Exploring Gender Differences in Children's Early Reading Development in the U.S.

Exploring Gender Differences in Children's Early Reading Development in the U.S.

Qualifying Paper

Submitted by

Laura Mesite

September, 2017

Acknowledgements

First, I would like to acknowledge my advisor, Dr. Gigi Luk, for all of her support chairing my committee and helping me shape my doctoral career. Next, I would like to acknowledge Dr. Meredith Rowe for her patience in helping me identify a suitable research topic and methodological approach among the numerous proposals and drafts that I asked her to read over the past two years. I would also like to acknowledge Dr. James Kim for sponsoring my access to the restricted-use data used in this analysis and providing me with the methods skills needed to build such complex models. In addition, I would like to acknowledge Dr. Dana Charles McCoy for providing considerable methods and moral support and helping me organize and clarify my thoughts. Lastly, I would like to dedicate this paper in memory of my Grandma Fran, Grandpa Chuck, and Katie.

Abstract

Even though boys have consistently scored lower than girls on reading assessments, relatively little is known about the nature of this gap. The present study explored this gender gap in reading in the U.S. using the Early Childhood Longitudinal Study, Kindergarten Class of 2010-11 (ECLS-K:2011; n=7,780). Hierarchical linear modeling was performed in order to explore whether this gap is present at Kindergarten entry and how it develops throughout second grade; whether the gap holds when controlling for theoretically motivated factors that may differ by gender; and whether the magnitude of the gap differs by sociocultural factors. Girls had significantly higher reading scores than boys, on average, at Kindergarten entry, and their reading scores grew at a faster rate than boys through the end of second grade. The gender gap at Kindergarten entry was fully explained by children's age of first word, disability status, independent reading frequency, externalizing behaviors, and approaches to learning, and the gender difference in reading growth was partially explained by children's externalizing behaviors and approaches to learning. The magnitude of the gender gap was found to differ by school geographic region, but only in the model containing controls. No other sociocultural factors were found to moderate the gender gap. These findings suggest that the gender gap is already present at school entry and grows throughout second grade; it does not disproportionally affect certain groups of students; and children's externalizing behaviors and approaches to learning may be of particular interest for future applied research.

Table of Contents

Background

Even though school-age boys have consistently scored significantly lower than girls on a variety of standardized reading assessments, across decades of research, in both the U.S. (Klecker, 2006) and international contexts (e.g., Chiu & McBride-Chang, 2006; Mullis, Martin, Kennedy, & Foy, 2007; Stoet & Geary, 2015), relatively little is known about when this gap emerges, how it develops over time, who it may affect disproportionately, and what factors may explain it. Much of the research on this topic has been descriptive and has involved identifying and comparing the gender gap across various countries and years of testing, discussing the practical significance of such gaps (if any), and/or suggesting the possible causes and policy/pedagogical solutions to this problem, often based on anecdotal evidence and/or gender stereotypes (Blackburn, 2003). Unfortunately, this has led to the implementation of policies and practices that may potentially be ineffective or even harmful (Blackburn, 2003; Disenhaus, 2015). Given the dearth of basic research on this topic needed to successfully inform applied research, policy, and practice, and particularly the lack of longitudinal research on this topic in the U.S., the purpose of this study is to conduct an in-depth investigation of the development of gender differences in reading ability from Kindergarten through second grade in a nationally-representative sample of children in the U.S.

The Gender Gap in Reading

The gender gap in reading has consistently been observed on large-scale, cross-sectional national and international assessments, such as the Program for International Student Assessment (PISA) reading literacy assessment, which is taken by 15-year-olds sampled from over 60 participating countries (e.g., Stoet & Geary, 2015); the Progress in International Reading Literacy Study (PIRLS), which is taken by fourth graders sampled from over 40 participating countries (Mullis et al., 2007; Mullis, Martin, Foy, & Drucker, 2012); and the National Assessment of Education Progress (NAEP) reading assessment, which is taken by a sample of fourth, eighth, and twelfth graders in the U.S. (Klecker, 2006). For instance, Stoet & Geary found that, on average, boys scored significantly lower than girls on the PISA reading assessment in all countries studied across all four years of testing (2000, 2003, 2006, and 2009). Furthermore, they found that although the magnitude of the gender gap differed by country, on average, it was estimated to be three times as large as the gender gap in mathematics, favoring boys, which tends to get more attention in educational research, particularly in the U.S. (Disenhaus, 2015). Similarly, in their analyses of both the 2006 and 2011 PIRLS, Mullis and colleagues found that, on average, boys scored significantly lower than girls in all countries studied except Luxembourg and Spain in 2006 and Columbia, Italy, France, Spain, and Israel in 2011. This lack of a statistically significant difference in fourth grade in a few countries is not particularly surprising given that the gender gap in reading is generally found to be larger in the upper grades than in the earlier grades (Klecker, 2006). For instance, in her analysis of the gender gap in NAEP reading performance during the six testing periods from 1992-2003, Klecker found that boys scored significantly lower than girls across all grade levels and years, on average, and that the gender gap tended to be larger in the upper grades than in fourth grade. Using the NAEP data explorer, I extended this analysis and found that the same trend held during the six additional years of testing between 2005 and 2015, with the smallest estimated gender difference, out of all years of testing, of 5 points (d=0.128 *SDs*) among fourth graders in 1998, and the largest difference of 16 points (d=0.438 *SDs*) among twelfth graders in 2002. As of 2015, the gender gap among fourth graders based on the NAEP data was estimated to be about 7 points (d=0.189 *SDs*) and 10 points (d=0.250 *SDs*) among twelfth graders. These results underscore the persistence of this gender gap in the U.S. across grade levels and years of testing.

In general, the effect sizes estimated for the gender gap in reading across different measures, countries, and years of testing tend to be in the small to medium range, which has lead some scholars (e.g., Hyde & Linn, 1988; White, 2013) to conclude that the gender gap in verbal ability and/or reading is of little-to-no practical significance and therefore interventions aimed at closing the gap are unnecessary. Given that college-educated men often earn more than college-educated women upon entering the workforce even if they have the same degree (Peter & Horn 2005), it is not surprising that some researchers are unconcerned with the practical significance of the gender gap in reading. Even so, the OECD

(2010) reported that the average gender gap in reading on the 2009 PISA across all countries was 39 points, which is equivalent to an average school year of literacy learning, which is seemingly a practically significant difference, especially considering that many of the countries had even larger gender gaps. Furthermore, a closer investigation of the magnitude of these reading gender gaps in relation to assessment item type illustrates that the gap may be much wider for certain kinds of literacy tasks (e.g., evaluating and interpreting versus retrieving information), texts (e.g., literary versus informational), and modes of assessment (e.g., paper versus digital) than others (e.g., Brozo et al., 2014; Mullis et al., 2012; Roe & Taube, 2003), and these larger gender differences may be of particular interest to practitioners aimed at preparing *all* students for the high-level literacy skills necessary for success in a technologically-advanced, global economic system. Furthermore, the gender gap in reading is estimated to be as large or larger than racial/ethnic achievement gaps in reading that are of much interest to educational researchers in the U.S. (Newkirk, 2000). In addition, many policies and practices aimed at closing the gender gap are already in effect (Brozo et al., 2014), so further research is necessary to guide the responsible development of such interventions. Moreover, countries with policies successful at closing the gender gap in math, favoring boys, tend to have much larger gender gaps in reading, favoring girls (Marks, 2008), and similarly, the gender gaps in math and reading are inversely correlated both across and within nations (Stoet & Geary, 2013).

This highlights the complexity of the problem, and suggests that further in-depth research is necessary to determine how gender equity can be achieved.

The "Boy Crisis" in Education and its Effects on Policy and Practice

The presence of such a large gender gap in reading, in addition to the tendency of boys to have lower grades (e.g., Voyer & Voyer, 2014), lower college attendance rates (e.g., Conger & Long, 2013), higher rates of school dropout (Kleinfeld, 2009), and higher learning disability incidences than girls (e.g., Quinn & Wagner, 2015), has led some researchers (e.g., Brozo et al., 2014; Sommers, 2000), organizations (e.g., the OECD, Save the Children, the Independent Women's Forum), and members of the popular press (e.g., Tyre, 2006; Perlman, 2016) to conclude that there is a "boy crisis" in education in which boys are systematically disadvantaged as compared to girls in certain aspects of schooling. Many of these individuals and organizations have called for international leaders to enact policies and promote practices aimed at closing the gender gap in reading, and many countries and schools have already begun implementing them (Brozo et al., 2014). For instance, in her account of her son's difficulties with literacy, Taylor (2004) recommends that practitioners incorporate a wider variety of text types into the classroom, give boys more autonomy over text selection, choose texts that appeal to boys' interests, start a boys-only book club, invite men to read books to the class, and engage in more interactive and hands-on teaching styles. Brozo (2002) paints boys as victims of the "feminized nature of schooling" and calls for many of these same practices to be implemented in addition to suggestions such as the “Guys rack” book shelf containing books “just for boys” (p. 91) and literacy units exploring what it means to be a man.

These kinds of practices and policies have been taken up by private and governmental organizations in many countries. For instance, in Germany, the Stiftung Lesen (Reading Foundation) implemented a campaign in many workplaces across five states promoting fathers reading aloud to their children; the Baden-Württemberg Foundation developed a program called “Kicking and Reading” where professional soccer players motivate adolescent boys to read; and a reading researcher at the University of Cologne developed a website containing books, practices, and activities deemed appealing to boys (Brozo et al., 2014, p. 589). Such efforts have not been as widespread in the United States, where most policies and practices aimed at closing the gender gap have taken place at the local, rather than state or national, levels (Disenhaus, 2015), however the Task Force on Gender Equity in Education in Maine issued a report recommending that literacy instructors choose texts that are of interest to students “even at the expense of canonical literature” (Maine Department of Education, 2007, p.49); include more non-fiction, action, sci-fi, humorous, sports-related, popular, and non-traditional (e.g., graphic novels, magazines) reading materials in the classroom; develop programs aimed at boys such as a “Guys Read” program, invite male role models to read to the class, and encourage parents to read to their sons as much as they read to their daughters in order to help close the gender reading gap. This illustrates that even though this issue is not at the forefront of educational research, policy, and practice in the United States nearly to the degree that it has been in other countries (e.g., Australia, the U.K., Canada), policies and practices stemming from this line of research have made their way to the U.S.

Importantly, these kinds of policies and practices may be ineffective at closing the gender gap in reading and potentially even harmful, particularly for students who do not conform to traditional notions of gender, due to the ways in which gender is essentialized (Blackburn, 2003; Disenhaus, 2015; Rowan, Knobel, Bigum, & Lankshear, 2002; Watson, Kehler, & Martino, 2010). For instance, Rowan and her colleagues (2002) reported that in one intervention where boys were placed in single-sex classrooms and taught only by male teachers using traditionally masculine texts, many of the boys reading scores increased and behavior problems diminished, however many others, many of whom were not stereotypically masculine, actually performed worse. This exemplifies the perspective held by the aforementioned scholars that policies and practices aimed at closing the gender gap in reading that reify traditional, hegemonic notions of masculinity actually further gender inequities in the classroom. For example, by implementing a “Guys Rack” in the classroom, boys may feel pressured to avoid reading books on the other bookshelves and girls may feel unwelcome to read “boys books”, thereby limiting students’ access to a diverse set of texts. Furthermore, McGeown, Goodwin, Henderson, and Wright (2012) found that students' gender roles[1] were more strongly associated with students' intrinsic reading motivation than their gender, such that children (both boys and girls) who had more feminine traits tended to be more intrinsically motivated to read than those who held more masculine traits. This illustrates, first of all, that a child's gender does not determine his or her traits or preferences, and therefore, treating all boys the same and all girls the same will not likely be effective. Secondly, it suggests that such gender roles may be more predictive of reading than gender, so boys who tend to be more feminine may not benefit (or worse, may be harmed) from such an intervention, while girls who tend to be more masculine may not receive needed support and may even be further discouraged when traditional gender roles are emphasized in the classroom.

Countries such as Ireland and Australia have taken such points into consideration in their attempts to close the gender gap in reading. For instance, in 2011, the Irish Department of Education and Skills implemented a plan aimed at improving literacy that promoted the use of a wide range of texts of interest to students, similar to other policies, however they also called for the implementation of critical literacy approach focused on the ways in which texts and reading lessons reify hegemonic gender structures and how to respond to or challenge these notions (Brozo et al., 2014). Similarly, in Australia, the government-funded "Success for Boys" program was implemented in 2006, which, despite its title,

[1] The authors use the term "gender identity" to describe this instead of "gender role", however since "gender identity" is often used interchangeably with gender and the Children's Sex Role Inventory (CSRI) Short Form (Boldizar, 1991) was used to measure it, I've used the term "gender role" to more clearly distinguish it from what I refer to as "gender" here (i.e., the child's biological sex).

focuses on determining "'which boys?' and 'which girls?' are at risk of failing in school" (Alloway, 2007, p. 591) by focusing on the interplay of gender, race/ethnicity, SES, and other social factors in order to "de-essentialize" boys as a group and provide resources to those students who need them the most. Even so, such approaches fail to take into account the utility of the basic research, particularly large-scale, quantitative, longitudinal research, to inform key aspects of the intervention, such as when to implement the intervention, what the potential mechanisms are behind these reading difficulties that need to be addressed, and how to identify the groups of students most at-risk for reading failure based on the interplay of their gender, race/ethnicity, SES, and other sociocultural factors[2].

The Unanswered Questions

Prior studies have attempted to address the aforementioned factors needed to effectively inform policies and practices aimed at closing the gender gap in reading, yet there are still many questions left unanswered regarding when to intervene, how, and with whom.

When to Intervene: Exploring the Development of the Gap

Many of the interventions that have been developed to ameliorate boys' reading difficulties have targeted adolescents, likely because reading gender gaps

[2] Of course, there is always some level of essentialization that goes on when trying to use quantitative research to predict students' outcomes. For instance, even in studying the intersection of race/ethnicity, SES, and gender by predicting the average reading development of a white, upper-middle class girl the prediction is still based on an average, and there are likely to be many students meeting those criteria who have vastly different reading trajectories due to unobserved factors (e.g., the child has a reading disability) or individual differences. Even so, in performing these kinds of analyses, we can make better predictions for individual student outcomes than if we were not to use them at all, and therefore they can be very helpful in anticipating which students may need more support.

in early childhood were not often detected in the past due to a dearth of large-scale assessment data in the early school years (Ready, LoGerfo, Burkham, & Lee, 2005). Nevertheless, multiple analyses of the Early Childhood Longitudinal Study-Kindergarten Class of 1998-99 (ECLS-K), the first comprehensive, nationally-representative study of the factors that influence American schoolchildren's development, have illustrated that the gender gap in reading is present in the fall of Kindergarten (Aikens & Barbarin, 2008; Husain & Millimet, 2009; McCoach, O'Connell, Reis, & Levitt, 2006; Ready et al., 2005; Robinson & Lubienski, 2011). Even so, all of these studies were conducted with the same sample, and therefore additional research is needed to replicate this finding. Furthermore, analyses of how the gender gap changes over the course of schooling using the ECLS-K have yielded mixed results. For instance, while McCoach and her colleagues found that girls had significantly faster rates of reading growth than boys in Kindergarten but not first grade, Aikens and Barbarin found that girls had significantly faster growth rates in Kindergarten and first grade, but not through third grade, and Rathbun and colleagues (2004) found that girls experienced significantly more reading growth from Kindergarten to third grade[3]. These at least suggest that girls growth rates are faster than boys, however some researchers have found the opposite pattern, for instance, Husain and Millimet (2009) found that boys gained ground in reading relative to girls during this time period, except

[3] Although the authors concluded that this difference was not "substantively meaningful" because it was smaller than one quarter standard deviation.

for the lowest achieving boys who lost ground, and Robinson and Lubienski reported that the overall gender gap narrowed between Kindergarten and fifth grade and widened again by eighth grade.

