American Psychologist

Gender Differences in Reading and Writing Achievement: Evidence From the National Assessment of Educational Progress (NAEP)

David Reilly, David L. Neumann, and Glenda Andrews

Online First Publication, September 20, 2018. http://dx.doi.org/10.1037/amp0000356

American Psychological Association

American Psychologist

0003-066X/18/$12.00

http://dx.doi.org/10.1037/amp0000356

Gender Differences in Reading and Writing Achievement: Evidence From the National Assessment of Educational Progress (NAEP)

David Reilly
Griffith University, Queensland, Australia

David L. Neumann and Glenda Andrews
Griffith University, Queensland, Australia, and Menzies Health Institute Queensland, Australia

A frequently observed research finding is that females outperform males on tasks of verbal and language abilities, but there is considerable variability in effect sizes from sample to sample. The gold standard for evaluating gender differences in cognitive ability is to recruit a large, demographically representative sample. We examined 3 decades of U.S. student achievement in reading and writing from the National Assessment of Educational Progress to determine the magnitude of gender differences (N = 3.9 million), and whether these were declining over time as claimed by Feingold (1988). Examination of effect sizes found a developmental progression from initially small gender differences in Grade 4 toward larger effects as students progress through schooling. Differences for reading were small-to-medium ($d = -.32$ by Grade 12), and medium-sized for writing ($d = -.55$ by Grade 12) and were stable over the historical time. Additionally, there were pronounced imbalances in gender ratios at the lower left and upper right tails of the ability spectrum. These results are interpreted in the context of Hyde's (2005) gender similarities hypothesis, which holds that most psychological gender differences are only small or trivial in size. Language and verbal abilities represent one exception to the general rule of gender similarities, and we discuss the educational implications of these findings.

Keywords: gender differences, reading, writing, literacy, sex differences

Supplemental materials: http://dx.doi.org/10.1037/amp0000356.supp

The question of whether males and females differ in cognitive abilities has been the focus of considerable research in recent decades. While there is a general consensus that males and females do not differ in general intelligence (Halpern, 2000), gender differences are commonly observed for more specific cognitive abilities such as visual–spatial ability (Voyer, Voyer, & Bryden, 1995) and language (Miller & Halpern, 2014). However, Hyde (2005) had proposed the gender similarities hypothesis (GSH), which claimed that males and females "are similar on most, but not all, psychological variables. That is, men and women, as well as boys and girls, are more alike than they are different" (p. 581). It holds that most gender differences are small or trivial (close to zero) in magnitude. One exception to this hypothesis may be the gender gap in reading achievement, which is found cross-culturally (Lynn & Mikk, 2009; Reilly, 2012) and exceeds the threshold proposed by Hyde and Grabe (2008, p. 170) for nontrivial gender difference effect sizes ($d \geq .10$). In a recent review, Hyde (2014) remarked that it is "difficult to reconcile" (p. 382) the magnitude of the gender gap observed in reading with other domains of verbal ability (e.g., vocabulary, anagrams), which Hyde and Linn (1988) claimed are typically much smaller.

While the issue of reading is received greater attention, there is a growing body of evidence that males and females also differ in writing ability (Camarata & Woodcock, 2006; Reynolds, Scheiber, Hajovsky, Schwartz, & Kaufman, 2015; Scheiber, Reynolds, Hajovsky, & Kaufman, 2015). Reynolds et al. (2015) noted that the issue of gender differences in writing skills has been overlooked because it is less frequently measured in educational assessments. In cases where writing ability is assessed, researchers should examine gender differences to determine if any meaningful differences occur. Moreover, researchers should compare the size of any differences to those observed with reading assessments when both domains are examined in the same sample.

David Reilly, School of Applied Psychology, Griffith University, Queensland, Australia; David L. Neumann and Glenda Andrews, School of Applied Psychology, Griffith University, Queensland, Australia, and Menzies Health Institute Queensland, Australia.

Correspondence concerning this article should be addressed to David Reilly, School of Applied Psychology, Griffith University, Southport, Queensland 4222, Australia. E-mail: davidreilly.com@gmail.com

David Reilly

Some researchers (e.g., Feingold, 1988) have claimed that as a response to societal changes in the status and roles of women, gender differences are declining (see the online supplemental materials for a more detailed discussion of these issues). Gender role attitudes in the United States have changed over time, giving boys the freedom to pursue language-arts fields just as an increasing number of girls now pursue science, technology, engineering, and mathematics fields. Feingold analyzed educational data from 1947 to 1980, showing a decline over time. More recently, Caplan and Caplan (1997, 2016) have questioned whether gender differences in verbal and language abilities even existed at all and were the product of selection bias in samples, while Hyde (2005) has claimed that most gender differences are either small or trivial in size. The current study examines whether historical patterns of gender differences in reading and writing are still present in modern samples and, if so, to determine their magnitude. It presents a meta-analysis of student reading and writing achievement drawn from the National Assessment of Educational Progress (NAEP), a large nationally representative sample of students from the United States conducted by the National Center for Educational Statistics (NCES). Before turning our attention to this dataset, we first present an overview of theoretical perspectives on gender differences in language ability.

Theoretical Perspectives on Gender Differences in Language Ability

In their pioneering text *The Psychology of Sex Differences*, Maccoby and Jacklin (1974) presented the first systematic review of the psychological literature on gender differences, arguing that gender differences in verbal ability and language were "well established" (p. 351) and showed a developmental progression toward larger gaps with increasing age. Much of the literature they reviewed focused exclusively on reading ability, rather than considering language proficiency more broadly with higher level tasks such as writing, spelling, and grammar usage. But a number of subsequent studies have also reported gender differences with the largest being spelling and use of grammar (Reilly, Neumann, & Andrews, 2016; Stanley, Benbow, Brody, Dauber, & Lupkowski, 1992).

Theoretical explanations for the emergence of gender differences in reading and language proficiency have been offered. These center around biologically based or sociocultural explanations for gender differences, or combinations of both (Eagly & Wood, 2013; Halpern & Tan, 2001): (a) differential rates of maturation, (b) gender differences in lateralization of brain function, (c) gender differences in variability, (d) gender differences in externalizing behavior and language competence, and (e) gender-stereotyping of reading and language as feminine traits. Each will be discussed in detail next.

Differential Rates of Maturation

Girls have a faster rate of maturation and may therefore be attaining greater proficiency than similarly aged boys (Dwyer, 1973), making reading easier and more enjoyable. Such an explanation holds that boys are merely delayed (developmental lag) and boys would attain an equivalent language proficiency given sufficient time. However, this claim is inconsistent with studies showing gender differences in reading that persist into adulthood (Kutner et al., 2007).

Gender Differences in Lateralization of Brain Function

Some researchers have claimed that lateralization of brain function for language may differ between males and females (Levy, 1969). It has been claimed that the regions responsible for language tasks are strongly lateralized to the left cerebral hemisphere in right-handed males, but that language regions in females are more likely to be distributed across both the left and right hemisphere (B. A. Shaywitz et al., 1995). Bilateral language function presumably affords some benefits, which could explain the female advantage observed on such tasks. However, empirical support for the Levy hypothesis is mixed (Kaiser, Haller, Schmitz, & Nitsch, 2009), with some neuroimaging studies showing gender differences in lateralization for language tasks (Burman, Bitan, & Booth, 2008; Clements et al., 2006), while others do not (Wallentin, 2009).

David L. Neumann

Gender Differences in Variability

One explanation for lower reading and language proficiency in males is the *greater male variability* effect, which states that males show greater variability in cognitive performance across all cultures (Feingold, 1992; Machin & Pekkarinen, 2008). Even if there were no gender differences in group means, the consequence of greater male variability is that males will be overrepresented at the extreme left tail of the ability distribution, which Hawke, Olson, Willcut, Wadsworth, and DeFries (2009) argued explains why gender ratios of poor readers favors females. Boys are also overrepresented in populations with reading impairment, dyslexia, attention disorders, and mental retardation suggesting that there may be a gender-linked neurological contribution (Halpern, Beninger, & Straight, 2011). While explanations for the greater male variability hypothesis in intelligence have been made by evolutionary psychologists (Geary, 2010), few specific evolutionary theories have been proposed for verbal and language abilities (Geary, Winegard, & Winegard, 2014), perhaps because these are more recent in an evolutionary sense.

