When Do Gender Differences in Academic Achievement Originate? Examining Preschoolers and Early School Children Longitudinally

Daniel Sewasew[1], Missaye Mengsite[2] & Ebabush Kassa[2]

[1] Unaffiliated Researcher, Silver Spring, Maryland, USA

[2] Department of Psychology, University of Gondar, Gondar, Ethiopia

Correspondence: Daniel Sewasew, Silver Spring, Maryland, USA. E-mail: sewasewdaniel@gmail.com

Abstract

The skewed emphases on central tendency and dispersion statistics often provide an estimated summary of scores and variances of the overall distribution. Studies may therefore overlook significant variations across these distributions' different percentiles. This study examined gender academic disparities in STEM and reading subjects of the USA sample (Early Childhood Longitudinal Study-Kindergarten, ECLS-K:2011). The Quantile Regression (QR) model was utilized and found academic gender differences across school subjects, students' academic grades, and proficiency levels. There were often more differences in the extreme tails of the distribution than around the mean. The gender gap started with the top students in kindergarten and quickly spread throughout the distribution and primary school grades. Boys at the extreme ends of the distribution had the lowest reading scores by a significant margin. However, boys consistently rank among the top students in math and science. Early-age attention and intervention are needed to avert subsequent-grade academic achievement inequalities.

Keywords: academic achievement, gender gaps, quantile regression model, longitudinal study, early-age students

1. Introduction

In studying academic gender inequalities, the statistical and computational limitations significantly impact research data interpretation, conclusions, and reporting. Several factors contribute to the gender differences in academic achievements, such as biological, sociocultural, psychological, institutional, and people (women) vs. things (men) in terms of occupational interests. However, an overemphasis on the popular conditional mean analyses for gender differences has made a sizable contribution to the one-sided study reports, as the mean often gives an incomplete picture of a single distribution (Baye & Monseur, 2016).

The computational inadequacies have been attributed partly to statistical restrictions. For example, the group difference story is mainly seen from the "mean" perspective, i.e., on average. The popular analytical approach (i.e., frequency statistics) emphasizes central tendency and dispersion statistics that most often provide an estimated summary of scores and variances of the overall distribution. However, studies may overlook significant variations across the different percentiles of the distributions. An outstanding prospective research opportunity would be to theorize group similarities and quantify and test its hypothesis, as this has routinely been done using a group difference hypothesis (Andraszewicz et al., 2015).

In disciplines like educational and behavioral sciences, conditional mean models (e.g., ordinary least square, OLS) are primarily utilized as the favorite method for data analysis. The analytical simplicity and optimality of observing the deviation from the regression line typically made the OLS the workhorse regression analysis model for many years (Benoit & Van den Poel, 2017). However, conditional mean models offer only incomplete statistics on the impact of the covariate on the dependent variable. Consequently, the statistical outputs (e.g., *p*-values, effect sizes) of group gaps and the imbalance of reporting and interpreting the findings of group differences have confirmed misleading evidence. Therefore, a non-significant group difference, that is, group similarity, should be as capturing and relevant as a group difference throughout distribution (Benoit & Van den Poel, 2017).

Computational limitations have far-reaching impacts on shaping the research questions/hypotheses and content under investigation. For example, in academic achievement, the gender similarity hypothesis could be tested not only by the conditional mean and variance analysis tools. On the other hand, if we need an academic competence-specific intervention for a small group or an individual level, quantile regression (QR) models at a

different portion of distribution would be an appropriate analytical approach (Baye & Monseur, 2016). Here, note that the contexts in which gender differences are created or erased should be stressed because the magnitude of effect sizes depends on the assessment accuracy and the diversity of the study population. Hence rough categories (i.e., small, medium, and large effect size) that provide a general guide should be substantiated by specific contexts—discussed in light of domain-related empirical benchmarks (Baye & Monseur, 2016; Hill et al., 2008).

The literature has shown that computational restrictions have serious consequences (e.g., Baye & Monseur, 2016; Hill et al., 2008; Penner & Paret, 2008). First, reporting academic gender gaps from central tendency statistics is one-sided because gender differences at the extreme tails of the distribution can differ from what is observed with central tendency indices. This could lead to an overly simplistic evaluation of the actual gender gap that might need serious intervention. Second, in academic gender differences, using only average performance-based measures could ruin the possibility of implementing competence-specific target group interventions (e.g., the least or the most proficient students). Third, the "rough" conditional mean estimation does not simplify the efforts to help students understand how a juxtaposition of one subject skill (e.g., math or science) with another (e.g., verbal) can be a relevant asset to specific careers (e.g., STEM; science, technology, engineering, and math).

Overall, even though a large volume of research has documented academic gender differences, given the current state of research the following are pressing study directions: (*i*) detecting gender gaps throughout the academic distribution longitudinally is highly relevant and informative, especially beginning from the kindergarten period, to know the genesis of any group differences; (*ii*) discussing contexts where gender differences appear or disappear will remain a practical approach in primary, secondary, and meta-analyses investigations.

1.1 Academic Achievement and Gender Differences

Although the gender disparity issue has received ample attention over the last several decades, the evidence of closing small or trivial gender gaps in academic achievements should be cautiously acknowledged, as an interpretation of gender differences is context-specific. Moreover, those small gaps could pave the way to the enlargement of gender differences and subsequently reason for occupation or career gender disparities (Penner & Paret, 2008).

