score well in math (Wang, Eccles, & Kenny, 2013). This same research showed that females are more likely than males to score well in both math and verbal domains, thus giving females "more options" than males in terms of what fields they may choose to pursue. These links between early scores in ability tilt and subsequent pursuits suggest that in addition to ability level, ability tilt should be considered when investigating female STEM underrepresentation.

1.4. Ability differences across time and across cultural contexts

Examining whether ability tilt differences between males and females have remained stable or changed over time and whether ability tilt is similar in different cultural contexts is important to assess given the link between tilt and long-term STEM outcomes. One cultural context in which females may particularly face biases and barriers is in India. Males outnumber females beginning at birth (Sen, 1992, 2003) and literacy rates favor males (UNESCO, 2014). Indian female representation in STEM careers remains low (Leggon, McNeely, & Yoon, 2015), and females tend to have low representation among the prestigious Institutes of Technology (Rao, 2015), though some have argued that highly educated females may be doing well in terms of high level STEM and business careers (Hewlett & Rashid, 2011). Makel, Wai, et al. (2016) showed that patterns across male-female math ability differences in the extreme right tail replicate across the U.S. and India, however, it has not yet been established whether male-female ability tilt (math vs. verbal) differences in the extreme right tail replicate across cultural contexts.

2. Present study

The current study examined math-verbal ability tilt in the extreme right tail at different ability levels, whether tilt changed over time across the last 35 years, and whether the pattern of math-verbal ability tilt is similar or different in the U.S. and India. Our basic research questions (RQs) are as follows:

- RQ1 : Are there sex differences in ability tilt in the right tail of cognitive abilities?
- RQ2 : Do sex differences increase as ability tilt increases (distance between math and verbal scores increases)?
- RQ3 : Do sex differences in ability tilt increase as ability selectivity increases (top 5%, top 1%, top 0.01% of academic ability)?
- RQ4 : Have sex differences in ability tilt changed over time?
- RQ5 : Do sex differences in ability tilt vary as a function of measure and cultural context?

3. Method

3.1. Participants

Data from the U.S. and India came from the Duke University Talent Identification Program (Duke TIP). To qualify for participation in the Duke TIP talent search, students must score in the top 5% on a within grade standardized test either on a composite score or relevant subtest. Students then take an above-level test. In the U.S., the above-level test is either the SAT or ACT; for the younger elementary aged students, the above-level test is the ACT EXPLORE test (hereafter referred to as EXPLORE). The full samples were as follows: SAT, 1981–2015, N = 1,343,890 (female = 673,756, male = 670,134), ACT, 1990–2015, N = 589,453 (female = 286,523, male = 302,930), and EXPLORE, 1996–2015, N = 119,922 (female = 57,002, male = 62,920).

For the Duke TIP India talent search, the above-level test is the ASSET test by Educational Initiatives. It is not a college entrance exam, but like in the U.S., 7th standard (7th grade) Indian students qualify for talent search participation by scoring at or above the 95th percentile on their regular grade-level tests. Then, in India, students took the version of the ASSET test designed and normed for typical Indian students in the 9th or 10th grade. Thus, the ASSET serves as an above level test with sufficient headroom capacity to capture the full spectrum right-tail of test scores in comparison to grade-level tests. Males outnumbered females in India roughly 1.74 to 1 in Indian talent search participation. From 2011 to 2015, there were N = 7119 Duke TIP Indian talent search participants who took the ASSET (female = 2595, male = 4523; and one student whose data were not included whose sex was not reported).

3.2. Data analysis approach

In this paper, we examined math-verbal ability tilt across multiple measures in the U.S. (SAT, ACT, EXPLORE) and India (ASSET), across multiple ability levels (full sample, top 1%, top 0.01%), and across time (SAT: 1981 to 2015; ACT: 1990 to 2015; EXPLORE: 1996 to 2015; ASSET: one time point grouping, 2011 to 2015). Given that the purpose of the analysis was to determine the relationship between *math/verbal ability tilt* and two independent variables (*sex* and *year*), a regression model was used (Faraway, 2014).

3.2.1. Dependent variable

This study modeled a dependent variable: *tilt*. *Tilt* was calculated by subtracting a student's verbal score from their math score (*tilt* = math − verbal). For the SAT this was simply SAT-Mathematics minus SAT-Verbal. For the ACT and EXPLORE tests, verbal composites were computed as an average of the Reading and English subtests (hereafter referred to as ACT-Verbal and EXPLORE-Verbal). For the ASSET test, tilt was determined by taking the difference between the ASSET-Math and ASSET-English (hereafter referred to as ASSET-Verbal).

3.2.2. Independent variables

Two independent variables were assessed in the model: *sex* and *year*.

Sex was coded as a binary variable where 1 indicated male and 0 indicated female. The variable *year* was coded in the same manner for all tests but was centered on different years. For the SAT, *year* was centered on 1981 (e.g., the year 1981 was coded as 0, 1982 as 1, and 1983 as 2). For the ACT, *year* was centered on 1990. For the EXPLORE, *year* was centered on 1996. Finally, for the ASSET test, *year* was centered on 2011.

3.2.3. Regression

The purpose of this research was to ascertain if sex differences in ability tilt existed among students with high ability tilt on different standardized tests (SAT, ACT, EXPLORE, and ASSET) at different levels of ability (top 5%, top 1%, and top 0.01%). A linear regression model was used in the analysis to assess the relationship between *sex* and *tilt*. The following model was used:

$$Y\,(tilt)_{it} = \alpha_i + \beta_1(sex)_i + \beta_2(year)_t + \beta_3(sex)_i * (year)_t + \varepsilon_{it}$$

Where the tilt of a given student is predicted by their sex, the year, and the interactions between their sex and the year. Model fitting was done using *R 3.3.1* (R Core Team, 2013).

