

Sex differences in reasoning abilities: Surprising evidence that male-female ratios in the tails of the quantitative reasoning distribution have increased

Joni M. Lakin

Auburn University

Accepted for publication in Intelligence

Keywords: Sex differences; Cognitive abilities; Quantitative reasoning; STEM; Secular trends

Author Note

Joni M. Lakin, Ph.D., Department of Educational Foundations, Leadership, and Technology, Auburn University.

The author gratefully acknowledges the helpful comments of James Gambrell, Jonathan Wai, David Lohman, Marianne Feller, two anonymous reviewers, and Wendy Johnson on earlier drafts of this article.

Correspondence concerning this article should be addressed to Joni Lakin, Department of Educational Foundations, Leadership, and Technology, 4032 Haley Center, Auburn University, AL 36849. Email: joni.lakin@auburn.edu

Abstract

Sex differences in cognitive abilities, particularly at the extremes of ability distributions, have important implications for the participation of men and women in highly valued and technical career fields. Although negligible mean differences have been found in many domains, differences in variability and high ratios of males to females in the tails of the ability distribution have been found in a number of studies and across domains. A few studies have also observed trends over time, most noting decreasing ratios of boys to girls in the highest levels of mathematics test performance. In this study, sex differences in means, variances, and ratios were evaluated in four cohorts (1984, 1992, 2000, and 2011) in verbal, quantitative, and nonverbal/figural reasoning domains as measured by the Cognitive Abilities Test. Samples included U.S. students in grades 3-11. Overall, the results were consistent with previous research, showing small mean differences in the three domains, but considerably greater variability for males. The most surprising finding was that, contrary to related research, the ratio of males to females in the upper tail of the quantitative reasoning distribution seemed to increase over time. Explanations for this finding are explored.

Sex differences in reasoning abilities: Surprising evidence that male-female ratios in the tails of the quantitative reasoning distribution have increased

Sex differences in cognitive abilities have long been a source of intense debate and concern among researchers. This is partly because of the implications of those differences for maintaining a strong workforce, particularly in mathematics-intensive fields (Ceci & Williams, 2006; Park, Lubinski, & Benbow, 2007; Wai, Lubinski, & Benbow, 2005), and partly because of the topic's inherently personal relevance and incendiary implications (Chipman, 1988; Eagly, 1995). Although research has established some areas of significant differences between men and women (e.g., mental rotation, mechanical reasoning, verbal memory, spelling; Feingold, 1988; Kimura, 2004), research indicates that men and women have equal or nearly equal mean ability in general intelligence and many broad cognitive abilities (Ceci, Williams, & Barnett, 2009; Hyde & Linn, 2006). Despite this, many researchers have observed that men are disproportionately represented in both the top and bottom extremes of the ability distribution (Ceci et al., 2009; Feingold, 1992), giving rise to the variability hypothesis. The variability hypothesis thus proposes that men are more variable in their abilities than women in many domains and that a wider distribution rather than mean differences accounts for the greater numbers of men at both the upper and lower extreme of the distribution of abilities (both general and broad; Feingold, 1992, 1994).¹ This hypothesis was proposed in early psychological research (see Feingold, 1992, and Johnson, Carothers, & Deary, 2008, for reviews) and has recently

¹ Some researchers have augmented the variability hypothesis to claim that cognitive ability distributions are non-normal, with a discernible subpopulation of individuals (disproportionately male) with birth defects causing low cognitive ability (e.g., Johnson et al., 2008). This theory predicts greater numbers of males at the lower tail than the upper tail, but does not substantially affect the overall observation that men are overrepresented at the two tails of the distribution or change the impact that overrepresentation at the high end of ability has on the workforce in science and engineering fields.

received more attention to explain male overrepresentation among elite math and science fields (Ceci et al., 2009).

In this study, I reviewed and extended a previous analysis (Lohman & Lakin, 2009) of male-female variability based on data from the Cognitive Abilities Test (CogAT; Lohman, 2011; Lohman & Hagen, 2001; Thorndike & Hagen, 1984, 1992) which measures verbal, quantitative, and nonverbal reasoning abilities for students in grades K-12 in the United States. The normative samples for each edition of the test were large and nationally representative, making the data appropriate for investigations of both means and variances. Lohman and Lakin (2009) examined male-female variance ratios and mean differences across grades 3-11 and three forms of the test administered in 1984, 1992, and 2000. However, they did not focus on changes in the distributions over time due to the limitations in identifying trends from only three data points. In this study, an additional cohort was added from the newest form of the CogAT that was administered to a nationally representative sample of U.S. school children in 2011. This study focused primarily on changes in the ability distributions over time both in the overall sample and in the proportions of males and females in the highest and lowest levels of ability.

Previous Research on Secular Trends in the Variability of Men and Women

Researchers have accumulated evidence of secular trends in sex differences at the means of distributions of various abilities, with considerable evidence pointing to decreasing mean differences between men and women in many domains through the 1970s and 1980s (Feingold, 1988; Friedman, 1989; Hyde, Fennema, & Lamon, 1990). One notable exception (at least in the late 1980s) was persistent male advantages at the highest levels of mathematical achievement (Feingold, 1988). This has led to a general belief that overrepresentation of men at the extremes of the ability distribution should also be diminishing. Group mean differences clearly influence

overrepresentation at the extremes of the distribution, but must be understood independently of variability differences to account for overrepresentation at the extremes.

In a study focused on domains comparable to those of CogAT, Hedges and Nowell (1995) analyzed five large, nationally representative datasets that were collected from the 1960s to the early 1990s as well as the NAEP longitudinal trend data from 1971-1992 (1978-1992 for mathematics data). Most of the samples were gathered from high school aged students. In the large scale datasets, they found mean effect sizes of $d = -0.18$ to 0.0 (favoring girls) for reading comprehension, 0.03 - 0.26 (favoring boys) for quantitative measures, and -0.22 and 0.04 differences for nonverbal reasoning measures. These differences were consistent with later research, but showed no trends over time across testing programs. Likewise, the variance ratios² for the testing programs (besides NAEP) were consistent over time and with other research. Specifically, males were found to be more variable in almost all domains, with slightly more variability in quantitative domains ($VR = 1.05$ to 1.25) than reading ($VR = 1.03$ to 1.16) and nonverbal reasoning ($VR=1.04$ and 1.15). In contrast, the NAEP testing program did show small trends in variance ratios over time. In their analyses of NAEP long-term trend data, Hedges and Nowell found that mean differences were consistent over time (approx. $d = 0.15$ from 1978-1992) and that variance ratios showed a slight, but consistent, increasing trend.³

Although studies of large, representative samples are critical to studying sex differences in the distribution of ability, complementary studies focused on the extremes of the distribution are also needed to understand the full distribution of the ability (Feingold, 1992) as well as to

² A variance ratio (calculated as the ratio of male variance to female variance) greater than 1.0 indicates that males were more variable than females. Feingold (1992) suggested that a variance ratio of 1.10 or greater would be of practical importance on these types of tests.

³ Using the NAEP Data Explorer (National Center for Educational Statistics, no date), the trend can currently be extended up to 2008. The results indicate that mean differences continue on a similar trajectory and variance ratios continue to be somewhat elevated compared to the early 1980s.

illuminate the male-female differences that exist among those competing in elite academic and professional fields (e.g., Benbow, Lubinski, Shea, & Eftekhari-Sanjani, 2000; Halpern, Benbow, Geary, Gur, Hyde, & Gernsbacher, 2007). Research has shown that increasingly elite levels of reasoning are associated with greater likelihood of engaging in STEM careers and creative occupational outcomes such as patents (Park et al., 2007; Wai et al., 2005). Thus, study of sex differences at high performance levels may illuminate sex differences in engagement with elite STEM fields.

In one of the few studies focused on longitudinal trends only on the right tail of ability, Wai, Cacchio, Putallaz, and Makel (2010) found that the differences in male-female ratios in the right tail have substantially decreased in mathematical ability between the early 1980s and early 1990s. Wai et al. (2010) analyzed an impressive dataset based on high-performing young students: the Duke University Talent Identification Program 7th Grade Talent Search (Duke TIP; Putallaz, Baldwin & Selph, 2005). The Duke Talent Search recruits students who score in the top 5% on various school-administered standardized tests, which were administered as part of their school's regular assessment programs, to take the SAT or ACT in grade 7 in order to compete for a place in an enrichment program. Only students who responded to the opportunity to compete (and therefore completed the SAT/ACT) were included. Both the ACT and SAT are designed for use in college admissions among high school juniors and seniors and thus are expected to differentiate well among younger students with significant academic talents in one or more domains.