Gender Differences in Externalizing Behavior and Language Competence

Other researchers have argued that gender differences in externalizing behavior may also partly explain a greater female language competence (Limbrick, Wheldall, & Madelaine, 2011). Clinicians identify more boys than girls with externalizing behavior and attention disorders (McGee, Prior, Williams, Smart, & Sanson, 2002) which have both been associated with reading and language impairment. For example, in a longitudinal sample from the United States, Rabiner and Coie (2000) reported that attention-impairment and externalizing behavior measured in kindergarten predicted later reading impairment in fifth grade. Other studies have followed children over longer time frames. In a longitudinal study of child development in Australia, Smart, Prior, Sanson, and Oberklaid (2001) found that externalizing behavior problems at age 7 predicted the severity of later reading and spelling difficulties at ages 13–14, even after controlling for intelligence and socioeconomic status. Although also present in girls, Smart et al. found that the association between externalizing behavior and reading impairment was significantly stronger in boys. Such an association is not necessarily causal, and may well be reciprocal in nature. Within the context of the educational environment, inattention and behavior problems may result in additional educational setbacks, as such problems can interfere with learning as well as lower academic motivation and rapport between teacher and student. But it is equally plausible that these conditions are related to a common neurobiological factor (Berninger, Nielsen, Abbott, Wijsman, & Raskind, 2008).

Gender-Stereotyping of Reading and Language as Feminine Traits

Kagan (1964) first observed that children readily classify social behaviors and even intellectual tasks as either masculine or feminine in nature, based on shared cultural beliefs about gender roles. Reading and language are generally regarded as feminine in nature (Plante, de la Sablonnière, Aronson, & Théorêt, 2013), and gender stereotypes about language are held by both males and females (Halpern, Straight, & Stephenson, 2011). The process by which a child acquires stereotypically masculine and feminine personality traits is termed *sex-typing* (Bem, 1981). Highly sex-typed individuals are motivated to keep their behavior and self-concept consistent with traditional gender norms (Martin & Ruble, 2010; Nash, 1979). The rigidity of sex-roles may translate into decreased reading interest and motivation for some boys if there is a perceived incompatibility between reading and masculine norms. Reading motivation is proposed as playing a strong role in later reading achievement, with boys reporting lower reading motivation and interest (Marinak & Gambrell, 2010; Mucherah & Yoder, 2008). Lowered reading motivation is reflected in the amount of leisure time spent on reading (Moffitt & Wartella, 1991), leading to differential levels of practice between boys and girls. Girls in elementary school also report more positive competence beliefs than boys for reading and language tasks (Eccles, Jacobs, & Harold, 1990).

Glenda Andrews

Large-Scale Assessments of Reading and Writing Achievement

One of the difficulties in evaluating research in the field of gender differences in cognitive ability comes from the use of sampling methods, and the potential for selection bias. It is not normally feasible to sample every male and female in a given population. Researchers thus often take a sample group of participants and then use statistics and probability to draw an inference about the underlying population. Hedges and Nowell (1995) note that this approach can be problematic for two reasons. First, as noted earlier, the greater male variability effect results in a greater number of male high and low achievers at the top and bottom of the ability distribution, respectively (Hawke et al., 2009; Machin & Pekkarinen, 2008). Greater variability may present a distorted picture of the underlying population which is magnified in highly selected samples (Becker & Hedges, 1988). Second, demographic factors such as socioeconomic status, ethnicity, and rural versus urban residence can greatly influence cognitive ability (Fernald, Marchman, & Weisleder, 2013; Hanscombe et al., 2012), which may further limit the generalizability of a convenience sample.

For this reason, the gold standard for research is to recruit a large sample that is *representative* of the population under investigation (Hedges & Nowell, 1995), in terms of gender, ethnicity, socioeconomic status, geographical region, and so forth. This approach increases confidence in the validity of any conclusions made about specific groups, such as males and females. Another reason why selection bias may be problematic in the context of gender differences in reading and writing is that when investigating specific subgroups (such as students that have been identified as poor readers), it is difficult to determine the underlying prevalence of males and females due to the issue of a gendered referral bias. Shaywitz, Shaywitz, Fletcher, and Escobar (1990) noted that more boys than girls are identified as poor readers by educational institutions, but when epidemiological studies investigate reading impairment in the community girls and boys approach an equal representation (Hawke et al., 2009; Jiménez et al., 2011). The implication here is that the prevalence of reading impairment in girls may simply just be underreported, and that there may be a referral bias for boys. In order to test such a claim, a study would need to recruit a large, nationally representative sample and administer a standardized reading assessment.

One such source is NAEP, which is conducted by NCES, part of the U.S. Department of Education. It has the added advantage that new waves of assessment have been conducted over several decades without major changes to the reading and writing frameworks so that temporal trends can be investigated. Before turning our attention to this analysis, we first review previous studies that have recruited nationally representative samples of males and females to investigate gender differences in reading and writing.

Gender Differences in Reading

Hedges and Nowell (1995) reported the largest study of gender differences in achievement scores ever conducted, across a wide range of content areas using nationally representative samples from the United States. These included student assessments of reading proficiency conducted by NAEP reported from 1971 to 1992. They found that girls showed significantly higher scores for tests of reading in each year of assessment, with effect sizes ranging from $d = -.18$ to $-.30$. Furthermore, they found that the performance of boys was more variable than that of girls with an average variance ratio (VR) of 1.12. This variability resulted in an overrepresentation of boys as poor readers. The researchers also examined data from a number of other data sets that recruited nationally representative stratified samples. Across these other data sets, Hedges and Nowell found similarly sized gender differences and greater male variance. They also reported that the ratio of boys to girls in the bottom 10% of reading comprehension (i.e., poor readers) ranged from 1.07 to 1.75, which paralleled that found in the NAEP data. Thus, there were both mean gender differences in reading ability and an overrepresentation of boys who are poor readers.

While pioneering at the time it was published, a serious limitation of Hedges and Nowell's (1995) analysis was that they only examined NAEP data from students near the end-point of their education, aged 17, and did not investigate whether gender differences were still present in *younger* students. Developmental differences are an impor-

tant consideration, as gender differences in literacy might emerge at earlier ages. As they also reported substantial variability across waves of assessment, the technique of meta-analysis ought to have been employed to *aggregate* findings across waves to determine the overall trend. Additionally, there is concern that with the passage of time such results may quickly become dated. A number of researchers such as Feingold (1988), or Caplan and Caplan (1997, 2016) have claimed that gender differences are disappearing (for a full overview see the online supplemental materials). NAEP collects assessments of reading periodically, and there are now numerous waves of data are unexamined. If gender differences in cognitive ability are disappearing, then such an effect should be observable in U.S. children and adolescents over a sufficiently large timeframe.

Evidence for gender differences in reading proficiency may also exist cross-culturally in large multinational assessments of student achievement (Lynn & Mikk, 2009; Reilly, 2012). One such source is the Programme for International Student Assessment (PISA) conducted by the Organisation for Economic Cooperation and Development (OECD) across member and partner nations. It seeks to assess student achievement in reading, mathematics, and science at age 15 (which is typically toward the end of compulsory schooling in most countries). Lynn and Mikk (2009) found appreciably sized gender differences across all nations in the 2000, 2003, and 2006 waves of PISA assessment, while Reilly (2012) reached a similar conclusion with the PISA 2009 dataset. There was also substantial variability across nations which researchers attribute to cultural factors such as national levels of gender equality (Guiso, Monte, Sapienza, & Zingales, 2008; Reilly, 2015).

Though gender differences in reading have been found by most studies that recruit sufficiently large and representative samples, it is also important to acknowledge that there are some rare exceptions. For example, Kaufman, Kaufman, Liu, and Johnson (2009) reported an analysis of the norming sample for the Kaufman Test of Educational Achievement–Brief Form. The authors did not find significant gender differences in reading for adults, though significant gender differences were found in children as subsequently reported by Scheiber et al. (2015) with this instrument. However, it is unclear whether this was the result of differences in test content across reading assessments, or if it was confounded by historical effects of educational inequality in their cross-sectional sample (adults aged 22–90). Thus it is crucial to identify under what contexts gender differences in reading may be found, but their existence is not a foregone conclusion.

Gender Differences in Writing

As noted earlier, there are a limited number of studies that have investigated gender differences in writing ability, and the number of studies recruiting representative samples are even fewer. Nowell and Hedges (1998) reported a more detailed analysis of NAEP writing data from the period 1984–1994, finding substantial gender differences in writing (ranging from $d = -.49$ to $-.55$), greater male variability, and that gender ratios for students falling in the bottom 10th percentile were between 2.6 and 3.3 males to every female (Nowell & Hedges, 1998, p. 38). At present, there has been no subsequent meta-analysis published investigating gender differences in NAEP writing assessments.