3.2.4. Assumptions

The assumption of homoscedasticity was assessed through examining the error plot. Normality was assessed through an examination of the qq-plot. Multivariable collinearity was assessed through examining the correlation matrix. Because the numerical distance between the two levels of a binary variable is statistically ill-defined (in this case male and female), the linearity of the relationship between the dependent variable and independent variables cannot be assessed in a precise manner (Faraway, 2014).

3.3. Method to determine cut scores for ability level

To determine cut scores for each ability level above the full sample (i.e., top 1%, top 0.01%), cutoffs were drawn from prior research and translated into current cut scores. In 1995 the SAT was recentered, so we used conversion tables to transform scores prior to 1995 so that they would be comparable to post-1995 scores (Educational Testing Service, 2016). Initial score benchmarks were drawn for the top 1% from Achter, Lubinski, Benbow, and Eftekhari-Sanjani (1999), and translated into current cut scores for the SAT (SAT-M 430+ or SAT-V 450+; female = 411,978, male = 448,787). The SAT percentiles for each of these cut scores in their respective distributions were used to find matching cut scores for the ACT (ACT-M 16+ or ACT-V 19+; female = 199,760, male = 220,055), EXPLORE (EXPLORE-M 16+ or EXPLORE-V 18+; female = 34,180, male = 40,655), and ASSET (ASSET-M 17+ or ASSET-V 51+; female = 1707, male = 3181). Initial score benchmarks were drawn for the top 0.01% from Lubinski et al. (2001), and translated into current cut scores for the SAT (SAT-M 700+ or SAT-V 700+; female = 1472, male = 3451). The SAT percentiles for each of these cut scores in their respective distributions was used to find matching cut scores for the ACT (ACT-M 27+ or ACT-V 32+; female = 1099, male = 2141), the EXPLORE (EXPLORE-M 25 or EXPLORE-V 25; female = 804, male = 1582), and ASSET (ASSET-M 37+ or ASSET-V 67+; female = 16, male = 21). Due to the extremely small samples at the top 0.01% level for ASSET, data were not used for comparison purposes.

3.4. Method used to display results graphically

One goal of the present study is to assess the nature of the relationship between *tilt* and *sex*. To do this, violin plots are used to graphically represent this relationship (Hintze & Nelson, 1998). As stated by Hintze and Nelson (1998), a violin plot extends the box plot proposed by Tukey (1977) by also showing the density of the distribution. A limitation of the box plot is that it cannot graphically display differences in distributions. A violin plot uses a density estimator to plot distributions. Further, this study uses a split violin plot wherein two densities are displayed side by side (male and female distributions). Creation of these plots was done through the use of *R 3.3.1* (R Core Team, 2013) and the R package *ggplot2* (Wickham, 2009).

4. Results

4.1. Assumptions

An examination of qq-plots suggested a roughly normal distribution of *tilt*. Of all test distributions examined, the distribution of *tilt* in ACT scores was the least normal, so further analyses were conducted within this sample specifically to assess whether linear regression was appropriate for all samples used in this study. Because the ACT sample is > 5000, a formal test for normality like the Shapiro-Wilk (Shapiro & Wilk, 1965) is inappropriate (Faraway, 2014). Faraway (2014) suggested that with a sufficiently large sample size, non-normality can be disregarded. For example, in the case of the ACT, the sample size reported for the analysis of *tilt* in ACT scores is sufficiently large (N = 589,409). In the case of homoscedasticity, the error plots were assessed (fitted values vs. residual plots). For all tests, there appeared to be a slight decrease in the error as the dependent variable increased but not to the extent that egregious heteroscedasticity could be inferred. Finally, for all tests, the correlation between the two independent variables (*sex* and *year*) was < 1%. This provides evidence that multivariable collinearity was not present. Given this evidence, a linear regression was appropriate for this analysis across all samples.

4.2. RQ1: Are there sex differences in ability tilt in the right tail of cognitive abilities?

Results from the regressions for the different tests provide evidence that *tilt* is associated with *sex*. All beta coefficients for *sex* were statistically significant in all models (see Tables 1 through 4). To interpret these coefficients, a reader should be reminded that the beta coefficient for *sex* must be considered with the coefficient for intercept in mind. A positive beta coefficient is associated with greater *tilt* for males toward math. Conversely a positive intercept coefficient is associated with greater *tilt* for females toward math whereas a negative coefficient

Table 1
Sex differences in ability tilt on the SAT test 9 (n = 1,343,890).

	Beta[a]	*SE*	*T*	*p*	LL[b]	UL[b]
Top 5% (n = 1343,876[c] [1.01])[d]						
Intercept	−11.35	0.22	−50.89	< 0.001	−11.78	−10.92
Year	0.59 (0.07)	0.01	54.87	< 0.001	0.57	0.60
Sex	17.58 (0.11)	0.32	55.05	< 0.001	16.95	18.21
Year * Sex	0.32 (0.04)	0.02	21.14	< 0.001	0.29	0.34
Top 1% (n = 863,146 [1.09])[d]						
Intercept	1.68	0.13	12.65	< 0.001	1.43	1.93
Year	< 0.01 (0.02)	< 0.01	13.94	< 0.001	< 0.01	< 0.01
Sex	29.73 (0.17)	0.18	161.57	< 0.001	29.37	30.08
Year * Sex	< 0.01 (< 0.01)	< 0.01	0.44	0.662	−0.01	0.01
Top 0.01% (n = 5019 [2.36])[d]						
Intercept	28.82	3.79	7.61	< 0.001	21.39	36.24
Year	0.03 (0.13)	< 0.01	4.51	< 0.001	0.02	0.04
Sex	97.57 (0.29)	4.56	21.57	< 0.001	88.23	106.11
Year * Sex	−0.02 (−0.08)	< 0.01	2.41	0.016	−0.03	0.01

[a] The first number indicates the beta coefficient, the number in parentheses is the standardized beta coefficient.
[b] The LL and the UL indicate the lower limit and the upper limit of the 95% confidence interval.
[c] 14 students selected a category other than male or female for their sex.
[d] The total number of male and female students; the number in brackets indicates the ratio of males to females.