Wai et al. (2010) found a sharp decline in the ratio of males to females at the highest levels of mathematical ability (as measured by the SAT) among seventh grade students. In the early 1980s, the male-female ratio among the top 1 in 10,000 performers (0.01%) was an

astounding 13.5:1 (13.5 boys for every 1 girl in the top 0.01%), but declined rapidly through the decade to remain stable at about 4 : 1 during the 1990s and 2000s. Wai et al. found that the ratio rapidly declined with less stringent cutoffs as well: In the top 1%, the ratio started at 1.4 in the early 1980s and declined to only 1.1 in the most recent cohort (2006-2010). For Wai et al.'s measures of verbal ability (the reading battery from the SAT), the ratio for the top 5 in 10,000 (.05%) appeared to decline modestly as well, from 1.2:1 in the early 1980s to 1:1 in the 2006-2010 cohort. For other measures of verbal and writing ability from the SAT and the ACT, the ratios (initially showing greater proportions of girls) appeared to decrease slightly (closer to parity in boys and girls) during the last 20 years. ACT-Science did not show any clear trends.

These findings are provocative, indicating that perhaps girls are gaining ground at the most elite levels of mathematics performance. However, there are limitations to their study. First, because their full sample (top 5%) was selected based on high performance in *any* tested area, the male-female ratio in the top 5% was influenced by male-female differences across the academic domain (something more akin to general ability), not exclusively math. Thus, we cannot assume that their estimates at that level of selectivity reflected the influence of a single domain. Second, and perhaps more importantly, their sample was based on students who volunteered to participate in additional testing for the opportunity to be selected for a summer enrichment program. This could have had a number of effects on the type of students who volunteer and their motivation levels relative to a random sample studied in the ordinary school context (Hedges & Nowell, 1995).

This study expanded on Wai et al.'s (2010) findings by exploring the changes in the proportion of boys and girls in the far left and right tails of the distribution using a different measure of reasoning abilities and with student samples that were administered the tests

universally and in lower stakes environments. It also continued Hedges and Nowell's (1995) analyses by considering another large and nationally representative sample that can further illuminate trends over time.

Previous Findings on the CogAT and the Current Study

A previous analysis of the CogAT (Lohman & Lakin, 2009) focused on international comparisons and changes in group differences across age cohorts. Lohman and Lakin (2009) found a slight trend of group mean differences on the verbal battery within test year with a small (-.1 to -.2) difference favoring girls at grade 3 and near-zero differences by grade 6 and above. Mean differences in the quantitative and nonverbal batteries varied across grades, but with no clear trends. With just three time points measured (Form 4 in 1984, Form 5 in 1992, and Form 6 in 2000), trends in ratios over time were difficult to discern and so the authors concluded that there were no clear trends over time. Another study (Keith, Reynolds, & Lakin, 2009) compared sex differences in variability using a latent trait model for the three domains and g and found greater male variability for males on latent variables as well. This study expanded on this previous work by adding an additional form with data on a nationally representative sample of U.S. school children assessed in 2011. The following research questions were addressed:

1. Did male-female mean differences in three reasoning domains change across the four form/time cohorts?
2. Did male-female variance ratios in three reasoning domains change across the four form/time cohorts?
3. Did the ratios of male-to-female students scoring in the tails of the distributions of the three reasoning domains (top and bottom 1%, 2%, and 5%) change across cohorts?

Methods

This study relied on the national standardization data from the 1984 (Form 4), 1992 (Form 5), 2000 (Form 6), and 2011 (Form 7) editions of the Cognitive Abilities Test (CogAT; Lohman, 2011; Lohman & Hagen, 2001; Thorndike & Hagen, 1984, 1992). For simplicity, the four forms are referred to as CogAT 4, CogAT 5, CogAT 6, and CogAT 7. For CogAT 4 through 6, only the data for levels A-G of the test (administered to students in grades 3-11) were included. The primary battery (grades K-2) was excluded because it uses a different set of tests that measure somewhat different abilities. Level H (grade 12) was also excluded because of the comparatively small sample size. For CogAT 7, the naming convention and grade organization of test levels was altered (Lohman, 2012a). For consistency with the other forms, the data for grades K-2 and 12 for CogAT 7 were omitted from this study.

The student samples used in the standardizations of CogAT 4 to 7 were drawn using a stratified random sample of public and private schools and a two-stage stratified random sampling plan of Catholic schools. The sampling units (school buildings) were sampled within strata defined by region of the country (four levels), school-district size (five levels), and school socioeconomic status (five levels). Around 400 schools were sampled for CogAT 4 to 6 while 250 were sampled for CogAT 7. Randomly selected schools within each stratification level were asked to participate. Within participating schools, all students in relevant grades were administered the test. For the analyses that follow, sample weights were used. These weights were based on the stratifying variables (region, size, and SES) to achieve a representative sample of U.S. schools. The students in the sample were found to be representative of that population, though weights did not adjust for individual characteristics (Lohman, 2012a; Lohman & Hagen,

2002). The sample sizes at each test level in the four standardization samples are shown in Table 1.

As Lohman and Lakin (2009) noted, these analyses of test forms generalize across time and changes in the demographic makeup of the United States. In Table 2, which shows the racial/ethnic composition of the norm samples, there appears to be a sharp decline in the proportion of non-Hispanic white students and a remarkable increase in the number of Hispanic students in the 2011 standardization sample. This was reflective of U.S. population shifts (U. S. Census Bureau, 2011). Such changes seemed likely to affect the overall performance of students in the CogAT norms sample, particularly as the U.S. Hispanic population consists of a considerable number of English learners with various levels of English proficiency (Federal Interagency Forum on Child and Family Statistics, 2011). However, such changes seemed unlikely to impact the *relative* performance of boys and girls which were the focus of this study.

It is also important to note that school administrators determined whether it was appropriate to administer the CogAT to students with or without accommodations. The definition and size of these populations have changed over time. English learners comprised 0.8% of the CogAT 5 sample, 4.0% of the CogAT 6 sample and 2.8% of the CogAT 7 sample.⁴ Of the students classified as English learners, just a fraction (around 18% for CogAT 6 and 7) received accommodations. Students with learning disabilities (as defined by the school district) comprised 3.6% of the CogAT 5 sample, 6.0% of the CogAT 6 sample and 7.0% of the CogAT 7 sample. In CogAT 6 and 7 data, 32 and 48% of these students, respectively, received at least one accommodation while taking the test.

⁴ Earlier technical manuals were less thorough in describing the sample. CogAT 4 did not report information on English learners and Special Education in the technical manuals. CogAT 5 reported partial information. CogAT 7 authors particularly sought English learner students for the norms group because the primary battery was designed to be less language loaded.

[Table 1]

[Table 2]

Measures

CogAT was designed to measure the full range of reasoning abilities that define general fluid reasoning (Gf). Each form and level of the CogAT consists of a verbal, quantitative, and nonverbal battery. The choice of batteries is supported by Carroll's (1993) factor analytic work which showed that the Gf factor is defined by three reasoning abilities: (a) *sequential reasoning*—verbal, logical, or deductive reasoning; (b) *quantitative reasoning*—inductive or deductive reasoning with quantitative concepts; and (c) *inductive reasoning*—typically measured with figural tasks. These correspond roughly with the three CogAT batteries.

CogAT 4 through 6 all used the same item formats in each of nine subtests. Three subtests measured verbal reasoning (Verbal Classification, Verbal Analogies, Sentence Completion); three measured quantitative reasoning (Quantitative Relations, Number Series, and Equation Building), and three measured figural/nonverbal reasoning (Figure Classification, Figure Analogies, Figure Analysis). For CogAT 7, all of the verbal and figural formats were retained, but two of the subtest formats for the quantitative battery were replaced. Specifically, Equation Building and Quantitative Relations (the latter of which had some verbal loading; both of which were somewhat speeded; Lohman & Hagen, 2002) were replaced with Number Analogies (a format used by the British version of the CogAT [called the CAT]) and a new format called Number Puzzles. The Number Analogies test applied the traditional analogy format to quantitative relationships (e.g., [2:4] [3:6] [4:?]). The Number Puzzles format required examinees to determine the numerical value(s) represented by one or more geometric shapes that will make one or more equations true (Lohman, 2012a). The new formats had no verbal loading

and were given in contexts that are less speeded than the formats they replaced. One of the substituted formats, Quantitative Relations, required students to identify which of two quantities or concepts (e.g., a quarter and a dollar) is greater. The other substituted format, Equation Building required students to combine a set of numbers and operations (e.g., $3 \ 5 \ 6 + x$) to mathematically yield one of the answer choices. Within each battery, the number of items and their difficulty remained approximately the same across forms.

