Intelligence 38 (2010) 412–423

Contents lists available at ScienceDirect

Intelligence

Sex differences in the right tail of cognitive abilities: A 30 year examination

Jonathan Wai*, Megan Cacchio, Martha Putallaz, Matthew C. Makel

Duke University, Talent Identification Program, 1121 West Main Street, Durham, NC 27701, United States

ARTICLE INFO

Article history:
Received 3 February 2010
Received in revised form 21 April 2010
Accepted 26 April 2010
Available online 9 June 2010

Keywords:
Sex differences
Cognitive abilities
STEM
Sociocultural factors
Historical examination

ABSTRACT

One factor in the debate surrounding the underrepresentation of women in science technology, engineering and mathematics (STEM) involves male–female mathematical ability differences in the extreme right tail (top 1% in ability). The present study provides male–female ability ratios from over 1.6 million 7th grade students in the right tail (top 5% in ability) across 30 years (1981–2010) using multiple measures of math, verbal, and writing ability and science reasoning from the SAT and ACT. Male–female ratios in mathematical reasoning are substantially lower than 30 years ago, but have been stable over the last 20 years and still favor males. Over the last two decades males showed a stable or slightly increasing advantage in science reasoning. However, more females scored in the extreme right tail of verbal reasoning and writing ability tests. The potential role of sociocultural factors on changes in the male–female ability ratios is discussed and the introduction of science reasoning as a potential new factor in the debate is proposed. The implications of continued sex differences in math and science reasoning is discussed within the context of the many important interlocking factors surrounding the debate on the underrepresentation of women in STEM.

1. Introduction

The underrepresentation of women pursuing high level careers in STEM has been researched and discussed for decades. Scholars have approached this topic from widely differing viewpoints. To provide a context for the present study, we first highlight some critical aspects of the issues involved in research on sex differences in high level STEM achievement, briefly review the relevant major studies, and then discuss what research evidence has been missing.

1.1. Some critical aspects of sex differences in high level STEM achievement

One prominent factor fueling the discussion has been apparent male–female mathematical ability differences, particularly greater male representation (Benbow, 1988; Benbow & Stanley, 1980, 1983; Hedges & Nowell, 1995; Johnson, Carothers, & Deary, 2008; Lohman & Lakin, 2009; Strand, Deary, & Smith, 2006), and the potential ramifications such differences might have for female underrepresentation.

Scholars who have focused on average test scores of mathematical ability advocate male–female similarities (Hyde, 2005; Hyde & Linn, 2006; Hyde, Lindberg, Linn, Ellis, & Williams, 2008). Authors who have focused on the male–female ratio of extremely high scorers on the SAT-Mathematics (SAT-M) in the seventh grade emphasize sex differences (Benbow, 1988; Benbow & Stanley, 1980, 1983). In terms of potential implications, we believe that the focus on the right tail is more relevant in that this is the intellectual or human capital pool from which the majority of high level STEM achievers are likely to be drawn from (Lubinski & Benbow, 2006).

Giving the SAT-M in the 7th grade allows individual differences in the extreme right tail of the distribution (i.e., the top 1% which includes over one third of the ability range) to be captured adequately (Benbow, 1988; Benbow & Stanley, 1980, 1983). For clarity, we use the terms right tail to refer to the top 5% in ability and extreme right tail to refer to the top 1% in ability. At more refined levels of ability, we simply denote the level to which we are referring.

* Corresponding author.
E-mail address: jon.wai@duke.edu (J. Wai).

0160-2896/$ – see front matter
doi:10.1016/j.intell.2010.04.006

Much of the debate surrounding the dearth of women in STEM was sparked by Benbow and Stanley (1980, 1983) who showed that of 40,000 students, the male–female ratio in the early 1980s on the SAT-M was 2.1 to 1 for scores ≥ 500 (top 0.5%), 4.1 to 1 for scores ≥ 600, and a remarkable 13 to 1 for those scoring ≥ 700 (top 0.01%).

1.2. A brief history of research on sex differences in cognitive abilities

There has been a substantial amount of discussion and research generated in the three decades since the 13 to 1 ratio was first published in *Science*. There have been periodic reports that this ratio has declined to around 3 to 1 (e.g., Spelke, 2005a), which along with research appearing in *Science* underscoring the lack of male–female differences at the mean (Hyde & Linn, 2006; Hyde et al., 2008), and anecdotal accounts of women winning the Nobel Prizes in STEM, has led to the impression that sex differences in math ability in the extreme right tail no longer exist or at least are no longer meaningful. This perception has developed despite the fact that no large empirical dataset has been examined in contemporary times confirming commensurate representation in math ability at the extreme right tail.

Although sex differences research has a rich history (Halpern, 2000; Johnson et al., 2008), foundational work by Hedges and Nowell (1995) set the stage for male–female population level ability research by examining sex ratios on multiple measures of ability (e.g., math, verbal and spatial) across six nationally representative datasets that spanned 1960–1992. They showed that in U.S. populations, males were more highly represented within the top 1 to 5%, stating that "The achievement of fair representation of women in science will be much more difficult if there are only one-half to one-seventh as many women as men who excel in the relevant abilities" (Hedges & Nowell, 1995, p. 269).

More recently, Strand et al. (2006) examined a nationally representative sample in the United Kingdom on multiple abilities, including a group in the top 4%, which spanned 2001–2003, finding a male advantage in math and non-verbal reasoning, and a female advantage in verbal reasoning. Lohman and Lakin (2009) linked current population level data in the U.S. with Strand et al. (2006) in the top 4%, showing that findings are consistent on multiple abilities in both the U.S. and the U.K. Additionally, Hyde et al. (2008) conducted meta-analytic research demonstrating male–female similarities at the center of the ability distribution on contemporary U.S. state assessments, but also examined male–female ratios in the right tail and extreme right tail of the distribution.

Across all these studies (Hedges & Nowell, 1995; Hyde et al., 2008; Lohman & Lakin, 2009; Strand et al., 2006), the analyses are not precisely aligned, but they are based on very large representative samples and they all demonstrate substantial sex differences in the right tail. However, other than Benbow and Stanley (1980, 1983), no studies have used measures necessary to capture individual differences in ability adequately or have explicitly sampled from the population that composes the extreme right tail. For example, Hyde et al. (2008) mention that in the assessments they used, there were not enough higher level items to capture differences in complex problem solving, illustrating that such measures had a ceiling that made it impossible to adequately detect male–female differences in the extreme right tail. It would seem necessary, then, to investigate whether male–female ability differences still exist today within a sample that is comparable to, yet independent of, Benbow and Stanley (1980, 1983) using tests that do not have a ceiling that prevents differences from being adequately captured.

1.3. Research that has been missing

Recent discussions of sex differences in the extreme right tail of mathematical ability (Ceci et al., 2009; Halpern et al., 2007; Hyde & Linn, 2006) have mentioned that the 13 to 1 ratio given by Benbow and Stanley (1983) has declined over time to 2.8 to 1. All cite the same personal communication (Monastersky, 2005) from Julian Stanley, an author on the original studies (Benbow & Stanley, 1980, 1983), who supplied this updated ratio but not a broader examination of the data used to generate it. Surprisingly, the debate has not relied on an historical examination over time of the male–female ratio in the extreme right tail.[1] Ceci and Williams (2010, p. 152), in their synthesis of the best research evidence surrounding the debate on the underrepresentation of women in STEM, note that the male–female ratios given by Benbow and Stanley (1980, 1983) are quite old and thus logically inquire: "Would the same results be found with today's children, who did not have to swim against the tide of sex biases and lack of female role models?" We are now able to address this query. Recently, Halpern et al. (2007, p. 13) stated that "There are no studies exploring the reasons for the decline." To our knowledge there have also been no studies exploring *when* the widely reported decline began, to *what degree* this decline has occurred, and *whether* this decline is continuing.