The academic gender gaps debate has long been acknowledged but is no longer the concern of only one gender group (i.e., females). Instead, the underachievement of students in certain subjects (e.g., boys in reading) is of increasing concern. Still, scant educational investigations have considered this concern regarding gender differences among low-achieving students, compared to those stressing boys' superior success among high-achieving students in math and science (Baye & Monseur, 2016). This high focus on accomplishment in math and science could be attributed to the significance of numerical skills in many STEM occupations (Halpern et al., 2007). Nevertheless, poor performance in language abilities is also a severe disadvantage to accessing the labor market (OECD, 2000).

The skewed emphasis of gender gap investigations on academic achievement has most often focused on older students, primarily late elementary and above grades. However, despite not being consistent (Lachance & Mazzocco, 2006), gender gaps have been documented earlier in development (Stoet & Geary, 2013) although these early gender gaps are fewer than those revealed in later periods of students' academic records. Such differences are vital since primary achievement cavities result in more differences in the later educational system and occupational careers (Penner & Paret, 2008). Nonetheless, the insignificant and trivial differences must also be theorized, statistically estimated, and justified. Hence, we claim that if any early gender gaps in academic achievement exist, they must be considered seriously. If these do not exist, appropriate theoretical backup, hypothesis-based testing, and empirical benchmark discussions (e.g., gender similarity) should be included in such arguments.

Recent empirical investigations (Baye & Monseur, 2016) based on the ten latest international large-scale assessments (PIRLS, PISA, TIMSS since 1995) for primary and secondary students and comprehensive analyses indicated that gender gaps vary by content area and students' educational and proficiency levels. Furthermore, the authors found out that the gender gaps at the extreme tails of the distribution were often more substantial than on the mean. For example, the most significant gap in reading was observed (0.58) for the weakest students (percentile 5). In contrast, in math, it was documented (-0.24) for the most proficient students (percentile 95), and in science there was a slight tendency for girls to be more proficient at the lower tail and for males to be more proficient at the upper tail. However, their fundamental analyses were based on only cross-sectional survey data. The discussion did not consider the gender similarity hypothesis for the non-significant results because of analysis limitation (i.e., frequentist approach). Moreover, Baye and Monseur settled on only reporting academic and gender relations thereby neglecting the role of social class and migration background (i.e., intersectional approach).

From a longitudinal perspective, three comprehensive analyses (Cimpian et al., 2016; Penner & Paret, 2008; Robinson & Lubienski, 2011) using a quantile regression model to examine gender gaps for the USA sample students revealed no math gender differences in kindergarten for males, except at the top of the distribution; by contrast, females throughout the distribution were found to lose ground in elementary school. In reading, differences favoring females generally narrow but widen among low-achieving students. Moreover, gender gaps were reported to be varied (Penner & Paret, 2008) across parental education (a male advantage for high parental education students) and migration background (Asian male and Latino girls surpassed their counterparts at the upper distribution). However, these authors did not document the effect size of gender gaps across the distribution, and their discussion has been confined only to the empirical benchmarks of gender differences.

Overall, the large volume of academic and gender-related investigations has been documented at a single time point of individual studies. However, only a few studies have investigated multi-content domains (usually either verbal or math, but possibly more than two) longitudinally, from kindergarten through elementary school. In the present study, the cross-time trends of gender gaps in variability and effect size throughout the distributions in three domains (math, science, and reading) are investigated. Also, this study compares both empirical benchmarks (gender differences) to give more meaning to the effect of gender differences in the academic achievement context.

1.2 The Current Study

The current study investigates females' and males' exclusive academic achievement tendencies in math, science, and reading. More precisely, the following research questions are addressed:

    a) If there is a gender difference in academic competence, at what age do these differences first originate, and where are the gaps most predominant in the achievement distribution?

    b) If gender differences exist, how do they relate/differ between the content area and the age/grade of the students?

    c) If any, to what level do gender differences at the extreme tails of the distribution diverge compared to gender mean differences, and in which direction?

2. Method

2.1 Data Collection and Sample

The "Early Childhood Longitudinal Study, Kindergarten Class of 2011" (ECLS-K:2011) data were collected by the U.S.A. Department of Education. The ECLS-K includes data on a public nationally representative sample of more than 18,174 kindergarten through grade five students in the academic years 2010 - 2016 (Tourangeau et al., 2015). For study purposes, the present researcher selected students who participated in kindergarten and primary school sample waves: fall and spring kindergarten (2010 -11); fall and spring first grade (2011-12); fall and spring second grade (2012 -13); spring third grade (2014); spring fourth grade (2015). Of the total students, 51.1% ($n = 9$, 288) were males, and 48.8% ($n = 8$, 847) were females. All waves of data are available from the National Center for Education Statistics (NCES), as are technical manuals describing the use of the data, sampling plans, and weights (http://nces.ed.gov/ecls/dataproducts.asp).