Items on each test form were developed through an extensive tryout process that included screening for difficulty, discrimination, and, on recent forms, for differential item functioning (DIF) using a Mantel-Haenszel procedure (Lohman, 2012a). In reviewing early forms, it was clear that earlier editions of the CogAT (CogAT 4 especially) were not subjected to the level of fairness review that has become common. The influence of cultural knowledge (American idioms, regional knowledge, etc.) and reading comprehension appeared to be greater on these early forms, though this is a subjective impression. Despite this, it seemed unlikely that a greater cultural bias would impact the *relative* performance of boys and girls.

For CogAT 6 (Lohman & Hagen, 2002), DIF analyses indicated that, across all levels, 3 verbal items favored males and 4 items favored females⁵; one quantitative item favored males (but only at one of four grade levels in which it appeared), none favored females; and no nonverbal items showed DIF. For CogAT 7, only verbal items were found to show DIF with roughly equal numbers favoring males and females (8 and 7, respectively, across all levels; less than 1% of items; Lohman, 2012a).

⁵ These items showed moderate DIF and were balanced to favor boys and girls. These seven items yielded 17 DIF results because items were repeated across levels. This was less than 1% of the 1,900 comparisons made. Only level G of CogAT 6 had more than 2 items showing DIF on the entire Verbal Battery. These items were never among the most challenging items for a given subtest and rarely among the easiest, making them unlikely to impact analyses of the tails of the distributions.

Items within each battery were independently scaled to create a unidimensional, cross-grade scale for each battery. CogAT 6 and 7 used a Rasch model (with fit statistics used in selecting items) to develop a unidimensional scale for each battery (Lohman, 2012a; Lohman & Hagen, 2002). For the verbal, quantitative, and nonverbal scores, K-R 20 reliabilities are typically around .95 (Lohman, 2012a; Lohman & Hagen, 2002; Thorndike & Hagen, 1987, 1997). Research has shown that scores on CogAT 6 correlate with IQ scores from individually administered ability tests about as well as the IQ scores from different individually administered tests correlate with each other (Lohman, 2003a, 2003b).

Analyses

Using the weighted standardization data from each form of the CogAT, the scale score distributions for each battery were used to calculate Cohen's *d* statistic (Cohen, 1988) for mean differences. Variance ratios were calculated by dividing the male score variance by female score variance. Finally, each distribution was divided into percentiles and the proportions of males and females at the top and bottom 1, 2, and 5% were calculated.⁶

Choice of methodology is important in the estimation of sex differences in mean and variances. Using observed scores rather than latent models supports comparison to the most relevant research to this study, but also is appropriate when the applications of the test scores (college admissions, choice of majors) is made by individuals using those observed scores. Although some studies have found that variability differences are eliminated in latent models (Irwing, 2012), others have confirmed their existence at the latent trait level (Keith et al., 2009). Thus, it was not expected that variability differences are an artifact of using observed scores, though additional research of latent trends would be of interest.

⁶ Though not reported in this paper, standard errors of proportions by stanine were calculated and used to evaluate the statistical significance of trends observed.

There were two longitudinal trends of potential interest. One was a change in the mean or variance differences as students get older (comparing grade cohorts within time cohorts) and the other was a change across time cohorts (secular trends), indicating trends over time across students in the same grades. Although both trends are important, the present study focused on secular trends in the male and female distributions across the 1984, 1992, 2000, and 2011 time cohorts.

Results

Cohen's d (1988) effect sizes for mean differences (male minus female) for the verbal, quantitative, and nonverbal batteries by test level and test form are reported in Table 3 and shown graphically in Figure 1. Across forms, for the verbal and nonverbal batteries, the mean differences for CogAT 7 fell within the range of the differences for earlier forms at all levels. Thus, the mean differences on these batteries, though variable across levels and form, did not form a clear pattern over time. The average mean difference for the verbal battery ranged from -.04 to -.14 SD across forms, indicating a small female advantage overall. The average mean difference for the nonverbal battery ranged from -.08 to .01 SD indicating mostly negligible differences between males and females. These results are consistent with previous research (see Ceci et al., 2009, for a review).

[Table 3]

[Figure 1]

For the quantitative battery, the differences were somewhat larger for all test levels for CogAT 7 compared to previous forms. The form with the second largest mean differences was CogAT 4, thus indicating no longitudinal trend across forms. The difference for CogAT 7 was

most pronounced at level D/12 ($d = .19$). Overall, a small male advantage was found (average d ranging from .05 to .15 across forms), consistent with previous research.

Variance Ratios

The male-female variance ratios are presented in Table 3 and plotted in the three panels of Figure 2. A variance ratio greater than 1.0 indicates that the scores for males were more variable than the scores for females. Feingold (1992) suggested that a variance ratio of 1.10 or greater would be of practical importance on these types of tests. For the verbal battery, the variance ratios for CogAT 7 fell within the range of the ratios for earlier forms at all levels. As with the mean differences, few trends across test levels (within form) were found. The variance ratios for the verbal battery ranged from 1.09 to 1.17 across forms (with no clear trend over time), indicating that males were 9-17% more variable than females (to a slight degree).

[Figure 2]

For the quantitative battery, the variance ratios were somewhat higher for all test levels for CogAT 7 compared to previous forms and formed a clear trend of increasing variance across CogAT 4 to 7 (though without a significant increase between CogAT 5 and 6). The male-female variance ratio started at 1.21 for CogAT 4 (meaning boys were 21% more variable) and grew to 1.53 (53% more variable) on CogAT 7, indicating substantially greater variability in quantitative reasoning scores for boys by 2011. The most marked differences were at levels C/11 and D/12 where boys were 60 and 75% more variable than girls in their quantitative scores, respectively. A disparity in variances between boys and girls of this size would be expected to significantly impact the number of boys and girls found at the extremes of the distribution, all else being equal.

For the nonverbal battery, the variance ratios were somewhat larger than for verbal but smaller than quantitative. Although there was considerable variability across test levels in Figure 2, when the variance ratios are averaged in Table 3, there appeared to be a slight increasing trend from boys being 15% more variable on CogAT 4 to 23% more variable on CogAT 6 and 7, echoing the findings from the quantitative battery.

Male-Female Ratios in the Two Tails of the Distribution

To directly compare CogAT data to previous studies on male-female differences in the tails of the distribution, percentile ranks were calculated by test level and used to calculate the ratio of males to females at the top 1, 2, and 5% as well as the bottom 1, 2, and 5% of the distribution in each of the three domains. See Tables 4 and 5, which include ratio estimates from directly comparable research (Machin & Pekkarinen, 2008; Wai et al., 2010). See also Figure 3.

[Table 4]

[Table 5]

[Figure 3]

As expected, the increasing male-female variance ratios for quantitative reasoning were reflected in quite extreme ratios of males to females within the top 1% of quantitative scores. For CogAT 4 (1984), the male-female ratio in the top 1% was about 2:1. By 2011, CogAT 7 showed ratios of nearly 3:1 in the top 1%. As seen in previous research, higher cutoffs resulted in greater disparities in the male-female ratio. However, Table 4 indicates that the ratios observed at the top 1% for all forms of the CogAT were considerably higher than the ratio estimated by Wai et al. Explanations for this discrepancy are explored in the discussion section. The ratio at the top 5% based on CogAT data was consistent with Machin and Pekkarinen's (2008) ratio estimate based on the Programme for International Student Assessment (PISA; governed by the Organization

for Economic Cooperation and Development or OECD) data for the same year for 15-year-olds in the United States.

At the bottom 5%, there was a similar trend of increasing ratios across all four forms on quantitative reasoning (from 1.1 on CogAT 4 to 2.1 on CogAT 7). The ratio found for the bottom 5% on PISA Mathematics by Machin and Pekkarinen (2008) was slightly lower than what was found on CogAT data (1.2 on PISA vs. 1.5 on CogAT). In contrast to the right tail, higher cutoffs (1 and 2%) were not consistently associated with higher ratios.