Table 2
Sex differences in ability tilt on the ACT test (n = 589,453).

	Beta[a]	*SE*	*T*	*p*	LL[b]	UL[b]
Top 5% (n = 589,453 [1.05])[c]						
Intercept	−1.93	0.01	260.48	< 0.001	−1.95	−1.91
Year	< 0.01 (0.06)	< 0.01	4.74	< 0.001	0.01	0.01
Sex	−0.28 (0.17)	0.02	11.45	< 0.001	−0.32	−0.24
Year * Sex	0.09 (0.04)	< 0.01	89.35	< 0.001	0.08	0.09
Top 1% (n = 429,746 [1.08])[c]						
Intercept	−2.52	0.01	270.30	< 0.001	−2.55	−2.50
Year	< 0.01 (< 0.01)	< 0.01	1.48	0.139	0.01	0.01
Sex	0.19 (0.02)	0.03	5.88	< 0.001	0.13	0.25
Year * Sex	0.08 (0.01)	< 0.01	63.61	< 0.001	0.01	0.01
Top 0.01% (n = 3697 [1.67])[c]						
Intercept	−7.21	0.20	36.80	< 0.001	−7.58	−6.80
Year	< 0.01 (0.06)	< 0.01	3.91	< 0.001	0.01	0.01
Sex	8.15 (0.49)	0.75	10.82	< 0.001	6.68	9.62
Year * Sex	−0.02 (< 0.01)	0.03	0.80	0.426	−0.08	0.04

[a] The first number indicates the beta coefficient, the number in parentheses is the standardized beta coefficient.
[b] The LL and the UL indicate the lower limit and the upper limit of the 95% confidence interval. Further, confidence intervals are rounded to 0.01 when the value is between 0.0049 and 0.
[c] The total number of male and female students; the number in brackets indicates the ratio of males to females.

Table 3
Sex differences in ability tilt on the EXPLORE test (n = 119,922).

	Beta[a]	*SE*	*T*	*p*	LL[b]	UL[b]
Top 5% (n = 119,922 [1.11])[c]						
Intercept	−2.33	0.03	−77.85	< 0.001	−2.39	−2.27
Year	0.04 (0.06)	< 0.01	15.64	< 0.001	0.04	0.04
Sex	1.13 (0.17)	0.04	27.53	< 0.001	1.05	1.21
Year * Sex	0.02 (0.04)	< 0.01	6.82	< 0.001	0.02	0.03
Top 1% (n = 79,148 [1.16])[c]						
Intercept	−2.74	0.04	−65.41	< 0.001	−2.82	−2.66
Year	0.03 (0.05)	< 0.01	9.11	< 0.001	0.03	0.03
Sex	1.48 (0.19)	0.06	25.79	< 0.001	1.36	1.60
Year * Sex	0.03 (0.05)	< 0.01	6.53	< 0.001	0.03	0.03
Top 0.01% (n = 4410 [1.20])[c]						
Intercept	−3.85	0.32	−12.14	< 0.001	−4.48	−3.23
Year	0.01 (0.01)	0.02	0.25	0.804	−0.03	0.05
Sex	3.69 (0.35)	0.42	8.85	< 0.001	2.87	4.51
Year * Sex	< 0.01 (< 0.01)	0.03	−0.16	0.871	−0.05	0.06

[a] The first number indicates the beta coefficient, the number in parentheses is the standardized beta coefficient.
[b] The LL and the UL indicate the lower limit and the upper limit of the 95% confidence interval.
[c] The total number of male and female students; the number in brackets indicates the ratio of males to females.

provides evidence for a greater *tilt* toward verbal. For example, as shown in Table 1, in the SAT (top 0.01%), the beta coefficient for *tilt* is 97.57 (*SE* = 4.52) and the intercept coefficient was 28.82 (*SE* = 3.79). This suggests that males and females in the top 0.01% are more tilted toward mathematics than verbal but that males are, on average, 97.57 points more tilted than females.

4.3. RQ2: Do sex differences increase as ability tilt increases?

The violin plots in Figs. 1 through 4 show that sex differences increase as ability tilt increases. Further, statistical evidence is provided through an examination of the regression coefficients. The regression coefficients for the top 0.01% for each testing measure are outside of the confidence intervals of the top 5% and top 1%. This provides strong evidence that the regression coefficients are statistically different and greater. For example, the confidence interval for the beta coefficient for the math tilt for *sex* in the regression for the top 0.01% of SAT takers is 88.23 to 106.11. Conversely, the confidence interval for the top 1% is 29.36 to 30.06. Clearly these confidence intervals do not overlap.

Table 4
Sex differences in ability tilt on the ASSET test in India (n = 7119).

	Beta[a]	*SE*	*T*	*p*	LL[b]	UL[b]
Top 5% (n = 7,116[c] [1.74])[d]						
Intercept	−28.44	0.35	−81.46	< 0.001	−29.13	−27.75
Year	−2.03 (−0.27)	0.13	−15.29	< 0.001	−2.28	−1.78
Sex	6.42 (0.30)	0.44	14.60	< 0.001	5.56	7.28
Year * Sex	−0.18 (−0.02)	0.17	−1.08	0.280	−0.51	0.15
Top 1% (n = 4887 [1.86])[d]						
Intercept	−28.54	0.46	−62.34	< 0.001	−29.44	−27.64
Year	−2.57 (−0.33)	0.17	−15.43	< 0.001	−2.90	−2.24
Sex	7.58 (0.32)	0.56	13.50	< 0.001	6.48	8.68
Year * Sex	−0.16 (−0.02)	0.21	−0.75	0.453	−0.57	0.25

[a] The first number indicates the beta coefficient, the number in parentheses is the standardized beta coefficient.
[b] The LL and the UL indicate the lower limit and the upper limit of the 95% confidence interval.
[c] 3 students selected a category other than male or female for their sex.
[d] The total number of male and female students; the number in brackets indicates the ratio of males to females.