Given that these analyses were all conducted using the same data set, the differences in results likely arose from the use of different methodological approaches to answer the question. For instance, McCoach and her colleagues (2006) and Aikens and Barbarin (2008) employed piecewise hierarchical linear models (HLM) with different linear splines used; Rathbun and colleagues (2004) employed a gain score regression analysis; Husain and Millimet (2009) utilized regression analyses as well as quantile treatment effects, and Robinson and Lubienski (2011) performed quantile regression analyses as well as a metric-free methodology that they developed. One issue with all of these approaches except for HLM is that they fail to take into account that, unlike standardized tests which all tend to be administered at the same time of the school year, the students within each wave of the ECLS-K were all assessed at different points throughout the semester and the spacing of the students' consecutive assessments were not equidistant. The ECLS-K researchers state that "such differences in ages, assessment dates and intervals, and children's grade levels may affect the results," although they note that "these differences may have a relatively small impact on analysis results for long time intervals" (Tourangaeu et al., 2009, p. 3-24). Another factor that may have led to the discrepancy in results is the differences in the analytic samples since students who were excluded may systematically differ from those included in the analysis. Although it is unclear what specific methodological factors led researchers to different conclusions regarding the development of gender differences over time despite using the same data set, additional research is needed that takes a sensible approach to longitudinal analysis to address the discrepancy in findings. Even so, taken together, these results suggest that interventions aimed at closing the gap should likely be directed at children in the first few years of schooling rather than adolescents since the gap is seemingly present even in the fall of Kindergarten, yet this decision should likely also be informed by an understanding of how this gap develops over time, which necessitates further research.

How to Intervene: Identifying Potential Causes of the Gap

Of course, it is not possible to identify the cause(s) of the gender gap in reading through observational research, but such research can provide insight into the kinds of factors associated with gender differences in reading achievement beyond theoretical arguments and qualitative observations. While much of the research, policy, and practices on gender gaps that adversely affect girls (e.g., STEM gap) is focused on external causes, such as stereotype threat, that can be intervened upon, much of the work on the gender gap in reading has focused on biological causes perceived to be immutable, thereby leading to the recommendation that practitioners adjust instruction to accommodate boys' natural inclinations, preferences, and learning approaches (Williams, 2013). For example, Gurian (2010) claims that based on gender differences in brain development, boys tend to be more impulsive, struggle to multitask, develop language more slowly, and have poorer memories than girls. While there is evidence supporting the existence of gender differences, on average, in overall brain volume and grey matter volume of certain brain regions (Ruigrok et al., 2014), and there is also evidence for gender differences in structural connectivity, even among prepubescent children (Ingalhalikar et al., 2014), it is not yet clear how these structural differences relate to functional and/or behavioral gender differences, so such claims, although not completely unfounded, are at least premature. Furthermore, even if one accepts the notion that solely biological factors (through prenatal hormone exposure, chromosomal differences, etc.) are responsible for the gender gap in reading, an opinion that's not widely held by researchers, it is unclear why this would necessitate changing the literacy environment to accommodate boys (e.g., through the "boys-only" approaches such as those advocated by Taylor, 2004 and Brozo, 2002) rather than encouraging them to change their behaviors, practices, beliefs, etc. to better match the classroom environment as is expected of girls in the math/science context (Williams, 2013).

Researchers such as Williams (2013) and Orr (2011) suggest that gender socialization and toxic masculinity, rather than biological factors, are likely responsible for gender differences in reading performance. For example, in her analysis of the ECLS-K, Orr found that children (both boys and girls) who reported engaging in more traditionally feminine activities had higher grades in Kindergarten. Even so, given the complexities of biological sex, which is likely a spectrum rather than a dichotomy (Ainsworth, 2015), it is unclear whether the children reported engaging in such activities due solely to differences in their gender socialization rather than biological differences (e.g., boys who reported engaging in more traditionally female activities may fall closer to the center of the spectrum of biological sex than those who reported engaging in more traditionally male activities, on average). There does seem to be some evidence supporting the notion that gender socialization and stereotypes play a role, however, considering that Hartley and Sutton (2013) found that young children believed that girls were better students than boys and that boys performed worse on a reading, writing, and math assessment when they were told ahead of time that boys tended to be worse students than girls and better on these assessments when they were told that boys and girls were equally good students. This suggests that environmental factors, such as gender stereotypes and socialization may play a role in the development of the gender gap, and should be studied further. Even so, given the many other potential sources of the gap, the consensus among most researchers is that the gap is likely caused by the interplay of a variety of both biological and environmental factors (e.g., Halpern, 2012; Stoet & Geary, 2013).

The factor that has gotten the most attention in the literature as a potential cause of the reading gender gap is reading motivation (e.g., Chiu & McBride-Chang, 2006; OECD, 2010). Given that many studies have replicated the finding that girls tend to report significantly higher levels of intrinsic reading motivation than boys (e.g., Marinak & Gambrell, 2010; Van Der Bolt & Tellegen, 1996;

Wigfield & Guthrie, 1997) and that reading motivation is correlated with reading performance (e.g., Guthrie, Wigfield, Metsala, & Cox, 1999), researchers have explored whether reading motivation mediates the relationship between gender and reading performance. For instance, in their HLM analysis of the predictors of PISA reading achievement for fifteen-year-old students within 43 countries in 2000, Chiu and McBride-Chang found that students' self-reported reading enjoyment mediated 42% of the gender gap in reading performance. They also tested whether SES, the approximate number of books in the home, and/or unusually low reading achievement mediated the gender gap, but they found no evidence of mediation for those variables. Importantly, they found that these results did not differ by country, which suggests that similar mechanisms underpin the gender gap, at least within the countries studied.

Similarly, using a more recent version of the PISA data, the OECD (2010) reported that students' reported reading enjoyment partially mediated the gender gap, and they also found that students' approaches to learning, as measured by their reported reading and summarizing strategies, partially mediated the gap, such that, on average, across the 34 OECD countries, almost 70% of the gender gap in reading performance was explained by both reading enjoyment and approaches to learning. Even so, the extent to which these factors mediated the gender gap differed by country, and in the United States, these factors were found to fully mediate the gender gap in reading performance such that there was no longer a statistically significant difference in performance between girls and boys, on average, when controlling for their reported reading engagement and approaches to learning. Ready and his colleagues (2005) reached similar conclusions about the role of approaches to learning in their analysis of whether children's classroom behavior mediates the gender gap in literacy learning during Kindergarten using the ECLS-K. They found that children's learning approaches, defined as their "attentiveness, task persistence, eagerness to learn, learning independence, flexibility, and organization" (p. 33), measured by teacher report, mediated over 70% of the gender gap in reading growth through Kindergarten. Additionally, they found no evidence that sociocultural factors mediated this gender gap, and although they found that the gap was partially mediated by self control, interpersonal skills, externalizing behavior (defined as "the extent to which [the] child argues, fights, gets angry, acts impulsively, and disturbs ongoing activities" p. 33), and internalizing behavior (defined as "the extent to which [the] child appears anxious, lonely, sad, or [has] low self-esteem" p. 33), these were not nearly as strong at mediating the gap as learning approaches. They did not include a measure of reading enjoyment in their analysis, so it is unclear how the results may have differed upon its inclusion. Taken together, these findings suggest that students' reading motivation and learning strategies at least partially, if not fully, mediate the gender gap in the U.S., and therefore applied research (e.g., implementing a randomized controlled trial) further exploring these relationships in order to determine whether they are causal rather than merely correlational, may be prudent. Nevertheless, additional factors, have been hypothesized to give rise to the gender gap in reading, such as gender differences in children's language development (see Halpern, 2012) and gender differences in the prevalence of learning disabilities (e.g., Quinn and Wagner, 2013), and therefore a more thorough investigation is called for that takes into account all of the key factors predictive of reading performance that are hypothesized to differ by gender in order to inform the development of effective interventions.

With Whom to Intervene: Interplay Among Gender and Other Factors

Although a surfeit of previous research has indicated the predictive relationship between sociocultural factors, such as race/ethnicity, SES, and school sector, and reading development (e.g., Aikens & Barbarin, 2008; Fryer & Levitt, 2006), it is unlikely that there are gender differences in such factors, and accordingly, prior research suggests that these factors do not mediate the gender gap in reading performance (Husain & Millimet, 2009; Ready et al., 2005). Even so, few studies have explored whether the gender gap is moderated by sociocultural factors. In their analysis of gender differences in American children's reading and math growth from Kindergarten through third grade using the ECLS-K, Husain and Millimet found evidence of the reading gender gap among all race/ethnicity groups, but they found that the magnitude of the gap at Kindergarten entry and throughout third grade differed by race/ethnicity such that the gaps for Black and Hispanic students widened significantly more than the gaps for white and Asian students over the course of schooling, and this effect was largest among Hispanic students. In addition, they found that the magnitude of the gap varied as a function of SES and school sector, such that the gender gap widened significantly more for lower SES students than for higher SES students as well as for public school as compared to private school students from Kindergarten through third grade. They also tested to see if the gap differed by school geographic region or location type and found no evidence for this in the full sample. These results suggest that the magnitude of gender gap differs by certain sociocultural factors such as race/ethnicity, SES, and school institutional status, but may not differ by other factors such as school region or location type. This suggests that interventions aimed at closing the gender gap in reading should be particularly targeted to Black and Hispanic children, those from lower SES backgrounds, and/or those who attend public schools, as those students are most affected, however additional research is needed to corroborate these findings.

The Present Study

In the present study, I seek to replicate as well as extend previous research on the gender gap in the context of American children's early reading development in order to inform research, policy, and practice by using the second, and most recent, cohort of the nationally-representative ECLS-K. In the current study, I aim to address the following research questions:

1) Are there gender differences in American children's reading ability at Kindergarten entry and/or in their reading growth rates through the end of second grade, on average?
2) If so, do these gender differences in children's reading ability at Kindergarten

and/or reading growth rates through second grade hold when controlling for other child-level factors that may differ by gender (e.g., disability status)?

3) Furthermore, does the size of these gender gaps in reading ability at Kindergarten and/or reading growth rates through second grade differ by sociocultural factors (e.g., race/ethnicity, SES)?

Hypotheses

Consistent with the findings of the various analyses conducted with the ECLS-K (Ready et al., 2005; McCoach et al., 2006; Husain & Millimet, 2009; Robinson & Lubienski, 2011), I anticipate that I will find evidence of gender differences in children's reading scores at Kindergarten entry, such that girls will have higher scores than boys, on average, even when controlling for demographic variables and other key predictors of reading ability. Additionally, based on prior research that utilizes similar analytic approaches (e.g., McCoach et al., 2006; Aikens & Barbarin, 2008), I anticipate that girls will exhibit a faster reading growth than boys, on average, between Kindergarten and second grade, and this will also hold when controlling for demographic variables and other key reading predictors. Furthermore, in line with the findings of Ready and his colleagues (2005) and the OECD (2010), I anticipate that such gender gaps (at Kindergarten entry and/or in reading growth) may be fully attenuated and no longer statistically significant when controlling for children's reading engagement (operationalized as child independent reading frequency) and learning approaches, and I expect that such gaps may be attenuated but retain statistical significance when controlling for externalizing behaviors. I have also included additional factors as covariates that have been illustrated to differ by gender and predict reading performance in previous research but have not been tested as possible mediators of the gender gap (e.g., age of first word), and I anticipate that some of these may also partially explain such gender gaps. In addition, I expect the magnitude of these gaps to differ by race/ethnicity, SES, and school sector but not by school region or location type consistent with the findings of Husain and Millimet (2009). For specific detail on hypotheses for each covariate and potential moderator, see Tables 1 and 2 in Appendix A.

Methods

ECLS-K:2011 Data

To address these questions, I performed a secondary analysis of the Early Childhood Longitudinal Study, Kindergarten Class of 2010-11 (ECLS-K:2011), which is a large, complex survey data set sponsored by the National Center for Education Statistics (NCES) within the Institute of Education Sciences (IES) of the U.S. Department of Education. The ECLS-K:2011 provides comprehensive information on U.S. children's cognitive, behavioral, social, emotional, and physical development as well as their home, classroom, and school environments from Kindergarten through fifth grade on a nationally-representative sample selected from both public and private schools, from a diverse racial/ethnic and socioeconomic backgrounds, and including both first time Kindergarteners and children who repeated Kindergarten. Given my focus on early reading development, I analyzed the Restricted-Use Kindergarten-Second Grade Data (ECLS-K:2011; Tourangeau et al., 2015), which contained six rounds of data collection (both fall and spring of Kindergarten, first, and second grades). During each of these waves, students were directly administered cognitive assessments during a one-on-one session with a tester. Children within each wave of data collection were assessed at different times during the semester (e.g., students in the spring of Kindergarten were assessed in March, April, May, or June), and data were only collected with a subsample of students in the fall of first and second grade in order to enable analyses of school versus summer learning. Parents completed phone interviews and teachers, school administrators, and before- and after-school childcare providers completed self-administered questionnaires during various waves of data collection in order to provide additional longitudinal information about the children's environment as well as factors about the children that were not directly assessed (e.g., children's classroom behavior).

Analytic Sample

Analysis of the ECLS-K:2011 requires the use of sampling weights to produce national-level estimates in order to account for both oversampling of certain groups of children (e.g., Asians, Native Hawaiians, and other Pacific Islanders) as well as non-response bias. For this reason, although the original ECLS-K:2011 sample consisted of 18,170[4] children from 970 schools across the

[4] All sample sizes in this manuscript have been rounded to the nearest ten in compliance with the ECLS-K:2011 restricted-use data policies.

U.S., my analytic sample was limited to 7,820 children from 820 schools as these were the children who had non-zero values on the child-level weight that best matched the data sources included in my analysis, W6C6P_6T0. The W6C6P_6T0 sampling weight was the most appropriate because it adjusted for nonresponse associated with child assessment, parent interview, and teacher questionnaire data from waves 1, 2, 4, and 6 (corresponding to fall and spring of Kindergarten and spring of first and second grade), and these represented the complete set of sources of the variables included in my analysis. I only analyzed data from the first, second, fourth, and sixth rounds of data collection, as advised by the ECLS-K:2011 researchers, in order to avoid biased reading performance which may be impacted by the unstructured summer break and to avoid reducing the sample size further by including those two fall time periods that only provided data on a subset of students. Lastly, even though my full analytic sample included 7,820 children, the number of subjects included in my final model, 5,590 children, was smaller due to item-level missingness, which is not adjusted for through the use of sampling weights. Even so, the ECLS-K:2011 researchers state that "analysis of potential bias due to item nonresponse is typically conducted for those items with a response rate less than 85%" (Mulligan, McCarroll, Flanagan, & Potter, 2015, p. 11), and given that all of the variables included in this analysis have a response rate above 85% except *A1STWRD*, which had a response rate of approximately 83%, the performance of listwise deletion due to item-level missingness was not believed to have biased the results.