Two other prominent studies have investigated gender differences in writing with large representative samples. Camarata and Woodcock (2006) presented data from the normative samples of the Woodcock-Johnson cognitive and achievement batteries, a large representative sample of males and females aged 5 through to 79. Females scored significantly higher in writing achievement, with an average effect size across the life span of $d = -.33$. More recently, Scheiber et al. (2015) analyzed a large nationally representative sample of adolescents and young adults completing the Kaufman Test of Educational Achievement–Second Edition Brief Form, which measures participants across reading, writing, and mathematics. While no difference was found in mathematics, females scored higher than males on the tests of reading and writing ability. The effect size for reading was small ($d = -.18$), but the effect size for writing ($d = -.40$) was twice as large as that for reading. Given the appreciable gender differences found in these samples, it seems justifiable to expect a similarly sized effect in NAEP data for writing tasks.

The Present Review

We sought to investigate whether the historical patterns of gender differences in reading and writing reported by Hedges and Nowell (1995) would be replicated for children growing up in more recent decades. Consistent with previous research, we hypothesized that gender differences in reading and writing achievement would be present. Based on the claim made by Feingold (1988) and Caplan and Caplan (2016) that gender differences in cognitive ability are decreasing, we also hypothesized that there would be a significant negative association between year of assessment and effect size, such that gender differences would show a decline over time. Given the large sample size employed by NAEP and that data from several decades of testing were available, the analysis would have strong statistical power to detect an effect. Hyde and Grabe (2008) have advocated that a threshold of evidence higher than statistical significance be adopted because although a very large sample size might yield statistically significant differences, the actual size of the effects might be trivial. Therefore we adopted the research practice recommended by Hyde and Grabe (2008, p.

170) and determined a priori that effect sizes smaller than $d = .10$ are characterized as trivial in size, even if they met the threshold for statistical significance. We used Cohen's (1988) recommendation that effect sizes around $d = .20$ be regarded as small, while around $d = .50$ medium.

Method

National Assessment of Educational Progress Data Source

The NAEP is a project of NCES, part of the U.S. Department of Education. The NAEP is used to track student achievement over time in fourth-, eighth- and 12th-grade at the state and national level of the United States. It measures student achievement in reading, mathematics, science and a variety of other subject areas. National and state performances are reported annually in a series of reports titled "The Nation's Report Card" (see http://nationsreportcard.gov/). This information is of use to parents, educators, and policymakers. However such reports only indicate that gender differences are statistically significant, without providing any context about the size of such differences or gender ratios of poor/advanced readers and writers.

NAEP data is also publically available so that it can be used by researchers to conduct secondary analysis, via the NAEP Data Explorer (http://nces.ed.gov/nationsreportcard/naepdata/). The sampling frame employed by NAEP is all students in the target grades (Grades 4, 8, and 12) in each of the 50 states of the United States, drawn from both public and private educational institutions. School and student responses are appropriately weighted to draw a nationally representative estimate of the target population that reflects student demographics such as socioeconomic status of school district, ethnicity, rural versus urban location and gender. For inclusiveness, the sampling frame also includes students with disabilities and English language learners, with the goal of reaching at least 85% of those identified as students with disabilities or English language learners. Additional information on the sampling methodology employed is available from the NAEP website (http://nces.ed.gov/nationsreportcard/about/samplesfaq.aspx).

Content for the reading assessment includes reading comprehension of a variety of different passages and genres (including information reports, stories, poetry and essays), as well as an understanding of vocabulary. Content for the writing assessment includes persuasive, informative, and narrative writing in response to stimuli material. Additional information on reading and writing frameworks in each grade level is available from the NAEP website.

Schedule of Assessment

Reading and writing assessments are conducted periodically, in adherence with the NAEP schedule. Reading assessments are given greater priority than writing and occur every 2 to 3 years (1988, 1990, 1992, 1994, 1998, 2000, 2002, 2003, 2005, 2007, 2009, 2011, 2013, 2015), with greater coverage given for students in Grades 4 and 8. Writing assessments occur approximately every 4 to 5 years (1998, 2002, 2007, 2011), and usually with a smaller sample size than the reading assessments. We also included archived data from the 1988, 1990, 1992, and 1996 writing assessments so that both dependent variables were assessed across the same time frame. All assessments from 1988 onward were included in the analysis.

Participants

National performance data for NAEP Reading assessments were examined from the period 1988–2015, with a combined total sample size of 3.035 million students. Testing data for the NAEP Writing assessments were examined for the period 1988–2011, with a combined total sample size of 934,800. Students provided deemed consent through their participation in each wave of assessment. This study used published archival data and did not recruit participants directly.

Meta-Analytic Procedure

Effect size statistics are presented as the mean difference between boys and girls in standardized units, commonly referred to as Cohen's *d* (Cohen, 1988). The meta-analysis employed a random effects model. Heterogeneity across samples was indicated by the I^2 statistic, representing the percentage of variation across samples attributed to genuine heterogeneity and not chance. We also investigated whether there were developmental differences in the magnitude of the gender gap across the three grade levels using subgroup analysis, and whether the year of testing was a potential moderator using metaregression (Kelley & Kelley, 2012). Full details of the methodology employed in our analysis are reported in the online supplemental materials.

Results

Gender Differences in Reading Achievement

Girls showed significantly higher reading scores than boys across every wave of assessment and in every grade, with an overall effect size of $d = -.27$, 95% confidence interval (CI) [–.29, −.25], $Z = -26.08$, $p < .001$ (see Figure 1). Gender differences significantly exceeded the predetermined cutoff ($d \leq .10$) advocated by Hyde and Grabe (2008) by a factor of 2.7. There was also significant heterogeneity in effect sizes, $Q(36) = 2594.45$, $p < .001$, $I^2 = 98.61$, indicating considerable variation across assessments. To better explain the variability in effect sizes, we

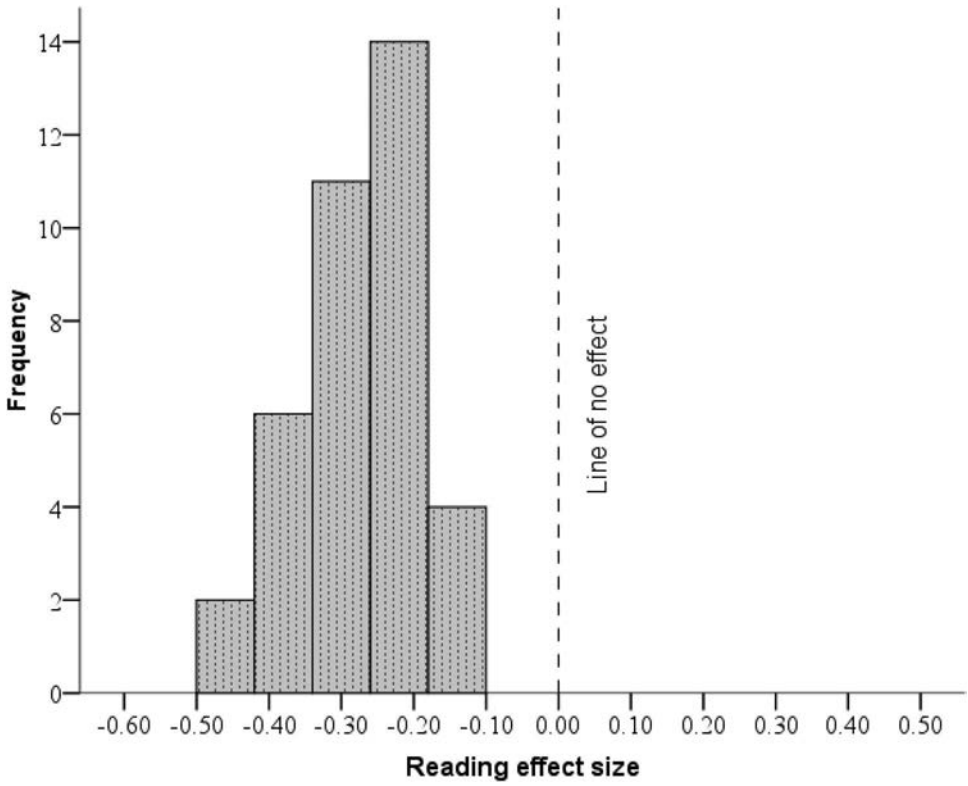

Figure 1. Histogram of effect sizes (Cohen's *d*) for the difference between boys and girls in reading achievement. All effect sizes fall to the left of the line of no effect and exceed Hyde's criterion.

investigated whether grade level or year of assessment were potential moderators of the gender difference.

Grade level. Table 1 presents comparisons between males and females in reading achievement across the three grade levels assessed by NAEP. There was a statistically significant difference between groups, $Q(2) = 148.49$, $p < .001$, with a tendency toward larger differences between boys and girls in older students. The initial gender difference in reading achievement was small in Grade 4 ($d = -.19$), but grew larger in Grade 8 ($d = -.30$) and Grade 12 ($d = -.32$).