2.2 Measures

2.2.1 Academic Achievement

Students' academic achievement was assessed with standardized tests. The reading, mathematics, and science assessments were administered individually to the sampled children by trained and certified child assessors. The reading test measured basic language skills (e.g., letter recognition, sound recognition, decoding multisyllabic words), vocabulary (vocabulary knowledge, receptive vocabulary), and passage comprehension (text interpretation using prior knowledge). As children's education levels progressed, the content of the tests changed from basic language skills to more complex comprehension skills in the respective grades. The math tests measured students' conceptual knowledge, procedural knowledge, and problem-solving skills. The mathematical content included number sense, properties, and operations, while measurement, geometry, algebra, and essential functions were added later. The science assessment included questions about physical, life, Earth and space sciences, and scientific inquiry (for more information, see Tourangeau et al., 2015). The NCES (2011) used a multistage panel review process to develop ECLS-K's verbal and math tests, which were based on the specifications of the NAEP (*National Assessment of Educational Progress*). All test items were exhaustively pilot-tested. In addition, their construct validity was examined with the *Woodcock-McGrew-Werder Mini-Battery of Achievement* (see Pollack, Najarian, Rock, & Atkins-Burnett, 2005; Tourangeau et al., 2015).

The NCES used item-response theory (IRT) methods to generate adaptive tests administered one-on-one to each

child in an untimed format. After a brief routing test, each student was administered a test that matched their performance (categorized as low, middle, or high performance). Such a tailored testing procedure minimized the risk of floor and ceiling effects and kept students motivated during the assessment. Technically, the NCES conducted one-parameter IRT (Rasch) analyses for scholastic competence. To compare students' competence across grades on a common scale, grade-specific test forms were equated using linking items (for details, see Tourangeau et al., 2015). We chose the latest revised version of ECLS-K IRT theta scores (math, science, and reading) as criterion-referenced achievement measures for the main analyses. The IRT reliability coefficients for math, science, and math achievement were consistently high kindergarten through grades in nine wave measurements (reading: $\alpha = .92/.94/.95/.95/.91/.90/.86/.87$; math: $.92/.93/.93/.93/ .93/.94/ .92/.91$; science:$.73/.83/.84/.86/.85/.83/.82$).

2.3 Statistical Analysis

The current study utilized free software called R for statistical computation (R Core Team, 2016). Specifically, the quantreg package (Koenker R, 2016) was used for quantile analysis. The data analysis part involves multi-level group estimation in academic competence: first, the QR is estimated in the conditional Median and at the different proficiency levels of the distribution (e.g., 5, 10, 25, 75, 90 percentiles) across a covariate (gender) using the quantreg package. When estimating the percentiles for boys and girls individually, QR employs the most negligible absolute value and then compares the percentiles for the two genders. As a result, QR reports conditional differences in percentiles instead of dependent differences in means as OLS does. More significantly, the models estimated in this research adopt the stance described as $y_i = X_i \beta + \varepsilon_i$. where y_i is the achievement score for student i, and X_i includes the independent variables, in this case, the constant and a dummy variable for being female (Koenker & Bassett, 1978). Subsequently, the group difference at the mean was also determined using OLS. Finally, the effect size was calculated for the different percentiles and the means. Precisely, the Hedges' g (Hedges & Friedman, 1993) method of evaluating the magnitude of group differences was utilized, as suggested (pooling the standard deviation) by extensive studies (e.g., PISA, OCED, 2009).

3. Results

The results section is presented in the following sequence: Figure 1 displays the general trend of academic achievements (math, science, and reading) using conditional means. Figure 2 shows academic achievements (math, science, and reading) using quantile regression at different distribution percentiles for each school subject separately. Table 1 presents a detailed view of the gender differences in conditional means, quantile regression, and Cohen's d (showing effect size magnitude).

Figure 1 displays the overall trend of academic achievements (math, science, and reading). As indicated below, girls generally outperform boys in reading from kindergarten to primary school, as their test scores remain positive. However, they underperform boys in math and science throughout kindergarten and primary school, except for fall kindergarten through grade 1.