Measures

Reading Ability. The outcome of interest was students' reading ability (*READ*) as measured by a one-on-one, two-stage cognitive assessment that consisted of questions on basic reading skills (print familiarity, letter recognition, beginning and ending sounds, rhyming words, word recognition), vocabulary, and reading comprehension (recalling facts, making inferences, and making judgments; Tourangeau et al., 2015). The first stage of the assessment was a routing section that contained common items and determined whether students would complete a low, middle, or high difficulty second stage assessment. During the Kindergarten assessments all students first completed a language screener (the Simon Says and Art Show tasks from the *pre*LAS 2000) and then completed the English basic reading skills (EBRS) section, which served as the first part of the reading routing section, prior to completing the reading assessment. Children who did not pass the language screener were administered the Spanish early reading skills (SERS assessment) if their home language was Spanish; otherwise they were not administered the reading assessment. During first grade, the procedure was similar except that students only completed the language screener if they had failed the screener in a previous round, and all children routed to the English version of the reading assessment completed a 30-item routing section, which directed students in the low- and middle-difficulty assessment groups to 18 EBRS items prior to the second-stage test while students directed to the high-difficulty group skipped the EBRS items. During second grade, a language screener was no longer used because 99.9% of children were routed through the assessment in English at this point. All children completed a 29-item routing section, which directed them to a low, middle, or high-difficulty second stage assessment. I used the item response theory-generated reading theta scores computed by the ECLS-K:2011 researchers because they allowed for comparisons across all waves of data collection despite differences in the items completed on the assessment.

Gender. The primary question predictor was student gender[5] (coded as *GIRL*=1 for girls), which was collected from the students' schools and checked against multiple parent interviews in order to ensure proper identification.

Time. Time was represented as the approximate number of months since Kindergarten entry (*MONTH*) that each assessment took place. To calculate this variable, I subtracted the first day of Kindergarten for each child (based on his or her approximate school start date during the Kindergarten year) from his or her approximate testing dates (and divided by 30.42 to convert to months) such that time was centered on each child's first month of Kindergarten. I then rounded these values to the nearest whole number in order to improve model fit and so that the intercept would not be an extrapolation. Table 3 in Appendix B presents the distribution of the *MONTH* variable, which indicates the months since

[5] The interviewer technically asked the parent for confirmation as to whether the child was male/female, and this variable was labeled as X_CHSEX_R, but I will refer to this variable as gender and use the terms girl/boy rather than female/male in order to use the terminology that is most consistent with the recent research on this topic. I am, of course, making the assumption here that the child's sex reported by the parent aligns with their presumed gender, which may not always be the case. Parents also had the option to refuse this response or say that they didn't know, and there were some cases in the full data set where such responses were reported, however, all of the students included in the analytic sample were identified as either "male" or "female".

Kindergarten entry that each reading assessment took place, as well as the wave of assessment with which each month corresponds. This coding of the time variable is fairly innovative as compared to prior studies because it takes into account the variability in testing dates within each wave and how this corresponds with the children's exposure to school. Many other studies that examine reading growth using the ECLS-K or ECLS-K:2011 treat each wave as if all of the students were tested at roughly the same time during the school year, but as is evident in Table 3, some of the children tested in wave 1, for instance, were tested during their first month of schooling, while others were tested during their fourth month of schooling, and some in wave 2 were tested in their sixth month of schooling. It is possible that the testing schedules were fairly evenly distributed by gender, race/ethnicity, etc., such that collapsing these testing dates by wave and studying growth rates that way would not have much of an effect on the results, however this operationalization of time allows us to directly answer our research questions as to whether there are gender differences, on average, at Kindergarten entry (during the first month of Kindergarten as opposed to during the fall of Kindergarten) and whether students' reading growth rates from the beginning of Kindergarten through second grade differ by gender, on average.

Covariates. I examined whether these gender gaps held when controlling for sociocultural factors (see potential moderators below) as well as a number of theoretically-motivated, child-level covariates measured in the fall of

Kindergarten[6] through parent interview or teacher questionnaire: *Language status, Kindergarten repeater status, Age of Kindergarten Entry, Number of Children's Books in the Home, Frequency of Child Independent Reading, Frequency of Parent Reading to Child, Age of First Word, Educational Expectations, Disability Status*[7]*, Approaches to Learning,* and *Externalizing Behaviors*. See Table 1 in Appendix A for a detailed description of each of these covariates and the rationale for their inclusion in the analysis.

Potential Moderators. I tested to see whether the magnitude of the gender gaps differs by various child- and school-level sociocultural factors measured in the fall of Kindergarten through parent interview, the Field Management System (FMS)[8], or the 2009-10 Common Core of Data (CCD) for public school children and 2009-10 Private School Survey (PSS) for private school children: *Race/ethnicity, Socioeconomic Status, School Sector, School Location Type*, and *School Region*. See Table 2 in Appendix A for a detailed description of each of these potential moderators and the rationale for their inclusion in the analysis.

Statistical Analyses

I analyzed these data by fitting a taxonomy of three-level HLMs with time of reading assessment (Level-1) nested within student (Level-2) nested within

---

[6] All of the variables were measured in the fall of Kindergarten except for *Age of First Word* and *Disability status*, which were measured during the spring of Kindergarten.
[7] Since specific learning disabilities in reading often go undiagnosed until second grade or later (Catts, 2017), I decided to compare students with or without a disability (broadly-defined) identified in Kindergarten.
[8] "The Field Management System includes information collected about the data collection effort, the study schools, school staff, and children from available administrative records or existing data sources (such as the Common Core of Data) or from conversations between data collection staff and school staff" (Tourangeau et al., 2015, p. 7-1).

school (Level-3) using the mixed command in Stata 14 (StataCorp, 2015) after first running the appropriate univariate and bivariate descriptive and inferential statistics. Although approximately 15% of students in the analytic sample (n=1,170) changed schools between the fall of Kindergarten and the spring of second grade, the analyses were conducted as if all students had remained in the same schools as they were enrolled in during the spring of Kindergarten in order to avoid fitting needlessly complex models (e.g., crossed random effects)[9]. In order to determine the appropriate shape of the growth trajectory, I inspected the raw growth curves and noted that they appeared to be either linear or slightly curvilinear[10]. As individual growth curves cannot be displayed due to the use of restricted-use data, Figure 1 in Appendix C presents aggregate observed growth curves by gender for students in a random subset of schools (n<10) in order to illustrate the general shape of these trajectories[11]. I then conducted preliminary analyses comparing an unweighted baseline HLM predicting reading ability with a linear functional form to one with a quadratic functional form and found the latter to be a significantly better fit to the data using a likelihood ratio test. For this reason, I included the quadratic term ($MONTHS^2$) in my model, although I did not include its random effect because there was inadequate information to reliably

---

[9] The ECLS-K:2011 researchers were consulted on this and responded that there was no need to exclude students who changed schools or take a different approach to modeling these data so long as the primary research question was "driven by non-school factors," as it was in this case.

[10] The slopes of the growth curves did not appear to be perfectly constant throughout the time period, likely due to differences in summer reading growth rates (see Alexander, Entwisle, & Olson, 2001), but since I was most interested in overall reading growth/*average* reading growth per month, I did not explicitly model differences in summer reading growth rates.

[11] Growth curves were collapsed across at least three students and data from at least three schools is presented in this figure in compliance with ECLS-K:2011 restricted-use data policies.

estimate the random effects for both growth terms simultaneously[12]. In these preliminary analyses, I also used likelihood ratio tests to determine the necessity of including the third level (school) in the model, as well as including a random effect for the linear term at the school level. The model used to address all three research questions was:

Level-1: At occation i for individual j in school k:

$$READ_{ijk} = \pi_{0jk} + \pi_{1jk}MONTH_{ijk} + \pi_{2jk}MONTH_{ijk}^2 + e_{ijk},\ e_{ijk} \sim N(0, \sigma^2)$$

Level-2: $\pi_{0jk} = \beta_{00k} + \beta_{01k}GIRL_{jk} + \beta_{02}M_{jk} + \beta_{03}GIRLxM_{jk} + \beta_{04}C_{jk} + r_{0jk}$

$$\pi_{1jk} = \beta_{10k} + \beta_{11k}GIRL_{jk} + \beta_{12}M_{jk} + \beta_{13}GIRLxM_{jk} + \beta_{14}C_{jk} + r_{1jk}$$

$$\pi_{2jk} = \beta_{20k} + \beta_{21k}GIRL_{jk} + \beta_{22}M_{jk} + \beta_{23}GIRLxM_{jk} + \beta_{24}C_{jk}$$

$$\left(r_{0jk}, r_{1jk}\right)' \sim N(0, T_\pi) \qquad T_\pi = \begin{bmatrix} \tau_{\pi 00} & \tau_{\pi 01} \\ \tau_{\pi 10} & \tau_{\pi 11} \end{bmatrix}, \tau_{\pi 10} = \tau_{\pi 01}$$

Level-3: $\beta_{00k} = \gamma_{000} + \gamma_{001}S_k + u_{00k}$

$$\beta_{01k} = \gamma_{010} + \gamma_{011}S_k$$

$$\beta_{10k} = \gamma_{100} + \gamma_{101}S_k + u_{10k}$$

$$\beta_{11k} = \gamma_{110} + \gamma_{111}S_k$$

$$\beta_{20k} = \gamma_{200} + \gamma_{201}S_k$$

$$\beta_{21k} = \gamma_{210} + \gamma_{211}S_k$$

$$(u_{00k}, u_{10k})' \sim N\left(0, T_\beta\right) \qquad T_\beta = \begin{bmatrix} \tau_{\beta 00} & \tau_{\beta 01} \\ \tau_{\beta 10} & \tau_{\beta 11} \end{bmatrix}, \tau_{\beta 10} = \tau_{\beta 01}$$

---

[12] Rabe-Hesketh and Skrondal (2008) note that "it is perfectly reasonable to allow only the lower-order terms of the polynomial used in the fixed part of the model to vary randomly between subjects" (p. 349).

At Level-1, π_{0jk} represents the average reading ability of U.S. children at Kindergarten entry, π_{1jk} represents the average linear growth rate per month, π_{2jk} represents the average quadratic growth rate per month, and e_{ijk} represents the occasion-level residual which is normally distributed with a mean of 0 and a variance of σ^2. At Level-2, the Level-1 intercept, π_{0jk}, linear growth term, π_{1jk}, and quadratic growth term, π_{2jk}, are each composed of an intercept (β_{00k}, β_{10k}, and β_{20k}, respectively), a slope associated with gender (β_{01}, β_{11}, β_{21}), a slope associated with each student-level moderating variable (*M*) included in the model (β_{02}, β_{12}, β_{22})[13], a slope associated with the interaction between gender and each of the moderating variables included in the model (β_{03}, β_{13}, β_{23}), and a slope associated with each student-level covariate (*C*) included in the analysis (β_{04}, β_{14}, β_{24}). At Level-2, the Level-1 intercept and linear growth term also include a student-level residual (r_{0jk} and r_{1jk}), which is drawn from a bivariate normal distribution with a mean vector of 0 and an unstructured covariance matrix, T_π. Lastly, at Level-3, each of the intercepts and slopes associated with gender in the Level-2 equations (β_{00k}, β_{10k}, β_{01k}, β_{11k}, β_{02k}, β_{21k}) are composed of an overall intercept (γ_{000}, γ_{100}, γ_{100}, γ_{110}, γ_{200}, γ_{210}) and a slope associated with each school-level moderating variable (*S*) included in the model (γ_{001}, γ_{011}, γ_{101}, γ_{111}, γ_{201}, γ_{211}). There is also a school-level residual for the

[13] Each additional student-level moderator included in the analysis would get its own coefficient, but this has been simplified here for illustrative purposes. This is also the case for the student-level covariates and school-level moderator variables and their interaction terms.

overall reading score as well as the linear growth term (u_{00k} and u_{10k}, respectively), which is drawn from a bivariate normal distribution with a mean vector of 0 and an unstructured covariance matrix, T_β.

To address my first research question, whether there are gender differences in reading ability at Kindergarten entry and/or in reading growth, I fit the model presented above without any covariates or potential moderators and evaluated the statistical significance and magnitude of the coefficients associated with gender: β_{01}, β_{11}, and β_{21}. To address my second research question, I first added in each of the covariates one at a time to investigate whether the magnitude and/or statistical significance of the coefficients associated with student gender differed when controlling for any of these other factors suggesting partial or full mediation. Next, I began building a model containing multiple controls by adding in all of the potential moderator variables as well as the covariates that were not found to explain the gender effects. If a covariate or moderator was no longer statistically significant upon the addition of another variable, I dropped it from the subsequent model. Then, in order to isolate the effects of each remaining covariate (previously found to either partially or fully explain the relationships of interest), I added them in one at a time until arriving at a model containing all of the statistically significant covariates, potential moderators, and their associated growth terms. To address my third research question, whether such gender gaps differ by sociocultural factors, I added the student- and school-level moderator variables, as well as their interactions with gender one at a time, to the appropriate levels of the model containing no covariates and evaluated whether or not they were statistically significant. I then tested for these interactions again in the model containing all statistically significant covariates, moderators, and their associated growth terms (the last model built in response to question 2) and retained any significant interactions to arrive at a final model.

Throughout the model building process, I considered tests to be statistically significant if they yielded a *p*-value less than $\alpha = 0.05$ or, equivalently, if the associated 95% confidence interval did not include the null value (in cases where the *p*-value was not provided). In order to determine if polytomous variables were statistically significant, I performed post-hoc Wald tests to see if the combination of dummy variables significantly contributed to the model. I did not correct for multiple comparisons because Gelman, Hill, and Yajima (2012) report that it is unnecessary in the case of HLM. All analyses (except sensitivity analyses) were conducted using the W6C6P_6T0 sampling weight; and therefore, the pseudo-maximum likelihood estimation method was used. Selya and her colleagues (2012) recommend calculating local effect sizes using Cohen's f^2 for multilevel models when both the independent and dependent variables of interest are continuous, yet since my primary research predictor, gender, was dichotomous, I calculated effect sizes (e.g., Kindergarten entry) using Cohen's *d* in a similar manner to Friedmann and his colleagues (2008)[14]. I performed diagnostics periodically in order to ensure that the HLM assumptions had been met, and I performed sensitivity analyses to determine whether the results differed when using unweighted data and/or using the data from all six time points (including the fall of first and second grade).

Results

Univariate and Bivariate Analyses

Table 4 in Appendix D presents a summary of the univariate descriptive statistics calculated for all of the variables included in the analysis. Both unweighted and weighted estimates are presented to illustrate the similarities and differences between them, although I will only report the weighted values in-text. The approximate response rates are also presented in order to report on the missingness for each item. For the majority of measures, the unweighted and weighted estimates were nearly identical, although there were slight differences between the unweighted and weighted estimates for *Race/ethnicity, School Location Type, School Sector,* and the *Number of Children's Books in the Home*.