Two other components are also missing to examine this issue fully. First, an examination of the male–female ratio using an additional measure of mathematical ability that captures individual differences adequately is necessary to ensure that results are not specific to the SAT-M. For example, Spelke (2005b) noted that the results from Benbow and Stanley (1980, 1983) were limited to this one measure, saying that what was needed included "a firmer yardstick for assessing and understanding gender differences in this talented population." The American College Test (ACT) is a logical choice of assessment because of its use in college admissions and talent searches. To our knowledge, no one has used the ACT-Mathematics (ACT-M) or the ACT-Science (ACT-S) subtests in a comparable analysis.[2] Second, mathematical ability, although a powerful variable, is not the sole predictor for success in STEM. Because other variables matter, a comparison of male–female ratios in the extreme right tail on other abilities relevant for careers in

[1] Goldstein and Stocking (1994) examined male–female ratios on the SAT-M and SAT-V from 1981–1992, but, perhaps because it was published as a book chapter, has been largely undiscovered or ignored by top researchers in the debate. Due to the limitation of the method used, this study failed to replicate the SAT-M findings from Benbow and Stanley (1980, 1983). Hedges and Nowell (1995) examined the 1960–1992 period and did not find any changes in abilities across time.

[2] Others have examined sex differences on the EXPLORE test (Swiatek, Lupkowski-Shoplik & O'Donoghue, 2000) and on the TIMSS (Penner, 2003).

STEM is also needed. For example, the ability to read and synthesize scientific papers, write compelling grant applications, and develop one's ideas requires verbal and writing ability. Thus, a certain level of verbal and writing ability would seem critical for high level careers in STEM. Thus, to fully inform the debate on the underrepresentation of females in STEM, a broad historical examination of sex differences including verbal and writing ability should also be conducted in the extreme right tail. Others have previously reported no male–female differences on the SAT-Verbal (SAT-V; Benbow, 1988; Benbow & Stanley, 1980, 1983), but have not examined historical data that could show changes over time.

2. The present study

The aims of the present study include: 1. determining whether male–female ratios on the SAT-M specifically, but also on all the other measures have remained stable or changed over time and, if they changed, when, to what degree, and whether such change continues, 2. determining whether sex ratios on the SAT can be generalized to the ACT, 3. determining whether there are male–female differences on the ACT-S, and 4. examining sex ratios in verbal and writing ability using multiple measures.

We provide a broad historical analysis of sex differences in the right tail on test scores by presenting results across a period of 30 years using a database of SAT and ACT scores from the Duke University Talent Identification Program 7th Grade Talent Search (Duke TIP; Putallaz, Baldwin & Selph, 2005). Study 1 examined mathematical ability and science reasoning.[3] Study 2 examined verbal reasoning and writing ability.

3. Study 1: Sex differences in mathematical ability and science reasoning

3.1. SAT-Mathematics sample

We used 1,173,350 ($M=587{,}832$; $F=585{,}518$) test scores from 1981–2010 of students primarily from a 16-state region in the South and Midwest United States who took the SAT-M in the seventh grade. To qualify for participation in the Duke TIP 7th Grade Talent Search, all participants had previously scored in the top 5% of ability for their grade on a standardized test either on a composite score or a subtest. Males and females participated in roughly equal numbers across the years (see Appendices A and B for numbers of males and females within each cell for not just the SAT-M, but all other subtests examined).

3.2. ACT-Mathematics and ACT-Science sample

We used 440,369 ($M=224$, 399; $F=215$, 970) test scores from a similar but independent population as the SAT sample. Because the ACT was adopted later in the Talent Search than the SAT, data were only available from 1990–2010. The ACT-M requires reasoning to solve practical problems in mathematics, whereas the ACT-S asks students to interpret, analyze, evaluate, reason, and problem-solve using scientific texts (ACT, 2005).

3.3. Mathematical reasoning ability results

Table 1a displays the male–female ratio in 5-year segments for students scoring at or above each respective level on the SAT-M. We used 5-year segments because some individual years had cells with no females and because it also helped decrease the noise related to individual year ratio fluctuations to uncover the stable trend. From 1981–1985, the male–female ratios at the ≥500, ≥600 and ≥700 levels were 2.61 to 1, 5.82 to 1, and 13.5 to 1, respectively, thus replicating previous findings (Benbow & Stanley, 1980, 1983; 2.1 to 1, 4.1 to 1, and 13 to 1 respectively). From 1986 to the present, the male–female ratio declined at several levels and time periods. As can be seen in Fig. 1, the ratio of students scoring ≥700 (top 0.01%) on the SAT-M began to fall immediately after 1981–1985, but has remained relatively stable for the last two decades at roughly 4 to 1, with the most recent time period (2006–2010) indicating a ratio of 3.83 to 1. Thus, we confirm a decline, although not to the level given by Stanley and commonly cited in the research literature. Among perfect scorers (800) on the SAT-M, the ratio was 6.58 to 1 for 2006–2010, showing that even at the utmost right tail, the male–female ratio is now well below 13 to 1.

Table 1b shows the male–female ratio for students scoring at or above each respective level on the ACT-M (scores≥24 are comparable to SAT-M scores≥700). The ACT-M ratios, although slightly smaller than those from the SAT-M, illustrate a similar trend for 1990–2010, thus replicating the general pattern from the SAT-M. However, as can be seen in Fig. 1 and Table 1b, rather than stabilizing at 4 to 1, there has been a continued slight declining of the ratio with the most recent ratio at 2.6 to 1, a figure closer to the 2.8 to 1 given by Stanley. During 1990–2010, there were only 8 males and 1 female to score a perfect 36 on the ACT-M in the Duke TIP sample. In general, males continue to have a larger representation in the highest ranges of mathematical ability before adolescence, but to a far lesser degree than in past years (see the drop from 13.5 to 1 to roughly 4 to 1 in Fig. 1). We should note, however, that both the SAT-M and ACT-M male–female ratios have been fairly stable across the last two decades.

3.4. Science reasoning results

Table 1c illustrates the male–female ratio from 1990–2010 for students scoring at or above each respective level on the ACT-S. There is evidence of a stable or possibly even slightly growing male advantage in science reasoning across the last two decades, with the most recent time period showing a strikingly similar male–female ratio (2.83 to 1) as that on the ACT-M at a comparable level (See Fig. 1, both are roughly 3 to 1 in the top 0.01%). Interestingly, among those with a perfect score of 36 on the ACT-S across the years 1990–2010, there

[3] We use the term "science reasoning" to describe the constructs being measured by the ACT-S in this 7th grade population. We acknowledge that it is not clear what this test measures in this population but, as we discuss later in this paper, we suspect it measures a combination of science reasoning, reading comprehension, as well as interest in, and familiarity with scientific content. Thus, the ACT-S may, at least in part, measure an ability to reason with scientific information within the context of a scientific framework.