Year of assessment. Next, we performed a metaregression on reading achievement, using the year of assessment as a predictor. There was no significant effect of year of assessment, $Z = .79$, $b = .0001$, 95% CI [−.001, .003], $p = .425$, which is inconsistent with the hypothesis of a declining gender difference over time.

Variance ratio. Consistent with previous research there was greater male variability present in every sample tested, although it sometimes felt just short of Feingold's threshold (VRs > 1.1) for individual years. Mean VR ratios were calculated for each grade. All grades exceeded Feingold's

Table 1
Gender Differences in National Assessment of Educational Progress Reading and Writing Achievement for Grades 4, 8, 12

				95% Confidence interval		Test of null (two-tail)			
Outcome	Grade	*k*	Cohen's *d*	Lower limit	Upper limit	*Z* value	*p*	VR	Heterogeneity
Reading	4	14	−.19[a,b]	−.21	−.18	−25.00	<.001	**1.11**	$Q(13) = 47.69$, $p < .001$, $I^2 = 72.74$
	8	13	−.30[a]	−.32	−.29	−39.07	<.001	**1.13**	$Q(12) = 157.06$, $p < .001$, $I^2 = 92.36$
	12	10	−.32[b]	−.34	−.30	−32.87	<.001	**1.22**	$Q(9) = 245.01$, $p < .001$, $I^2 = 96.33$
Writing	4	7	−.42[a,b]	−.47	−.37	−17.55	<.001	1.01	$Q(6) = 94.96$, $p < .001$, $I^2 = 93.68$
	8	9	−.62[a]	−.66	−.58	−29.94	<.001	1.06	$Q(8) = 96.74$, $p < .001$, $I^2 = 93.68$
	12	9	−.55[b]	−.59	−.51	−26.40	<.001	1.07	$Q(8) = 256.73$, $p < .001$, $I^2 = 96.88$

Note. *k* = number of assessments conducted for each grade; VR = mean variance ratio. Boldface represents VRs that exceed Feingold's threshold. Three planned contrasts between grades were conducted with a Bonferroni correction applied to control family-wise Type I error rate. Contrast C between Grade 8 and 12 did not significantly differ with Bonferroni correction for reading and writing.
[a] Grade 4 versus 8 significant. [b] Grade 4 versus 12 significant.

critical value, with progressively higher variance in older students.

Gender ratios for poor and gifted readers. In order to evaluate the combined effect of mean gender differences and greater male variability on poor and gifted readers, we examined the gender ratios of readers falling below the "basic" proficiency standard defined by NAEP as well as those exceeding the "advanced" proficiency standard. The analysis examined the risk ratio of males to females attaining these levels. Equivalent proportions of boys and girls at a particular achievement level would be indicated by a risk ratio of 1.00. Higher risk ratios (i.e., >1.00) would indicate an overrepresentation of boys attaining this standard, while lower risk ratios (i.e., <1.00) would reflect an overrepresentation of girls at a particular standard.

The weighted risk ratio for poor readers was 1.39, 95% CI [1.34, 1.44], $Z = 19.82$, $p < .001$. The subgroup analysis is reported in Table 2. As can be seen, more boys than girls were poor readers, which reached a ratio of 1.54 times as many boys as girls falling below the minimum standard of literacy by Grade 12. The effect was reversed for advanced readers, with more girls than boys achieving the advanced literacy standard. Additionally the concentration of males in the lower left tail of the distribution was higher than the concentration of females at the upper right. The weighted risk ratio for advanced readers was 0.55, 95% CI [0.52, 0.59], $Z = -17.28$, $p < .001$, with subgroups also reported in Table 2. Expressed in a metric that may be more intuitive for nonstatisticians, by the time students reach Grade 12 there are almost twice as many girls than boys that reach the advanced standard of reading proficiency. Moderator analysis showed a slight tendency toward smaller gender gaps in poor readers over time ($Z = -2.55$, $p = .010$), but larger gender gaps in advanced readers over time ($Z = 2.31$, $p = .021$).

Gender Differences in Writing Achievement

Next we examined the gender difference in writing achievement for the period 1988–2011. Overall, the gender difference between males and females in writing was larger than that found for reading, $d = -.54$, 95% CI [–.57, –.51], $Z = -36.14$, $p < .001$ (see Figure 2). Gender differences in writing exceeded the predetermined cutoff ($d \leq .10$) advocated by Hyde and Grabe (2008) by a factor of 5.4. There was also significant heterogeneity in effect sizes, $Q(24) = 974.07$, $p < .001$, $I^2 = 97.54$, indicating considerable variation across assessments. In order to better explain the variability in effect sizes, we investigated whether grade level or year of assessment were potential moderators of the gender gap. An additional factor introducing heterogeneity may be the changes in writing frameworks (new frameworks were introduced in 1988, 1998, and 2011) and the marked variability in sample sizes for more recent assessments.

Grade level. Table 1 presents comparisons between males and females in writing achievement across the three grade levels assessed by NAEP. The difference between grades was statistically significant, $Q(2) = 42.01$, $p < .001$, with a tendency toward a smaller initial gender difference in writing proficiency for students in Grade 4. The initial gender difference in writing was medium-sized in Grade 4 ($d = -.42$), but grew larger in Grade 8 ($d = -.62$) and Grade 12 ($d = -.55$).

Year of assessment. We performed a metaregression on writing achievement, using the year of assessment as the predictor. There was no significant effect of year, $Z = -1.85$, $b = -.004$, 95% CI [–.001, .001], $p = .063$, indicating stability in effect sizes across historical time.

Variance ratio. In examining the variance ratios presented in Table 1, there was minimal support for greater male variability with all grades falling short of Feingold's threshold.

Gender ratios for poor and gifted writers. In order to evaluate the joint effect of greater male variability and mean gender differences on poor and gifted writers, we examined the gender ratios of readers below the 'basic' achievement as well as those exceeding the "advanced" achievement level. Writing proficiency levels attained were not pub-

Table 2
Risk Ratio for Poor and Advanced Proficiency Level Readers and Writers, Across Grade Levels

		Poor readers					Advanced readers				
			95% Confidence interval		Test of null			95% Confidence interval		Test of null	
Outcome	Grade	Risk ratio	Lower limit	Upper limit	*Z*	*p*	Risk ratio	Lower limit	Upper limit	*Z*	*p*
Reading	4	1.22[a,b]	1.20	1.24	25.55	<.001	.70[a,b]	.68	.73	−19.64	<.001
	8	1.45[a,b,c]	1.43	1.48	44.84	<.001	.47[a,b]	.45	.49	−36.82	<.001
	12	1.54[b,c]	1.50	1.58	34.76	<.001	.52	.49	.55	−21.94	<.001
Writing	4	2.01	1.55	2.60	5.31	<.001	.28	.20	.39	−7.40	<.001
	8	2.27	1.89	2.72	44.84	<.001	.27	.21	.36	−9.66	<.001
	12	2.21	1.84	2.65	8.51	<.001	.39	.28	.54	−5.60	<.001

Note. Three planned contrasts between grades were conducted, with a Bonferroni correction applied to control family-wise Type I error rate.
[a] Grade 4 versus 8 significant. [b] Grade 4 versus 12 significant. [c] Grade 8 versus 12 significant.

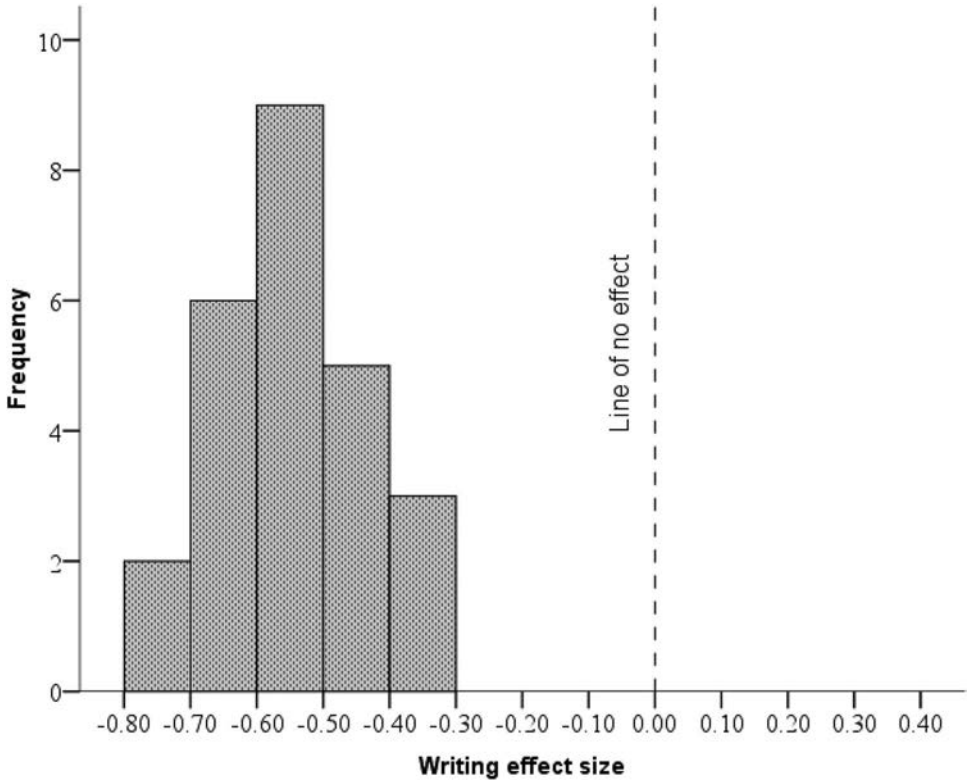

Figure 2. Histogram of effect sizes for the difference between boys and girls in writing achievement. All effect sizes fall to the left of the line of no effect and exceed Hyde's criterion for nontrivial gender differences.

lished for the archived reports (1988–1996), and were therefore excluded from analysis.