4.4. RQ3: Do sex differences in ability tilt increase as ability selectivity increases?

As shown in Tables 1 through 4, in all cases (save one), the beta coefficient for *sex* increased as selectivity increased. This suggests that, as ability selectivity increases (top 5%, top 1%, top 0.01%), males were increasingly tilted toward mathematics than females. This provides evidence that the presence of ability tilt in a population is associated with ability selectivity. This is illustrated in the split violin plots in Figs. 1 through 4; the shapes of the violin plots change as ability selectivity increases. In particular, the distribution of tilt shifts more extreme. These plots provide a visual confirmation that ability tilt increases as selectivity increases.

The only case where the coefficient for sex did not increase as ability selectivity increased was for the ACT from the top 5% to top 1%. An examination of the violin plot does suggest a difference in the distribution of tilt. A further examination of coefficients suggests that the sex differences are explained through the interaction of sex and year given the relatively larger coefficient of 0.08.

4.5. RQ4: Have sex differences in ability tilt changed over time?

The effect of *year* on ability tilt is dependent upon ability selectivity. For students in the top 5%, the effect of *year* was a significant predictor as well as a practically significant predictor. For example, in the case of the SAT, students in the top 5% of ability had a coefficient of 0.59 (*SE* = 0.01). After 30 years, this means that tilt would increase by roughly 17.7 points overall. For students in the top 1% and top 0.01%, there are significant coefficients but they are not likely practically significant. Again, using the SAT as an example, the coefficient was 0.03 (*SE* < 0.01). In the same 30 year time frame, this would equate to less than a 1 point increase in tilt. Thus, there is evidence that sex differences in ability tilt meaningfully changed in our sample between 1980 and 2016, but not for students in the top 1% and top 0.01%.

4.6. RQ5: Do sex differences in ability tilt vary as a function of measure and cultural context?

The relationship between *sex* and *tilt* does not vary as a function of measure when the results for the SAT, ACT, and EXPLORE are examined (see Tables 1 through 4). In all cases, there was a pattern of males

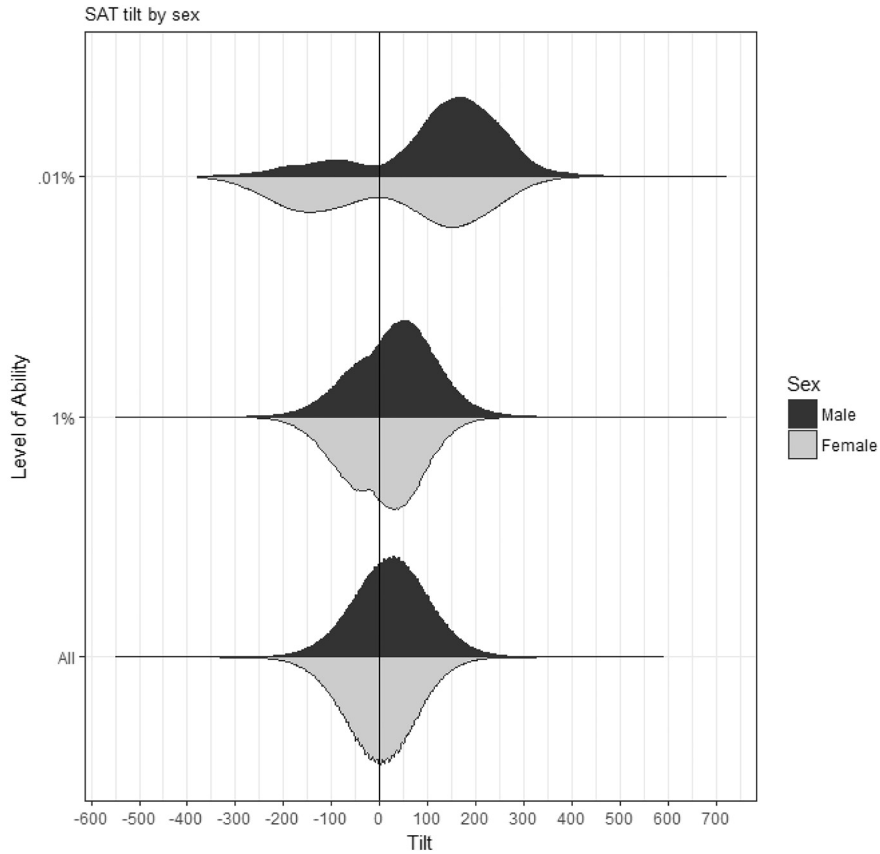

Fig. 1. Split violin plots of the distribution of academic tilt by level of ability and sex. Tilt is the student's SAT math score minus their SAT verbal score. If the value on the x axis is negative this indicates a tilt favoring verbal ability, whereas if the value on the x axis is positive this indicates a tilt favoring math ability.