For the verbal battery, Table 4 highlights a result that was not especially clear in the preceding analyses of mean differences and variance ratios. Specifically, there appeared to be a substantial drop in the ratio of males to females from CogAT 4 to the later forms, where near parity was observed. The similarity between the observations in this study and that of Wai et al. (2010) of a similar drop on SAT-V scores may indicate that this was a generalizable result.

For the nonverbal battery as well, a trend was apparent in the bottom 5% of increasing male-female ratios, similar to, but smaller than, the trend for quantitative reasoning. This effect was not apparent in the right tail. It was consistent with the slight trend of increasing variance ratios observed in Table 3.

Discussion

Consistent with previous research, girls had slightly higher mean scores on the verbal battery across forms. Small differences in variance (with boys being 9-17% more variable) were found which resulted in nearly equal numbers of males and females at the highest score levels and more males at the lowest score levels.

Previous studies offer estimates of male-female ratios which are useful in contrasting the results of CogAT to other testing programs. In the verbal domain, the longitudinal findings in

this study were consistent with Wai et al.'s (2010) in that the CogAT data showed a drop in male-female ratios in the early 1980s to move from a slight advantage to males to a ratio of parity or slightly favoring females by the 1990s and 2000s. As with Wai et al., these changes were consistent across increasingly stringent cutoffs (5, 2, and 1% in this case; see Table 4). The overall size of the ratios observed at the top 1% level was also similar to Wai et al. (1.06 vs. 0.94 in the late 2000s). Hedges and Nowell (1995) did not find a trend in mean differences over time on NAEP and other large-scale testing programs, but their estimates of near-parity were similar. When contrasting the results found in this study to Machin and Pekkarinen (2008), it appeared that at the top 5% level, CogAT ratios were much closer to parity than the PISA data (1.04 vs. 0.60, respectively). Thus, only the PISA data showed ratios that strongly favor girls. Overall, it appeared that there are few, if any, differences in male-female ratios at the high end of the verbal domain.

At the bottom 5% level (see Table 5), CogAT estimates were again more conservative than Machin and Pekkarinen (2008; 1.67 compared to their 2.45 estimate). Hedges and Nowell's (1995) data seemed to comport with CogAT findings (range of male-female ratios = 1.07 to 1.75), indicating that boys were overrepresented at the lowest verbal reasoning levels.

Nonverbal Domain

For nonverbal reasoning, mean differences were nearly non-existent. However, greater male variability was again observed, with males showing 15-23% greater variance than females. This resulted in consistent male-female ratios (with 1.2 to 1.3 males for every female) at both the top and bottom of the distribution of nonverbal reasoning scores. There was a slight trend of increasing variance ratios which primarily affected the left tail of the distribution.

The only other study that reported estimates for a nonverbal domain, Hedges and Nowell (1995), found similar ratios at the bottom of the distribution, with male-female ratios of 1.26-1.92 for CogAT at the bottom 5% and ratios of 1.0 and 1.49 for Hedges and Nowell at the bottom 10%. At the high end, CogAT yielded ratios that were consistently around 1.33 and higher than those reported by Hedges and Nowell (.67 and 1.0). The source of this discrepancy was unclear, given the differences in the tests used, age of the sample, and the time elapsed between Hedge and Nowell's samples and those of CogAT.

Quantitative Domain

For the quantitative battery, the ratios of males to females in the top and bottom stanines for quantitative reasoning was clearly larger for CogAT 7 and appeared to form part of an increasing trend across the four forms. However, the distributional change for CogAT 7 was not clearly linked to either changes in variance or mean differences, but appeared to involve changes in both.

On the one hand, the variance ratios for quantitative reasoning showed a small, but consistent, increasing trend across forms. This supported the conclusion that the increasing ratios of males to females at the right and left tail is due to increasing variability of males rather than growing mean differences. A small, consistent increasing trend is a good indication of a steady shift in the population being tested in terms of their true quantitative reasoning skills. The important fact that the trend found in the CogAT data was in direct opposition to the trend demonstrated Wai et al. (2010) will be explored later.

In contrast to the variance ratios, the mean differences in Figure 1 do not indicate a consistent trend. CogAT 7 showed mean differences favoring males that were markedly larger than the previous forms (though still below the cutoff for "small" effect sizes; Cohen, 1988),

with CogAT 4 having the second largest mean differences. Jumps in the mean differences as seen for CogAT 7 would be more consistent with changes in the test than with a sudden shift in quantitative reasoning among schoolchildren.

In the quantitative domains, CogAT-based findings were starkly different from Wai et al. (2010). Rather than a rapid decline in the male-female ratios from the early 1980s, the CogAT data revealed an increasing trend in the ratios and, overall, showed ratios in the top 1% that were much larger than Wai et al. found. By the late 2000s, the ratio for the top 1% of quantitative scores for CogAT was 2.77 while Wai et al. reported a ratio of 1.10 for SAT-M. PISA data seemed to confirm the CogAT estimates in the right tail where Machin and Pekkarinen (2008) reported ratios for the top 5% that were closer to those obtained on CogAT (1.72 vs. 2.02, respectively). The NAEP long-term trend data analyzed by Hedges and Nowell (1995) also confirmed these finding through the 1990s by noting a slight, but consistent, increasing trend in variance ratios for the mathematics test. They did not estimate male-female ratios for the right tail of the distribution for that test.

All of the studies confirmed that, unlike the verbal domain, the differences in male-female ratios in the quantitative domain are magnified as increasingly stringent cutoffs are used. For example, on CogAT 7, the male-female ratio for the top 5% of quantitative scores was 2.02 while at the top 1% it was 2.77. Wai et al. (2010) found even more striking differences, especially in earlier years (surging up to 13.5:1 with the most stringent cutoffs in the early 1980s).⁷ Hedges and Nowell (1995) found in the 1960s Project Talent data (the only dataset in their study with sufficient score ceilings to support such estimates) that male-female ratios were 1.3 in the top 10%, 1.5 in the top 5%, and 7.0 in the top 1%. Although the magnitude of changes

⁷ Differences appear stable in the 1996-2010 year range, which was confirmed by their analyses of 5th and 6th grade samples in Wai and Putallaz (2011).

in the male-female ratios varied considerably across studies, such a trend will have direct implications for understanding the low number of women observed in elite mathematics fields.

At the other end of the distribution, CogAT data for quantitative reasoning also indicated an increasing ratio of males to females among the lowest scoring students. The point estimates from CogAT 4 to CogAT 6 (1.1-1.5 for the bottom 5%) were consistent with estimates from Hedges and Nowell (1995; ranging from 1.33 to 1.90 in the bottom 10% across studies) and from Machin and Pekkarinen (2008) (1.21 in the bottom 5%). The estimate from CogAT 7, however, of 2.07 in the top 5% was markedly larger, indicating that the trend of increasing ratios was apparent at both ends of the CogAT 7 distribution.

Differences in the ratios at the top 1% level for the CogAT data compared to Wai et al. (2010) may be a function of sampling methods, which for their study starts at the 5% level with similar numbers of males and females likely due to the selection process being based on high scores in any academic domain (they did not indicate that any effort was made to identify equal numbers of boys and girls). Because the CogAT samples came from the entire range of abilities, there was no such selection pressure on the ratios of talented males and females. Wai et al. also estimated their percentiles using sample statistics from selected talent search participants (Lubinski et. al., 2001), so the true population percentiles could be different from those reported in the study. In fact, because Wai et al. used a more liberal either/or cut-off approach to identification (Lohman, 2012b), in other words allowing students to qualify on the basis of any of a number of tests they completed, measurement error may have played a significant role in selection, making it more likely that their sample was closer to the top 10% than the top 5%. If this was the case, and the results reported by Wai et. al. for the top 1% actually reflected the top 5% population percentile, their ratios were quite comparable to those from CogAT.

An important qualification to comparing the CogAT findings to other research was that the distributions presented here combine the performance of students in grades 3 to 11 (roughly ages 9-17) while Wai et al. (2010) studied only 7th grade students and Machin and Pekkarinen (2008) studied only 15-year-old students. An inspection of variance ratios and mean differences among 15-year-olds (level F/15) and 7th grade students (level E/13) finds that these grade levels were not exceptional compared to other grade levels. Thus, the difference between this study's findings based on the combined group and other studies based on single age/grade groups was not likely to be due to differences in the age groups studied.