1: Gender differences in math, science, and reading competence across grades using conditional mean analysis
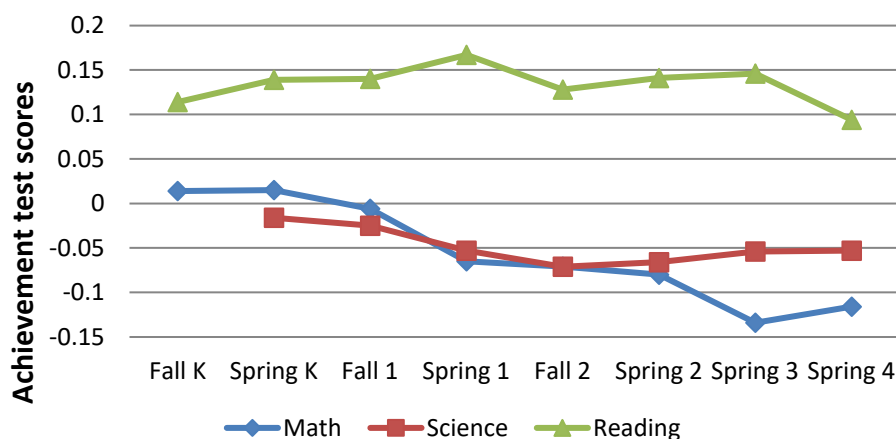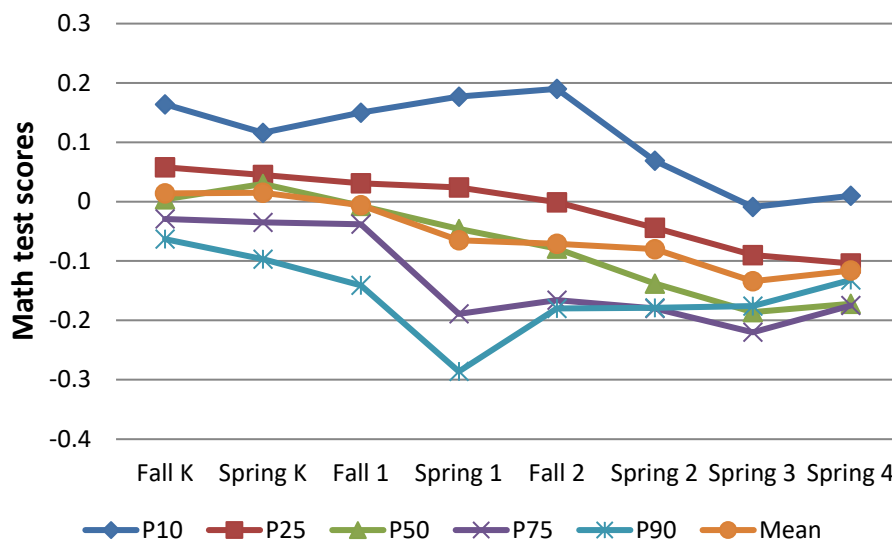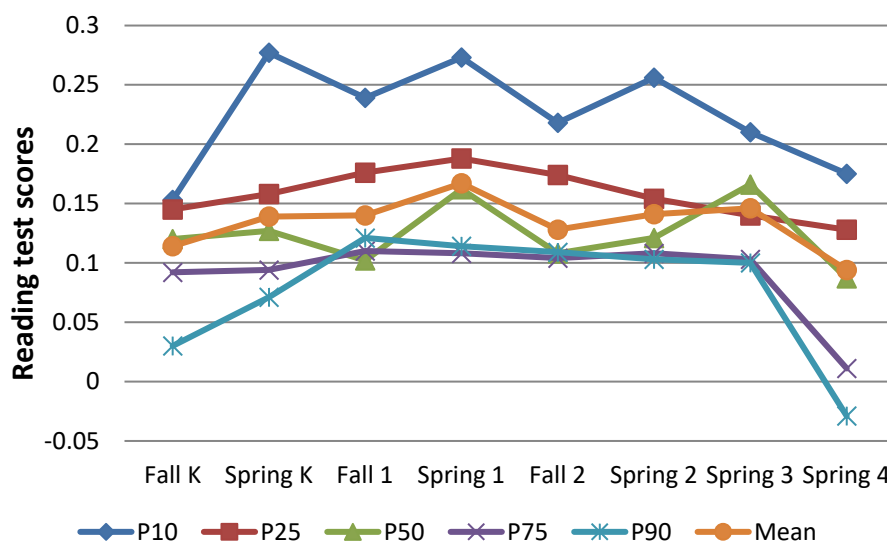

Figure 1. $n =16,936$. Gender differences based on the conditional mean analyses of math, reading, and science domains using standardized test scores. Positive values indicate higher scores for females

Figure 2 portrays separate academic achievement test scores (math, reading, and science) at different distribution percentiles. Overall, in math and science, females performed below the median in the 10th and 25th percentiles of the distribution. In contrast, boys score above the 50th percentile of the distribution (75th and 90th percentiles). However, in reading, females outperform boys in almost all of the percentile distributions (detail presented in table 1 analysis).

1: Gender differences in math competence at different proficiency levels across grades

2: Gender differences in reading competence at different proficiency levels across grades

3: Gender differences in science competence at different proficiency levels across grades
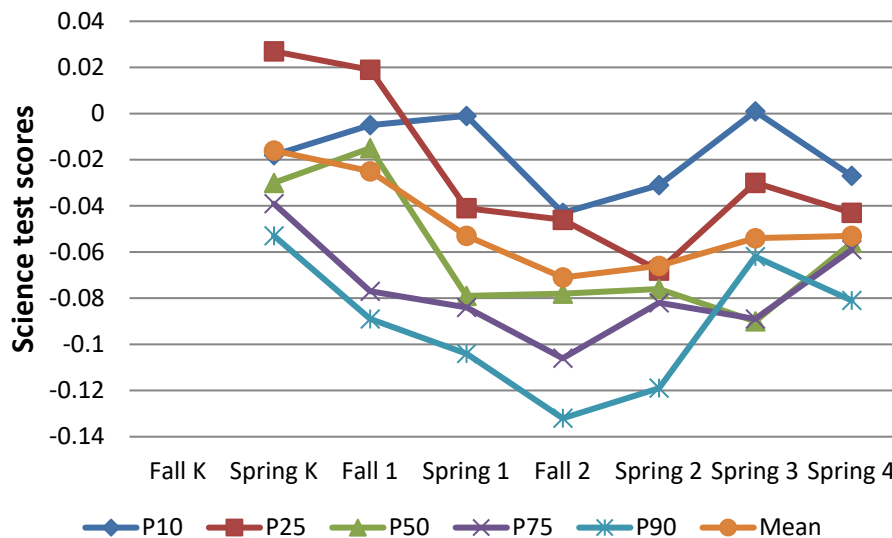
Figure 2. *n* =16, 936. Gender differences at the different percentile of the distribution based on a quantile regression model with standardized test scores for math, reading, and science domains. Positive values indicate higher scores for females