Table 1
Male to female ratios in math ability and science reasoning in the top 5% across 30 years.

a.	SAT-Math male to female ratio						
	≥200	≥300	≥400 Top 1%	≥500[a] 0.5%	≥600	≥700[b] 0.01%	800
1981–1985	0.93	0.98	1.43	2.61	5.82	13.50	–
1986–1990	0.94	0.97	1.31	2.15	4.67	7.60	–
1991–1995	1.00	1.02	1.26	1.95	3.04	3.87	–
1996–2000	1.05	1.05	1.14	1.55	2.56	4.13	4.00
2001–2005	1.03	1.04	1.13	1.57	2.51	3.55	5.60
2006–2010	1.00	1.01	1.10	1.54	2.50	3.83	6.58

b.	ACT-Math male to female ratio							
	≥1	≥12	≥16	≥18[a]	≥20	≥24[b]	≥28	≥32
1990–1995	0.96	0.97	1.15	1.46	1.97	3.14	3.25	2.75
1996–2000	1.04	1.04	1.22	1.59	1.99	3.06	8.63	9.00
2001–2005	1.04	1.04	1.20	1.49	1.98	2.92	4.29	14.00
2006–2010	1.10	1.10	1.21	1.50	1.82	2.60	3.99	3.87

c.	ACT-Science male to female ratio							
	≥1	≥12	≥16	≥20[a]	≥24	≥28	≥30[b]	≥32
1990–1995	0.96	0.96	1.00	1.24	1.75	2.42	2.98	4.27
1996–2000	1.04	1.04	1.05	1.29	1.87	3.42	3.72	6.73
2001–2005	1.04	1.04	1.06	1.26	1.94	3.64	4.42	4.67
2006–2010	1.10	1.10	1.11	1.51	1.82	2.76	2.83	5.13

Each cell was computed by taking the total number of males over the total number of females. Dashes "—" were placed in cells with insufficient data to compute ratios with confidence. For the ACT, intermediate values were omitted (e.g., ≥4 and ≥8 for the ACT-M and S) if they did not provide any relevant information. There were insufficient data to compute ratios with confidence for those scoring 36 on the ACT-M and S. Benchmarks for intellectual levels similar to 500 (the top 0.5%) and 700 (the top 0.01%) on the SAT-M were calculated using within distribution z-score approximations based on the overall SAT-M means and standard deviations. These comparable benchmarks are denoted by a superscript "a" and "b" within each panel.

have been 18 males and only 1 female. Therefore, it appears that males not only have a larger representation in the highest ranges of mathematical ability before adolescence, but maybe even a greater representation on the ACT-S, generalizing findings beyond mathematical ability to science reasoning. The ACT-S male–female ratio has also been fairly stable across the last two decades.

4. Study 2: Sex differences in verbal reasoning and writing ability

4.1. SAT-Verbal sample

As in Study 1, we used 1,173,350 (M = 587,832; F = 585,518) test scores from 1981–2010.

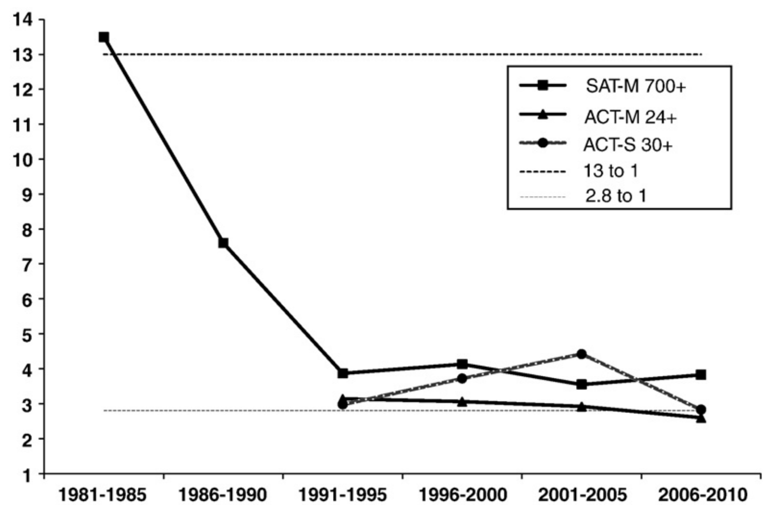

Fig. 1. Male to female ratio in math ability and science reasoning in the top 0.01% across 30 years. Includes male–female ratio data from the SAT-M, ACT-M, and ACT-S from participants in the top 0.01% of ability from 1981–2010. Also in Fig. 1 are lines indicating the 13 to 1 ratio mentioned by Benbow and Stanley (1980, 1983) and the 2.8 to 1 ratio mentioned by Stanley.

Table 2
Male to female ratios in verbal reasoning and writing ability in the top 5% across 30 years.

a.	SAT-Verbal male to female ratio						
	≥200	≥300	≥400	≥500[a]	≥600	≥700[b]	800
1981–1985	0.93	0.97	1.11	1.21	1.33	1.00	–
1986–1990	0.94	0.96	1.04	1.14	1.39	1.00	–
1991–1995	1.00	0.96	1.00	1.18	1.22	0.78	–
1996–2000	1.05	1.03	1.00	0.98	1.04	1.10	1.00
2001–2005	1.03	1.01	0.97	0.98	1.02	1.23	1.67
2006–2010	1.00	0.99	0.94	0.95	0.99	0.87	0.83

b.	SAT-test of standard written English male to female ratio								
	≥20	≥25	≥30	≥35	≥40	≥45[a]	≥50	≥55	≥60[b]
1981–1985	0.93	0.89	0.86	0.82	0.78	0.76	0.71	0.67	0.70
1986–1990	0.94	0.87	0.82	0.78	0.74	0.72	0.69	0.67	0.56
1991–1994	0.99	0.92	0.85	0.79	0.73	0.68	0.66	0.65	0.64

c.	ACT-English male to female ratio							
	≥1	≥12	≥16	≥20	≥22[a]	≥24	≥28	≥32[b]
1990–1995	0.96	0.93	0.81	0.69	0.63	0.60	0.53	0.70
1996–2000	1.04	1.00	0.89	0.76	0.72	0.69	0.63	0.70
2001–2005	1.04	0.99	0.88	0.77	0.73	0.68	0.62	0.62
2006–2010	1.10	1.05	0.94	0.84	0.78	0.75	0.68	0.58

d.	ACT-Reading male to female ratio								
	≥1	≥8	≥12	≥16	≥20	≥24[a]	≥28	≥32	≥34[b]
1990–1995	0.96	0.96	0.93	0.87	0.83	0.81	0.84	0.74	0.82
1996–2000	1.04	1.04	1.01	0.93	0.88	0.85	0.86	0.87	0.87
2001–2005	1.04	1.03	1.01	0.93	0.89	0.87	0.81	0.70	0.66
2006–2010	1.10	1.10	1.08	1.02	0.98	0.99	0.98	0.94	0.81

e.	SAT-Writing male to female ratio					
	≥200	≥300	≥400	≥500[a]	≥600	≥700[b]
2006	0.99	0.95	0.84	0.75	0.73	0.57
2007	0.99	0.95	0.81	0.71	0.69	0.61
2008	0.98	0.93	0.78	0.68	0.64	0.42
2009	1.03	0.97	0.81	0.72	0.57	0.47
2010	1.05	1.00	0.84	0.68	0.68	0.76

Each cell was computed by taking the total number of males over the total number of females. Dashes "—" were placed in cells with insufficient data to compute ratios with confidence. For the ACT, intermediate values were omitted (e.g., ACT-R≥4) if they did not provide any important information. There were insufficient data to compute ratios with confidence for those scoring 36 on the ACT-E and R and 800 on the SAT-W. Benchmarks for intellectual levels similar to 500 (the top 0.5%) and 700 (the top 0.01%) on the SAT-V were calculated using within distribution z-score approximations based on the overall SAT-V means and standard deviations. These comparable benchmarks are denoted by a superscript "a" and "b" within each panel.