The weighted risk ratio for poor writers was 2.19, 95% CI [2.00, 2.40], $Z = 17.06$, $p < .001$, with subgroup analysis reported for grades in Table 2. As can be seen by the table, there were twice as many boys falling into the category of poor writing than girls. However, the effect is reversed for advanced readers with more girls achieving the advanced standard for written expression, with a weighted risk ratio of 0.30, 95% CI [.25, .36], $Z = 13.29$, $p < .001$. In other words, by the time students reach Grade 12 there are over 2.54 times as many girls than boys that attain the advanced standard of writing proficiency. Moderator analysis showed no change in gender ratios for poor writers over time ($Z = .15$, $p = .881$), or advanced ($Z = 1.80$, $p = .071$).

Discussion

Annual reporting of NAEP data had noted girls performed significantly higher than boys, but failed to provide estimates of how large these differences were. By calculating an effect size, we can hold evidence of gender differences to a much higher standard than mere statistical significance by examining whether the effect is *practically* significant. While a focus on mean gender differences is important, we also considered its combined effect with greater male variability on the gender ratios at the lower left (poor readers/writers) and upper right (advanced readers/writers) tails of the ability distribution. Both measures (effect size for mean gender differences, gender ratios of low/high achievers) provide a more comprehensive perspective than simply examining probability values. Further, the detailed records kept for NAEP testing data offered a window into the past to examine how boys and girls have fared in reading and writing achievement over historical time (both developmentally across grades, and cohort effects across historical time).

Reading Proficiency

Girls significantly outperformed boys in reading ability across all grades, with a tendency toward larger effect sizes in high school than primary school (Grades 8, $d = -.30$ and 12, $d = -.32$). These exceed Hyde's criterion by a factor of 3, and fall in the small-to-medium effect size category proposed by Cohen (1988). They are also comparable to effect sizes for American students in international assessments such as PISA (Reilly, 2012) and the Progress in International Reading Literacy assessment (PIRLS; Mullis, Martin, Gonzalez, & Kennedy, 2003), where small to medium effect sizes were found. There was also no evidence of a decline in the magnitude of effect sizes over time as had

been hypothesized, though it is possible that this might be detectable over a longer passage of historical time. However to compare these results to other standardized tests of reading and writing would be problematic, and introduce a methodological confound in test content, level of difficulty, and sampling size.

How ought we interpret the practical impact of such gender differences in reading? Rosenthal and Rubin (1982a) developed the binomial effect size display (BESD) to illustrate the practical impact of such differences for nonstatisticians (such as parents and educators), especially for students falling near the middle of the distribution. This metric shows the percentage of males and females that meet or exceed an average score. Represented in the BESD format, the likelihood of being average or higher in reading ability for a student at the end of high school increases from 42.1% for boys to 57.8% for girls, a not insubstantial amount.

We can also contextualize this by considering the size of gender differences for other types of cognitive ability, such as science, technology, engineering, and mathematics achievement. While considerable research focuses on the gender gap in mathematics and science, the effect sizes for reading are more substantial—over twice the size as that found in comparable NAEP assessments of mathematics (McGraw, Lubienski, & Strutchens, 2006; Reilly, Neumann, & Andrews, 2015). Thus, is it is important to acknowledge female strengths *as well* as those areas where males perform higher. Claims by researchers such as Caplan and Caplan (1997, 2016) that cognitive differences are disappearing are therefore premature, but neither does it support a claim that boys and girls are radically different in reading literacy and would benefit from gender-segregated instruction as is claimed by same-sex advocates.

However, effect size statistics only represent the *typically performing* girl or boy. When we examined students that fall below the minimum proficiency standard, far more boys than girls fall into this category across all grades and more importantly by the end of high school by a factor of 1.5. This imbalance in a representative sample contradicts the claim made by Shaywitz et al. (1990) that a greater diagnosis of reading impairment in boys is merely the result of a gendered referral bias. A completely different pattern was found for advanced readers though, with far more girls attaining this level of proficiency (by a factor of almost 2). The pattern of results shows that at all levels of the ability distribution, girls significantly outperform boys in reading achievement. Hyde (2005) had proposed the GSH, arguing that most mean gender differences are small or trivial in magnitude. A limitation of that hypothesis though is that it focuses *exclusively* on mean gender differences and effect sizes, while ignoring evidence from the upper and lower tails of the ability distribution. Taken together, it would appear there are gender differences in reading favoring girls across all levels of ability distribution, with these being small (and more similar) in the middle of the distribution but much larger (and impactful) at the tails. Furthermore, these gender differences are found in younger students, as well as older ones. We did not find strong support for the greater male variability hypothesis because the larger number of low scoring boys was offset by the higher number of high scoring girls.

Writing Proficiency

As was expected from previous studies (e.g., Hedges & Nowell, 1995; Reynolds et al., 2015), girls significantly outperformed boys in writing ability across all grades and assessment waves. The magnitude of effect sizes was higher than that found for reading, with effect sizes falling into the medium size range by Cohen's (1988) conventions. Comparisons between boys and girls were slightly smaller in Grade 4 ($d = -.42$), but this gender difference widened for older students (Grades 8, $d = -.62$; Grades 12, $d = -.55$). Represented in the BESD format, the likelihood of being average or higher in writing ability for a student at the end of high school increases from 36.7% for boys to 63.3% for girls (i.e., a minority of boys attain this standard, but the majority of girls do). Furthermore, when examining the association between effect size and year of assessment, there was no decline in the magnitude of effect sizes as predicted.

At the lower end of the ability distribution, boys were greatly overrepresented by a factor of 2 or more which grew slightly larger for older students. Just as with reading, there was a reversal of gender ratios for students attaining an advanced writing proficiency, with girls greatly overrepresented by a factor of 2 or more. These results are consistent with the position held by Reynolds et al. (2015) who argued that a gender difference in writing is an exception to the GSH and cannot be easily dismissed as a small or trivial difference.

Why might the effect size for writing be larger than the effect size for reading? Writing represents a more challenging task, and larger gender differences are typically found as the complexity of the task increases. While reading is a passive task, writing is a generative task that draws on other components of verbal and language abilities that typically show larger gender differences. For example, it requires careful organization of ideas and the production of material that is clearly expressed, and grammatically accurate. Research shows that females score significantly higher on standardized tests of spelling and of grammar, with medium-sized effects (Stanley et al., 1992). Finding the right words to express a particular concept or nuance is also a demanding task for writers, and draws on verbal fluency (where females also show significantly higher performance than males). Halpern and Tan (2001) noted that effect sizes for verbal fluency fall in the medium to large range. All of these verbal skills can be improved with sufficient practice

and instruction, however, highlighting the importance of these basic skills in a crowded educational curriculum.

Gender Similarities Hypothesis

Hyde (2005) has proposed the GSH, which claims that most—but not all—psychological gender differences are small or trivial in size. Zell, Krizan, and Teeter (2015) used the technique of metasynthesis to test this claim, finding that most effect sizes were small. However, they also noted a number of important exceptions (see Zell et al., 2015, Table 3). The gender differences observed in the present study for reading (small-to-medium) and writing (medium-sized) also represent exceptions that may have been overlooked because a meta-analysis on the NAEP dataset had yet to be published for modern samples. The identification of new areas where meaningful gender differences remain does not invalid the GSH, but does serve as a prompt for further investigation.