Another important qualification in comparing the results of the various studies was a possible ceiling effect on CogAT scores. Wai et al.'s (2010) data was based on tests that were designed for much older examinees and thus their tests would be expected to allow greater variability among highly able 7th grade students. The CogAT forms, although they show sufficient ceiling for their intended uses, did not have such extensive ranges at the top and bottom of the distribution (particularly beyond the 1st and 99th percentiles). Thus, it seems likely that the ratios reported from CogAT forms may be lower-bound estimates of the differences in variability between males and females.

Potential Explanations for Changes in the Quantitative Reasoning Distribution

It seemed likely that changes in test design—specifically, the introduction of two new item formats—could account for the greater mean differences observed for quantitative reasoning on CogAT 7 and may partially explain the higher male-female ratios in the tails of the distribution for that form. There were a number of possible mechanisms. One was that the Quantitative Relations format which was dropped from CogAT 7 required some verbal load while the new formats (Number Puzzles and Number Analogies) did not. This substitution could

have the effect of increasing mean differences given that girls often have an advantage in verbal domains (Ceci et al., 2009; Feingold, 1988). Likewise, the Equation Building format, which was also replaced on CogAT 7, assumed automatized basic computation skills (from addition through division). Both the reliance on computational skills and the stronger tie to the school curriculum (compared to the formats which replaced Equation Building) may also have advantaged girls on earlier forms (Friedman, 1989).

It seemed likely, therefore, that changes in the test design have had an impact on the distribution of males and females on CogAT 7. However, the consistent trend of increasing variance ratios from CogAT 4 to CogAT 7 could not be similarly explained by test design and indicated that something more systematic was affecting quantitative reasoning to increase both the variance ratios and the male-female ratios in the top and bottom stanines of the distribution. These changes most likely reflected some combination of changes in society in terms of educational opportunity and personal changes in the educational choices of boys and girls. A number of researchers have offered such explanations for decreases in mean differences and ratios (e.g., increased mathematics course-taking by girls; Ceci et al., 2009; Wai et al., 2010). However, it was difficult to imagine what changes in the cultural context of the United States and/or its schools could account for an *increase* in the number of boys relative to girls scoring at the highest levels on the quantitative battery of the CogAT. Mechanisms offered by other researchers to explain observed decreases in sex differences over the same time period would seemingly affect quantitative reasoning skills as measured by the CogAT in the same way.

Two important contrasts between the CogAT data and the Duke TIP data may offer potential explanations for the discrepancies in trends between this study and Wai et al. (2010).

One was the type of items being used in the assessments and the other was the motivation level of students.

First, the CogAT test formats call on skills that were less clearly related to the tasks and skills that are taught in the typical K-12 curriculum. This may have allowed examinees to use more inventive strategies in solving problems. In contrast, ACT and SAT tasks were typically much more curriculum-related, which would permit the application of well-rehearsed strategies. Researchers have argued that girls have an advantage (or less of a disadvantage, in some cases) when the content being assessed is more directly tied to school curriculum and tasks (Ceci et al., 2009; Else-Quest, Hyde, & Linn, 2010; Gallagher & DeLisi, 1994; Fennema, Carpenter, Jacobs, Franke, & Levi., 1998). Thus, systematic changes in the course taking patterns of girls (Feingold, 1988; Finn, 1999) would be more likely to affect those tests. Historical changes in the SAT to make the test more directly tied to instruction (Lawrence, Rigol, van Essen, & Jackson, 2004) would also exacerbate this effect and could therefore explain the decrease in male-female ratios Wai et al. observed.

Another important distinction between the two studies is the level of motivation of the examinees. When administered for standardization purposes, performance on the CogAT is usually “low stakes” for the students because schools do not use the standardization results for making educational decisions. In contrast, for the students participating in the Duke Talent Search, the results are relatively high-stakes and motivating to students because high scores result in the opportunity to take part in enrichment programs.

That said, differences in motivation were not enough to explain differences in ratio trends, because there was no reason to expect this has changed over time for CogAT or Duke TIP. However, motivation becomes critical when interests are also considered. Specifically,

Ackerman and Heggestad (1997; see also Ackerman & Kanfer, 2004) found that personality and conative traits (like interest) were more strongly associated with “typical” test performance (where typical can be equated to a low-stakes testing environment) than “maximal” test performance (or high-stakes testing). Given the well-established finding of sex differences in vocational interests (with women preferring people to things and men vice versa; Ceci et al, 2009; Lubinski & Benbow, 1992, 2006), one possible explanation for the disparity in the trends is that, although girls rapidly increased in the number of mathematics courses taken, raising their performance on high-stakes achievement-oriented tests, the persistent differences in interests or preferences for more verbally oriented domains may have continued to affect their performance when taking low-stakes ability tests under typical (rather than maximal) motivation.

Although one can speculate further about the causes of these shifts (for example, differences in time spent on video games that may build relevant reasoning skills, Else-Quest et al., 2010; Terlecki & Newcombe, 2005), the data considered in this study could not directly test hypotheses about *why* these changes have occurred. Future research is needed to confirm this paradoxical trend in quantitative reasoning and mathematics achievement and explore potential mechanisms. If indeed girls are gaining ground in mathematical achievement (as Wai et al. found), but not in quantitative reasoning (as this study found), it is also important to explore what impact there would be on increasing the participation of women in STEM career fields.

Limitations

No single dataset can provide conclusive evidence on whether male-females differences (means and/or variability) in cognitive ability are declining, in part because of natural limitations of data collection and in part because of the sheer variety of broad and specific cognitive abilities that are studied and the variety of psychometric approaches to measuring those abilities. For

instance, in terms of natural data limitations, the analyses in this study would have been improved by more frequent assessment over the time period under consideration and a wider age range to include young adults (where some claim that sex differences fully reveal themselves; Irwing, 2012). But these are often limitations of using test norms data, which offer other benefits such as being based on large and representative samples. An important limitation to this dataset was that there were changes to the CogAT test development over time, always intended to improve the assessment of cognitive abilities in children, but having the unfortunate result of making longitudinal comparisons of test scores somewhat less reliable. For example, the introduction of two new formats on CogAT 7 seemed to introduce larger mean differences on the Quantitative Battery which partially confounded a longitudinal trend of increasing ratios of males to females in the tails of the quantitative distribution.

An important future area of research is whether latent trait models replicate sex differences that appear in observed score distributions, such as those addressed in this study. On the one hand, observed score analyses are critical because we often make decisions (e.g., schools making educational placement decisions) on the basis on observed scores, not latent traits. On the other hand, observed score analyses do not account for measurement error and hierarchical relationships of abilities. Thus, both approaches to sex differences research are needed. As noted before, the research on sex differences in latent trait models have shown both differences and similarity to the trends identified using observed score analyses (Irwing, 2012; Keith et al., 2009).

Conclusions

This study sought to explore changes in the distribution of cognitive abilities in males and females ages 9 to 17 from 1984-2011 in verbal, quantitative, and nonverbal/figural reasoning

domains. The data revealed small changes in verbal and nonverbal reasoning and more pronounced changes in quantitative reasoning which warrant further investigation and replication in other datasets. The increasing ratio of males at the highest levels of quantitative reasoning was especially surprising and contrary to previous research. Such findings have important implications for the societal goal of increasing the representation of women in elite STEM fields, because quantitative reasoning skills have been shown to be important predictors of participation in such careers (Park et al., 2007; Wai et al., 2005).

It should be noted, as is almost always the case, that the magnitude of differences *between* males and females in the three reasoning domains under study here, even in the tails of the distributions, paled in comparison to the magnitude of differences *within* male and female populations (a point also emphasized by Hyde, 1981). Girls made up a large proportion of students scoring at the highest levels of quantitative reasoning on all forms of the CogAT as did boys on the verbal reasoning tasks. Thus, a student's sex was of little relevance in predicting their individual likelihood of success in a particular field. That said, when considering the widely held goal of increasing the participation of females in the most elite levels of mathematics and science domains, these population-level differences may have an important impact and thus warrant study and concern.