4.2. SAT-test of standard written English (SAT-TSWE) sample

There were 384,833 (M = 187,968; F = 196,865) students with scores on the TSWE from 1981 through 1994 (after which the test was discontinued). The SAT-TSWE measured students' ability to use the conventionalized, edited language required in college (Educational Testing Service, 1975).

4.3. SAT-Writing (SAT-W) sample

There were 207,224 (M = 103,866; F = 103,358) students who took the SAT-W from 2006–2010. Introduced in 2005, the SAT-W measures writing ability (Mattern, Camara, & Kobrin, 2007).

4.4. ACT-English and Reading (ACT-E and ACT-R) sample

For each test, we used 440,369 (M = 224, 399; F = 215, 970) test scores from 1990–2010. The ACT-E measures students' ability to use the conventions of standard written English (ACT, 2005). The ACT-R measures students' reading comprehension ability (ACT, 2005).

4.5. Verbal reasoning and writing ability results

Table 2a demonstrates that across nearly all levels and time periods the male–female ratio has remained a roughly stable 1 to 1 for the SAT-V. The male–female ratio of perfect scorers in the most recent time period was 0.83 to 1, slightly favoring females (expressed as a female to male ratio it would be 1.2 to 1). Table 2b and c shows a consistent advantage for females on the ACT-E and SAT-TSWE that has been slightly increasing over time. For the ACT-E, across the years 1990–2010, there were 4 males and 8 females (a male–female ratio of 0.5 to 1, or a female–male ratio of 2 to 1) who earned a perfect score. For the SAT-TSWE, the male–female ratio of perfect scorers in the 1991–1994 period was 0.64 to 1 (a female–male ratio of 1.56 to 1). Table 2d illustrates an advantage for females on the ACT-R in the earlier time periods, but there seems to be little to no male–female difference on the ACT-R from 2006–2010. Across the years 1990–2010, there were 58 males and 74 females (a male–female ratio of 0.78 to 1, or a female–male ratio of 1.28 to 1) who earned a perfect score on the ACT-R indicating a slight female advantage. Table 2e demonstrates that for writing ability, females also show a slight advantage and that this advantage has also been increasing over time. There have only been 1 male and 3 females who achieved perfect scores on the SAT-W from 2006–2010.

Overall, females seem to have a small but clear and possibly increasing advantage on measures of conventions of standard written English and verbal reasoning ability (SAT-TSWE and ACT-E), as well as on the one measure of writing ability (SAT-W). On tests of vocabulary and reading comprehension (SAT-V and ACT-R), there appear to be little to no male–female differences. These findings are consistent with previous research (Cole, 1997; Halpern et al., 2007; Hedges & Nowell, 1995; Strand et al., 2006), with the female advantage in writing being among the most robust (peaking in 2008 at a 2.38 to 1 female–male ratio for scores ≥ 700). In summary, there has been a stable 1 to 1 male–female ratio on the SAT-V and ACT-R. However, females currently have a higher representation in the highest ranges of verbal reasoning (ACT-E) and writing ability (SAT-W) before adolescence.

5. Discussion

We inform the debate on the underrepresentation of women in STEM by examining the male–female ratio in the right tails of mathematical reasoning, science reasoning, verbal reasoning, and writing across three decades. The male–female ratio in the top 0.01% of mathematical ability on the SAT-M rapidly declined from 13.5 to 1 in the early 1980s to roughly 4 to 1 in the early 1990s (see Fig. 1). Whereas the SAT-M continued to show a fairly stable ratio from 1990–2010, the ACT-M showed a slightly declining ratio across the same span. Further, the mathematical reasoning data are augmented by the science reasoning data, which showed that males have a stable or slightly increasing advantage over females.

5.1. Limitations of this study

Our sample is not a random sample of the general population, but we believe it to be a reasonably representative sample of the extreme right tail population of both males and females. Additionally, our sample is larger and more geographically representative than the sample examined by Benbow and Stanley (1980, 1983). Using data from 1960–1992, Hedges and Nowell (1995) demonstrated that high male–female ratios (5 to 1 among the top 3% and 7 to 1 among the top 1%) were found in the right tail of ability distributions of nationally representative samples, and concluded that the high male–female ratios uncovered in studies such as Benbow and Stanley (1980, 1983), "need not be attributed to differential selection by sex" (Hedges & Nowell, 1995, p. 45). As a contemporary follow up to this, we examined SAT and ACT data for U.S high school students from 1996–2009 (see Appendix C). The pattern of findings, even though based on national samples, were strikingly similar to our findings among 7th graders across all subtests. For example, for perfect scorers on the SAT (800) in 2009, the male–female ratio (above the top 1%) was 2.22 to 1 for the SAT-M, 0.93 to 1 for the SAT-V, and 0.79 to 1 for the SAT-W. For the ACT in the 2001–2005 period, for scores ≥ 33, the male–female ratio was 2.6 to 1 for the ACT-M (top 2%), 2 to 1 for the ACT-S (top 1%), 0.75 to 1 for the ACT-E (top 3%), and 0.95 to 1 for the ACT-R (top 3%). This analysis strongly suggests that the Johns Hopkins (Benbow & Stanley, 1980, 1983) and Duke TIP databases are reasonably reflective of national sex differences.

The content of measures of academic achievement may have an influence on male–female differences (e.g., Lohman & Lakin, 2009). The SAT underwent content changes in 1994 and 2005 (Kobrin & Melican, 2007; Lawrence, Rigol, Van Essen, & Jackson, 2003) during which time sex ratios remained relatively stable. In 1994, for the SAT-M, the major change was the introduction of the student-produced response section and the use of calculators. For the SAT-V, antonyms were removed and instead of reading comprehension, critical reading was introduced (Lawrence et al., 2003). In 2005, for the SAT-M, third year college-preparatory math content was added and quantitative comparison items were removed. For the SAT-V, analogy items were removed and paragraph items were added to the critical reading section. Also in 2005, the SAT-W was introduced (Mattern et al., 2007). According to the College Board (e.g., Kobrin & Melican, 2007), the constructs measured by the SAT are comparable across test versions. We believe these test changes are unlikely to have significantly affected the general trends of the ratios reported here and likely cannot account for the decrease in the SAT-M ratios over time. These changes have not likely had an impact on the ratio trends partly because the SAT is designed for college bound high school students, and especially when given to a younger population as an out of level testing measure, the measured ability constructs have likely not changed meaningfully across test versions. We believe this is due at least in part to the fact that problems designed to measure achievement in older students may better measure reasoning ability in younger students (Benbow, 1988). In addition, since our findings on both the SAT-M and ACT-M are quite similar (roughly 4 to 1 and 3 to 1, respectively, at the top 0.01% level), the male–female difference is not necessarily tied to the ability measure used. The slightly lower male–female ratio on the ACT-M may be found because the ACT is overall a more verbal measure (Koenig, Frey, & Detterman, 2008). Finally, and maybe most importantly, the male–female ratio drop began well before the first major content changes on the SAT in 1994 (see Fig. 1 and Table 1a).

5.2. Linking our results to other data

Ceci and Williams (2010; Ceci et al., 2009) suggested that there has been conflicting evidence on the ratio of males to females in the right tail of mathematical ability. Although we agree that the data are not perfectly aligned, we think that with the addition of our sample, we are able to bridge current and past datasets as well as population and extreme right tail data to show a consistent overall pattern (Steen, 1988). By examining the top 5% of ability, we demonstrate that both the population level right tail data (Hedges & Nowell, 1995; Hyde et al., 2008; Lohman & Lakin, 2009; Strand et al., 2006) and extreme right tail data (Benbow & Stanley, 1980, 1983) in the U.S. and U.K. are not in conflict with one another but show reasonably similar trends. For example, in the top 4% (Lohman & Lakin, 2009; Strand et al., 2006) and top 1% (Hyde et al., 2008) in mathematical ability for population level right tail data, the male–female ratio typically found is about 2 to 1. Consistent with these reports, our data show a male–female ratio of 1.5 to 1 (an SAT-M score of about 500 indicates the top 0.5%).