Detailed results are presented in Table 1. In math, overall, the conditional mean analysis showed an increasing trend of gender differences, with minor differences close to zero during kindergarten. However, nontrivial gender differences were found after students joined formal schooling, particularly in grades three and four (SDs = -0.134, -0.116; *d* = -0.175, -0.15, respectively; see Table 1, refer horizontally). A further close analysis of students' competence throughout the achievement distribution is needed, as the overall comparison of gender differences can obscure significant differences between genders. During the kindergarten periods (i.e., Fall and Spring), boys do less at the lower end of the distribution than girls (range gap from SDs = 0.004 to 0.164; *d* = 0.00 to 0.16). That means we observed that girls score higher across the achievement distribution of the 1st through median percentiles for boys and girls. However, at the upper tails of the distribution (i.e., 75th percentiles and above), boys score higher than girls (range gap from SDs = -0.029 to -0.097; *d* = -0.03 to -0.12). Although the effect sizes were small in kindergarten, reasonably comparable effect sizes were obtained in the upper and lower tails. Hinting at the gender gap but not at the mean indicates that investigating mean differences cannot spot the significant gender gaps. which previously existed before students started schooling, justifying the necessity of examining gender differences across the distribution.

At the beginning of formal schooling (i.e., fall 1st grade), a similar overall trend was observed: at the lower tail, females are at an advantage; at the upper tail, males have the advantage (gap SDs = 0.150 and -0.141; *d* = 0.19 and -0.11, respectively); and at the median and mean, there is an area with no gender differences. However, a more pronounced and significant male dominance across a large portion of the distribution (including the median and above) is shown in the spring for 1st graders, whereas at the bottom of the distribution, females are at an advantage (gap SDs = -0.286 and 0.177; *d* = -0.32 and 0.22, respectively). In grade 2 for the fall semester, results show a significant advantage for boys, with all points above the 10th percentile and especially from above the median (range gap from SDs = -0.180 to 0.190). For the spring term, students showed an outstandingly constant gender difference across the distribution from the 10th percentile up (range gap from SDs = -0.180 to 0.069; *d* = -0.21 to 0.16). In grade 3, the division widened, and boys took all the advantage from girls across the distribution and scored significantly above the 10th percentile (range gap from SDs = -0.009 to 0.220; *d* = 0.06 to -0.27). Finally, a similar pattern of boys' significant advantage was observed in fourth grade from above the 10th percentile (range gap from SDs = 0.010 to -0.175; *d* = 0.05 to -0.22). Overall, considerable gender gaps were exhibited, favoring females' math competence at the bottom of the quartile beginning for kindergarten entry; however, they were underrepresented in the top percentiles of the distribution across time. In the last three consecutive grades (grades 2, 3, and 4), boys who were advantaged across a large portion of the distribution were consistently documented from above the 10th percentile. However, the differences at the top of the distribution did not vary substantially from those across the rest. On the other hand, the effect size result for math essentially revealed less than reading; however, the effect size differs consistently with the student's competence level. At the bottom of the distribution,

effect sizes favored females, whereas steadily, at the top of the distribution, males were more proficient than females. The most significant gap in math was observed in the spring of first grade (-0.32) for the most proficient students (percentile 90).

Despite the small effect size of the gender gap in the conditional mean, the overall gender differences in reading achievement favoring females were stable and significant from kindergarten throughout elementary school years (range gap from SDs = 0.094 to 0.167; d = 0.15 to 0.22). Further quantile regression analyses were run to trace the gender gap across the distribution's different percentiles, presented below. Overall, consistent significant gender gaps supporting females' advantage throughout the distribution in the fall of kindergarten, stable through the spring of fourth grade, were observed. The gender difference at the beginning of kindergarten is about 0.15 SDs (d =0.18) at the 10th percentile and shows a decreasing trend of gender gaps towards the upper tails of the distribution with the lowest 0.03 SDs (d =0.05) at the 90th percentiles. After the fall semester of kindergarten, the gender gaps at all the respective quantiles show increasing gender differences until the spring of third grade, declining afterward. The large gender gap was observed at the bottom of the quantiles, with the highest 0.27 SDs (d =0.40) at the 10th percentiles of spring first graders and the lowest 0.17 SDs (d =0.37) of spring fourth graders. In the other portion of the distribution, the gap reached about 0.07 SDS to 0.18 SDs (d = 0.10 to 0.26) interquartile ranges. In the spring of fourth grade, a decline in the gender gap was exhibited compared to the previous school years (0.01 to 0.17 SDs; d = 0.04 to 0.37 range) and across the quantiles, with unique males, a significant advantage in the upper percentile of the distribution (-0.02 SD; d = -0.05).

In summary, gender differences have been shown favoring females throughout the distribution beginning from kindergarten. At the bottom percentiles of the distribution, significant reading competence gender gaps were exhibited. At the top quantiles, small gender gaps were witnessed, and the advantage was shifted to males in the upper elementary grades. Comparisons of the effect size were also analyzed. In the case of reading, although the gender gaps were comparably higher throughout the ability distribution compared to math, they were especially large at the lower tail with effect sizes sometimes almost twice as large as at the upper bottom. The most significant gap in reading was observed in the spring of second grade (0.44) for the weakest students (percentile 10).