References

- Ackerman, P.L., & Heggestad, E.D. (1997). Intelligence, personality, and interests: Evidence for overlapping traits. *Psychological Bulletin, 121*, 219-245.
- Ackerman, P.L., & Kanfer, R. (2004). Cognitive, affective, and conative aspects of adult intellect within a typical and maximal performance framework. In D.Y. Dai & R.J. Sternberg (Eds.), *Motivation, emotion, and cognition: Integrative perspectives on intellectual functioning and development* (pp. 119-141). Mahwah, NJ, US: Lawrence Erlbaum.
- 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.
- Carroll, J. B. (1993). *Human Cognitive Abilities: A Survey of the Factor-Analytic Studies*. New York: Cambridge University Press.
- Ceci, S. J., & Williams, W. M., (Eds.). (2006). *Why aren't more women in science? Top researchers debate the evidence*. Washington, DC: American Psychological Association.
- Ceci, S. J., Williams, W. M., & Barnett, S. M. (2009). Women's underrepresentation in science: Sociocultural and biological considerations. *Psychological Bulletin, 135*, 218-261.
- Chipman, S. F. (1988). Far too sexy a topic. [Review of J. S. Hyde and M. C. Linn. (Eds.) *The psychology of gender: Advances through meta-analysis*. Baltimore MD: John Hopkins University Press.], *Educational Researcher, 17*, 46-49.
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences*. Hillsdale, NJ: Erlbaum.
- Eagly, A. H. (1995). The science and politics of comparing women and men. *American Psychologist, 50*, 145-158.

- Else-Quest, N.M., Hyde, J.S., & Linn, M.C. (2010). Cross-national patterns of gender differences in mathematics: A meta-analysis. *Psychological Bulletin, 136*(1), 103-127.
- Federal Interagency Forum on Child and Family Statistics. (2011). *America's children: Key national indicators of well-being*. Washington, DC: U.S. Government Printing Office.
- Feingold, A. (1988). Cognitive gender differences are disappearing. *American Psychologist, 43*, 95-103.
- Feingold, A. (1992). Sex differences in variability in intellectual abilities: A new look at an old controversy. *Review of Educational Research, 62*, 61-84.
- Feingold, A. (1994). Gender differences in variability in intellectual abilities: A cross-cultural perspective. *Sex Roles: A Journal of Research, 30*, 81-92.
- Fennema, E., Carpenter, T., Jacobs, V., Franke, M., & Levi, L. (1998). Longitudinal Study of Gender Differences in Young Children's Mathematical Thinking. *Educational Researcher, 27*, 6–11.
- Finn, J. D. (1999, Spring). Opportunity offered—opportunity taken: Course-taking in American high schools. *ETS Policy Notes, 9* (1). Princeton, NJ: Educational Testing Service.
Retrieved from <https://www.ets.org/Media/Research/pdf/PICPNV9N1.pdf>
- Friedman, L. (1989). Mathematics and the gender gap: A meta-analysis of recent studies on sex differences in quantitative tasks. *Review of Educational Research, 59*, 185–213.
- Gallagher, A., & DeLisi, R. (1994). Gender differences in Scholastic Aptitude Test–Mathematics problem solving among high-ability students. *Journal of Educational Psychology, 86*, 204–211.

- 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.
- Hedges, L. V., & Nowell, A. (1995). Sex differences in mental test scores, variability, and numbers of high-scoring individuals. *Science*, 269, 41-45.
- Hyde, J. S. (1981). How large are cognitive gender differences? A meta-analysis using omega-square and d. *American Psychologist*, 36, 892-901.
- Hyde, J. S., Fennema, E., & Lamon, S. J. (1990). Gender differences in mathematics performance: A meta-analysis. *Psychological Bulletin*, 107, 139-155.
- Hyde, J. S., & Linn, M. C. (2006). Gender similarities in mathematics and science. *Science*, 314, 599–600.
- Irwing, P. (2012). Sex differences in g: An analysis of the US standardization sample of the WIAS-III. *Personality and Individual Differences*, 53, 126-131. Doi: 10.1016/j.paid.2011.05.001
- 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(6), 518-531.
- Keith, T.Z., Reynolds, M.R., Lakin, J.M. (2009). *Sex Differences in Latent Means and Variances of Cognitive Abilities: Evidence from the CogAT*. Paper presented at the annual meeting of the International Society for Intelligence Research, Madrid, Spain.
- Kimura, D. (2004). Human sex differences in cognition, fact, not predicament. *Sexualities, Evolution, & Gender*, 6(1), 45-53.

- Lawrence, I., Rigol, G., van Essen, T., & Jackson, C. (2004). A historical perspective on the content of the SAT. In R. Zwick (Ed.), *Rethinking the SAT: The Future of Standardized Testing in University Admissions* (pp. 57-74). New York, NY: Routledge.
- Lohman, D. F. (2003a). *The Wechsler Intelligence Scale for Children III and the Cognitive Abilities Test (Form 6): Are the general factors the same?* Retrieved July 4, 2007, from <http://faculty.education.uiowa.edu/dlohman>
- Lohman, D. F. (2003b). *The Woodcock-Johnson III and the Cognitive Abilities Test (Form 6): A concurrent validity study.* Retrieved July 4, 2007, from <http://faculty.education.uiowa.edu/dlohman>.
- Lohman, D.F. (2011). *Cognitive Abilities Test, Form 7.* Rolling Meadows, IL: Riverside Publishing.
- Lohman, D.F. (2012a). *Cognitive Abilities Test, Form 7, Research and Development Guide.* Rolling Meadows, IL: Riverside Publishing.
- Lohman, D. F. (2012b). [Decision strategies](#). In S. L. Hunsaker (Ed.), *Identification: The Theory and Practice of Identifying Students for Gifted and Talented Education Services.* (pp. 217-248). Mansfield Center, CT: Creative Learning Press.
- Lohman, D.F., & Hagen, E. (2001). *Cognitive Abilities Test (Form 6).* Itasca, IL: Riverside.
- Lohman, D. F., & Hagen, E. (2002). *Cognitive Abilities Test (Form 6) research handbook.* Itasca, IL: Riverside.
- Lohman, D.F., & Lakin, J.M. (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.

Machin, S. & Pekkarinen, T. (2008). Global sex differences in test score variability. *Science*, 322, 1332–1333.

National Center for Educational Statistics (no date). *NAEP Data Explorer*. Retrieved from <https://nces.ed.gov/nationsreportcard/lttdata/>

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.

Putallaz, M., Baldwin, J., & Selph, H. (2005). The Duke University Talent Identification Program. *Journal of High Ability Studies*, 16, 41–54.

Terlecki, M. S., & Newcombe, N. S. (2005). How important is the digital divide? The role of computer and videogame usage to gender differences in mental rotation. *Sex Roles*, 53, 433–441.

Thorndike, R. L., & Hagen, E. P. (1984). *Cognitive Abilities Test (Form 4)*. Itasca, IL: Riverside.

Thorndike, R. L., & Hagen, E. P. (1992). *Cognitive Abilities Test (Form 5)*. Itasca, IL: Riverside.

U. S. Census Bureau. (2011). *Overview of race and Hispanic origin: 2010*. Retrieved from <http://www.census.gov/prod/cen2010/briefs/c2010br-02.pdf>

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(4), 412-423.

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.

Table 1

Sample Sizes by Test Level for Cognitive Abilities Test (CogAT) Forms 4 to 7

Level		Form			
CogAT 4 to 6 (by grade)	CogAT 7 ^a	4	5	6	7
A (3)	9 (3)	7,087	11,343	14,152	6,141
B (4)	10 (4)	13,798	16,953	14,309	6,120
C (5)	11 (5)	13,596	17,717	15,146	6,555
D (6)	12 (6)	14,585	16,055	13,407	5,601
E (7)	13/14 (7-8)	13,264	13,352	12,454	9,669
F (8-9)	15/16 (9-10)	24,819	20,883	18,237	7,912
G (10-11)	17/18 (11)	22,575	13,633	11,234	3,295

Note. ^a CogAT 7 used slightly different organizations of grades compared to CogAT 4 to 6.