Educational Implications

What might be the educational implications of such a gender gap in reading and writing proficiency during primary and high school years? While information in the classroom environment is often presented verbally, students are expected to read textbooks and literary material as independent reading. Difficulties in reading would be a serious impediment, particularly if reading takes boys longer or if they are unable to gain a deep understanding of the text. While it has been known for some time that boys are overrepresented in reading disabilities and formal diagnoses of dyslexia (Berninger et al., 2008; Rutter et al., 2004), this pattern of results suggests a more general reading deficiency for the typical male student. Written communication is also important during the high school years, as this format is commonly adopted in the format of essays or laboratory reports. While boys tend to perform better than girls on standardized tests, girls tend to achieve significantly higher grades during schooling (Duckworth & Seligman, 2006; Voyer & Voyer, 2014), and it is possible that the assessment format (exam vs. written assignment) may be a contributing factor. While the issue of gender differences in reading ability has been the focus of much research, gender differences in writing ability may have been previously underestimated by researchers and educators. The magnitude of the gender gap in writing ability is sufficiently large that it may warrant educational interventions and further research on etiology. It may also be reflective of a more general language deficiency (rather than just an issue with reading), as other studies have also reported pronounced gender differences in grammar and language usage (e.g., Stanley et al., 1992).

While the existence of a gender gap in reading and writing during compulsory schooling is troubling, the educational implications for students considering pursuing tertiary education are potentially compounded. In a review of gender inequalities in education, Buchmann, DiPrete, and McDaniel (2008) note that women enroll in college and universities at a much higher rate than their similarly aged male peers, achieve higher grades on average than males, and have a higher rate of degree completion (Buchmann & DiPrete, 2006). This pattern is mirrored across most OECD countries and is not confined to the United States (OECD, 2016). The transition from secondary to tertiary education can be difficult for many students, because it involves independent learning and considerable hours of study outside of classroom contact time. The ability to read textbooks and assigned readings is a crucial part of learning. Although especially poor readers are less likely to pursue tertiary studies, the gender gap in reading appears to be present in average students as well, though smaller. In addition, the ability to communicate verbally in a written format takes on increasing importance, as producing reports and essays are a common form of student assessment. Systemic gender gaps in writing might leave male students significantly underprepared for tertiary admission, and offer a partial explanation for why females on average achieve higher grades in their tertiary studies (Voyer & Voyer, 2014) and have higher completion rates (Buchmann et al., 2008).

Parents, educators, and policymakers may wonder what to make of gender differences in reading and writing, and what changes might be made to address them in the interests of equality of educational outcomes. It would be a mistake to take evidence from this study to argue that boys and girls learn in fundamentally different ways, require different styles of teaching, or would benefit from same-sex schooling. Scientific literature is clear about the negative effects of highlighting gender in this way (Halpern, Eliot, et al., 2011), and how treating a particular demographic group (i.e., just boys) can serve to undermine their confidence and motivation to improve. While attention has been paid in the past to early intervention for reading, educational interventions for writing may be warranted, and a greater focus on writing tasks in the curriculum to provide additional opportunities to practice writing skills and provide feedback to students. These should be offered broadly to *all* students—while the findings of this study suggest that boys would benefit from these initiatives, these results also suggest that many girls would similarly benefit.

Future Directions for Research and Limitations

While this study documents the existence and magnitude of gender differences in reading and writing, it cannot shed any light on their etiology and *which* biological and socialization factors most contribute to their development (Eagly & Wood, 2013). Most researchers advocate a biopsychosocial model of gender differences (Halpern, 2000), but some tentative conclusions may be drawn from the generalizability of gender differences in language outcomes across sam-

ples, historical time periods, and cultures. Our study replicates findings reported by Hedges and Nowell (1995) for earlier decades, and the magnitude of gender differences remains stable over historical time. Additionally, we found a developmental effect, such that gender differences were found in much younger students but increased with additional years of schooling. International studies of reading achievement find that greater female performance is found universally across all nations (Guiso et al., 2008; Reilly, 2012), which would at least be consistent with biological factors. Yet there is also substantial variability across nations in the size of the gender gap, with sociocultural factors such as nation's level of gender equality and gender norms also making a strong contribution (Reilly, 2015). Like many gender researchers, we advocate a broad biopsychosocial model of factors that contribute to gender differences (Halpern, 2000), rather than any single cause.

Despite the strengths that a large demographically representative sample like NAEP offers, the chief limitation is that it is limited to students from the United States, and to the way in which reading and writing skills are measured. Educational practices and frameworks clearly differ from country to country. International assessments of students' reading achievement such as PIRLS and PISA (Lynn & Mikk, 2009; Reilly, 2012) have found that the gender difference in reading is universal (i.e., all countries find girls significantly and meaningfully outperform boys). However, it is unclear whether a similarly sized gender difference in writing skills would exist internationally. Additionally due to limitations of the dataset (such as lack of subgroup sample sizes in the publicly available data), it was not possible to investigate Gender × Ethnicity interactions or socioeconomic status differences, a limitation shared by other analyses of NAEP data (e.g., Reilly et al., 2015). One factor that cannot be controlled in this dataset though is student dropout rates. As compulsory schooling extends in most U.S. states up to age 16, there should not be any meaningful attrition between Grades 4 and 8. While high school completion rates have been steadily increasing over the historical time period examined, the gender differences reported for Grade 12 do not include young adults that leave before this time (and presumably may have poorer reading and writing proficiency). As more girls complete high school than boys, this may underestimate the extent of gender differences in reading and writing in the general population. Additionally, it only presents a snapshot of students enrolled in U.S. schools—children that are unable to attend formal schooling due to other factors such as intellectual disability would likewise not be measured.

Conclusion

Gender differences in reading and writing achievement were found across all levels of the ability spectrum. Girls outperformed boys in mean reading and writing achievement, and contrary to our hypothesis these gender differences do not appear to be declining over the time period analyzed (1988–2015). Furthermore, there were pronounced differences in gender ratios for poor readers/writers, with boys greatly overrepresented. This pattern was reversed for those students attaining an advanced proficiency standard, with significantly more girls than boys. Our study also examined gender differences in younger students than those reported by Hedges and Nowell (1995), finding a developmental effect toward larger gaps as students progress through their schooling. These findings hold educational implications for students' academic success during primary and high school, as well as academic readiness to embark on further college studies. A challenge for researchers is to identify the precise nature of gender differences in reading and writing so that educators can design targeted interventions to improve children's reading and writing skills.

References

Archer, J., & Lloyd, B. (2002). *Sex and gender* (2nd ed.). Cambridge, United Kingdom: Cambridge University Press. http://dx.doi.org/10.1017/CBO9781139051910

Auster, C. J., & Ohm, S. C. (2000). Masculinity and femininity in contemporary American society: A reevaluation using the Bem Sex-Role Inventory. *Sex Roles, 43,* 499–528. http://dx.doi.org/10.1023/A:1007119516728

Becker, B. J., & Hedges, L. V. (1988). The effects of selection and variability in studies of gender differences. *Behavioral and Brain Sciences, 11,* 183–184. http://dx.doi.org/10.1017/S0140525X00049256

Bem, S. L. (1981). Gender schema theory: A cognitive account of sex typing. *Psychological Review, 88,* 354–364. http://dx.doi.org/10.1037/0033-295X.88.4.354

Berninger, V. W., Nielsen, K. H., Abbott, R. D., Wijsman, E., & Raskind, W. (2008). Gender differences in severity of writing and reading disabilities. *Journal of School Psychology, 46,* 151–172. http://dx.doi.org/10.1016/j.jsp.2007.02.007

Borenstein, M., Hedges, L. V., Higgins, J. P. T., & Rothstein, H. R. (2009). *Introduction to meta-analysis*. West Sussex, United Kingdom: Wiley. http://dx.doi.org/10.1002/9780470743386

Buchmann, C., & DiPrete, T. A. (2006). The growing female advantage in college completion: The role of family background and academic achievement. *American Sociological Review, 71,* 515–541. http://dx.doi.org/10.1177/000312240607100401

Buchmann, C., DiPrete, T. A., & McDaniel, A. (2008). Gender inequalities in education. *Annual Review of Sociology, 34,* 319–337. http://dx.doi.org/10.1146/annurev.soc.34.040507.134719

Burman, D. D., Bitan, T., & Booth, J. R. (2008). Sex differences in neural processing of language among children. *Neuropsychologia, 46,* 1349–1362. http://dx.doi.org/10.1016/j.neuropsychologia.2007.12.021

Camarata, S., & Woodcock, R. (2006). Sex differences in processing speed: Developmental effects in males and females. *Intelligence, 34,* 231–252. http://dx.doi.org/10.1016/j.intell.2005.12.001

Caplan, J. B., & Caplan, P. J. (2016). *Thinking critically about research on sex and gender* (3rd ed.). New York, NY: Routledge.