5.3. Where do sex differences begin to appear in the top 5%?

Because our data are consistent with population-level data in that substantial sex differences in the right tail are evident, our data are also likely a reasonable picture of the top 5% of cognitive abilities. Therefore, we may be able to explicate the point at which male–female differences begin to appear on each of the specific abilities in the top 5%. As Table 1 illustrates, male–female ratio differences appear around the cut for the top 1% in ability (e.g., about SAT-M ≥ 400). Whether mathematical ability or science reasoning in the top 1% are required for success in high level STEM careers is an open empirical question that our data cannot address. However, our data can at least provide some clues as to where in the right tail of ability male–female representation might have the potential to have an impact.

5.4. Cognitive abilities remain an important factor

Regarding the query by Ceci and Williams (2010, p. 152) as to whether the same results from the early 1980s would be found with today's children, our results reveal that the same results are not found. Yet, even though the sex biases and lack of female role models have declined, we continue to find male–female differences in the extreme right tail of mathematical ability as well as male–female differences in the extreme right tail of science reasoning (see Fig. 1 and Table 1). Thus, the declined yet continued greater representation of males in the upper tails of mathematical ability and science reasoning should remain an important factor in the debate on the underrepresentation of women in STEM. The continuity of the male–female ratios in math and science over the past two decades suggests that diminished biases and/or an increased number of role models likely do not fully explain current ratios, although they may, at least in part, continue to play a role.

5.5. Results are not just isolated to the SAT-M

Our findings on the ACT-M also validate the SAT-M as a measure of mathematical reasoning ability. This suggests that links made with the SAT in other research (e.g., Lubinski & Benbow, 2006) may also hold for the ACT as well because certain specific ability constructs being measured are quite similar (e.g., the SAT-M and ACT-M both measure mathematical ability) in this 7th grade population. Future research might also benefit by using the ACT-S to determine if it provides any incremental validity above and beyond mathematical, verbal, and spatial ability (as well as other factors) in the prediction of longitudinal educational and vocational outcomes in STEM.

5.6. Science reasoning: A new factor in the debate?

Although the ACT-S is described as measuring science reasoning, we could find no recent construct validation support. It is not uncommon for a test to measure something different than its name (or the testing company) implies (Kelley, 1927; Koenig et al., 2008; Lubinski, 2004). Regardless, it is remarkable that we find a male advantage in the extreme right tail on this measure and that among all perfect scorers, 18 have been males and only one has been a female. Given the either sex equivalence or female advantage in reading comprehension, it is unlikely that the ACT-S measures solely reading comprehension (although that is clearly a component). The male advantage may result in part from early sex differences in familiarity with and interest in scientific content. This is particularly relevant because the test was given to students before they had been formally introduced to the relevant scientific content. It is possible that visits to science museums and extracurricular science classes are more common among boys and this may partly explain these results (Linn & Pulos, 1983), although, as the ability ratio change in our data illustrate, it is important to test whether this is still the case today. Additionally, a new genetic analysis by Haworth, Dale, and Plomin (2008) of 9 year olds' science ratings by teachers indicates a fairly strong genetic component and a modest non-shared environmental component, with somewhat greater variance for boys, despite little overall mean differences between the sexes, suggesting, among other things, that there are male–female science ability, achievement, interest, and/or familiarity differences well before the 7th grade.

Similar to Benbow's (1988) argument regarding the SAT-M as a particularly powerful measure of mathematical ability for 7th grade students, the ACT-S may function as a more powerful measure of science reasoning. We acknowledge that we are uncertain as to the specific constructs the ACT-S measures in this younger population, and believe that future research aimed at understanding the reasons for this male–female ability/interest/familiarity difference will help place these results in a broader theoretical context. However, Hedges and Nowell (1995) found male–female differences in science achievement among high school students in the top 5% that ranged from 2.5 to 1 to 7.2 to 1. Although the content and cognitive requirements of the science measures utilized were different from the ACT-S, male–female differences were still apparent and in line with our findings. Regardless of why we find a male–female difference, the fact that one exists is interesting and should now play a part in the conversation about the underrepresentation of women in STEM.

5.7. Females are higher on verbal reasoning and writing ability

Our examination of verbal reasoning and writing ability revealed markedly different trends from those found for mathematical ability and science reasoning. Unlike the SAT-V and ACT-R, the female advantage on the ACT-E, SAT-TSWE, and SAT-W slightly increased over time demonstrating that females outnumber males in verbal reasoning and writing ability in the right tail. These findings are generally aligned with much of the research literature. For example, Hedges and Nowell (1995) also found that females had an advantage in the upper 5% of verbal reasoning and writing ability. Even among perfect scorers, our data demonstrate that males are not always more highly represented than females in the right tail and show that females continue to have an advantage on a verbal reasoning and a writing ability measure.

5.8. Other factors in the debate

Mathematical reasoning, science reasoning, verbal reasoning and writing are all likely important for careers in STEM (Halpern et al., 2007; Hyde, 2005; Hyde & Linn, 2006; Hyde et al., 2008), although not necessarily to the same degree. Because females have an advantage on verbal reasoning and writing ability in the right tail, it is likely that the underrepresentation of women in STEM is not due to any sex differences in general ability. It may be that there are more males than females in the right tail who have a mathematical (and/or science reasoning) ability tilt (one score higher than the others; Ceci et al., 2009; Halpern et al., 2007; Park, Lubinski, & Benbow, 2007), or relatively higher spatial ability (Wai, Lubinski, & Benbow, 2009; Webb, Lubinski, & Benbow, 2007), that make them more likely to pursue a career in STEM where such abilities are critical components. Such ability profile differences have been found to covary with different motivational propensities in both education and the world of work, which may in turn have implications for commitment to developing expertise in high level STEM arenas (Ackerman, 1996; Ackerman & Heggestad, 1997; Schmidt, Lubinski, & Benbow, 1998; Webb et al., 2007). Additionally, non-cognitive factors (Ceci et al., 2009; Halpern et al., 2007; Hyde & Linn, 2006; Hyde et al., 2008; Ferriman et al., 2009) likely play an important role in influencing career pursuits. One of these non-cognitive factors may include stereotype or signaling threat (Ceci et al., 2009; Halpern et al., 2007; Murphy et al., 2007). Another includes the number of hours worked (Hewlett & Luce, 2006; Leslie, 2007; Lubinski & Benbow, 2006; Mason & Goulden, 2004).