Table 2

Percentages by Ethnicity for CogAT 4 to 7

Ethnicity	Form (Year)			
	CogAT 4 (1984)	CogAT 5 (1992)	CogAT 6 (2000)	CogAT 7 (2010)
White (not Hispanic)	80.6	68.1	65.0	55.7
Black	11.6	15.4	16.3	13.3
Hispanic	6.4	7.5	11.5	17.6
Asian/Pacific Islander	1.0	4.2	3.6	5.3
Native American	0.5	1.6	2.5	4.8

Table 3

Cohen's d Measures of Effect Size and Variance Ratios (VR) of Males to Females by Battery, Test Form, and Level of CogAT

Level	Verbal							
	CogAT 4		CogAT 5		CogAT 6		CogAT 7	
	d	VR	d	VR	d	VR	d	VR
A/9	-0.23	1.22	-0.12	1.16	-0.15	1.05	-0.14	0.96
B/10	-0.11	1.23	-0.04	1.22	-0.14	1.14	-0.08	1.07
C/11	-0.03	1.21	-0.05	1.15	-0.15	1.09	0.02	1.23
D/12	-0.01	1.20	-0.10	1.08	-0.07	1.15	-0.07	1.05
E/13	0.02	1.15	-0.11	0.98	-0.04	1.15	-0.07	1.07
F/15	0.01	1.11	-0.07	1.01	-0.10	1.08	-0.05	1.24
G/17	0.02	1.08	-0.04	1.04	-0.09	1.08	0.12	1.32
Total	-0.05	1.17	-0.08	1.09	-0.11	1.10	-0.04	1.13
Level	Quantitative							
	d	VR	d	VR	d	VR	d	VR
	d	VR	d	VR	d	VR	d	VR
A/9	-0.02	1.14	0.02	1.28	0.09	1.25	0.13	1.38
B/10	0.06	1.17	0.07	1.23	0.04	1.25	0.14	1.44
C/11	0.10	1.22	0.05	1.34	-0.01	1.27	0.15	1.60
D/12	0.07	1.19	0.01	1.32	0.04	1.31	0.19	1.75
E/13	0.05	1.29	0.01	1.27	0.06	1.39	0.15	1.45
F/15	0.06	1.19	0.06	1.37	0.03	1.34	0.12	1.55
G/17	0.07	1.27	0.12	1.49	0.08	1.56	0.22	1.51
Total	0.06	1.21	0.05	1.33	0.05	1.34	0.15	1.53
Level	Nonverbal							
	d	VR	d	VR	d	VR	d	VR
	d	VR	d	VR	d	VR	d	VR
A/9	-0.03	1.10	-0.05	1.10	-0.01	1.18	-0.01	1.18
B/10	0.00	1.18	0.01	1.17	-0.06	1.17	0.01	1.16
C/11	0.03	1.17	-0.04	1.18	-0.10	1.18	0.05	1.20
D/12	-0.02	1.14	-0.08	1.17	-0.06	1.32	0.07	1.35
E/13	-0.01	1.18	-0.12	1.14	-0.11	1.20	-0.04	1.10
F/15	0.00	1.14	-0.05	1.19	-0.12	1.25	-0.03	1.33
G/17	0.07	1.17	-0.01	1.31	-0.07	1.32	0.07	1.26
Total	0.01	1.15	-0.05	1.18	-0.08	1.23	0.01	1.23

Table 4

Ratios of Males to Females in the Top 5, 2, and 1% of Scores on Three CogAT Batteries and Related Samples

Cohorts ^a	Verbal/Reading domains					Quantitative/Math domains					Nonverbal		
	CogAT	PISA ^b	CogAT	CogAT	SAT ^c	CogAT	PISA ^b	CogAT	CogAT	SAT ^c	CogAT		
	5%	5%	2%	1%	1%	5%	5%	2%	1%	1%	5%	2%	1%
1984	1.15	--	1.24	1.25	1.11	1.51	--	1.90	2.08	1.43	1.39	1.50	1.52
1992	0.92	--	0.97	0.99	1.00	1.72	--	2.03	2.34	1.26	1.22	1.28	1.32
2000	0.92	0.60	0.88	0.98	0.97	1.73	1.72	2.07	2.37	1.13	1.21	1.30	1.45
2011	1.04	--	1.08	1.06	0.94	2.02	--	2.33	2.77	1.10	1.33	1.39	1.39

a Years correspond to CogAT Forms 4 to 7.

b From Machin and Pekkarinen's (2008) data from the 2003 PISA Cohort.

c From Wai et al. (2010). To obtain comparable cohorts, I used the 1981-1985, 1991-1995, 2001-2005, and 2006-2010 cohorts.

Table 5

Ratios of Males to Females in the Bottom 5, 2, and 1% of Scores on Three CogAT Batteries and Related Samples

Cohorts ^a	Verbal/Reading domains				Quantitative/Math domains				Nonverbal		
	CogAT	PISA ^b	CogAT	CogAT	CogAT	PISA ^b	CogAT	CogAT	CogAT		
	5%	5%	2%	1%	5%	5%	2%	1%	5%	2%	1%
1984	1.84	--	1.58	1.34	1.10	--	1.00	1.08	1.26	1.25	1.19
1992	1.78	--	1.63	1.71	1.36	--	1.46	1.44	1.62	1.58	1.53
2000	1.89	2.45	1.64	1.53	1.50	1.21	1.47	1.47	1.67	1.56	1.57
2011	1.67	--	1.69	1.56	2.07	--	1.65	1.34	1.92	1.66	1.49

a Years correspond to CogAT Forms 4 to 7.

b From Machin and Pekkarinen's (2008) data from the 2003 PISA Cohort.

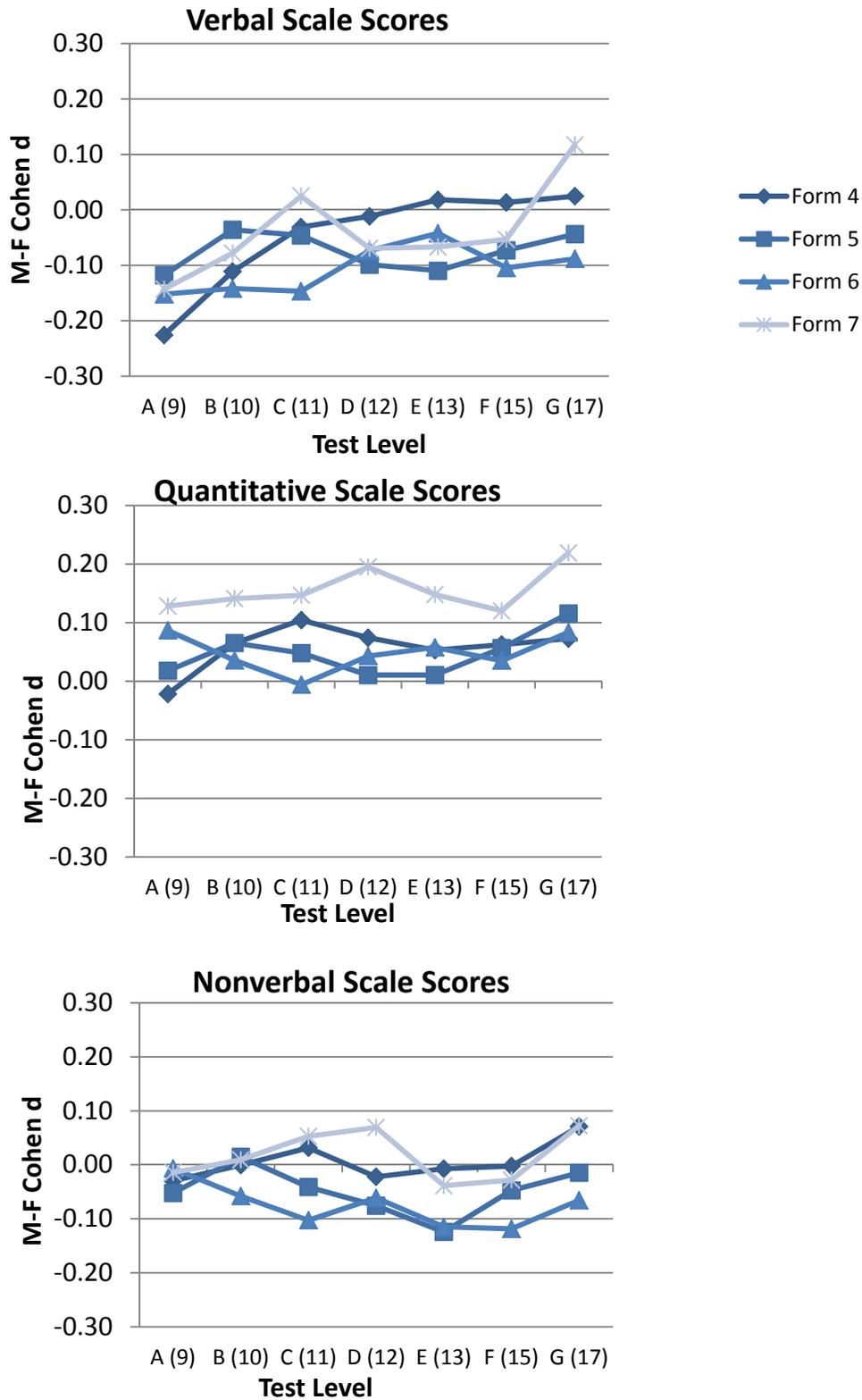


Figure 1. Cohen's d effect sizes across levels and forms. Positive d effects indicate a male advantage at the mean.