Turning to the science achievement scores, the mean result revealed consistent statistical significance with small variability (range gaps from SDs = -0.16 to -0.071) in favor of males, after students began formal schooling; there was almost zero effect size from kindergarten through the elementary school years (from d = -0.02 to -0.07). However, we understand that gender differences differ considerably throughout the distribution (see Table 1). For instance, from the 1st percentile to the median in the kindergarten period, there were no significant gender differences in average science competence for boys and girls. However, at the 75th percentile and above, boys scored significantly higher than girls (-0.039 and -0.053 SDs; d = -0.05 and -0.05, respectively). A similar trend was observed at the beginning of first grade: that is, there were no gender differences in the lower tails (10th through median) but significant gender differences in science competence favoring the boys in the upper tails (75th and 90th). In the spring of first grade, beginning from the median through the upper tails, boys scored significantly higher in science competence than girls; however, below the median, the gender differences were small and insignificant. By the fall and spring of second grade, the significant male advantage was apparent at the 25th percentile and reached its strongest level at the 90th percentile (-0.132 SD; d = -0.09). Both third and fifth grades revealed consistent gender differences through the distribution from the 25th percentile up. Overall, from kindergarten through grade four, males at the top of the distribution (75th percentile and above) consistently exhibit higher science competence than females. Hence, students at the upper tails of science achievement still show a pattern of the most significant male advantage. Regarding effect size analyses, science results appear like math; however, in both the upper and lower tails of the distribution, the effect size was close to zero, mainly favoring males.

To summarize, in STEM subjects (math and science), the mean gender gaps are evidenced as early as the spring of first grade. Across the achievement distributions, the gender differences emerged as early as the fall and spring of kindergarten at the top of the achievement distribution and spread throughout the distribution during the elementary school years. In reading, consistent mean gender differences are obtained as early as the fall of kindergarten through fourth grade. Besides, females show a gain in reading at all proficiency levels of the achievement distribution across time. The effect size on the means differs for different types of outcomes and different levels of education. The gender gaps in reading are higher than gender differences in math and science. In reading, gender gaps increase with age, whereas in math and science, gender gaps increase slightly.

Table 1. Descriptive statistics and gender differences for math, reading, and science scores by data collection point

Domain	Waves of data collection							
	Fall kindergarten	Spring kindergarten	Fall 1st grade	Spring 1st grade	Fall 2nd grade	Spring 2nd grade	Spring 3rd grade	Spring 4th grade
Math								
Mean	-0.49	0.45	0.83	1.65	1.85	2.45	3.05	3.42
SD	0.93	0.77	0.85	0.85	0.85	0.82	0.77	0.78
Differences by gender (female-male) in standard deviation units								
Mean	0.014	0.015	-0.006	-0.065**	-0.071**	-0.080**	-0.134**	-0.116**
Cohen's *d*	-0.015	-0.019	0.007	0.076	0.083	0.098	0.175	0.149
Quantiles								
10%	0.164**	0.116**	0.150**	0.177**	0.190**	0.069**	-0.009	0.010
25%	0.058*	0.045*	0.031	0.024	-0.001	-0.044*	-0.090**	-0.104**
Median	0.004	0.030*	-0.007	-0.046**	-0.079*	-0.138**	-0.186**	-0.172**
75%	-0.029	-0.035**	-0.038	-0.189**	-0.166**	-0.180**	-0.220*	-0.175**
90%	-0.063**	-0.097**	-0.141**	-0.286**	-0.180**	-0.179**	-0.176**	-0.132**
N	15596	17144	5223	15104	4730	13831	12867	12081
Reading								
Mean	-0.54	0.45	0.85	1.59	1.80	2.19	2.6	2.88
SD	0.85	0.79	0.80	0.76	0.67	0.64	0.66	0.62
Differences by gender (female-male) in standard deviation units								
Mean	-0.114**	-0.139**	-0.140**	-0.167**	-0.128**	-0.141**	-0.146**	-0.094**
Cohen's *d*	-0.134	-0.178	-0.177	-0.221	-0.191	-0.222	-0.223	-0.151
Quantiles								
10%	0.153**	0.277**	0.239**	0.273**	0.218**	0.256**	0.210**	0.175**
25%	0.145**	0.158**	0.176**	0.188**	0.174**	0.154**	0.140**	0.128**
Median	0.120**	0.127**	0.102**	0.162**	0.108**	0.121**	0.166**	0.087**
75%	0.092**	0.094**	0.110**	0.108**	0.104**	0.108**	0.103**	0.011**
90%	0.030	0.071*	0.121**	0.114**	0.109**	0.103**	0.100**	-0.029*
N	15,669	17,185	5,194	15,115	4,725	13,837	12,866	12,074
Science								
Mean		-0.01	0.28	0.85	1.05	1.59	2.18	2.58
SD		0.90	0.98	0.89	1.08	0.94	0.80	0.84
Differences by gender (female-male) in standard deviation units								
Mean	-	-0.016	-0.025	-0.053**	-0.071*	-0.066**	-0.054**	-0.053**
Cohen's *d*	-	0.018	0.025	0.054	0.066	0.070	0.068	0.063
Quantiles								
10%	-	-0.018	-0.005	-0.001	-0.043	-0.031	0.001	-0.027
25%	-	0.027	0.019	-0.041	-0.046	-0.068**	-0.030**	-0.043*
Median	-	-0.030	-0.015	-0.079**	-0.078*	-0.076**	-0.090**	-0.056**
75%	-	-0.039**	-0.077*	-0.084**	-0.106**	-0.082**	-0.089**	-0.059**
90%	-	-0.053**	-0.089*	-0.104**	-0.132**	-0.119**	-0.062**	-0.081**
N		16, 936	5,180	15, 072	4,724	13,819	12,856	12,069

Note. A positive value indicates a higher score for females; a negative value indicates a higher score for males at different percentiles, 10, 25, Median, 75,90, and the mean score. Significance difference between gender scores: *$p < .05$, **$p < .01$, ***$p < .0001$.