Caplan, P. J., & Caplan, J. B. (1997). Do sex-related cognitive differences exist, and why do people seek them out? In P. J. Caplan, M. Crawford, J. S. Hyde, & J. T. E. Richardson (Eds.), *Gender differences in human cognition* (pp. 52–80). New York, NY: Oxford University Press. http://dx.doi.org/10.1093/acprof:oso/9780195112917.003.0003

Clements, A. M., Rimrodt, S. L., Abel, J. R., Blankner, J. G., Mostofsky, S. H., Pekar, J. J., . . . Cutting, L. E. (2006). Sex differences in cerebral laterality of language and visuospatial processing. *Brain and Language, 98,* 150–158. http://dx.doi.org/10.1016/j.bandl.2006.04.007

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Erlbaum.

Conger, D., & Long, M. C. (2010). Why are men falling behind? Gender gaps in college performance and persistence. *The Annals of the American Academy of Political and Social Science, 627,* 184–214. http://dx.doi.org/10.1177/0002716209348751

Duckworth, A. L., & Seligman, M. E. P. (2006). Self-discipline gives girls the edge: Gender in self-discipline, grades, and achievement test scores. *Journal of Educational Psychology, 98,* 198–208. http://dx.doi.org/10.1037/0022-0663.98.1.198

Dwyer, C. A. (1973). Sex differences in reading: An evaluation and a critique of current theories. *Review of Educational Research, 43,* 455–467. http://dx.doi.org/10.3102/00346543043004455

Eagly, A. H., & Wood, W. (1999). The origins of sex differences in human behavior: Evolved dispositions versus social roles. *American Psychologist, 54,* 408–423. http://dx.doi.org/10.1037/0003-066X.54.6.408

Eagly, A. H., & Wood, W. (2013). The nature–nurture debates: 25 years of challenges in understanding the psychology of gender. *Perspectives on Psychological Science, 8,* 340–357. http://dx.doi.org/10.1177/1745691613484767

Eccles, J. S., Jacobs, J. E., & Harold, R. D. (1990). Gender role stereotypes, expectancy effects, and parents' socialization of gender differences. *Journal of Social Issues, 46,* 183–201. http://dx.doi.org/10.1111/j.1540-4560.1990.tb01929.x

Else-Quest, N. M., & Hyde, J. S. (2018). *The psychology of women: Half the human experience* (9th ed.). Thousand Oaks, CA: Sage.

Feingold, A. (1988). Cognitive gender differences are disappearing. *American Psychologist, 43,* 95–103. http://dx.doi.org/10.1037/0003-066X.43.2.95

Feingold, A. (1992). Sex differences in variability in intellectual abilities: A new look at an old controversy. *Review of Educational Research, 62,* 61–84. http://dx.doi.org/10.3102/00346543062001061

Fernald, A., Marchman, V. A., & Weisleder, A. (2013). SES differences in language processing skill and vocabulary are evident at 18 months. *Developmental Science, 16,* 234–248. http://dx.doi.org/10.1111/desc.12019

Geary, D. C. (2010). *Male, female: The evolution of human sex differences* (2nd ed.). Washington, DC: American Psychological Association. http://dx.doi.org/10.1037/12072-000

Geary, D. C., Winegard, B., & Winegard, B. (2014). Reflections on the evolution of human sex differences: Social selection and the evolution of competition among women. In V. Weekes-Shackleford & T. K. Shackelford (Eds.), *Evolutionary perspectives on human sexual psychology and behavior* (pp. 393–412). New York, NY: Springer. http://dx.doi.org/10.1007/978-1-4939-0314-6_20

Gignac, G. E., & Szodorai, E. T. (2016). Effect size guidelines for individual differences researchers. *Personality and Individual Differences, 102,* 74–78. http://dx.doi.org/10.1016/j.paid.2016.06.069

Guiso, L., Monte, F., Sapienza, P., & Zingales, L. (2008). Diversity: Culture, gender, and math. *Science, 320,* 1164–1165. http://dx.doi.org/10.1126/science.1154094

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

Halpern, D. F., Beninger, A. S., & Straight, C. A. (2011). Sex differences in intelligence. In R. J. Sternberg & S. B. Kaufman (Eds.), *The Cambridge handbook of intelligence* (pp. 253–272). New York, NY: Cambridge University Press. http://dx.doi.org/10.1017/CBO9780511977244.014

Halpern, D. F., Eliot, L., Bigler, R. S., Fabes, R. A., Hanish, L. D., Hyde, J., . . . Martin, C. L. (2011). Education: The pseudoscience of single-sex schooling. *Science, 333,* 1706–1707. http://dx.doi.org/10.1126/science.1205031

Halpern, D. F., Straight, C. A., & Stephenson, C. L. (2011). Beliefs about cognitive gender differences: Accurate for direction, underestimated for size. *Sex Roles, 64,* 336–347. http://dx.doi.org/10.1007/s11199-010-9891-2

Halpern, D. F., & Tan, U. (2001). Stereotypes and steroids: Using a psychobiosocial model to understand cognitive sex differences. *Brain and Cognition, 45,* 392–414. http://dx.doi.org/10.1006/brcg.2001.1287

Hanscombe, K. B., Trzaskowski, M., Haworth, C. M. A., Davis, O. S. P., Dale, P. S., & Plomin, R. (2012). Socioeconomic status (SES) and children's intelligence (IQ): In a U.K.-representative sample SES moderates the environmental, not genetic, effect on IQ. *PLoS ONE, 7,* e30320. http://dx.doi.org/10.1371/journal.pone.0030320

Hawke, J. L., Olson, R. K., Willcut, E. G., Wadsworth, S. J., & DeFries, J. C. (2009). Gender ratios for reading difficulties. *Dyslexia, 15,* 239–242. http://dx.doi.org/10.1002/dys.389

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

Hyde, J. S. (2005). The gender similarities hypothesis. *American Psychologist, 60,* 581–592. http://dx.doi.org/10.1037/0003-066X.60.6.581

Hyde, J. S. (2014). Gender similarities and differences. *Annual Review of Psychology, 65,* 373–398. http://dx.doi.org/10.1146/annurev-psych-010213-115057

Hyde, J. S., & Grabe, S. (2008). Meta-analysis in the psychology of women. In F. Denmark & M. A. Paludi (Eds.), *Psychology of women: A handbook of issues and theories* (pp. 142–173). Westport, CT: Praeger.

Hyde, J. S., & Linn, M. C. (1988). Gender differences in verbal ability: A meta-analysis. *Psychological Bulletin, 104,* 53–69. http://dx.doi.org/10.1037/0033-2909.104.1.53

Jiménez, J. E., de la Cadena, C. G., Siegel, L. S., O'Shanahan, I., García, E., & Rodríguez, C. (2011). Gender ratio and cognitive profiles in dyslexia: A cross-national study. *Reading and Writing, 24,* 729–747. http://dx.doi.org/10.1007/s11145-009-9222-6

Kagan, J. (1964). The child's sex role classification of school objects. *Child Development, 35,* 1051–1056.

Kaiser, A., Haller, S., Schmitz, S., & Nitsch, C. (2009). On sex/gender related similarities and differences in fMRI language research. *Brain Research Brain Research Reviews, 61,* 49–59. http://dx.doi.org/10.1016/j.brainresrev.2009.03.005

Kaufman, A. S., Kaufman, J. C., Liu, X., & Johnson, C. K. (2009). How do educational attainment and gender relate to fluid intelligence, crystallized intelligence, and academic skills at ages 22–90 years? *Archives of Clinical Neuropsychology, 24,* 153–163. http://dx.doi.org/10.1093/arclin/acp015

Kelley, G. A., & Kelley, K. S. (2012). Statistical models for meta-analysis: A brief tutorial. *World Journal of Methodology, 2,* 27–32. http://dx.doi.org/10.5662/wjm.v2.i4.27

Kutner, M., Greenberg, E., Jin, Y., Boyle, B., Hsu, Y.-c., & Dunleavy, E. (2007). *Literacy in everyday life: Results from the 2003 National Assessment of Adult Literacy*. Washington, DC: National Center for Education Statistics.

Levy, J. (1969). Possible basis for the evolution of lateral specialization of the human brain. *Nature, 224,* 614–615. http://dx.doi.org/10.1038/224614a0

Limbrick, L., Wheldall, K., & Madelaine, A. (2011). Why do more boys than girls have a reading disability? A review of the evidence. *Australasian Journal of Special Education, 35,* 1–24. http://dx.doi.org/10.1375/ajse.35.1.1

Lynn, R., & Mikk, J. (2009). Sex differences in reading achievement. *Trames, 13,* 3–13. http://dx.doi.org/10.3176/tr.2009.1.01

Maccoby, E. E., & Jacklin, C. N. (1974). *The psychology of sex differences*. Stanford, CA: Stanford University Press.