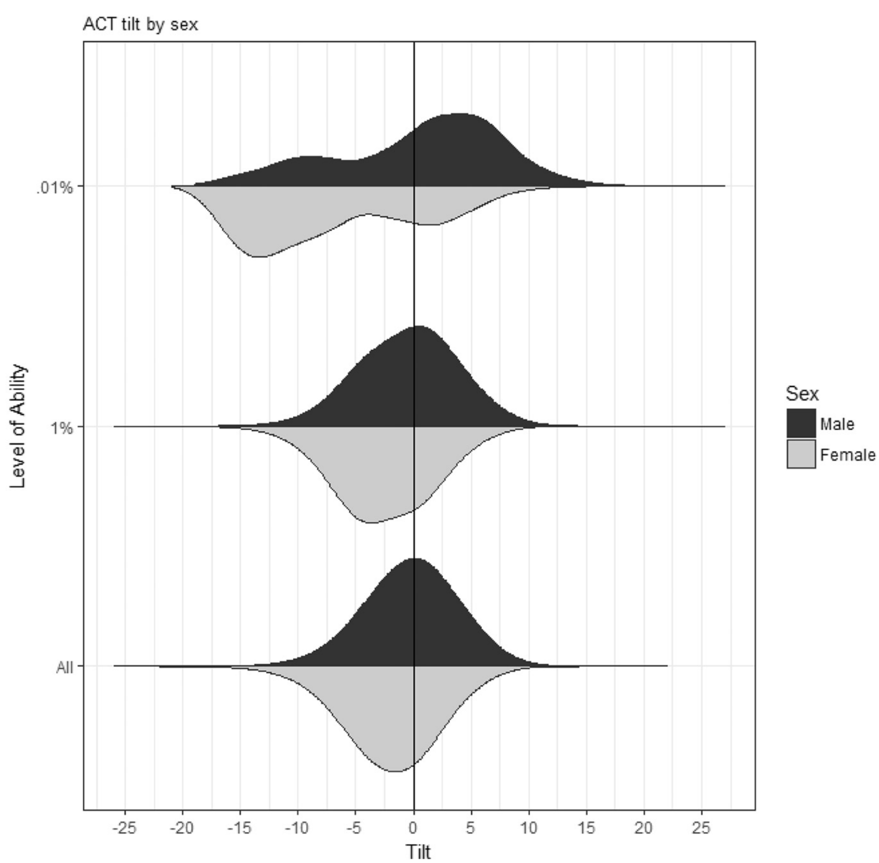

Fig. 2. Split violin plots of the distribution of academic tilt by level of ability and sex. Tilt is the student's ACT-Math score minus their ACT-Verbal score. ACT-Verbal was an average of the Reading and English subtests. If the value on the x axis is negative this indicates a tilt favoring verbal ability, whereas if the value on the x axis is positive this indicates a tilt favoring math ability.

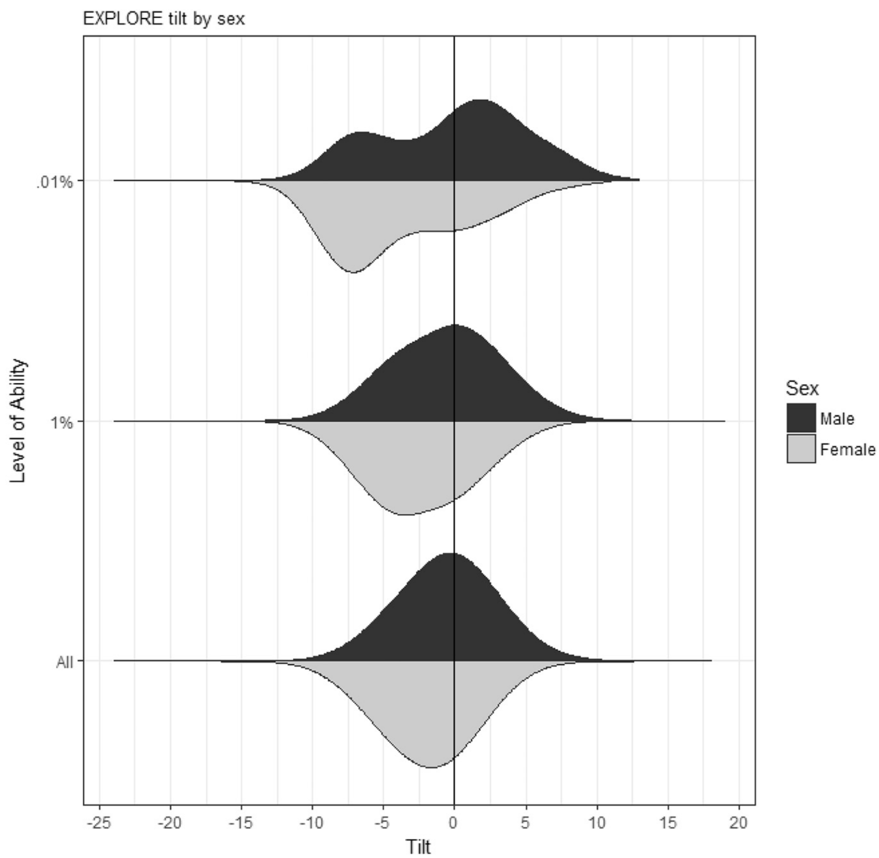

Fig. 3. Split violin plots of the distribution of academic tilt by level of ability and sex. Tilt is the student's EXPLORE-Math score minus their EXPLORE-Verbal score. Explore-Verbal was an average of the Reading and English subtests. If the value on the x axis is negative this indicates a tilt favoring verbal ability, whereas if the value on the x axis is positive this indicates a tilt favoring math ability.