As illustrated in the table, the population of U.S. Kindergarten students in 2010 was roughly equally distributed by gender, with approximately 49% girls and 51% boys. Additionally, the majority of students were White (roughly 52%), 25% were Hispanic, 13% were Black, 5% were Asian, and 6% were of another race or

---

[14] The numerator will be the HLM-estimated mean gender difference in reading ability at the time point of interest (e.g., 21 months since Kindergarten entry) and the denominator will be the pooled standard deviation at that wave of data collection (e.g., from wave 4 since it corresponds with 21 months).

more than one race. The majority of students attended public (89%) rather than private (11%) schools, and the regions and Location types of these schools were roughly evenly distributed across all categories except that students more frequently attended schools in the South (38%) and less frequently attended schools in towns (11%), respectively. Furthermore, the majority of children were reported to engage in independent reading during Kindergarten either 3 to 6 times a week (36%) or every day (36%) during a typical week, with 23% reported to read once or twice a week and 5% reported never to do so, and the majority of parents reported that a family member read to their child every day (51%) during a typical week in the Kindergarten year, with approximately 34% reported to have read 3 to 6 times per week, 14% reported to have read once or twice a week, and 1% reported to never have read to their child. In addition, a majority of children were reported to have spoken their first word at 10 to 12 months of age (42%), with 21% reported to have done so at 0 to 9 months of age (21%), 14% reported to have done so at 13 to 15 months, 10% at 16 to 18 months, 7% at 19 to 24 months, and 6% after 24 months or not at all. Twenty percent of the Kindergarten children were reported to have a disability, 16% reported a non-English primary home language, and 5% were reported to have repeated Kindergarten. A majority of parents expected their children to earn a Bachelor's degree (48%), with 17% expecting their children to earn a M.D., Ph.D., or other advanced degree, 16% expecting them to earn a Master's degree or equivalent, 15% expecting them to receive some postsecondary education, and 5% expecting them to earn a high school diploma or less. The mean age of Kindergarten entry was 66 months (SD=4.50 months), the average SES was 0.06 units (SD=0.82), and children were reported to have 93 children's books in their homes, on average (SD=143 books). On average, teachers reported that children had high levels of approaches to learning (M=3.00, SD=0.67) and lower levels of externalizing problem behavior (M=1.58, SD=0.61) during the fall of Kindergarten.

Table 5 in Appendix D presents the weighted means and tabulations for each variable by gender and indicates whether they differ significantly from each other. As indicated in the table, I found no evidence of gender differences in the children's *SES*, *Number of Children's Books in the Home*, *School Sector*, or *Language Status*. I did find that girls had higher reading scores than boys, on average, at each of the waves of data collection ($p<0.01$ for all comparisons). Furthermore, I found that, among students in the weighted sample, there were more boys than girls in the bottom 10th percentile of reading ability at each wave. For instance, during the first wave of data collection (fall of Kindergarten), approximately 450 boys scored in the bottom 10th percentile as compared to 330 girls, and during the last wave of data collection (spring of second grade), about 480 boys scored in the bottom 10th percentile as compared to 300 girls. Conversely, I found that there were more girls than boys in top tenth percentile at all waves except for wave 1. For instance, during the first wave of data collection, approximately 370 girls scored in the top 10th percentile as compared to 400 boys, but during the second wave, about 400 girls scored in the top 10th percentile as compared to 380 boys and in last wave, about 430 girls scored in the top 10th percentile as compared to 350 boys. Taken together, these descriptive findings suggest that young boys in the U.S. do not only perform lower on reading ability, on average, but they also are more likely to struggle with reading and less likely to be high achieving readers than girls.

In addition to scoring higher on reading ability, girls also had higher reported levels of positive learning approaches in Kindergarten (*diff*=0.28 points, $p<0.001$) than boys and had a higher frequency of engaging in independent reading daily (*diff*=6.22 percentage points, $p<0.001$) and producing their first word at 0 to 9 months of age (diff=1.60, $p<0.05$), on average. Boys, however, had higher reported externalizing problem behaviors (*diff*=0.24 points, $p<0.001$), were more frequently reported to have a disability (*diff*=5.42 percentage points, $p<0.001$), tended to be older at Kindergarten entry (*diff*=0.38 months, $p<0.01$), and were more frequently reported to have repeated Kindergarten (*diff*=1.01 percentage points, $p<0.01$) than girls, on average. Furthermore, boys were more frequently reported to never engage in independent reading (*diff*=2.83 percentage points, $p<0.001$) or engage in it once or twice per week (*diff*=5.30 percentage points, $p<0.001$) than girls, and were more frequently reported to be read to once or twice per week (*diff*=1.60 percentage points, $p<0.05$) or 3 to 6 times a week (*diff*=1.82 percentage points, $p<0.05$) than girls, on average. Finally, boys were more frequently reported to have spoken their first word at 13 to 15 months (*diff*=1.56 percentage points, $p<0.01$), 19 to 24 months (*diff*=1.10 percentage points, $p<0.01$), or after 24 months or not at all (*diff*=2.11 percentage points, $p<0.001$) than girls, and their parents more frequently expected them to earn a high school diploma or less (*diff*=0.71 percentage points, $p<0.05$) or attend some postsecondary school (*diff*=1.96 percentage points, $p<0.001$) than girls, on average. These findings suggest that the gender gap in reading ability and/or growth may be attenuated once these gender differences are controlled for. Even so, many of the variables are highly correlated, as illustrated in Table 6, which presents the correlations among all of the continuous variables to be included in the analysis. Given that almost all of the variables were significantly correlated with each other, they may explain some of the same variation in reading ability, so further multivariate analyses were conducted using HLMs in order to isolate the relationships of interest.

Question 1. Gender Differences in the Uncontrolled Model

A taxonomy of selected models from the HLM analysis is presented in Table 7 in Appendix E. In Model A, which addresses my first research question regarding the presence of a gender difference in reading ability at Kindergarten entry and/or in reading growth when not controlling for other variables, boys are estimated to enter Kindergarten with an average reading score of -0.794 units on the theta scale, and this score is estimated to grow linearly by 0.170 units and decrease quadratically by 0.002 units, on average, every month. Girls, however, are estimated to enter Kindergarten with a reading ability score that is 0.054 units higher than boys' (for an average reading score of -0.740 units, $z=2.49$, $p=0.013$), and they are estimated to grow linearly by 0.005 units (*z*=3.02, *p*=0.003) and decrease quadratically by 0.0001 units (*z*=-2.60, *p*=0.009) more than boys every month. Given that “except in the linear case, it’s not easy to understand the shape of a polynomial without plotting it” (Rabe-Hesketh & Skrondal, 2008, p. 349), the estimated average reading growth curves by gender from Model A are presented in Figure 2 in Appendix F. As illustrated in the figure, girls are estimated to have significantly higher reading scores than boys from Kindergarten through 2nd grade, on average, and the gender gap appears to widen to its largest point at the end of first grade (at around 21 months since Kindergarten entry) to about 0.110 units, and reduces to about 0.099 units by the end of second grade (at around 33 months since Kindergarten entry). Even so, this is equivalent to an effect size of about 0.064 standard deviations at Kindergarten entry, 0.151 standard deviations by the end of first grade, and 0.159 standard deviations by the end of second grade, which suggests that the gender gap continues to grows larger in effect from Kindergarten through the end of second grade. This suggests that girls in the U.S. enter Kindergarten with stronger reading skills than boys, on average, and this gender gap grows throughout early elementary school.

Question 2. Investigating Gender Differences when Adding Controls

In Models B through H, I explored whether the gender difference in reading ability at school entry, as well as in reading growth, held when controlling for various student- and school-level factors. As illustrated in Model B, the gender gap at Kindergarten entry (*b*=0.069, *z*=3.22, *p*=0.001), as well as the gender differences in linear (*b*=0.005, *z*=2.73, *p*=0.006) and quadratic (*b*=-0.0001, *z*=-2.66, *p*=0.008) reading growth, remained stable when controlling for *Race/ethnicity*, *SES*, *School Location Type, School Region, School Sector, Language status, Kindergarten repeater status, Age of Kindergarten Entry,* and the *Number of Children's Books in the Home,* even though all of these control variables significantly predicted reading ability at Kindergarten entry and many also significantly predicted linear and/or quadratic reading growth (those that did not were dropped from the model). Even though the gender gap at Kindergarten entry had attenuated upon the addition of *Frequency of Parent Reading to Child* and *Educational Expectations* to the model individually without other controls, the gender gap at Kindergarten entry (*b*=0.059, *z*=2.79, *p*=0.005) as well as the gender differences in linear (*b*=0.005, *z*=2.79, *p*=0.005) and quadratic (*b*=-0.0001, *z*=-2.68, *p*=0.007) reading growth, remained stable upon adding these two covariates in Model C. Even so, upon adding *Age of First Word* in Model D, the gender difference at Kindergarten entry attenuated slightly and was no longer statistically significant (*b*=0.040, *z*=1.66, *p*=0.096), although the linear (*b*=0.005, *z*=2.79, *p*=0.005) and quadratic (*b*=-0.0001, *z*=-2.80, *p*=0.005) growth terms associated with gender remained stable. Similarly, in Model E, upon the addition of *Disability Status*, the gender difference at Kindergarten entry was further attenuated and not statistically significant (*b*=0.023, *z*=0.95, *p*=0.342), and the linear (*b*=0.005, *z*=2.81, *p*=0.005) and quadratic (*b*=-0.0001, *z*=-2.85, *p*=0.004) growth terms associated with gender remained stable. Likewise in Model F, upon the addition of *Frequency of Child Independent Reading*, the gender difference at Kindergarten entry was further attenuated and not statistically significant (*b*=-0.010, *z*=-0.40, *p*=0.690), and the linear (*b*=0.005, *z*=2.78, *p*=0.005) and quadratic (*b*=-0.0001, *z*=-2.50, *p*=0.012) growth terms associated with gender remained stable. Interestingly, *Frequency of Parent Reading* was no longer predictive of reading ability at Kindergarten entry (or reading growth through second grade), when controlling for the other variables in the model, so this covariate was dropped from all subsequent models. In Model G, upon the addition of *Externalizing Behaviors*, the gender difference at Kindergarten entry was further attenuated and not statistically significant (*b*=-0.021, *z*=-0.87, *p*=0.384), and the gender differences in linear (*b*=0.004, *z*=1.88, *p*=0.060) and quadratic (*b*=-0.0001, *z*=-1.74, *p*=0.083) reading growth were also attenuated and were no longer statistically significant. Finally, in Model H, upon the addition of *Approaches to Learning*, the gender difference at Kindergarten entry became statistically significant in the opposite direction such that boys were predicted to have higher scores at Kindergarten entry than girls, controlling for all other variables in the model (*b*=-0.079, *z*=-3.20, *p*=0.001), and the gender difference in linear (*b*=0.005, *z*=2.32, *p*=0.020) reading growth was once again statistically significant and in the same direction and magnitude as in Models A-F. Even so, the gender difference in quadratic reading growth was slightly attenuated in comparison to Models A-F and only marginally significant (*b*=-0.0001, *z*=-1.96, *p*=0.05).

These results illustrate that, while the gender difference in reading ability at Kindergarten entry remains stable when controlling for sociocultural factors (e.g., *Race/ethnicity*, *SES*) as well as other factors that may differ by gender (e.g., *Kindergarten repeater status, Age of Kindergarten Entry*), it is attenuated and becomes non-statistically significant upon controlling for *Age of First Word*, *Disability Status*, *Frequency of Child Independent Reading*, and *Externalizing Behaviors*. Furthermore, when also controlling for *Approaches to Learning,* this relationship actually reversed, and boys were then predicted to enter school with higher reading scores, on average[15]. The gender difference in reading growth, however, remained stable when controlling for all of these factors except for *Externalizing Behaviors* and *Approaches to Learning*. These findings suggest that the gender gap at Kindergarten entry is fully explained in Model H, however the gender gap in reading growth through second grade is only partially explained by *Externalizing Behaviors* and *Approaches to Learning*. Even so, before settling on a final model and assuming that these relationships hold for all subgroups of students, further analyses are necessary to establish this.

Question 3. Exploring Whether Gender Gaps Differ in Size by Sociocultural Factors

I first tested whether the gender difference in reading ability at school entry or in reading growth differed by any of the potential moderators listed above when

[15] This was also the case when only controlling for *Approaches to Learning* and not including any other covariates in the model, but that model was not included in the taxonomy for the sake of brevity.

not controlling for any other variables, and I found no evidence that they differed in magnitude by any of the potential moderator variables, as none of the interaction terms were statistically significant. Nevertheless, I tested for these interactions once more when controlling for all of the statistically significant factors from Model H and found no evidence that the magnitude of the marginal gender gaps differed by any of the potential moderator variables except *School Region*. As illustrated in the final model, Model I, the gender difference in reading ability at Kindergarten entry was found to differ by *School Region* when controlling for all of the variables included in Model H, although there was no evidence that the magnitude of the coefficients associated with gender differences in linear and quadratic reading growth differed by *School Region*, as the latter were not statistically significant and therefore were not included in Model I. Contrary to the conclusions based on Model H (which assumes a main effect of *School Region*), I found no evidence of a gender difference in reading ability at Kindergarten entry among children who attended school in the Northeast (b=0.008, z=0.140, p=0.885) when controlling for their *Approaches to Learning*, among other variables; however, there was evidence of the reversed gender gap, among children who attended school in the South (b=-0.068, z=-2.06, p=0.039), Midwest (b=-0.121, z=-3.40, p=0.001), and West (b=-0.118, z=-3.36, p=0.001), such that boys were predicted to have higher reading scores at Kindergarten entry when controlling for all other variables in the model.

Figures 3 and 4 in Appendix F depict the results of the final model (Model

I) with predicted average reading trajectories displayed for prototypical girls and boys from each region when holding all other variables in the model constant[16]. As illustrated in Figure 3, the marginal gender gap from Kindergarten through second grade, favoring boys, was relatively large among children who attended schools in the West but it was smaller, and not statistically significant, for students in the North. Similarly, as illustrated in Figure 4, the reversed gender gap from Kindergarten through second grade was relatively large and statistically significant among children who attended schools in the Midwest, but in the South, the difference was statistically significant at Kindergarten entry, but not by the end of second grade. Taken together, these results suggest that the raw reading gender gap does not vary by sociocultural factors, although when controlling for children's *Approaches to Learning* in addition to other covariates, the prevalence/magnitude of the marginal gender gap, favoring boys, varies by *School Region*.

Regression Diagnostics and Sensitivity Analyses

Figures 5-12 in Appendix G present the regression diagnostics associated with the final model[17]. Students' predicted growth trajectories were fairly similar to the observed reading growth trajectories[18], which suggests that this model is an

[16] These prototypical lines needed to be displayed in two separate figures to enhance readability; the decision to display them in two separate figures was not based on substantive differences among the four regions.
[17] Regression diagnostics were performed throughout the model building process, and the assumptions held for each model, but I only present the diagnostics for the final model here due to space limitations.
[18] I was unable to present individual growth curves in the paper due to restricted-use data policies, but they were compared against the fitted data.

appropriate fit to the data. Figure 5, which presents box plots of the predicted random intercepts at the school and child levels as well as the residuals at the occasion level, illustrates that there was much more variation at the school and child levels than at the occasion level, which is to be expected; however there are quite a few outliers at both the child and occasion levels. Even so, given that the occasion-level residuals and predicted random slopes and intercepts at both the child and school levels appear to approximate a Normal distribution, as illustrated in Figures 6 through 8, this is not likely a cause for concern. In addition to checking the functional form and distributions of the residuals, as illustrated in Figures 9 and 10 , I also visually inspected scatter plots of the residuals on each predictor to check for homoscedasticity. Overall the assumption appears to hold, although there does seem to be substantially less variability in school-level intercepts among schools located in towns in comparison to other location types and there also appears to be less variability in child-level intercepts among children who repeated Kindergarten. There appear to be floor and ceiling effects, respectively, in the distributions of *Externalizing Behaviors* and *Approaches to Learning*. Finally, as illustrated in Figure 11, the homoscedasticity assumption appears to hold at the observation-level, although, as previously discussed, there appear to be a few children with unexpectedly low scores at wave 2. Even so, these outliers as well as the issues with homoscedasticity for certain variables are minor, and therefore, I conclude that these assumptions hold in the population.