4. Discussion

Using quantile regression models and effect size, we analyzed extensive representative data (ECLS-K: 2011) from kindergarten through elementary school years to examine gender gaps throughout the math, science, and reading distributions. Specifically, we analyzed portions in the achievement distributions where gender gaps were revealed and pinpointed the grade levels when these differences first appeared.

In math, we found that across the achievement distributions, gender differences emerged as early as the fall and spring of kindergarten at the top of the achievement distribution and spread throughout the distribution during the primary schooling years. This result replicated the previous finding using another cohort of the ECLS-K: 1999, that is, the underrepresentation of girls at the top of the distribution (Cimpian et al., 2016; Edossa et al., 2022; Penner & Paret, 2008; Robinson & Lubienski, 2011). Considering effect size analyses, the present study reveals that females inconsistently overtake males at the bottom of the distributions. In contrast, consistently at the top of the distribution, males are more proficient than females. The most significant gap in math is observed in the spring of first grade (-0.32) for the most proficient students (percentile 90). A similar effect size of gender gaps is reported on the PISA 2003 survey (-0.24) for the most proficient students (percentile 95; Baye & Monseur, 2016). The current findings are beneficial as they give/extend replication evidence that the male advantage in math across time among higher-scoring students is present at an early kindergarten age. Also, these findings help explain the familiar achievement gaps frequently revealed at the end of high school (Penner & Paret, 2008).

In terms of reading skills, this study has revealed gender differences in favor of females beginning from kindergarten and continuing, throughout the distribution. Among lower-scoring students, more significant reading competence gender gaps were observed. Among higher-scoring students, however, small gender gaps were witnessed, and the advantage was shifted to males in the upper primary grades. The effect size of gender gaps in some portions of the distribution were about twice as large as at the upper tail. The most significant gap in reading was observed in the spring of second grade (0.44) for the weakest students (percentile 10), to the disadvantage of males. Similar result patterns are reported in the cross-sectional large-scale international investigations by Baye and Monseur (2016) with the lowest students (percentile 5) of the enormous reading gap (0.58) on the PISA 2012 survey as well as the longitudinal designs using SDs as the unit of measurement (see Penner & Paret, 2008; Robinson & Lubienski, 2011). The significant and consistent gender gaps in reading competence at the mean and across most of the distribution should be taken seriously. It hints that most low-performing students are males, and their circumstances are far worse than that of females.

In science, we found that from kindergarten through grade four, at the top of the distribution, males consistently exhibit higher science competence than females. Hence, the upper tails of science achievement still have the most significant male advantage, like math. Moreover, even though one can barely find comparable longitudinal studies to support the present findings, many individual and large-scale cross-sectional studies revealed similar trends (Halpern & LaMay, 2000; Robinson & Lubienski, 2011). Our fundamental analyses of the competence distribution gave us further information beyond the conventional reports of the mean difference.

4.1 Conclusion

Academic gender differences are found across school subjects, students' academic grades, and proficiency levels. Overall, there are often more differences in the extreme tails of the distribution than around the mean. The gender disparity in this study began early among kindergarten high achievers and expanded immediately throughout the distribution and through primary school grades. Boys at the extreme ends of the distribution had the lowest reading scores by a significant margin. However, boys consistently rank among the top students in math and science. Early-age attention and interventions are needed to avert subsequent-grade academic achievement inequalities.

4.2 Implications

The present study's findings unpacked many relevant content, methods, practice, and policy implications. Content-wise, this study highlights females among higher-scoring students in STEM subjects (i.e., math and science). In contrast, males among lower-scoring students require serious consideration beginning from kindergarten to back up their enduring goals and visions in reading. This could be the initial point at which low-achieving students start to lag their higher-achieving peers. Therefore, although the strength of gender gaps in achievement studies might be considered a small effect size, it should not be underestimated. Rather, as suggested by the results reported here, these data should be cautiously interpreted, considering the lower proportion of females in STEM studies in college and in matching certified careers. Moreover, the differential position of gender gaps at the various portions of the achievement distribution across other subject areas necessitates complete attention to how both genders are grouped, tracked, retained, and selected in education systems (Baye & Monseur,

2016).

Data and Reproducibility

The data used for this study were obtained from the USA "Early Childhood Longitudinal Study, Kindergarten Class of 2011" (ECLS-K:2011), a project aimed at assessing educational, physical, and social development among children over time (Tourangeau et al., 2015). All ECLS-K, kindergarten through fifth-grade public-use files are available from the National Center for Educational Statistics (http://nces.ed.gov/ecls/dataproducts.asp).