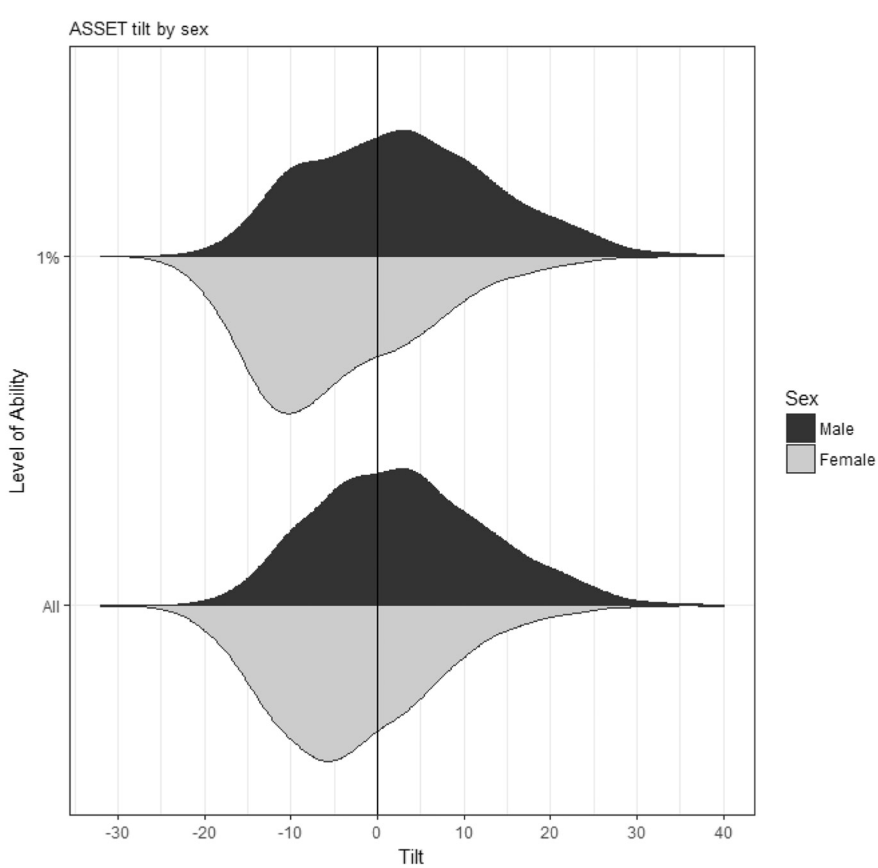

Fig. 4. Split violin plots of the distribution of academic tilt by level of ability and sex. Tilt is the student's ASSET-Math score minus their ASSET-Verbal score (i.e., the ASSET-English test). The ASSET test is a sample from India. If the value on the x axis is negative this indicates a tilt favoring verbal ability, whereas if the value on the x axis is positive this indicates a tilt favoring math ability. Due to the extremely small samples at the top 0.01% level for ASSET, data were not used for comparison purposes.

having higher *tilt* than females. Further, there was a pattern of increasing *tilt* as selectivity increased. This suggests that the presence of *tilt* is not associated with the measure used. For the ASSET test in India, the pattern of *tilt* seen in the SAT, ACT, and EXPLORE test was observed. This suggested that *tilt* is not associated with cultural context. A visualization of these findings can be seen in the split violin plots in Figs. 1 through 4.

5. Discussion

The present study builds upon prior work examining sex differences in the right tail of cognitive abilities (Makel, Wai, et al., 2016; Wai et al., 2010) as well as in ability pattern (Kell et al., 2013; Makel, Kell, et al., 2016; Park et al., 2007) by examining the additional role of sex differences in the right tail of math-verbal cognitive ability tilt in the U.S. and India. Overall, it appears that there are sex differences in ability tilt and such differences increase as ability tilt increases. As ability selectivity increased, tilt also became more extreme. In general, there was no evidence that ability tilt changed over time across multiple tests, something found by prior researchers using different samples and measures across a different time span (Hedges & Nowell, 1995). Findings broadly replicated across the U.S. and India.

Given that in the U.S. top 0.01% on math ability on the SAT and ACT there are currently about 2.5 males for every female (Makel, Wai, et al., 2016), and that we found that there are more males than females with positive math tilt and more females with positive verbal tilt in the top 0.01% (ACT, SAT, EXPLORE), ability tilt favoring males on math ability in the extreme right tail of cognitive abilities may play a role in the underrepresentation of women in STEM. Also, moving from the top 5% to the top 1%, to the top 0.01% showed that as ability selectivity increased, math-verbal tilt ratios favoring sex differences also increased.

Prior research showed that even within the extreme right tail of abilities, more math ability (Wai et al., 2005) and math > verbal ability tilt (Coyle et al., 2014, 2015; Park et al., 2007) matters for STEM major choice and eventual high level STEM careers. The current findings contribute to the empirical evidence of relevant factors in the discussion surrounding female representation in high level STEM careers (Ceci et al., 2014; Halpern et al., 2007) by showing that in addition to math ability, math > verbal ability tilt has been fairly unchanged across the last 35 years. Data from this paper, when connected to this body of prior work, suggests that math abilities (in relation to verbal abilities) likely remain a factor in contributing to the explanation of the underrepresentation of women in high level STEM careers.

5.1. Limitations and future directions

Multiple measures were utilized in this study to determine whether the broad pattern of ability tilt and changes over time were measure and sample specific or potentially more robust through replication pattern. Broadly, in the right tail of cognitive abilities, it appears that sex differences in ability tilt exist, and that over time, ability tilt is fairly stable. However, it's unclear whether findings on measures over time might have been caused by factors such as potential ceiling effects on measures (e.g., Wai, Putallaz, & Makel, 2012), the removal of certain items to reduce gender differences (e.g., Loewen, Rosser, & Katzman, 1988), or the differing content across the measures examined given revamping of tests (e.g., Kobrin & Melican, 2007; Lohman & Lakin, 2009).

Because the U.S. and Indian samples in this study had no overlap in the tests they took, we could not develop a direct concordance across these samples, limiting what we can state about the degree of tilt differences across these two cultural contexts (Makel, Wai, et al., 2016). The broad pattern across the U.S. and India regarding math-verbal ability tilt replicated in pattern, though differed in degree to some extent. Compared to the U.S. samples, the Indian sample was relatively smaller, though the general patterns appeared to replicate across these two contexts.

It should be noted it is likely that the magnitude of the selectivity level moderator (top 5%, top 1%, top 0.01%) is underestimated. In the analysis, the ratio of male to females was skewed toward males. This skew increased as selectivity increased. This unbalanced design can lead to the true effect being greater than what is reported. In other words, it is possible that the magnitude of tilt is greater than what is reported in this research.