In particular, research on preferences (Ceci & Williams, 2010; Ceci et al., 2009; Eccles & Jacobs, 1986; Ferriman et al., 2009; Hakim, 2006; Halpern et al., 2007) shows that, on average, females are more drawn to *people* while males are more drawn to *things* (Lippa, 1998; Su, Rounds, & Armstrong, 2009) and appears to be an especially important potential factor in the underrepresentation of women in STEM. For example, Eccles and Jacobs (1986) have discussed how such preferences may direct more females into the life sciences and fewer females into the physical sciences. Some researchers have argued that preferences are much more significant than abilities (Ceci & Williams, 2010; Ceci et al., 2009), and Hakim (2006, p. 279) has even stated that, "There are no sex differences in cognitive ability." However, our data clearly show that there are sex differences in cognitive abilities in the extreme right tail with some favoring males and some favoring females. We agree that preferences are likely important in explaining the underrepresentation of women in high level STEM careers. However, we have shown that male–female differences on the SAT-M (as well as the ACT-M and ACT-S), although significantly declined, still remain and do not appear to be further declining or disappearing (as suggested by Feingold, 1988) and, in some cases, are even increasing. Sex differences favoring males in math ability and science reasoning may have declined from the early 1980s to present, possibly in response to fewer barriers, more encouragement and role models provided to females, but the fact that they are still substantial and have remained relatively stable for two decades may at least partly account for the dearth of women in STEM careers because these types of gatekeeper tests are used by graduate admission committees in selecting applicants into STEM fields (e.g., the Graduate Record Examination-Quantitative). Thus, sex differences in abilities in the extreme right tail should not be dismissed as no longer part of the explanation for the dearth of women in math-intensive fields of science. Studies have shown that ability differences alone (measured as scores on the SAT-M) within the top 1% make a difference in earning low base rate achievements such as a STEM PhD, publication, patent, or tenure at a top university (Park et al., 2007; Park, Lubinski, & Benbow, 2008; Wai, Lubinski, & Benbow, 2005). Male–female differences in representation in the highest ranges of mathematical ability may thus still have relevance as an explanatory factor, albeit a substantially smaller one, than previous researchers have argued (Benbow 1988; Benbow & Stanley, 1980, 1983; Benbow, Lubinski, Shea, & Eftekhari-Sanjani, 2000). Consequently, the declined ability differences may have implications for declined future achievement differences between males and females. However, our finding of male–female differences in science reasoning also provides support for the potential importance of abilities. Of course, this does not mean that biases and lack of role models have been eliminated entirely, and we do not know, and cannot predict, whether the ratios will remain stable or change in the future.

5.9. Potential explanations and future directions

Our findings are not inconsistent with previous explanations focusing on either biological (Benbow, 1988; Benbow & Stanley, 1980, 1983) or social or cultural (Guiso, Monte, Sapienza, & Zingales, 2008; Hyde & Linn, 2006; Hyde et al., 2008; Penner, 2008; Penner & Paret, 2008) aspects, but are likely best explained via frameworks that examine multiple perspectives simultaneously (Ceci & Williams, 2010; Ceci et al., 2009 Halpern et al., 2007). It is extremely likely that sociocultural factors played a role in the rapid decline from a 13.5 to 1 ratio in the early 1980s to a 4 to 1 ratio by the early 1990s in the top 0.01% of SAT-M scores. Some sociocultural factors that may have introduced this change include the increased educational opportunities available to girls in the form of more encouragement to participate in special programs and mentoring to encourage them to participate in upper level math courses (Ceci et al., 2009; Halpern et al., 2007). Stanley hypothesized that the male–female SAT-M ratio has decreased because "women have had the opportunity to take their math earlier" (quoted in Monastersky, 2005, p. 45). Thus, it is possible that the change was due to the increased availability of accelerated math courses starting in the early 1980s and hence could be attributed, at least in part, to exposure and achievement (Wai, Lubinski, Benbow, & Steiger, in press). We should, of course, continue to equally encourage both women and men to pursue careers that they are passionate about (Pinker, 2008; Webb, Lubinski, & Benbow, 2002), including high level careers in STEM. Future research should be conducted to ascertain the reasons for the male–female ratio decline in mathematical ability (Halpern et al., 2007), and what role sociocultural or other factors play in impacting the current generation of males and females.

Acknowledgments

This research was partly supported by a Duke University Dean's Summer Research Fellowship awarded to the second author. Portions of this paper were presented at the 53rd annual conference of the National Association for Gifted Children, Charlotte, NC, November, 2006. We thank Terence Costello for his assistance with data management as well as Stephen J. Ceci, Rick Courtright, Doug Detterman, David Lubinski, Richard Lynn, Maya Wai and especially Wendy Johnson for reviewing earlier versions of this manuscript.

Appendix A. Number of males and females for math ability and science reasoning.

a.	SAT-Math						
	≥200	≥300	≥400	≥500	≥600	≥700	800
1981–1985	40334M	36137M	17695M	4642M	664M	54M	0M
	43367F	36689F	12379F	1779F	114F	4F	0F
1986–1990	74531M	66524M	32765M	8920M	1575M	152M	4M
	79544F	68278F	24993F	4158F	337F	20F	0F
1991–1995	93662M	86174M	46399M	13851M	2645M	271M	4M
	93787F	84660F	36857F	7090F	870F	70F	0F
1996–2000	133570M	130356M	100938M	32903M	4753M	363M	12M
	127788F	123871F	88611F	21288F	1860F	88F	3F
2001–2005	141867M	138313M	107463M	38103M	5936M	600M	28M
	137673F	133200F	95249F	24205F	2367F	169F	5F
2006–2010	103868M	101738M	80673M	28516M	4911M	628M	79M
	103359F	100752F	73540F	18460F	1962F	164F	12F

b.	ACT-Math								
	≥1	≥12	≥16	≥18	≥20	≥24	≥28	≥32	36
1990–1995	42636M	41843M	23120M	9134M	2448M	270M	39M	11M	2M
	44329F	43277F	20118F	6237F	1245F	86F	12F	4F	1F
1996–2000	53347M	53070M	34193M	13488M	4424M	615M	69M	9M	0M
	51435F	50966F	27971F	8466F	2222F	201F	8F	1F	0F
2001–2005	59294M	58928M	35782M	12800M	4622M	854M	90M	14M	0M
	57284F	56790F	29909F	8606F	2333F	292F	21F	1F	0F
2006–2010	69122M	68858M	51845M	23235M	10954M	2388M	311M	58M	6M
	62922F	62590F	42826F	15512F	6019F	919F	78F	15F	0F

c.	ACT-Science								
	≥1	≥12	≥16	≥20	≥24	≥28	≥30	≥32	36
1990–1995	42636M	41878M	32538M	13731M	2811M	435M	152M	47M	2M
	44329F	43528F	32656F	11096F	1602F	180F	51F	11F	0F
1996–2000	53347M	51743M	41484M	17760M	3921M	605M	223M	74M	4M
	51435F	49904F	39371F	13734F	2099F	177F	60F	11F	0F
2001–2005	59294M	56605M	43107M	17070M	3047M	375M	115M	56M	8M
	57284F	54632F	40537F	13512F	1574F	103F	26F	12F	1F
2006–2010	69122M	64285M	50182M	25987M	4115M	503M	187M	77M	4M
	62922F	58619F	45065F	17188F	2266F	182F	66F	15F	0F

This appendix includes sample sizes of males and females for each cell computed in Table 1. The first column for each table (e.g., SAT-M≥200) indicates the total number of males and females in the sample.