Table 8 in Appendix G presents selected models (corresponding to Models A, H, and I from Table 7) from the sensitivity analyses exploring whether the results differed when the data were unweighted but limited to the same sample and/or when all six waves of data were included and the sample was limited to children who had non-zero values on the W6C6P_6T0 sampling weight[19]. As illustrated by the first three models in Table 8, the unweighted results are fairly similar to the weighted results in terms of point estimates and substantive conclusions. The main difference is that the quadratic term associated with the gender difference in reading growth was statistically significant in both Model H-nw and Model I-nw (*p*=0.006) rather than just marginally significant (*p*=0.050) as in the weighted models. As illustrated by the final three models in Table 8, the results including all six waves of data produce slightly different results, particularly regarding gender differences in reading growth. For instance, in Model A-6, girls were identified as having higher reading scores than boys, on average, at Kindergarten entry, but there was no evidence of gender differences in reading growth through second grade. Furthermore, in Model H-6 and Model I-6 there was also no evidence of a gender difference in reading ability at Kindergarten entry controlling for all other variables in the model, and in Model I-6, the difference in magnitude of the gender gap at Kindergarten entry by region was only marginally significant ($\chi^2(3)=6.53$, $p=0.089$). Even so, these differences

[19] This was the appropriate weight to use in this case because it corresponded to children who had reading scores at all 6 time points as well as the necessary parent interview and teacher survey data.

can likely be explained by the relatively small sample size and improper functional form given the summer reading slide (Alexander et al., 2001).

Discussion

The present study explored gender differences in American children's reading development from Kindergarten through the end of second grade in order address the gaps in the body of knowledge needed to inform appropriate and effective interventions aimed at closing the gender gap in reading. Specifically, the study set out to both replicate and extend previous research on when the gender gap in reading emerges and how it develops throughout early elementary school in order to investigate whether interventions aimed at closing the gender gap should be implemented earlier. In addition, an in-depth exploration of potential mediators and moderators of the gender gap was also undertaken in order to help inform the kinds of factors that may be of interest to applied researchers as potential components of an intervention as well as to identify the groups of students who may be particularly affected by the gender gap and therefore may need additional supports. In order to investigate these questions, a taxonomy of three-level HLMs was constructed using the ECLS-K:2011 data set.

When to Intervene: Findings and Implications

In response to my first research question, consistent with prior research (e.g., Husain & Millimet, 2009; McCoach, O'Connell, Reis, & Levitt, 2006; Ready et al., 2005), I found that girls had significantly higher reading scores than boys at Kindergarten entry, on average. The effect size (d=0.064 *SDs*) was rather small, however, I also found that the gender gap grew in effect from Kindergarten throughout second grade, where it was estimated to be a slightly more substantial (*d*=0.159 *SDs*). This finding is consistent with Rathbun and colleagues (2004) who found that the gender gap grew between Kindergarten and third grade, and it is not completely inconsistent with those of McCoach and colleagues (2006) or Aikens and Barbarin (2008) who also found that the gender gap grew in early elementary school, however their conclusions were likely more targeted (to specific grades) due to their use of a piecewise rather than a linear or quadratic model. Even so, these results differ from those of Husain and Millimet (2009) and Robinson and Lubienski (2011), who found that the gender gap narrowed during elementary school. The major differences between these two sets of studies is that those that found that the gap widens during early elementary school (including myself) focused exclusively on the early school years and employed HLM or regression-based approaches while those that found that the gap narrows used quantile regression approaches and/or their own methodologies. One advantage to my approach using HLM is that it enabled me to operationalize time in such a way that takes into account the relation between the child's approximate testing date and their first day of school in order to more authentically explore the estimated gender difference at Kindergarten *entry* as opposed to the fall of Kindergarten as well as capture growth in a more accurate manner. Even so, given that all of the aforementioned studies, including my own, vary slightly in their methodological approaches and therefore in their conclusions, additional work is necessary to replicate these findings and in particular to tease apart why there may be discrepancies in results when employing certain approaches rather than others. In addition, future studies should examine how these gender gaps develop over a longer time course, in order to examine whether the gender gap grows to the small, yet practically significant, sizes estimated by the NAEP data (e.g., d=0.189 *SDs* by fourth grade and d=0.250 *SDs* by twelfth grade). Additionally, studies should further investigate the development of gender gaps in specific reading tasks, particularly those which have been found to be much larger than the gender gap across all reading skills, such as evaluating and interpreting versus retrieving information (e.g., Mullis et al., 2012), since large gaps in specific skills may be obscured by studying literacy as a unitary construct.

How to Intervene: Findings and Implications

Consistent with prior research (e.g., Chiu and McBride-Chang, 2006), I found no evidence that sociocultural factors (e.g., race/ethnicity, language status) or the number of books in the children's home explained the gender gap either in at Kindergarten entry or in reading growth through second grade. The age of Kindergarten entry and whether or not the child repeated Kindergarten were also not found to meditate these relationships of interest. Parents' educational expectations for their children and the amount that they reported reading to their children per week were also not found to explain these relationships of interest when controlling for other variables, and the frequency of parents reading to their children was actually not found to be predictive of reading at either Kindergarten entry or growth through second grade when controlling for the frequency of child independent reading. Given that the gender difference in reading at Kindergarten entry was no longer statistically significant when controlling for age of first word, disability status, independent reading frequency, externalizing behaviors, and approaches to learning, both independently and when combined, this suggests that these variables fully explain this gender gap. Even so, assuming that these results are correct and the gender gap does increase over time, applied researchers, practitioners, and policy makers are likely also interested in factors that explain the gender gap in reading *development* throughout the school years, since these are the factors that may prevent the gap from widening, provided that the relationship is indeed causal and such factors can be modified through intervention. The results of my analysis suggest that children's externalizing behaviors fully explain the gender difference in reading growth through second grade, and children's approaches to learning may also partially explain this gap. This suggests that while the gender difference at Kindergarten entry may also be explained by factors such as children's disability status and their frequency of independent reading in addition to their classroom behaviors, the widening of this gap over time seems to be explained by gender differences in students' classroom behaviors.

These findings are partially consistent with previous research given that Ready and colleagues (2005) found that children's Kindergarten classroom behaviors explained the gender gap in early reading development, however they found that children's approaches to learning more strongly explained this relationship than their externalizing behaviors. Nevertheless, these findings suggest that children's classroom behaviors may be important to consider for interventions aimed at closing the reading gender gap, although it is unclear given the study design whether this relationship is indeed causal or merely correlational, so applied research on this topic is necessary to determine this. Furthermore, approaches to learning and classroom behavior can be measured in many different ways, and therefore a deeper understanding of each of these constructs is necessary to inform interventions. For instance, is it the metacognitive strategies that children use when reading, as defined by the OECD (2010) that are driving the mediation for "approaches to learning" or the construct including attentiveness, task persistence, and flexibility as defined by Tourangeau and colleagues (2015)? Additionally, despite all of the prior research that suggests that students' intrinsic reading motivation mediates the gender gap (e.g., Chiu & McBride-Chang, 2006; OECD, 2010), I found that only the gender gap at Kindergarten entry, but not in reading growth, was explained by children's frequency of independent reading. Even so, this is not a direct measure of students' reading motivation, and therefore it is possible that the gender gap in reading growth would have been explained if a different measure were used. The ECLS-K:2011 does include a measure of reading engagement beginning in third grade, so this can be directly addressed in future analyses of reading development in the upper elementary grades.

With Whom to Intervene: Findings and Implications

In contrast to the findings of Husain and Millimet (2009), I found no evidence that the magnitude of the gender gap in reading either at Kindergarten entry or in growth through second grade differed by race/ethnicity, SES, and/or school sector. Consistent with the authors, I found no evidence that these gender gaps were moderated by school location type or by school region when not controlling for other variables. Even so, I found that the magnitude of the marginal gender gap at Kindergarten entry differed as a function school region when controlling for sociocultural factors as well as the aforementioned mediators of the gender gap such that the gap was actually estimated to favor boys from Kindergarten to second grade for children in the West and Midwest and in Kindergarten but not second grade in the South, and was not estimated to differ by gender during these early school years in the Northeast. Given that these differences were not detected in the uncontrolled model, this suggests that students within particular regions of the U.S. are not necessarily more affected by the gender gap than others and therefore this gender difference by region is not particularly practically interesting. It may suggest that the effects of an intervention targeting classroom behaviors would differ by region, controlling for sociocultural factors, but this isn’t relevant at this stage given that much more research would be necessary to establish causality and determine which aspects of classroom behavior explain the gap prior to intervening.

Given that these results stand in opposition to those of Husain and Millimet (2009), who take a different methodological approach in their analysis, further research is necessary to resolve these discrepancies. From a practical standpoint, however, the groups of students that Husain and Millimet identified as needed extra reading supports (e.g., Black and Hispanic, low-SES and/or public school boys) are similar to the students that I would argue need extra support based on my findings. Although the gender gap was not found to differ by SES, race/ethnicity, or school sector, when exploring the interplay of each of these factors and gender during reading development, a similar picture arises as to which students may need extra supports. Figures 12-14 in Appendix H illustrate estimated growth curves for boys and girls by race/ethnicity, SES, and school sector, respectively, when controlling for sociocultural factors (Model B). As illustrated in Figure 12, Black, Hispanic, and White boys are estimated to have the lowest levels of reading ability, on average, from Kindergarten through second grade when controlling for SES, language status, and all other covariates included in the model. In Figure 13, low SES boys are estimates to have the lowest reading levels of reading ability, on average, from Kindergarten through second grade when controlling for race/ethnicity, school sector and other sociocultural factors, and in Figure 14, private school boys are estimated to begin school with lower reading scores holding all else constant, however they grow at a faster rate than public school students such that public school boys are estimated to have the lowest reading abilities, on average, by the end of second grade. Consistent with the recommendations of Husain and Millimet, this analysis suggests that boys of color (and possibly also white boys), low-SES boys, and/or boys who attend public school have lower levels of reading ability, on average, as compared to their peers. This suggests that advocates for closing the racial/ethnic, SES, and other achievement gaps in reading would be prudent to take gender into account in both their research and possibly also in their development of interventions.

Limitations and Future Directions

One key limitation of this study is that the measure of reading ability administered in the ECLS-K:2011 is reported as a unitary construct, and therefore it is not possible to explore which specific reading skills differ by gender in order to determine if such gaps are of practical significance, and, if so, develop a targeted intervention. It would be difficult to develop a longitudinal study that would be able to measure the same kinds of skills/reading tasks over time, especially during early elementary school, so further cross-sectional work may be useful to achieve this. Similarly, the measures of children’s approaches to learning and externalizing behaviors suffer from ceiling and floor effects, respectively, and the scores on these broad, teacher-reported constructs were developed by taking a simple average, which means that there’s likely measurement error. Factor analysis may help with the measurement error and to determine whether this broad construct is indeed one construct/measuring the same construct for all students. Another limitation was the lack of a reading motivation measure; this should be explored in future longitudinal research.

There were many limitations to using the HLM framework within Stata, as well, for instance, there is no built-in way to calculate effect sizes, so it was difficult to evaluate and communicate the magnitude of the gender gap. Furthermore, rather than just studying the development of the gender gap, on average, it would have been useful to see how the gap varies as a function of reading ability (similar to Robinson & Lubienski, 2011), however I would have had to utilize a quantile regression approach which would not allow me to operationalize time in the same way. In addition, it would have been more intuitive to be able to build-in factor models and test mediation using the structural equation modeling framework, however, latent growth curve modeling requires that the students were assessed at even intervals, which was not the case in the ECLS-K:2011. Another major limitation is that the analytic sample may have differed from the full sample in systematic ways given that there was some item-level missingness that resulted in a 28% reduction of the analytic sample despite the use of sampling weights; therefore it is possible that these results are no longer representative of children in the U.S. (who entered Kindergarten in 2010). Lastly, this research is observational, so although I have found evidence for relationships among factors, I cannot make such claims based on these analyses. These relationships may be of interest, however, to applied researchers who can, in turn, study whether these relationships are indeed causal and/or if such approaches are effective at closing the gap.

Conclusion

In conclusion, there is evidence of a gender difference in American children's reading abilities at Kindergarten entry, favoring girls, and this gender gap appears to grow in effect through the end of second grade. This gap is fairly small in magnitude, however, it is expected to grow through schooling, and may be larger, and of much more practical importance, for certain reading tasks than others. Although the origin of this reading gap remains unclear, the findings of this study, coupled with previous research findings, suggest that children's classroom behavior may be an important mediator of this relationship both at Kindergarten entry and in their development of reading skills through second grade. Further work is necessary to explore the role of reading motivation in the development of this gap throughout school, as it has been implicated in prior studies but was not included in this data set. Measures of children's gender roles and/or gender stereotypes related to reading were also not included in this data set and would be useful to explore further, especially in combination with these other factors. Lastly, I found no evidence that the magnitude of the gender gap differed by sociocultural factors, except for school region when controlling for mediators of the gender gap, yet the combination of gender with these other factors suggests that certain students, namely boys except for Asian boys who have a low-SES background and/or attend public school, have a less favorable reading trajectory than their peers from Kindergarten through second grade, on average, which may also have implications for reading instruction/intervention. Although there are still many unanswered questions as to the role of gender in students' early reading development, this study provides scaffolding as to the necessary next steps in basic and applied research to inform practice and policies aimed at ensuring that all students have access to the opportunities that strong literacy skills afford.

References

Aikens, N. L., & Barbarin, O. (2008). Socioeconomic differences in reading trajectories: The contribution of family, neighborhood, and school contexts. *Journal of Educational Psychology*, *100*(2), 235.

Ainsworth, C. (2015). Sex redefined. *Nature*, *518*(7539), 288.

Alexander, K. L., Entwisle, D. R., & Olson, L. S. (2001). Schools, achievement, and inequality: A seasonal perspective. *Educational evaluation and policy analysis*, *23*(2), 171-191.

Alloway, N. (2007). Swimming against the tide: Boys, literacies, and schooling: An Australian story. *Canadian Journal of Education/Revue canadienne de l'éducation*, 582-605.

Boldizar, J. P. (1991). Assessing sex typing and androgyny in children: The Children's Sex Role Inventory. *Developmental Psychology*, *27*(3), 505.

Boyle, C. A., Boulet, S., Schieve, L. A., Cohen, R. A., Blumberg, S. J., Yeargin-Allsopp, M., Visser, S., & Kogan, M. D. (2011). Trends in the prevalence of developmental disabilities in US children, 1997-2008. *Pediatrics, 127*(6), 1034-1042.

Brent, D., May, D. C., & Kundert, D. K. (1996). The incidence of delayed school entry: A twelve-year review. *Early Education and Development*, *7*(2), 121-135.

Brozo, W. G. (2002). *To be a boy, to be a reader: Engaging teen and preteen boys in active literacy*. Order Department, International Reading Association, 800 Barksdale Road, PO Box 8139, Newark, DE 19714-8139.

Brozo, W. G., Sulkunen, S., Shiel, G., Garbe, C., Pandian, A., & Valtin, R. (2014). Reading, Gender, and Engagement. *Journal of Adolescent & Adult Literacy*, *57*(7), 584–593.