Further, it is important to note that in an examination of scores in the top 0.01% of test takers that underlying statistical assumptions of hypothesis testing are not always ideal. Extreme value theory is the statistical theory governing the behavior of distributions at the extreme ends of the normal distribution (de Haan & Ferreira, 2007). Given the population, the normal distribution used in the analysis is likely to misestimate the true effect. In essence, it is possible that a generalized pareto distribution with a finite maximum is a more appropriate distribution to use in analysis (de Haan & Ferreira, 2007).

To mirror prior investigations of ability tilt in gifted (e.g., Lubinski et al., 2001; Park et al., 2007) and general population (e.g., Coyle et al., 2014, 2015) samples, we compared math to verbal tilt only. The ACT, EXPLORE, and ASSET all include science and writing measures, and none of these measures include spatial abilities (e.g., Kell, Lubinski, Benbow, & Steiger, 2013; Wai, Lubinski, & Benbow, 2009). Measures including spatial abilities may be worth investigating in future research given their potential link to later STEM outcomes.

5.2. Conclusion

Overall, we found that ability tilt did not change broadly. Prior research shows that within general population samples and right tail ability samples more math ability and math > verbal ability tilt in adolescence is related to the earning of STEM PhDs, STEM publications, STEM patents, and ending up in a STEM occupation many years later and more verbal ability and verbal > math ability tilt in adolescence is related to the earning of verbal and humanities outcomes many years later. Our findings in this study confirm adolescent sex differences in ability tilt in the right tail broadly. Such male-female ability tilt differences should therefore be taken into consideration when examining the underrepresentation of women in math or STEM careers and men in verbal or humanities careers. When combined with research on sex differences in interests (Su et al., 2009), these ability tilt patterns may become more relevant. Such trends should be monitored in the future.

References

Achter, J. A., Lubinski, D., Benbow, C. P., & Eftekhari-Sanjani, H. (1999). Assessing vocational preferences among gifted adolescents adds incremental validity to abilities: A discriminant analysis of educational outcomes over a 10-year interval. *Journal of Educational Psychology, 91*, 777–786. http://dx.doi.org/10.1037/0022-0663.91.4.777.

Benbow, C. P., & Stanley, J. C. (1980). Sex differences in mathematical ability: Fact or artifact? *Science, 210*, 1262–1264. http://dx.doi.org/10.1126/science.7434028.

Benbow, C. P., & Stanley, J. C. (1983). Sex differences in mathematical reasoning ability: More facts. *Science, 222*, 1029–1031.

Borkenau, P., McCrae, R. R., & Terracciano, A. (2013). Do men vary more than women in personality? A study in 51 cultures. *Journal of Research in Personality, 47*, 135–144. http://dx.doi.org/10.1016/j.jrp.2012.12.001.

Ceci, S. J., Ginther, D. K., Kahn, S., & Williams, W. M. (2014). Women in academic science: A changing landscape. *Psychological Science in the Public Interest, 15*, 75–141. http://dx.doi.org/10.1177/1529100614541236.

Ceci, S. J., & Williams, W. M. (2010). *The mathematics of sex: How biology and society conspire to limit talented women and girls.* New York, NY: Oxford University Press.

Core Team, R. (2013). *R: A language and environment for statistical computing.* Vienna, Austria: R Foundation for Statistical Computing. http://www.R-project.org/.

Coyle, T. R. (2018). Non-*g* residuals of group factors predict ability tilt, college majors, and jobs: A non-*g* nexus. *Intelligence, 67*, 19–25. http://dx.doi.org/10.1016/j.intell.2017.12.003.

Coyle, T. R., Purcell, J. M., Snyder, A. C., & Richmond, M. C. (2014). Ability tilt on the SAT and ACT predicts specific abilities and college majors. *Intelligence, 46*, 18–24. http://dx.doi.org/10.1016/j.intell.2014.04.008.

Coyle, T. R., Snyder, A. C., & Richmond, M. C. (2015). Sex differences in ability tilt: Support for investment theory. *Intelligence, 50*, 209–220. http://dx.doi.org/10.1016/j.intell.2015.04.012.

Educational Testing Service (2016). Equivalence tables: Convert individual and mean scores from the original scale to the recentered scale. Retrieved from: https://research.collegeboard.org/programs/sat/data/equivalence

Faraway, J. J. (2014). *Linear models with R.* Boca Raton, FL: CRC press.

de Haan, L., & Ferreira, A. (2007). *Extreme value theory: An introduction.* New York City, NY: Springer Science & Business Media, LLC.

Halpern, D. F., Benbow, C. P., Geary, D., Gur, R., Hyde, J. S., & Gernsbacher, M. A. (2007). The science of sex differences in science and mathematics. *Psychological Science in the Public Interest, 8*, 1–51. http://dx.doi.org/10.1111/j.1529-1006.2007.00032.x.

Hedges, L. V., & Nowell, A. (1995). Sex differences in mental test scores, variability, and numbers of high-scoring individuals. *Science, 269*, 41–45. http://dx.doi.org/10.1126/science.7604277.

Hewlett, S. A., & Rashid, R. (2011). *Winning the war for talent in emerging markets: Why women are the solution.* Boston, MA: Harvard Business Review Press.

Hintze, J. L., & Nelson, R. D. (1998). Violin plots: A box plot-density trace synergism. *The American Statistician, 52*(2), 181–184.

Kell, H. J., Lubinski, D., & Benbow, C. P. (2013). Who rises to the top? Early indicators. *Psychological Science, 24*, 648–659.

Kell, H. J., Lubinski, D., Benbow, C. P., & Steiger, J. H. (2013). Creativity and technical innovation: Spatial ability's unique role. *Psychological Science, 24*, 1831–1836. http://dx.doi.org/10.1177/0956797613478615.

Kobrin, J. L., & Melican, G. J. (January, 2007). Comparability of scores on the new and prior versions of the SAT reasoning test. *College Board Research notes (RN-31)*.

Leggon, C., McNeely, C. L., & Yoon, J. (2015). Advancing women in science: Policies for progress. In W. Pearson, L. M. Frehill, & C. L. McNeely (Eds.). *Advancing women in science: An international perspective* (pp. 307–340). NY: Springer.