Appendix B. Number of males and females for verbal reasoning and writing ability.

a.	SAT-Verbal						
	≥200	≥300	≥400	≥500	≥600	≥700	800
1981–1985	40334M	31900M	11109M	1683M	137M	3M	0M
	43367F	32962F	9996F	1392F	103F	3F	0F
1986–1990	74531M	54132M	16758M	2396M	210M	6M	0M
	79544F	56589F	16186F	2094F	151F	6F	0F
1991–1995	93662M	66932M	21703M	3446M	278M	7M	0M
	93787F	69528F	21786F	2920F	227F	9F	0F
1996–2000	133570M	125481M	81602M	22027M	2396M	128M	3M
	127788F	121884F	81993F	22390F	2293F	116F	3F
2001–2005	141867M	133194M	84933M	21767M	2532M	130M	5M
	137673F	131449F	87333F	22100F	2473F	106F	3F
2006–2010	103868M	97891M	65060M	20026M	2651M	117M	5M
	103359F	99085F	69064F	21026F	2691F	134F	6F

(continued on next page)

Appendix B. *(continued)*

b.	SAT-test of standard written English								
	≥20	≥25	≥30	≥35	≥40	≥45	≥50	≥55	≥60
1981–1985	40334M	35808M	29917M	21966M	13704M	7386M	3016M	828M	87M
	43367F	40188F	34815F	26700F	17574F	9765F	4272F	1240F	124F
1986–1990	74531M	61784M	49752M	35736M	21481M	11204M	4536M	1177M	100M
	79544F	71285F	60513F	45928F	29203F	15669F	6582F	1747F	178F
1991–1994	73066M	62372M	50028M	35723M	21717M	11034M	4420M	1136M	123M
	73917F	68152F	58553F	45210F	29736F	16143F	6735F	1736F	191F

c.	ACT-English								
	≥1	≥12	≥16	≥20	≥22	≥24	≥28	≥32	36
1990–1995	42636M	39645M	25685M	10366M	5327M	2567M	340M	16M	1M
	44329F	42671F	31841F	15034F	8400F	4252F	640F	23F	2F
1996–2000	53347M	48097M	32484M	13248M	6826M	3554M	588M	54M	2M
	51435F	48333F	36416F	17360F	9486F	5163F	940F	77F	0F
2001–2005	59294M	52950M	34899M	14250M	7374M	3686M	650M	63M	1M
	57284F	53533F	39577F	18563F	10145F	5449F	1041F	101F	0F
2006–2010	69122M	60554M	41057M	19527M	9914M	5091M	1013M	154M	0M
	62922F	57848F	43454F	23248F	12742F	6824F	1494F	264F	6F

d.	ACT-Reading									
	≥1	≥8	≥12	≥16	≥20	≥24	≥28	≥32	≥34	36
1990–1995	42636M	42363M	38696M	27205M	13060M	4651M	1208M	225M	102M	21M
	44329F	44187F	41603F	31300F	15783F	5756F	1446F	303F	124F	30F
1996–2000	53347M	53122M	49073M	31915M	16749M	5928M	1484M	283M	109M	18M
	51435F	51306F	48568F	34339F	19052F	6939F	1722F	327F	125F	11F
2001–2005	59294M	59046M	55213M	33352M	15871M	5619M	1425M	212M	68M	4M
	57284F	57160F	54532F	35781F	17800F	6463F	1760F	303F	103F	13F
2006–2010	69122M	68642M	63982M	43179M	21686M	7387M	2267M	374M	92M	15M
	62922F	62599F	59384F	42489F	22155F	7498F	2303F	397F	113F	20F

e.	SAT-Writing						
	≥200	≥300	≥400	≥500	≥600	≥700	800
2006	22787M	21266M	12512M	2957M	266M	8M	0M
	23073F	22329F	14924F	3917F	362F	14F	0F
2007	21041M	19735M	11569M	2428M	232M	11M	0M
	21286F	20774F	14371F	3441F	335F	18F	0F
2008	20467M	18602M	10250M	2407M	260M	8M	1M
	20918F	20023F	13186F	3530F	407F	19F	0F
2009	21005M	19094M	10359M	2193M	178M	9M	0M
	20333F	19584F	12720F	3038F	311F	19F	1F
2010	18566M	17356M	10193M	2244M	249M	19M	0M
	17748F	17372F	12138F	3286F	367F	25F	2F

This appendix includes sample sizes of males and females for each cell computed in Table 2. The first column for each table (e.g., SAT-V≥200) indicates the total number of males and females in the sample.

Appendix C. Male–female ratios on the SAT and ACT for college bound students.

a.	SAT-Math			SAT-Verbal			SAT-Writing		
	≥700 Top 6%	≥750 3%	800 <1%	≥700 5%	≥750 2%	800 <1%	≥700 4%	≥750 1%	800 <1%
1996–2000	1.97	2.47	–	1.01	1.03	–	–	–	–
2001–2005	1.83	2.20	–	1.04	1.05	–	–	–	–
2006–2009	1.77	2.00	–	0.96	0.95	–	0.74	0.73	–
2009 only	1.75	1.95	2.22	0.99	0.96	0.93	0.76	0.76	0.79

b.	ACT-Math		ACT-Science		ACT-English		ACT-Reading	
	≥28 Top 9%	≥33 2%	≥28 6%	≥33 1%	≥28 11%	≥33 3%	≥28 15%	≥33 3%
1997–2000	1.74	2.00	1.88	2.00	0.79	0.50	0.96	0.90
2001–2005	1.73	2.60	1.83	2.00	0.80	0.75	0.97	0.95

Dashes "–" were placed in cells where there was no data available. All percentiles are based on the SAT in 2009 and the ACT from 2007–2009. Values in Appendix Ca are drawn from a pool of 8,707,700 males and 10,063,730 females who took the SAT from 1996–2009. In 1996 about 75,000 more females than males took the SAT and in 2009 about 100,000 more females took the test. Values in Appendix Cb are drawn from a pool of 4,222,429 males and 5,485,592 females who took the ACT from 1997–2005 and are calculated by taking the ratio of percentages of males to females earning a particular score as raw Ns were not available. Data were also not available by gender for the ACT from 2006–2009. In 1997 about 120,000 more females than males took the ACT and in 2005 about 145,000 more females took the test. Data for the SAT were adapted from Archived Data and SAT Reports for the years 1996–2009 downloaded from http://professionals.collegeboard.com/data-reports-research/sat/archived and http://professionals.collegeboard.com/data-reports-research/sat/data-tables. Data for the ACT were adapted from ACT National Scores for the years 1997–2005 downloaded from http://www.act.org/news/data.html.

References

Ackerman, P. L. (1996). A theory of adult intellectual development: Process, personality, interests, and knowledge. *Intelligence*, *22*, 227−257.

Ackerman, P. L., & Heggestad, E. D. (1997). Intelligence, personality, and interests: Evidence for overlapping traits. *Psychological Bulletin*, *121*, 218−245.

ACT (2005). Your guide to the ACT. Downloaded on October 25, 2009 from http://www.act.org/aap/pdf/YourGuidetoACT.pdf

Benbow, C. P. (1988). Sex differences in mathematical reasoning ability among the intellectually talented: Their characteristics, consequences, and possible explanations. *Behavioral and Brain Sciences*, *11*, 169−183.

Benbow, C. P., & Stanley, J. C. (1980). Sex differences in mathematical ability: Fact or artifact? *Science*, *210*, 1262−1264.

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

Benbow, C. P., Lubinski, D., Shea, D. L., & Eftekhari-Sanjani, H. (2000). Sex differences in mathematical reasoning ability: Their status 20 years later. *Psychological Science*, *11*, 474−480.

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

Ceci, S. J., Williams, W. M., & Barnett, S. M. (2009). Women's underrepresentation in science: Sociocultural and biological considerations. *Psychological Bulletin*, *135*, 218−261.

Cole, N. S. (1997). *ETS gender study: How females and males perform in educational settings.* Princeton, NJ: Educational Testing Service.

Eccles, J., & Jacobs, J. (1986). Social forces shape math attitudes and performance. *Signs: Journal of Women in Culture and Society*, *11*, 367−389.

Educational Testing Service (1975). The test of standard written English: A preliminary report. Downloaded on October 25, 2009 from http://www.eric.ed.gov/ERICDocs/data/ericdocs2sql/content_storage_01/0000019b/80/31/7d/52.pdf

Feingold, A. (1988). Cognitive gender differences are disappearing. *American Psychologist*, *43*, 95−103.