Machin, S., & Pekkarinen, T. (2008). Assessment: Global sex differences in test score variability. *Science, 322,* 1331–1332. http://dx.doi.org/10.1126/science.1162573

Marinak, B. A., & Gambrell, L. B. (2010). Reading motivation: Exploring the elementary gender gap. *Literacy Research and Instruction, 49,* 129–141. http://dx.doi.org/10.1080/19388070902803795

Martin, C. L., & Ruble, D. N. (2010). Patterns of gender development. *Annual Review of Psychology, 61,* 353–381. http://dx.doi.org/10.1146/annurev.psych.093008.100511

McGee, R., Prior, M., Williams, S., Smart, D., & Sanson, A. (2002). The long-term significance of teacher-rated hyperactivity and reading ability in childhood: Findings from two longitudinal studies. *Journal of Child Psychology and Psychiatry, and Allied Disciplines, 43,* 1004–1017. http://dx.doi.org/10.1111/1469-7610.00228

McGraw, R., Lubienski, S. T., & Strutchens, M. E. (2006). A closer look at gender in NAEP mathematics achievement and affect data: Intersections with achievement, race/ethnicity, and socioeconomic status. *Journal for Research in Mathematics Education, 37,* 129–150.

Miller, D. I., & Halpern, D. F. (2014). The new science of cognitive sex differences. *Trends in Cognitive Sciences, 18,* 37–45. http://dx.doi.org/10.1016/j.tics.2013.10.011

Moffitt, M. A. S., & Wartella, E. (1991). Youth and reading: A survey of leisure reading pursuits of female and male adolescents. *Reading Research and Instruction, 31,* 1–17. http://dx.doi.org/10.1080/19388079209558075

Mucherah, W., & Yoder, A. (2008). Motivation for reading and middle school students' performance on standardized testing in reading. *Reading Psychology, 29,* 214–235. http://dx.doi.org/10.1080/02702710801982159

Mullis, I. V. S., Martin, M. O., Gonzalez, E., & Kennedy, A. M. (2003). *PIRLS 2001 International Report: IEA's study of reading literacy in primary schools in 35 countries*. Chestnut Hill, MA: TIMSS & PIRLS International Study Center.

Nash, S. C. (1979). Sex role as a mediator of intellectual functioning. In M. A. Wittig & A. C. Petersen (Eds.), *Sex-related differences in cognitive functioning: Developmental issues* (pp. 263–302). New York: Academic Press.

National Center for Education Statistics. (2016). Digest of Educational Statistics. Retrieved from: https://nces.ed.gov/programs/digest/

Nowell, A., & Hedges, L. V. (1998). Trends in gender differences in academic achievement from 1960 to 1994: An analysis of differences in mean, variance, and extreme scores. *Sex Roles, 39,* 21–43. http://dx.doi.org/10.1023/A:1018873615316

Organisation for Economic Cooperation and Development (OECD). (2016). *Education at a Glance 2016: OECD Indicators*. Paris: Author.

Plante, I., de la Sablonnière, R., Aronson, J. M., & Théorêt, M. (2013). Gender stereotype endorsement and achievement-related outcomes: The role of competence beliefs and task values. *Contemporary Educational Psychology, 38,* 225–235. http://dx.doi.org/10.1016/j.cedpsych.2013.03.004

Rabiner, D., & Coie, J. D., & The Conduct Problems Prevention Research Group. (2000). Early attention problems and children's reading achievement: A longitudinal investigation. *Journal of the American Academy of Child & Adolescent Psychiatry, 39,* 859–867. http://dx.doi.org/10.1097/00004583-200007000-00014

Reilly, D. (2012). Gender, culture, and sex-typed cognitive abilities. *PLoS One, 7*(7), e39904. http://dx.doi.org/10.1371/journal.pone.0039904

Reilly, D. (2015). *Gender differences in reading from a cross-cultural perspective: The contribution of gender equality*. Paper presented at the International Convention of Psychological Science, Amsterdam, Netherlands. http://dx.doi.org/10.13140/RG.2.2.18218.72647

Reilly, D., Neumann, D. L., & Andrews, G. (2015). Sex differences in mathematics and science: A meta-analysis of National Assessment of Educational Progress assessments. *Journal of Educational Psychology, 107,* 645–662. http://dx.doi.org/10.1037/edu0000012

Reilly, D., Neumann, D. L., & Andrews, G. (2016). Sex and sex-role differences in specific cognitive abilities. *Intelligence, 54,* 147–158. http://dx.doi.org/10.1016/j.intell.2015.12.004

Reynolds, M. R., Scheiber, C., Hajovsky, D. B., Schwartz, B., & Kaufman, A. S. (2015). Gender differences in academic achievement: Is writing an exception to the gender similarities hypothesis? *The Journal of Genetic Psychology, 176,* 211–234. http://dx.doi.org/10.1080/00221325.2015.1036833

Rosenthal, R., & Rubin, D. B. (1982a). Further meta-analytic procedures for assessing cognitive gender differences. *Journal of Educational Psychology, 74,* 708–712. http://dx.doi.org/10.1037/0022-0663.74.5.708

Rosenthal, R., & Rubin, D. B. (1982b). A simple, general purpose display of magnitude of experimental effect. *Journal of Educational Psychology, 74,* 166–169. http://dx.doi.org/10.1037/0022-0663.74.2.166

Rutter, M., Caspi, A., Fergusson, D., Horwood, L. J., Goodman, R., Maughan, B., . . . Carroll, J. (2004). Sex differences in developmental reading disability: New findings from 4 epidemiological studies. *Journal of the American Medical Association, 291,* 2007–2012. http://dx.doi.org/10.1001/jama.291.16.2007

Scheiber, C., Reynolds, M. R., Hajovsky, D. B., & Kaufman, A. S. (2015). Gender differences in achievement in a large, nationally representative sample of children and adolescents. *Psychology in the Schools, 52,* 335–348. http://dx.doi.org/10.1002/pits.21827

Shaywitz, B. A., Shaywitz, S. E., Pugh, K. R., Constable, R. T., Skudlarski, P., Fulbright, R. K., ... (1995). Sex differences in the functional organization of the brain for language. *Nature, 373,* 607–609. http://dx.doi.org/10.1038/373607a0

Shaywitz, S. E., Shaywitz, B. A., Fletcher, J. M., & Escobar, M. D. (1990). Prevalence of reading disability in boys and girls: Results of the Connecticut Longitudinal Study. *Journal of the American Medical Association, 264,* 998–1002. http://dx.doi.org/10.1001/jama.1990.03450080084036

Smart, D., Prior, M., Sanson, A., & Oberklaid, F. (2001). Children with reading difficulties: A six year follow-up from early primary school to secondary school. *Australian Journal of Psychology, 53,* 45–53. http://dx.doi.org/10.1080/00049530108255121

Spence, J. T., & Buckner, C. (2000). Instrumental and expressive traits, trait stereotypes, and sexist attitudes: What do they signify? *Psychology of Women Quarterly, 24,* 44–62. http://dx.doi.org/10.1111/j.1471-6402.2000.tb01021.x

Stanley, J. C., Benbow, C. P., Brody, L., Dauber, S., & Lupkowski, A. (1992). Gender differences on eighty-six nationally standardized aptitude and achievement tests. In N. Colangelo, S. G. Assouline & D. L. Ambroson (Eds.), *Talent development: Proceedings from the 1991 Henry B. and Jocelyn Wallace National Research Symposium on Talent Development* (pp. 42–65). Unionville, NY: Trillium Press.

Voyer, D., & Voyer, S. D. (2014). Gender differences in scholastic achievement: A meta-analysis. *Psychological Bulletin, 140,* 1174–1204. http://dx.doi.org/10.1037/a0036620

Voyer, D., Voyer, S., & Bryden, M. P. (1995). Magnitude of sex differences in spatial abilities: A meta-analysis and consideration of critical variables. *Psychological Bulletin, 117,* 250–270. http://dx.doi.org/10.1037/0033-2909.117.2.250

Wallentin, M. (2009). Putative sex differences in verbal abilities and language cortex: A critical review. *Brain and Language, 108,* 175–183. http://dx.doi.org/10.1016/j.bandl.2008.07.001

Zell, E., Krizan, Z., & Teeter, S. R. (2015). Evaluating gender similarities and differences using metasynthesis. *American Psychologist, 70,* 10–20. http://dx.doi.org/10.1037/a0038208

Received January 10, 2018
Revision received May 22, 2018
Accepted May 29, 2018 ■