Lehre, A., Lehre, K. P., Laake, P., & Danbolt, N. C. (2008). Greater intrasex phenotype variability in males than in females is a fundamental aspect of the gender differences in humans. *Developmental Psychobiology, 51*, 198–206. http://dx.doi.org/10.1002/dev.20358.

Loewen, J. W., Rosser, P., & Katzman, J. (1988). Gender bias in SAT items. *Paper presented at the annual meeting of the American Educational Research Association. New Orleans, LA*.

Lohman, D. F., Gambrell, J., & Lakin, J. (2008). The commonality of extreme discrepancies in the ability profiles of academically gifted students. *Psychology Science Quarterly, 50*, 269–282.

Lohman, D. F., & Lakin, J. (2009). Consistencies in sex differences on the cognitive abilities test across countries, grades, test forms, and cohorts. *British Journal of Educational Psychology, 79*, 389–407. http://dx.doi.org/10.1348/000709908X354609.

Lubinski, D., Webb, R. M., Morelock, M. J., & Benbow, C. P. (2001). Top 1 in 10,000: A 10-year follow-up of the profoundly gifted. *Journal of Applied Psychology, 86*, 718–729. http://dx.doi.org/10.1037/0021-9010.86.4.718.

Makel, M. C., Kell, H. J., Lubinski, D., Putallaz, M., & Benbow, C. P. (2016). When lightning strikes twice: Profoundly gifted, profoundly accomplished. *Psychological Science, 27*, 1004–1018. http://dx.doi.org/10.1177/0956797616644735.

Makel, M. C., Wai, J., Peairs, K., & Putallaz, M. (2016). Sex differences in the right tail of cognitive abilities: An update and cross cultural extension. *Intelligence, 59*, 8–15. http://dx.doi.org/10.1016/j.intell.2016.09.003.

Miller, D. I., & Wai, J. (2015). The bachelor's to Ph.D. STEM pipeline no longer leaks more women than men: A 30-year analysis. *Frontiers in Psychology: Developmental, 6*(37), http://dx.doi.org/10.3389/fpsyg.2015.00037.

Moss-Racusin, C. A., Dovidio, J. F., Brescoll, V. L., Graham, M. J., & Handelsman, J. (2012). Science faculty's subtle gender biases favor male students. *Proceedings of the National Academy of Sciences, 109*, 16474–16479. http://dx.doi.org/10.1073/pnas.1211286109.

National Academy of Sciences (2010). *Rising above the gathering storm, revisited: Rapidly approaching category 5.* Washington, D.C.: National Academy of Sciences.

Park, G., Lubinski, D., & Benbow, C. P. (2007). Contrasting intellectual patterns predict creativity in the arts and sciences: Tracking intellectually precocious youth over 25 years. *Psychological Science, 18*, 948–952. http://dx.doi.org/10.1111/j.1467-9280.2007.02007.x.

Rao, Y. (August 20, 2015). More girls now make it to IIT. *Times of India*. Retrieved from http://timesofindia.indiatimes.com/home/education/news/More-girls-now-make-it-to-IIT/articleshow/48551282.cms.

Ritchie, S. J., Cox, S. R., Shen, X., Lombardo, M. V., Reus, L. M., Alloza, C., ... Deary, I. J. (1992). *Sex differences in the adult human brain: Evidence from 5,216 UK Biobank participants.* Preprint first posted online April 4, 2017 https://doi.org/10.1101/123729.

Sen, A. (1992). Missing women. *British Medical Journal, 304*, 304–305. http://dx.doi.org/10.1136/bmj.304.6827.587.

Sen, A. (2003). Missing women-revisited. *British Medical Journal, 327*, 1297–1298. http://dx.doi.org/10.1136/bmj.327.7427.1297.

Shapiro, S. S., & Wilk, M. B. (1965). An analysis of variance test for normality (complete samples). *Biometrika, 52*(3/4), 591–611.

Su, R., Rounds, J., & Armstrong, P. I. (2009). Men and things, women and people: A meta-analysis of sex differences in interests. *Psychological Bulletin, 135*, 859–884. http://dx.doi.org/10.1037/a0017364.

Tukey, J. W. (1977). *Exploratory data analysis.* Reading, MA: Addison-Wesley.

UNESCO (2014). Country profiles: India. Retrieved from: http://uis.unesco.org/country/in

Wai, J., Cacchio, M., Putallaz, M., & Makel, M. C. (2010). Sex differences in the right tail of cognitive abilities: A 30-year examination. *Intelligence, 38*, 412–423. http://dx.doi.org/10.1016/j.intell.2010.04.006.

Wai, J., Lubinski, D., & Benbow, C. P. (2005). Creativity and occupational accomplishments among intellectually precocious youth: An age 13 to age 33 longitudinal study. *Journal of Educational Psychology, 97*, 484–492. http://dx.doi.org/10.1037/0022-0663.97.3.484.

Wai, J., Lubinski, D., & Benbow, C. P. (2009). Spatial ability for STEM domains: Aligning over fifty years of cumulative psychological knowledge solidifies its importance. *Journal of Educational Psychology, 101*, 817–835. http://dx.doi.org/10.1037/a0016127.

Wai, J., Putallaz, M., & Makel, M. C. (2012). Studying intellectual outliers: Are there sex differences, and are the smart getting smarter? *Current Directions in Psychological Science, 21*, 382–390. http://dx.doi.org/10.1177/0963721412455052.

Wang, M. T., Eccles, J. S., & Kenny, S. (2013). Not lack of ability but more choice: Individual and gender differences in choice of careers in science, technology, engineering, and mathematics. *Psychological Science, 24*, 770–775. http://dx.doi.org/10.1177/0956797612458937.

Wickham, H. (2009). *ggplot2: Elegant graphics for data analysis.* New York, NY: Springer-Verlag.