Ferriman, K., Lubinski, D., & Benbow, C. P. (2009). Work preferences, life values, and personal views of top math/science graduate students and the profoundly gifted: Developmental changes and gender differences during emerging adulthood and parenthood. *Journal of Personality and Social Psychology*, *97*, 517−532.

Goldstein, D., & Stocking, V. B. (1994). TIP studies of gender differences in talented adolescents. In K. A. Heller, & E. A. Hany (Eds.), *Competence and responsibility: Proceedings of the Third European Conference of the European Council for High Ability* (pp. 190−203). Goettingen, Germany: Hogrefe and Huber.

Guiso, L., Monte, F., Sapienza, P., & Zingales, L. (2008). Culture, gender, and math. *Science*, *320*, 1164−1165.

Halpern, D. F. (2000). *Sex differences in cognitive abilities (3rd Ed.).* Mahwah, NJ: Lawrence Erlbaum Associates.

Halpern, D. F., Benbow, C. P., Geary, D. C., Gur, R. C., 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.

Hakim, C. (2006). Women, careers, and work-life preferences. *British Journal of Guidance & Counseling*, *34*, 279−294.

Haworth, C. A., Dale, P., & Plomin, R. (2008). A twin study into the genetic and environmental influences on academic performance in science in nine-year-old boys and girls. *International Journal of Science Education*, *30*, 1003−1025.

Hedges, L. V., & Nowell, A. (1995). Sex differences in mental test scores, variability, and numbers of high-scoring individuals. *Science*, *269*, 41−45.

Hewlett, S. A., & Luce, C. B. (2006, December). Extreme jobs: The dangerous allure of the 70 hour workweek. *Harvard Business Review*, 49−59.

Hyde, J. S. (2005). The gender similarities hypothesis. *American Psychologist*, *60*, 581−592.

Hyde, J. S., & Linn, M. C. (2006). Gender similarities in mathematics and science. *Science*, *314*, 599−600.

Hyde, J. S., Lindberg, S. M., Linn, M. C., Ellis, A. B., & Williams, C. C. (2008). Gender similarities characterize math performance. *Science*, *321*, 494−495.

Johnson, W., Carothers, A., & Deary, I. J. (2008). Sex differences in variability in general intelligence: A new look at the old question. *Perspectives on Psychological Science*, *3*, 518−531.

Kelley, T. (1927). *Interpretation of educational measurements.* New York: World Book.

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

Koenig, K. A., Frey, M. C., & Detterman, D. K. (2008). ACT and general cognitive ability. *Intelligence*, *36*, 153−160.

Lawrence, I. M., Rigol, G. W., Van Essen, T., & Jackson, C. A. (2003). A historical perspective on the content of the SAT. *College Board Research Report No. 2003-3.*

Leslie, D. W. (2007, March). The reshaping of America's academic workforce. TIAA-CREF Institute, #87. Retrieved January 10, 2010 from www.tiaa-crefinstitute.org

Linn, M. C., & Pulos, S. (1983). Male–female differences in predicting displaced volume: Strategy use, aptitude relationships, and experience influences. *Journal of Educational Psychology*, *75*, 86−96.

Lippa, R. A. (1998). Gender-related individual differences and the structure of vocational interests on the people-things dimension. *Journal of Personality and Social Psychology*, *74*, 996−1009.

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.

Lubinski, D. (2004). Introduction to the special section on cognitive abilities: 100 years after Spearman's (1904) "'general intelligence,' objectively determined and measured". *Journal of Personality and Social Psychology*, *86*, 96−111.

Lubinski, D., & Benbow, C. P. (2006). Study of mathematically precocious youth after 35 years: Uncovering antecedents for the development of math–science expertise. *Perspectives on Psychological Science*, *1*, 316−345.
Mason, M. A., & Goulden, M. (2004). Marriage and baby blues: Redefining gender equity and the academy. *Annals of American Political and Social Scientist*, *596*, 86−103.
Mattern, K., Camara, W., & Kobrin, J. L. (2007, January). SAT writing: An overview of research and psychometrics to date. *College Board Research Notes* (pp. 32).
Monastersky, R. (2005, March 4). Primed for numbers. The Chronicle of Higher Education, 51, p. A1. Retrieved October 20, 2009, from http://chronicle.com/article/Primed-for-Numbers/26462
Murphy, M. C., Steele, C. M., & Gross, J. J. (2007). Signaling threat: How situational cues affect women in math, science, and engineering settings. *Psychological Science*, *18*, 879−885.
Park, G., Lubinski, D., & Benbow, C. P. (2007). Contrasting intellectual patterns for creativity in the arts and sciences: Tracking intellectually precocious youth over 25 years. *Psychological Science*, *18*, 948−952.
Park, G., Lubinski, D., & Benbow, C. P. (2008). Ability differences among people who have commensurate degrees matter for scientific creativity. *Psychological Science*, *19*, 957−961.
Penner, A. M. (2003). International gender x item difficulty interactions in mathematics and science achievement tests. *Journal of Educational Psychology*, *95*, 650−655.
Penner, A. M. (2008). Gender differences in extreme mathematical achievement: An international perspective on biological and social factors. *American Journal of Sociology*, *114*, S138−S170.
Penner, A. M., & Paret, M. (2008). Gender differences in mathematics achievement: Exploring the early grades and the extremes. *Social Science Research*, *37*, 239−253.
Pinker, S. (2008). *The sexual paradox: Men, women, and the real gender gap.* New York: Scribner.
Putallaz, M., Baldwin, J., & Selph, H. (2005). The Duke University talent identification program. *Journal of High Ability Studies*, *16*, 41−54.
Schmidt, D. B., Lubinski, D., & Benbow, C. P. (1998). Validity of assessing educational–vocational preference dimensions among intellectually talented 13-year-olds. *Journal of Counseling Psychology*, *45*, 436−453.
Spelke, E. S. (2005a). Sex differences in intrinsic aptitude for mathematics and science? A critical review. *American Psychologist*, *60*, 950−958.
Spelke, E. S. (2005b). The science of gender and science: Pinker vs. Spelke, a debate. Edge: The Third Culture. Downloaded March 8, 2010 from: http://www.edge.org/3rd_culture/debate05/debate05_index.html
Steen, L. A. (1988, April 29). The science of patterns. *Science*, *240*, 611−616.
Strand, S., Deary, I. J., & Smith, P. (2006). Sex differences in cognitive abilities test scores: A UK national picture. *British Journal of Educational Psychology*, *76*, 463−480.
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.
Swiatek, M. A., Lupkowski-Shoplik, A., & O'Donoghue, C. C. (2000). Gender differences in above-level EXPLORE scores of gifted third through sixth graders. *Journal of Educational Psychology*, *92*, 718−723.
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.
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.
Wai, J., Lubinski, D., Benbow, C. P., & Steiger, J. H. (in press). Accomplishment in science, technology, engineering, and mathematics (STEM) and its relation to STEM educational dose: A 25-year longitudinal study. *Journal of Educational Psychology*.
Webb, R. M., Lubinski, D., & Benbow, C. P. (2007). Spatial ability: A neglected dimension in talent searches for intellectually precocious youth. *Journal of Educational Psychology*, *99*, 397−420.
Webb, R. M., Lubinski, D., & Benbow, C. P. (2002). Mathematically facile adolescents with math/science aspirations: New perspectives on their educational and vocational development. *Journal of Educational Psychology*, *94*, 785−794.