Catts, H. W. (2017). Early identification of reading disabilities. *Theories of Reading Development*, *15*, 311.

Chiu, M. M., & McBride-Chang, C. (2006). Gender, context, and reading: A comparison of students in 43 countries. *Scientific Studies of Reading*, *10*(4), 331–362.

Conger, D., & Long, M. C. (2013). Gender gaps in college enrollment: The role of gender sorting across public high schools. *Educational Researcher*, *42*(7), 371-380.

Disenhaus, N. (2015). Boys, Writing, And The Literacy Gender Gap: What We Know, What We Think We Know. *Graduate College Dissertations and Theses*. Paper 330.

Friedmann, P. D., Rose, J. S., Swift, R., Stout, R. L., Millman, R. P., & Stein, M. D. (2008). Trazodone for Sleep Disturbance After Alcohol Detoxification: A Double-Blind, Placebo-Controlled Trial. *Alcoholism: Clinical and Experimental Research*, *32*(9), 1652-1660.

Fryer Jr., R. G., & Levitt, S. D. (2006). The black-white test score gap through third grade. *American Law and Economics Review*, *8*(2), 249-281.

Gelman, A., Hill, J., & Yajima, M. (2012). Why we (usually) don't have to worry about multiple comparisons. *Journal of Research on Educational Effectiveness*, *5*(2), 189-211.

Gurian, M. (2010). *Boys and girls learn differently! A guide for teachers and parents: Revised 10th anniversary edition*. John Wiley & Sons.

Guthrie, J. T., Wigfield, A., Metsala, J. L., & Cox, K. E. (1999). Motivational and cognitive predictors of text comprehension and reading amount. *Scientific studies of reading*, *3*(3), 231-256.

Halpern, D. F. (2012). *Sex differences in cognitive abilities,* 4th edition. New York, NY: Psychology press.

Hartley, B. L., & Sutton, R. M. (2013). A stereotype threat account of boys' academic underachievement. *Child Development*, *84*(5), 1716-1733.

Husain, M., & Millimet, D. L. (2009). The mythical "boy crisis"? *Economics of Education Review*, *28*(1), 38–48.

Huttenlocher, J., Haight, W., Bryk, A., Seltzer, M., & Lyons, T. (1991). Early vocabulary growth: Relation to language input and gender. *Developmental psychology*, *27*(2), 236.

Hyde, J. S., & Linn, M. C. (1988). Gender differences in verbal ability: A meta-analysis. *Psychological Bulletin*, *104*(1), 53–69.

Ingalhalikar, M., Smith, A., Parker, D., Satterthwaite, T. D., Elliott, M. A., Ruparel, K., ... & Verma, R. (2014). Sex differences in the structural connectome of the human brain. *Proceedings of the National Academy of Sciences*, *111*(2), 823-828.

Keenan, K., & Shaw, D. (1997). Developmental and social influences on young girls' early problem behavior. *Psychological bulletin*, *121*(1), 95.

Klecker, B. M. (2006). The gender gap in NAEP fourth-, eighth-, and twelfth-grade reading scores across years. *Reading Improvement*, *43*(1), 50.

Kleinfeld, J. (2009). The state of American boyhood. *Gender Issues*, *26*(2), 113-129.

Maine Department of Education. (2007). Final Report: Task Force on Gender Equity in Education in Maine. Retrieved from https://www1.maine.gov/education/gender_task_force/final_report.pdf

Marks, G. N. (2008). Accounting for the gender gaps in student performance in reading and mathematics: evidence from 31 countries. *Oxford Review of Education*, *34*(1), 89–109.

Marinak, B. A., & Gambrell, L. B. (2010). Reading Motivation: Exploring the Elementary Gender Gap. *Literacy Research and Instruction*, *49*(2), 129–141. https://doi.org/10.1080/19388070902803795

McCoach, D. B., O’Connell, A. A., Reis, S. M., & Levitt, H. A. (2006). Growing readers: A hierarchical linear model of children’s reading growth during the first 2 years of school. *Journal of Educational Psychology*, *98*(1), 14.

McGeown, S., Goodwin, H., Henderson, N., & Wright, P. (2012). Gender differences in reading motivation: does sex or gender identity provide a better account?. *Journal of Research in Reading*, *35*(3), 328-336.

McKenna, M. C., Kear, D. J., & Ellsworth, R. A. (1995). Children's attitudes toward reading: A national survey. *Reading research quarterly*, 934-956.

Mulligan, G. M., McCarroll, J. C., Flanagan, K. D., & Potter, D. (2015). Findings From the Second-Grade Rounds of the Early Childhood Longitudinal Study, Kindergarten Class of 2012-11 (ECLS-K:2011) (NCES 2015-077). U.S. Department of Education, Washington, DC: National Center for Education Statistics. Retrieved [March 10, 2017] from http://nces.ed.gov/pubsearch.

Mullis, I. V., Martin, M. O., Foy, P. & Drucker, K. (2012). PIRLS 2011 international results in reading. TIMSS & PIRLS International Study Center. Boston, MA: Lynch School of Education, Boston College.

Mullis, I.V., Martin, M.O., Kennedy, A.M. & Foy, P. (2007). PIRLS 2006 international report: IEA's progress in international reading literacy study in primary schools in 40 countries. Boston, MA: Lynch School of Education, Boston College.

Newkirk, T. (2000). Misreading masculinity: Speculations on the great gender gap in writing. *Language Arts*, *77*(4), 294-300.

OECD (2010), *PISA 2009 Results: Learning to Learn – Student Engagement, Strategies and Practices (Volume III)http://dx.doi.org/10.1787/97 89264083943-en*

Orr, A. J. (2011). Gendered capital: Childhood socialization and the "boy crisis" in education. *Sex Roles*, *65*(3-4), 271–284.

Perlman, E. (2016, July 18). School gender gap crisis: boys 'twice as likely to fall behind girls'. *Newsweek*. Retrieved from http://www.newsweek.com/gender-gap-crisis-england-boys-fall-behind-girls-language-skills-early-years-481323

Peter, K., & Horn, L. (2005). Gender Differences in Participation and Completion of Undergraduate Education and How They Have Changed Over Time. Postsecondary Education Descriptive Analysis Reports. NCES 2005-169. *US Department of Education*.

Quinn, J. M., & Wagner, R. K. (2015). Gender differences in reading impairment and in the identification of impaired readers: Results from a large-scale study of at-risk readers. *Journal of learning disabilities*, *48*(4), 433-445.

Rabe-Hesketh, S., & Skrondal, A. (2008). *Multilevel and longitudinal modeling using Stata*. STATA press.

Rathbun, A., West, J., & Hausken, E. G. (2004). From Kindergarten Through Third Grade Children's Beginning School Experiences. NCES 2004 007. *US Department of Education*.

Ready, D. D., LoGerfo, L. F., Burkam, D. T., & Lee, V. E. (2005). Explaining girls' advantage in kindergarten literacy learning: Do classroom behaviors make a difference? *The Elementary School Journal*, *106*(1), 21–38.

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.

Roe, A. & Taube, K. (2003). Reading achievement and gender differences. In S. Lie, P. Linnakylä & A. Roe (Eds.), Northern lights on PISA, (pp. 21–37). Oslo, Norway: Department of Teacher Education and School Development, University of Oslo.

Rowan, L., Knobel, M., Bigum, C., & Lankshear, C. (2002). Boys, literacies and schooling. *Buckingham and Philadelphia*. Retrieved from: http://www.mheducation.co.uk/openup/chapters/0335207561.pdf

Ruigrok, A. N., Salimi-Khorshidi, G., Lai, M. C., Baron-Cohen, S., Lombardo, M. V., Tait, R. J., & Suckling, J. (2014). A meta-analysis of sex differences in human brain structure. *Neuroscience & Biobehavioral Reviews*, *39*, 34-50.

Selya, A. S., Rose, J. S., Dierker, L. C., Hedeker, D., & Mermelstein, R. J. (2012). A practical guide to calculating Cohen's f2, a measure of local effect size, from PROC MIXED. *Frontiers in psychology*, *3*.

Sommers, C. H. (2000). *The war against boys: How misguided feminism is harming our young men*. New York: Simon & Schuster.

StataCorp. 2015. Stata Statistical Software: Resease 14. College Station, TX: StataCorp LP.

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), e57988.

Taylor, D. L. (2004). “Not Just Boring Stories”: Reconsidering the Gender Gap for Boys. *Journal of Adolescent & Adult Literacy*, *48*(4), 290–298. https://doi.org/10.2307/40016916

Tourangeau, K., Nord, C., Lê, T., Sorongon, A. G., and Najarian, M. (2009). *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, Institute of Education Sciences, U.S. Department of Education. Washington, DC.

Tourangeau, K., Nord, C., Lê, T., Wallner-Allen, K., Vaden-Kiernan, N., Blaker, L. & Najarian, M. (2015). *Early Childhood Longitudinal Study, Kindergarten Class of 2010–11 (ECLS-K:2011) User’s Manual for the ECLS-K:2011 Kindergarten-Second Grade Data File and Electronic Codebook, Restricted Version* (NCES 2015-049). U.S. Department of Education. Washington, DC: National Center for Education Statistics.

Tyre, P. (2006, January 29). Education: boys falling behind girls in many areas. *Newsweek*. Retrieved from http://www.newsweek.com/education-boys-falling-behind-girls-many-areas-108593

Van Der Bolt, L., & Tellegen, S. (1996). Sex differences in intrinsic reading motivation and emotional reading experience. *Imagination, Cognition and Personality*, *15*(4), 337-349.

Voyer, D., & Voyer, S. D. (2014). Gender differences in scholastic achievement: A meta-analysis. *Psychological Bulletin*, *140*(4), 1174.

Watson, A., Kehler, M., & Martino, W. (2010). The Problem of Boys' Literacy Underachievement: Raising Some Questions. *Journal of Adolescent & Adult Literacy*, *53*(5), 356–361. https://doi.org/10.1598/JAAL.53.5.1

White, B. (2007). Are Girls Better Readers than Boys? Which Boys? Which Girls? *Canadian Journal of Education / Revue Canadienne de L'éducation*, *30*(2), 554–581. https://doi.org/10.2307/20466650

Wigfield, A., & Guthrie, J. T. (1997). Relations of children's motivation for reading to the amount and breadth or their reading. *Journal of educational psychology*, *89*(3), 420.

Williams, J. A. (2012). Girls Can Be Anything: But Boys Will Be Boys: Discourses of Sex Difference in Education Reform Debates. *Nev. LJ*, *13*, 533.

Appendix A. Detailed descriptions of Covariates and Moderators

Table 1.

Detailed descriptions of covariates included in the analysis

Variable name	Description	Measurement	Hypothesis/Rationale for inclusion
Language status[1]	Whether English is a primary language used in the child's home (obtained from fall K parent interview)	Dichotomous variable: 0: *ENGLISH* (English is a primary home language) 1: *NON_ENGLISH* (English is not a primary home language)	Since some students completed their reading assessments in Spanish rather than English (in the earlier waves of data collection), it was important to control for language status to ensure that this did not affect the estimates or conclusions. Given that this variable was not of theoretical interest either as a potential mediator or moderator, I included it as a covariate rather than in the "potential moderators" section even though it represents a sociocultural factor.
Kindergarten Repeater status	Whether or not the child repeated Kindergarten (obtained from fall K parent interview or teacher questionnaire)	Dichotomous variable: 0: *NO REPEAT* (Child did not repeat Kindergarten) 1: *KREPEAT* (Child repeated Kindergarten)	Since some students repeated Kindergarten, the gender differences in reading development observed may have been due to gender differences in Kindergarten retention, and therefore it was important to control for this to ensure that this did not affect the estimates or conclusions.
Age of Kindergarten Entry	Age of the child at Kindergarten entry	Continuous variable (*AGEENT)*: child's age in months on September 1st during the year that the child (first) entered Kindergarten ECLS-K:2011-generated composite using child's date of birth and year of Kindergarten as reported on the fall K parent interview	Since prior research suggests that boys are more likely than girls to be delayed from entering Kindergarten (e.g., Brent, May, & Kundert, 1996), it was important to control for the age at which the child enters Kindergarten in order to investigate whether this gender difference may explain the reading gender gap at school entry and/or in growth rates through the end of second grade.
Number of Children's Books in the	Number of children's books in the child's	Continuous variable (*NUMBOOKS*): approximate number of	I did not expect there to be any gender differences in children's home literacy environment

Home	home (obtained during fall K parent interview)	children's books in the child's home (including library books)	(operationalized here as number of books in the child's home), based on the findings of Chiu and McBride-Chang (2006), however this is an important predictor of reading development, so I decided to include it as a covariate.
Frequency of Child Independent Reading	How often the child reads independently outside of school in a typical week (obtained from the fall K parent interview)	Polytomous variable: 0: *R3-6* (Child reads three to six times per week) 1:*RNEVER* (Child never reads independently) 2: *R1or2* (Child reads once or twice per week) 3:*REVERY* (Child reads every day)	Since there are no measures of reading engagement or motivation collected in the ECLS-K:2011 until third grade, I used the reported frequency of child independent reading as a proxy for this. Since girls may have higher levels of reading interest than boys (McKenna, Kear, & Ellsworth, 1995), on average, and gender differences in reading motivation/interest are thought to mediate the gender gap in reading ability (Chiu & McBride-Chang, 2006), the gender difference in reading at Kindergarten entry and/or in reading growth may be attenuated upon the addition of this measure to the model.
Frequency of Parent Reading to Child	How often the parent or any other family member reads to the child in a typical week (obtained from the fall K parent interview)	Polytomous variable: 0: *P_READ_3-6* (Family member reads to child three to six times per week) 1:*P_READ_NEVER* (Family member never reads to child) 2: *P_READ_1or2* (Family member reads to child once or twice per week) 3:*P_READ_EVERYDAY* (Family member reads to child every day)	Children who tend to read more independently may also have parents who read more frequently to their children, so it was important to control for this in order to isolate the association between child independent reading frequency and reading ability.
Age of First Word	Age at which child spoke his or her first word (obtained during	Polytomous variable: 0: *A10to12* (Child first spoke at 10-12 months)	Since prior work suggests that girls have faster early vocabulary growth than boys (e.g., Huttenlocher, Haight,

	spring K parent interview)	1: *A01to9* (Child first spoke before 10 months) 2: *A13to15* (Child first spoke at 13-15 months) 3: *A16to18* (Child first spoke at 16-18 months) 4: *A19to24* (Child first spoke at 19-24 months) 5: *A25orNever* (Child first spoke after 25 months or never learned to speak)	Bryk, Seltzer, & Lyons, 1991) among other slight advantages in verbal ability (e.g., Hyde & Linn, 1988), I expected that such differences in language development might account for the gender difference in reading ability at Kindergarten entry (especially since many of the completed items are likely pre-reading, language-based skills) and/or in reading growth.
Educational Expectations[2]	Parent expectations for child's educational degree attainment (obtained during fall K parent interview)	Polytomous variable: 0: *BACHELORS* (child expected to earn Bachelor's degree) 1: *HSORLESS* (child expected to complete high school or less) 2:*POSTSEC* (child expected to complete some postsecondary education) 3: *MASTERS* (child expected to complete Master's degree or equivalent) 4: *MDPHD* (child expected to complete Ph.D., M.D., or other advanced degree)	I expected that any gender differences in parents' expectations for their child's educational attainment would be so slight that they would not likely explain away the reading gender gap, however, I was interested in exploring whether such a gender difference exists, and if so, whether the coefficients of interest (related to the reading gender gap) would be attenuated upon its inclusion in the model.
Disability status	Child disability status (obtained during spring K parent interview)	Dichotomous variable: 0: *NODISABILITY* (child does not have a disability) 1: *DISABILITY* (child has a disability) ECLS-K:2011-generated composite variable	Since previous research suggests that boys are more likely than girls to be diagnosed with developmental disabilities (e.g., Boyle et al., 2011), and many of these disabilities, such as autism, ADHD, and learning disabilities tend to be associated with poorer language and literacy outcomes, it was important to control for

		coded as 1 if the parent answered yes to any of the interview questions about child diagnosis or therapy services (except in cases related to mild vision and hearing diagnoses, such as the need to wear glasses).	child disability status to ensure that any gender differences observed in reading development were not explained by gender differences in the prevalence of such disabilities.
Approaches to Learning[3]	Child's approaches to learning (obtained during fall K teacher questionnaire)	Continuous variable (*APPROACH)*: higher scores indicate that the child more often exhibited positive approaches to learning (e.g., eagerness to learn, persistence in completing tasks, works independently) ECLS-K:2011-generated composite–Mean rating on seven ECLS-K-generated items (Tourangeau et al., 2015)	I hypothesized that both the reading gender gap at Kindergarten entry and the gap in growth through second grade would be partially, if not fully, explained by children's approaches to learning, consistent with the findings of Ready and his colleagues (2005).
Externalizing Behaviors[3]	Child's externalizing problem behavior (obtained from fall K teacher questionnaire)	Continuous variable (*EXTERN)*: higher scores indicate that the child exhibits externalizing problem behaviors (e.g., breaking the rules, fighting) more often ECLS-K:2011-generated composite–Mean rating on five items based on items from the Social Skills Rating System (NCS Pearson as cited in Tourangeau et al., 2015)	Given that previous research indicates that boys are more likely to engage in externalizing problem behaviors than girls (e.g., Offord et al. as cited in Keenan & Shaw, 1997), it was important to control for such behaviors when investigating gender differences in reading development to ensure that such differences are not solely due to differences in classroom behavior. Consistent with Ready and colleagues (2005), I expect that the gender gaps of interest may be partially, but not fully, explained by this variable.