References

Andraszewicz, S., Scheibehenne, B., Rieskamp, J., Grasman, R., Verhagen, J., & Wagenmakers, E.-J. (2015). An Introduction to Bayesian Hypothesis Testing for Management Research. *Journal of Management*, *41*(2), 521-543. https://doi.org/10.1177/0149206314560412

Baye, A., & Monseur, C. (2016). Gender differences in variability and extreme scores in an international context. *Large-Scale Assessments in Education*, *4*(1). https://doi.org/10.1186/s40536-015-0015-x

Benoit, D. F., & Van den Poel, D. (2017). A Bayesian Approach to Quantile Regression. *Journal of Statistical Software*, *76*(7). https://doi.org/10.18637/jss.v076.i07

Cimpian, J. R., Lubienski, S. T., Timmer, J. D., Makowski, M. B., & Miller, E. K. (2016). Have Gender Gaps in Math Closed? Achievement, Teacher Perceptions, and Learning Behaviors Across Two ECLS-K Cohorts. *AERA Open*, *2*(4), 233285841667361. https://doi.org/10.1177/2332858416673617

Edossa, A. K., Lockl, K., & Weinert, S. (2022). Developmental Relationship Between Metacognitive Monitoring and Reading Comprehension. *Journal of Educational and Developmental Psychology*, *13*(1), 1. https://doi.org/10.5539/jedp.v13n1p1

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), 1-51. https://doi.org/10.1111/j.1529-1006.2007.00032.x

Halpern, D., & LaMay, M. L. (2000). The Smarter Sex: A Critical Review of Sex Differences in Intelligence. *Educational Psychology Review*. https://doi.org/doi.org/10.1023/A:1009027516424

Hedges, L. V., & Friedman, L. (1993). Gender Differences in Variability in Intellectual Abilities: A Reanalysis of Feingold's Results. *Review of Educational Research*, *63*(1), 94-105. https://doi.org/10.3102/00346543063001094

Hill, C. J., Bloom, H. S., Black, A. R., & Lipsey, M. W. (2008). Empirical Benchmarks for Interpreting Effect Sizes in Research. *Child Development Perspectives*, *2*(3), 172-177. https://doi.org/10.1111/j.1750-8606.2008.00061.x

Koenker, R. (2016). *Quantreg: An R package for quantile regression and related methods. R package version 5.29*, ResearchGate. Retrieved from https://cran.r-project.org/package=quantreg

Koenker, R., & Bassett, G. (1978). Regression Quantiles. *Econometrica*, *46*(1), 33. https://doi.org/10.2307/1913643

Lachance, J. A., & Mazzocco, M. M. M. (2006). A longitudinal analysis of sex differences in math and spatial skills in primary school age children. *Learning and Individual Differences*, *16*(3), 195-216. https://doi.org/10.1016/j.lindif.2005.12.001

OCED. (2009). *PISA data analysis manual SAS® (2nd ed.). Paris: OECD.* Retrieved from http://www.oecd.org/pisa/pisaproducts/pisa2012technicalreport.htm

OECD. (2000). Literacy in the Information age: Final Report of the International Adult Literacy Survey. *Paris: OECD; Canada: Minister of Industry.* Retrieved from https://www.researchgate.net

Penner, A. M., & Paret, M. (2008). Gender differences in mathematics achievement: Exploring the early grades and the extremes. *Social Science Research*, *37*(1), 239-253. https://doi.org/10.1016/j.ssresearch.2007.06.012

Pollack, J. M., Najarian, M., Rock, D. A., & Atkins-Burnett, S. (2005). *Early Childhood Longitudinal Study, Kindergarten Class of 1998?99 (ECLS-K). Psychometric Report for the Fifth Grade. NCES 2006?036.* ED Pubs, P. Retrieved from https://eric.ed.gov/?id=ED489079

R Core Team. (2016). *R: The R Project for Statistical Computing.* Retrieved from https://www.r-project.org/

Robinson, J. P., & Lubienski, S. T. (2011). The Development of Gender Achievement Gaps in Mathematics and

Reading During Elementary and Middle School: Examining Direct Cognitive Assessments and Teacher Ratings. *American Educational Research Journal*, *48*(2), 268-302. https://doi.org/10.3102/0002831210372249

Stoet, G., & Geary, D. C. (2013). Sex Differences in Mathematics and Reading Achievement Are Inversely Related: Within- and Across-Nation Assessment of 10 Years of PISA Data. *PLoS ONE*, *8*(3). https://doi.org/10.1371/journal.pone.0057988

Tourangeau, K., Nord, C., Le, T., Sorongon, A. G., & Najarian, M. (2009a). *Early Childhood Longitudinal Study, Kindergarten Class of 1998-99 (ECLS-K): Combined User's Manual for the ECLS-K Eighth-Grade and K-8 Full Sample Data Files and Electronic Codebooks. NCES 2009-004*. National Center for Education Statistics. Retrieved from https://eric.ed.gov/?id=ED511826

Tourangeau, K., Nord, C., Le, T., Sorongon, A. G., & Najarian, M. (2009b). *Early Childhood Longitudinal Study, Kindergarten Class of 1998-99 (ECLS-K): Combined User's Manual for the ECLS-K Eighth-Grade and K-8 Full Sample Data Files and Electronic Codebooks. NCES 2009-004*. National Center for Education Statistics. Retrieved from https://eric.ed.gov/?id=ED511826

Copyrights