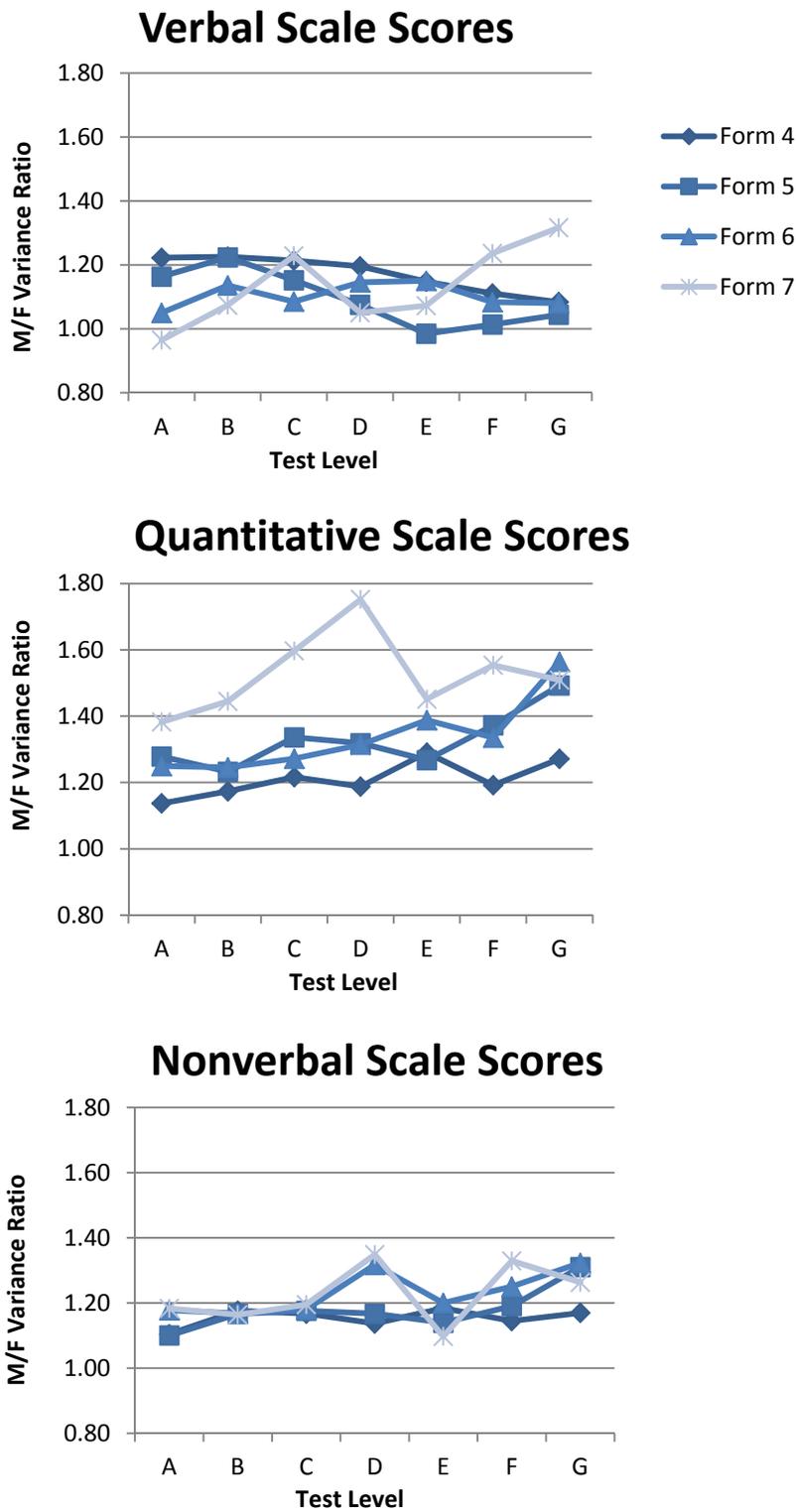


Figure 2. Variance ratios across levels and forms. Values greater than 1.00 indicate greater male variability.

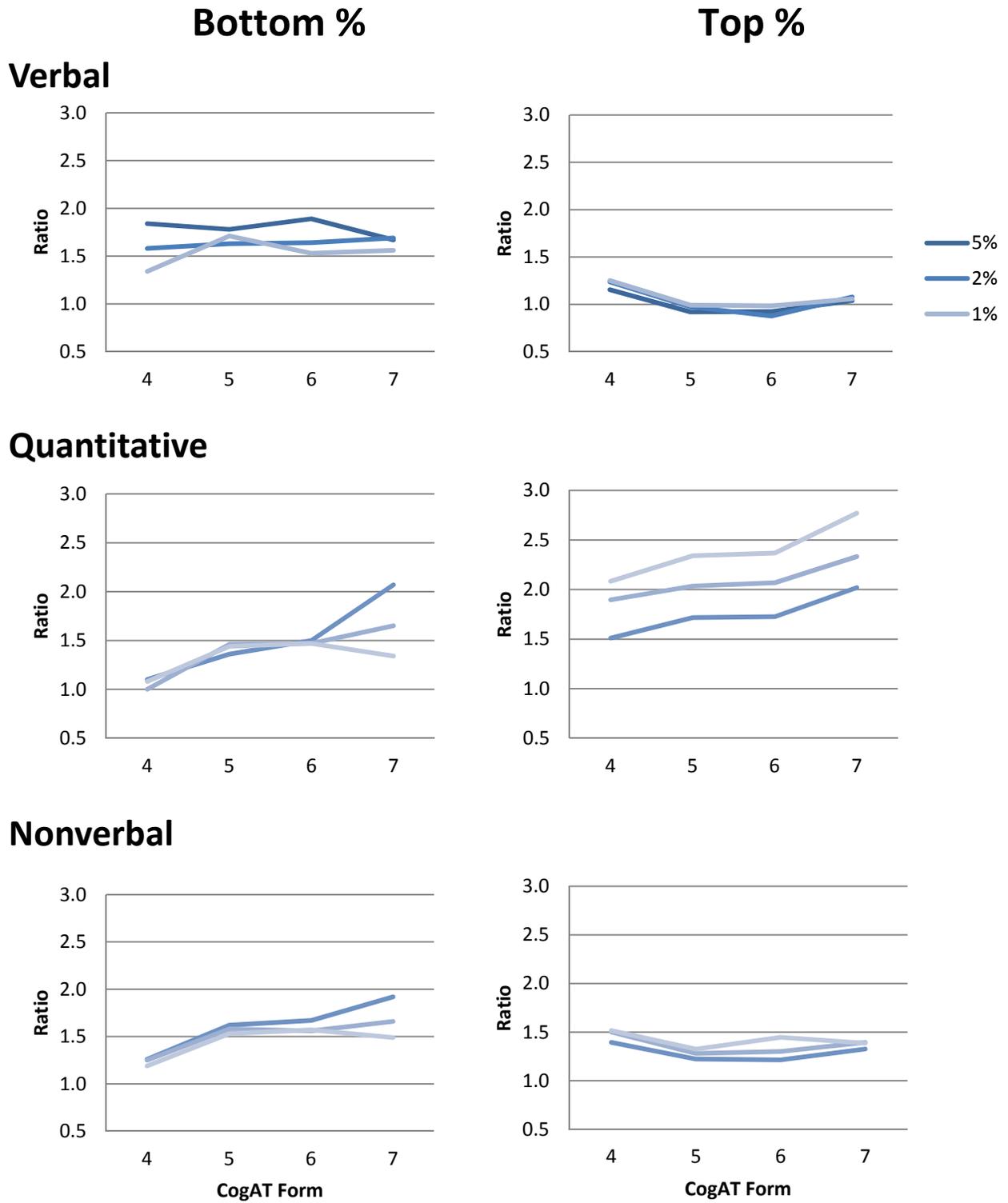


Figure 3. Ratios of males to females at the bottom (left column) and top (right column) 1, 2, and 5%. Values greater than 1.00 indicate greater numbers of males.