[1]The original language status composite variable generated by the ECLS-K:2011 researchers from the fall or spring of Kindergarten parent interview data had a third category (other than English vs. Non-English) which represented children whose parents reported that both English and a non-English language were spoken at home, but they couldn't choose a primary language because they used both equally. I collapsed this group of students (only 340 in the analytic sample) into the reference category since they were exposed to English as often as to another language at home and were likely exposed primarily to English at school and in their communities as well.

[2] The original educational expectations variable had seven categories, but I combined the "less than a high school diploma" category with the "high school diploma" category to form the "high school diploma or less" category, and I combined the "vocational or technical school after high school" category with the "two or more years of college" category to form the "some postsecondary education" category.

[3]Teachers were also asked to complete items measuring children's self-control, attentional focus, and internalizing problem behaviors, but since these measures were all highly correlated with either *Externalizing Behaviors* or *Approaches to Learning*, and the latter were the items most frequently used in previous research on this topic, I only included these items in the analysis to avoid (multi)collinearity.

Table 2.

Detailed descriptions of potential moderators included in the analysis

Variable name	Description	Measurement	Hypothesis/Rationale for inclusion
Race/ethnicity	Child race/ethnicity (derived from fall or spring of K parent interview or the Field Management System)[1]	Polytomous variable: 0: *WHITE* (White, non-Hispanic) 1: *BLACK* (Black or African American, non-Hispanic) 2: *HISPANIC* (Hispanic or Latino) 3: *ASIAN* (Asian, non-Hispanic) 4. *OTHER* (e.g., Pacific Islander, Native American, Two or More Races, all non-Hispanic)	Based on the findings of Husain and Millimet (2009), I anticipated the gender gap in reading ability at Kindergarten entry to be smaller for Black and Hispanic students than for white and Asian students. Additionally, I expected gender differences in reading growth to vary as a function of race/ethnicity resulting in a smaller gender gap at the end of second grade for white and Asian students than for Black and Hispanic students.
Socioeconomic status	Child's socioeconomic status (derived from parent interview in fall or spring of K)	Continuous variable (*SES*): on a scale ranging from -2.33 to 2.6 units in the analytic sample, with higher values representing a higher SES background ECLS-K:2011-generated composite of parental education, occupational prestige, and household income with missing values imputed.	I hypothesized that the gender gap in reading ability at Kindergarten entry would be smaller among lower SES than higher SES children, and that the rates of reading growth by gender would also differ by SES resulting in a smaller gender gap at the end of second grade for higher SES than lower SES students, consistent with the findings of Husain and Millimet (2009).
School Sector	Sector of the child's school–public or private (ECLS-K:2011-generated composite from fall of K Field Management System data)	Dichotomous variable: 0: *PUBLIC* (public school) 1: *PRIVATE* (Catholic, other religious, or other private school)	Consistent with the findings of Husain and Millimet (2009), I anticipated that the reading gender gap at Kindergarten entry would be smaller among public school students than among private school students, but that gender differences in reading growth would vary as a function

			of school sector such that the gap would be much larger among public than private school children by the end of second grade.
School Location Type	Location type of the child's school (obtained from the 2009-10 CCD for public school children and 2009-10 PSS for private school children)	Polytomous variable: 0: *CITY* (territory inside an urbanized area and inside a principal city) 1: *RURAL* (census-defined rural territory) 2: *TOWN* (territory inside an urban cluster) 3: *SUBURB* (territory outside a principle city)	Based on the findings of Husain and Millimet (2009), I did not expect the gender gaps to differ significantly by locale.
School Region	Geographic region of the child's school (obtained from the Common Core of Data (2009-10 CCD) for public school children and Private School Survey (2009-10 PSS) for private school children)	Polytomous variable: 0: *NORTHEAST* (CT, ME, MA, NH, RI, VT, NJ, NY, PA) 1: *MIDWEST* (IL, IN, MI, OH, WI, IA, KS, MN, MO, NE, ND, SD) 2: *SOUTH* (DE, DC, FL, GA, MD, NC, SC, VA, WV, AL, KY, MS, TN, AR, LA, OK, TX) 4: *WEST* (AZ, CO, ID, MT, NV, NM, UT, WY, AK, CA, HI, OR, WA)	Based on the findings of Husain and Millimet (2009), I did not expect the gender gaps to differ significantly by school region.

[1]The original ECLS-K:2011-generated race/ethnicity variable had eight categories, but I collapsed "Hispanic, race-specified" and "Hispanic, no race specified" into the *HISPANIC* category and combined the "Native Hawaiian/Pacific Islander, non-Hispanic," "American Indian/Alaskan Native, non-Hispanic," and "Two or more races, non-Hispanic" categories to form the *OTHER* category.

Appendix B: Distribution of the *MONTHS* variable

Table 3.

Distribution of the child reading assessment dates over the course of the study with frequency and wave of observations presented for each month

Months since Kindergarten Entry	Frequency	Wave
0	340	1
1	5090	1
2	5910	1
3	3230	1
4	500	1
5	<10	1
6	230	2
7	5260	2
8	6860	2
9	4140	2
10	240	2
11	<10	2
18	290	4
19	4520	4
20	6550	4
21	3240	4
22	90	4
30	60	6
31	3090	6
32	7110	6
33	3140	6
34	80	6
Total	**59,980**	

Note: All sample sizes have been rounded to the nearest ten in compliance with the ECLS-K:2011 restricted-use data policies.

Appendix C. Visualization of Raw Data

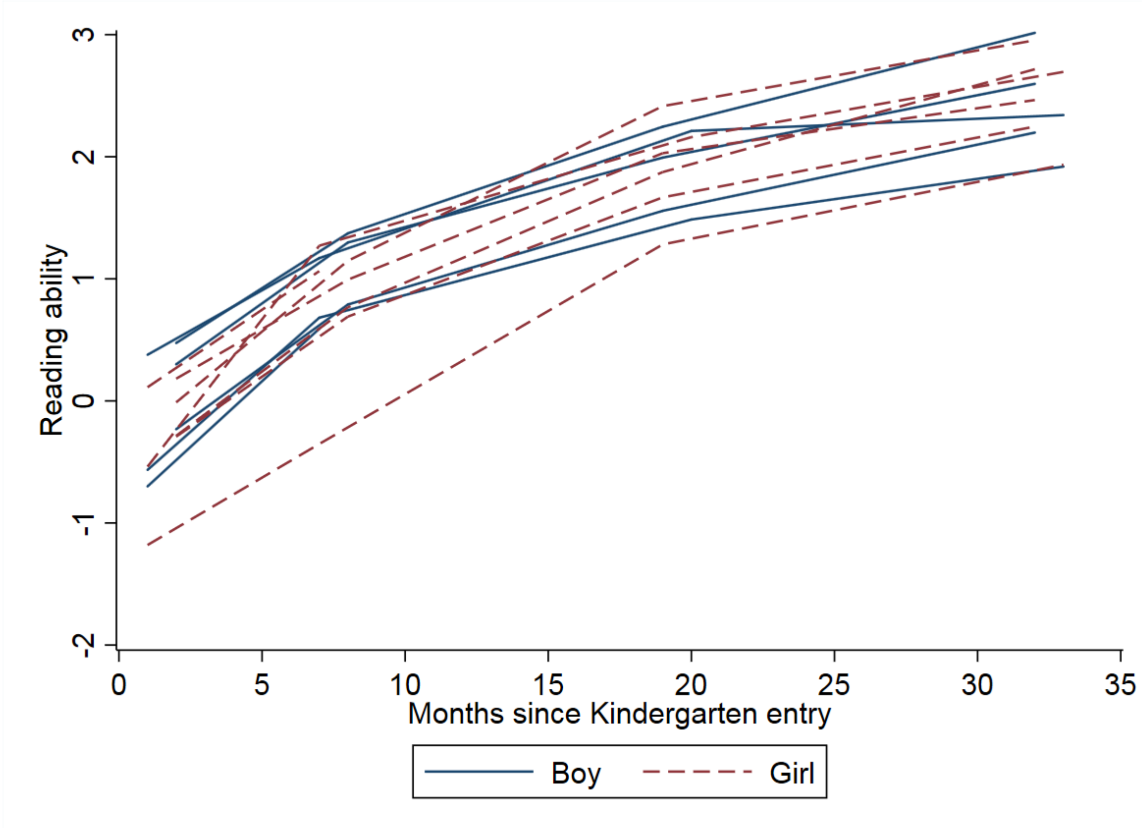

Figure 1. Students' aggregate observed reading growth trajectories (n<10 subjects per line) differentiated by gender for students in a random subset of schools (n<10)

Appendix D: Univariate and Bivariate Analyses

Table 4.

Descriptive Statistics for all Variables to be Included in the Analysis: Both Weighted and Unweighted Percentages Reported for all Categorical Variables and Weighted and Unweighted Means and Standard Deviations as well as Minimum and Maximum Values Reported for all Continuous Variables; the Number of Observations and Approximate Response Rate also Reported for Each Variable.

Variable	Weighted Values	Unweighted Values	*N*	Approx. Response Rate %
	%	%		
Gender:			7,820	100.00
Boy	51.41	51.04		
Girl	48.59	48.96		
***Race/ethnicity*:**				
White	51.94	55.75	7,820	100.00
Black	13.44	9.94		
Hispanic	24.54	22.05		
Asian	4.47	6.57		
Other	5.61	5.69		
***School Region*:**			7,820	100.00
Northeast	15.95	15.90		
Midwest	22.16	22.55		
South	37.92	38.54		
West	23.97	23.02		
***School Location Type*:**			7,690	98.34
City	32.28	28.16		
Suburb	33.64	36.73		
Town	11.01	8.29		
Rural	23.07	26.82		

Variable	Weighted Values	Unweighted Values	*N*	Approx. Response Rate %
***School Sector*:**			7,820	100.00
Public	89.10	86.87		
Private	10.90	13.13		
Frequency of Child Independent Reading:			7,270	92.97
Never	4.78	4.68		
Once or twice a week	23.18	23.15		
3 to 6 times a week	35.67	36.33		
Every day	36.36	35.84		
Frequency of Parent Reading to Child:			7,280	93.09
Never	1.26	1.11		
Once or twice a week	13.8	12.64		
3 to 6 times a week	33.79	34.04		
Every day	51.15	52.20		
Age of First Word:			6,500	83.12
0 to 9 months	20.52	20.39		
10 to 12 months	42.22	41.54		
13 to 15 months	14.18	14.67		
16 to 18 months	9.98	10.14		
19 to 24 months	7.32	7.42		
After 24 months or not at all	5.79	5.83		

Variable	**Weighted Values**	**Unweighted Values**			***N***	**Approx. Response Rate %**
Disability status:					7,000	89.51
Does not have disability	79.60	79.80				
Has a disability	20.40	20.20				
***Educational Expectations*:**					7,270	92.97
High school diploma or less	4.62	4.24				
Some postsecondary	14.66	12.89				
Bachelor's degree	47.82	49.06				
Master's degree or equivalent	15.83	16.66				
Ph.D., M.D., or other advanced degree	17.07	17.16				
Kindergarten Repeater status:					7,800	99.74
Did not repeat Kindergarten	94.95	95.53				
Repeated Kindergarten	5.05	4.47				
Language status:					7,810	99.87
English	84.18	84.14				
Non-English	15.82	15.86				
	Mean (SD)	***Mean (SD)***	***Min***	***Max***		

Variable	Weighted Values	Unweighted Values			*N*	Approx. Response Rate %
Reading Ability:						
Fall K	-0.43 (0.86)	-0.49 (0.84)	-3.09	2.98	7,770	99.36
Spring K	0.54 (0.76)	0.49 (0.76)	-2.69	2.98	7,810	99.87
Spring 1st Grade	1.67 (0.74)	1.63 (0.74)	-1.82	3.88	7,800	99.74
Spring 2nd Grade	2.27 (0.63)	2.22 (0.62)	-0.22	3.83	7,800	99.74
Socioeconomic status	0.06 (0.82)	-0.08 (0.77)	-2.33	2.60	7,810	99.87
Number of Children's Books in the Home	93.39 (142.61)	87.40 (133.63)	0	4000	7,260	92.84
Age of Kindergarten entry (in months)	66.42 (4.50)	66.23 (4.55)	39.10	88.47	7,800	99.74
Approaches to Learning	3.00 (0.67)	2.97 (0.67)	1	4	7,610	97.31
Externalizing Behaviors	1.58 (0.61)	1.55 (0.59)	1	4	7,420	94.88

Note: Sample sizes have been rounded to the nearest ten in compliance with the ECLS-K:2011 restricted-use data policies. Weighted analyses were conducted using the W6C6P_6T0 sampling and corresponding replicate weights.

Table 5.

Exploring Gender Differences in Children's Reading Ability at Each Wave as Well as in Each of the Other Predictor Variables to be Included in the Analysis Using Weighted Sample (N = 7,820)

Variable	**Boys**	**Girls**
***Reading ability*:**		
Fall K:		
Mean	-0.53	-0.46**
SD	(0.86)	(0.83)
Spring K:		
Mean	0.44	0.54***
SD	(0.78)	(0.72)
Spring 1st Grade:		
Mean	1.57	1.68***
SD	(0.75)	(0.72)
Spring 2nd Grade:		
Mean	2.16	2.28***
SD	(0.64)	(0.60)
Socioeconomic status		
Mean	-0.08	-0.07
SD	(0.78)	(0.77)
$\log_2$(*Number of children's books in the home*)		
Mean	5.70	5.70
SD	(1.54)	(1.58)
***Age of Kindergarten entry* (in months)**		
Mean	66.42**	66.04
SD	(4.57)	(4.53)
Approaches to Learning		
Mean	2.83	3.11***

SD	(0.67)	(0.64)
Externalizing Behaviors		
Mean	1.70***	1.46
SD	(0.64)	(0.54)
***Race/ethnicity* (%):**		
White	27.02*	24.92
Black	6.79	6.65
Hispanic	12.66	11.88
Asian	2.15	2.32
Other	2.80	2.81
***School Region* (%):**		
Northeast	8.27	7.68
Midwest	11.27	10.89
South	19.91*	18.01
West	11.97	12.00
***School Location Type* (%):**		
City	16.08	16.20
Suburb	17.38	16.25
Town	6.14*	4.88
Rural	11.73	11.34
***School Sector*:** Private (%)	5.44	5.47
***Frequency of Child Independent Reading* (%):**		
Never	3.80***	0.97
Once or twice a week	14.24***	8.94
3 to 6 times a week	17.98	17.70
Every day	15.07	21.29***
***Frequency of Parent Reading to Child* (%):**		
Never	0.66	0.61
Once or twice a week	7.70*	6.10

3 to 6 times a week	17.80*	15.98
Every day	24.92	26.23
***Age of First Word* (%):**		
0 to 9 months	9.46	11.06*
10 to 12 months	20.78	21.44
13 to 15 months	7.87**	6.31
16 to 18 months	5.22	4.76
19 to 24 months	4.21**	3.11
After 24 months or not at all	3.95***	1.84
***Disability status*:** Disability (%)	12.91***	7.49
***Educational Expectations* (%):**		
High school diploma or less	2.66*	1.95
Some postsecondary	8.31***	6.35
Bachelor's degree	24.07	23.74
Master's degree or equivalent	7.95	7.89
Ph.D., M.D., or other advanced degree	8.01	9.06
Kindergarten Repeater status: Yes(%)	3.03**	2.02
Language status: Non-English (%)	8.30	7.52

Note: Comparisons for continuous variables were conducted with independent samples t-tests assuming equal variances, and comparisons for categorical variables were conducted using Pearson's χ^2 *tests of independence corrected for the survey design using the Rao and Scott (1984) second-order correction followed by post-hoc adjusted Wald tests. All analyses were conducted using the W6C6P_6T0 sampling and corresponding replicate weights.*

* $p < 0.05$, indicated on the higher of the two numbers

** $p < 0.01$, indicated on the higher of the two numbers

*** $p < 0.001$, indicated on the higher of the two numbers

Table 6.

Exploration of Bivariate Relationships Among Continuous Variables using Pearson Correlation Coefficients with Unweighted Sample Data (n=6,850)

	Fall K READ	*Spring K READ*	*Spring 1st READ*	*Spring 2nd READ*	*SES*	Log_2NUMBOOKS	*AGEENT*	*APPROACH*
Fall K READ	-							
Spring K READ	0.80	-						
Spring 1st READ	0.71	0.81	-					
Spring 2nd READ	0.65	0.73	0.85	-				
SES	0.43	0.38	0.40	0.42	-			
Log_2NUMBOOKS	0.32	0.31	0.32	0.33	0.47	-		
AGEENT	0.14	0.14	0.11	0.11	0.04	0.06	-	
APPROACH	0.38	0.40	0.42	0.40	0.19	0.15	0.11	-
EXTERN	-0.15	-0.21	-0.22	-0.21	-0.11	-0.06	-0.02	-0.58

Note: All relationships statistically significant with $p < 0.001$ except for correlation between age of K entry and externalizing behaviors, which is not statistically significant at the $\alpha = 0.05$ level and the correlation between age of K entry and SES, which is statistically significant with $p = 0.002$.

Appendix E. Taxonomy of Selected Models

Table 7.

Taxonomy of Selected Hierarchical Linear Models exploring Gender Differences in Reading Development from Kindergarten Through Second Grade in U.S. Children. Unstandardized estimated regression coefficients and (robust standard errors) are presented below. All analyses were conducted using the W6C6P_6T0 sampling weight.

	(1) Model A b/se	(2) Model B b/se	(3) Model C b/se	(4) Model D b/se	(5) Model E b/se	(6) Model F b/se	(7) Model G b/se	(8) Model H b/se	(9) Model I b/se
Fixed Effects									
INTERCEPT	-0.79391***	-3.02000***	-2.99749***	-3.13897***	-3.12388***	-2.99805***	-2.82356***	-3.75497***	-3.79719***
	(0.02155)	(0.18006)	(0.18231)	(0.18041)	(0.18105)	(0.17941)	(0.18854)	(0.22827)	(0.22865)
MONTH	0.16966***	0.20570***	0.20743***	0.20481***	0.20411***	0.20395***	0.21648***	0.22482***	0.22481***
	(0.00160)	(0.00516)	(0.00517)	(0.00584)	(0.00590)	(0.00615)	(0.00707)	(0.00686)	(0.00686)
MONTH^2	-0.00241***	-0.00248***	-0.00251***	-0.00247***	-0.00247***	-0.00248***	-0.00284***	-0.00284***	-0.00284***
	(0.00004)	(0.00004)	(0.00005)	(0.00006)	(0.00006)	(0.00007)	(0.00010)	(0.00009)	(0.00009)
GIRL	0.05372*	0.06989**	0.05921**	0.03982	0.02289	-0.00985	-0.02138	-0.07889**	0.00767
	(0.02156)	(0.02153)	(0.02121)	(0.02392)	(0.02408)	(0.02471)	(0.02457)	(0.02462)	(0.05314)
GIRLxMONTH	0.00519**	0.00454**	0.00477**	0.00522**	0.00534**	0.00548**	0.00373	0.00457*	0.00457*
	(0.00172)	(0.00171)	(0.00171)	(0.00187)	(0.00190)	(0.00197)	(0.00198)	(0.00197)	(0.00197)
GIRLxMONTH^2	-0.00012**	-0.00011**	-0.00012**	-0.00014**	-0.00014**	-0.00013*	-0.00009	-0.00010	-0.00010
	(0.00004)	(0.00004)	(0.00004)	(0.00005)	(0.00005)	(0.00005)	(0.00005)	(0.00005)	(0.00005)
BLACK		-0.03803	-0.06085	-0.02067	-0.02337	-0.05390	-0.03265	-0.01729	-0.01658
		(0.04738)	(0.04588)	(0.05373)	(0.05417)	(0.05380)	(0.05742)	(0.05353)	(0.05356)
HISPANIC		-0.09634**	-0.11711***	-0.12434**	-0.12160**	-0.12246**	-0.12703**	-0.11769**	-0.11419**
		(0.03353)	(0.03360)	(0.04125)	(0.04170)	(0.04192)	(0.04373)	(0.04235)	(0.04252)
ASIAN		0.23924***	0.20704***	0.28523***	0.27282***	0.27670***	0.25552***	0.24120***	0.24077***
		(0.06227)	(0.06211)	(0.06728)	(0.06623)	(0.06415)	(0.06966)	(0.06947)	(0.06953)
OTHER		0.05844	0.03502	0.04828	0.03549	0.02368	0.03530	0.03431	0.03468
		(0.05282)	(0.05273)	(0.05502)	(0.05461)	(0.05339)	(0.05425)	(0.04980)	(0.04963)
BLACKxMONTH		0.00005	0.00026	-0.00138	-0.00151	-0.00120	-0.00155	-0.00165	-0.00165
		(0.00124)	(0.00123)	(0.00134)	(0.00135)	(0.00133)	(0.00139)	(0.00137)	(0.00137)
HISPANICxMONTH		0.00300**	0.00320**	0.00357**	0.00344**	0.00336*	0.00366**	0.00353**	0.00353**
		(0.00112)	(0.00114)	(0.00131)	(0.00130)	(0.00131)	(0.00135)	(0.00134)	(0.00134)
ASIANxMONTH		-0.00458**	-0.00402*	-0.00559**	-0.00520**	-0.00537**	-0.00564**	-0.00552**	-0.00552**
		(0.00169)	(0.00170)	(0.00194)	(0.00191)	(0.00187)	(0.00201)	(0.00201)	(0.00201)
OTHERxMONTH		-0.00051	-0.00025	-0.00072	-0.00070	-0.00052	-0.00061	-0.00059	-0.00059
		(0.00137)	(0.00138)	(0.00145)	(0.00146)	(0.00146)	(0.00147)	(0.00145)	(0.00144)
SES		0.31785***	0.29305***	0.29187***	0.28660***	0.28650***	0.28974***	0.24278***	0.24371***
		(0.01790)	(0.01801)	(0.01912)	(0.01928)	(0.01892)	(0.02007)	(0.01982)	(0.01981)
SESxMONTH		-0.00695***	-0.00671***	-0.00616***	-0.00617***	-0.00613***	-0.00737***	-0.00681***	-0.00681***
		(0.00140)	(0.00146)	(0.00150)	(0.00151)	(0.00151)	(0.00154)	(0.00154)	(0.00154)
SESxMONTH^2		0.00012**	0.00012**	0.00010*	0.00010*	0.00010**	0.00013***	0.00013***	0.00013***
		(0.00004)	(0.00004)	(0.00004)	(0.00004)	(0.00004)	(0.00004)	(0.00004)	(0.00004)
SUBURB		0.02154	-0.02489	0.16976**	0.14337*	0.10794*	0.09727	-0.02473	-0.02281
		(0.13194)	(0.13264)	(0.06481)	(0.06790)	(0.05495)	(0.06286)	(0.21652)	(0.21596)
TOWN		0.38194***	0.27418***	0.55861***	0.43589***	0.50888***	0.42618***	0.42272**	0.40041**
		(0.04902)	(0.07200)	(0.07429)	(0.06812)	(0.06541)	(0.06820)	(0.13110)	(0.13060)
RURAL		-0.15281	-0.20638	-0.09192	-0.07273	-0.11832*	-0.11719*	-0.20086	-0.19292
		(0.09382)	(0.11430)	(0.05356)	(0.05199)	(0.04961)	(0.05177)	(0.12388)	(0.12356)
SUBURBxMONTH		0.00704***	0.00772***	0.00224	0.00232	0.00253	0.00232	0.00278*	0.00278*
		(0.00145)	(0.00166)	(0.00216)	(0.00223)	(0.00273)	(0.00221)	(0.00114)	(0.00115)
TOWNxMONTH		-0.02041***	-0.02035***	-0.03659***	-0.03689***	-0.03820***	-0.03649***	-0.03643***	-0.03645***
		(0.00305)	(0.00331)	(0.00441)	(0.00433)	(0.00444)	(0.00436)	(0.00423)	(0.00421)
RURALxMONTH		0.00984	0.00932	0.01458***	0.01474***	0.01525***	0.01455***	0.01488***	0.01489***

	(0.00702)	(0.00724)	(0.00246)	(0.00245)	(0.00261)	(0.00247)	(0.00229)	(0.00229)
MIDWEST	0.11355**	0.13246**	0.15448***	0.15611***	0.14654***	0.16484***	0.11477*	0.17921**
	(0.04386)	(0.04454)	(0.04223)	(0.04184)	(0.04088)	(0.04162)	(0.04562)	(0.05471)
SOUTH	0.08925*	0.09887*	0.08036*	0.07262	0.05456	0.07425	0.02520	0.06100
	(0.04415)	(0.04408)	(0.03950)	(0.03920)	(0.03860)	(0.04008)	(0.04349)	(0.05099)
WEST	0.07313	0.07287	0.09384*	0.07692	0.05713	0.06677	0.02196	0.08295
	(0.04591)	(0.04568)	(0.04408)	(0.04375)	(0.04293)	(0.04455)	(0.05601)	(0.06242)
NON_ENG	-0.10465*	-0.13977**	-0.12583*	-0.14615**	-0.13414**	-0.15682**	-0.18781***	-0.18930***
	(0.04403)	(0.04393)	(0.04986)	(0.04916)	(0.04907)	(0.05100)	(0.04916)	(0.04925)
NON_ENGxMONTH	0.00404***	0.00446***	0.00436**	0.00445**	0.00408**	0.00408**	0.00423**	0.00423**
	(0.00121)	(0.00123)	(0.00137)	(0.00137)	(0.00138)	(0.00140)	(0.00140)	(0.00140)
KREPEAT	0.52864***	0.54551***	0.54057***	0.56203***	0.57726***	0.59147***	0.53086***	0.53066***
	(0.05610)	(0.05612)	(0.05992)	(0.06073)	(0.05960)	(0.06177)	(0.05969)	(0.05967)
KREPEATxMONTH	-0.04498***	-0.04527***	-0.04504***	-0.04505***	-0.04505***	-0.04573***	-0.04537***	-0.04538***
	(0.00471)	(0.00473)	(0.00530)	(0.00533)	(0.00531)	(0.00538)	(0.00538)	(0.00538)
KREPEATxMONTH^2	0.00075***	0.00075***	0.00078***	0.00078***	0.00077***	0.00079***	0.00080***	0.00080***
	(0.00012)	(0.00012)	(0.00013)	(0.00014)	(0.00014)	(0.00014)	(0.00014)	(0.00014)
AGEENT	0.02653***	0.02746***	0.02884***	0.02940***	0.02847***	0.02827***	0.01978***	0.01987***
	(0.00239)	(0.00242)	(0.00250)	(0.00248)	(0.00246)	(0.00260)	(0.00252)	(0.00252)
AGEENTxMONTH	-0.00057***	-0.00058***	-0.00055***	-0.00054***	-0.00053***	-0.00054***	-0.00048***	-0.00048***
	(0.00006)	(0.00007)	(0.00007)	(0.00007)	(0.00007)	(0.00008)	(0.00008)	(0.00008)
PRIVATE	-0.03240	-0.03737	-0.21519*	-0.24813**	-0.32012***	-0.27948***	-0.36616**	-0.37854**
	(0.03378)	(0.04080)	(0.08522)	(0.08520)	(0.07917)	(0.05845)	(0.12668)	(0.14058)
PRIVATExMONTH	-0.00435	-0.00389	0.00473	0.00469	0.00602	0.00626	0.00700	0.00699
	(0.00964)	(0.00936)	(0.01044)	(0.01043)	(0.01064)	(0.01049)	(0.01076)	(0.01076)
PRIVATExMONTH^2	0.00034**	0.00032**	0.00032**	0.00032**	0.00032**	0.00027*	0.00027*	0.00027*
	(0.00011)	(0.00011)	(0.00012)	(0.00012)	(0.00012)	(0.00012)	(0.00012)	(0.00012)
Log_2NUMBOOKS	0.07271***	0.06524***	0.06308***	0.06514***	0.05888***	0.06010***	0.04847***	0.04827***
	(0.00653)	(0.00657)	(0.00744)	(0.00730)	(0.00740)	(0.00751)	(0.00702)	(0.00700)
P_READ_NEVER		0.10208	0.12167	0.12232	0.16497			
		(0.08504)	(0.10677)	(0.10184)	(0.10526)			
P_READ_1TO2		-0.04985	-0.05917	-0.06791*	-0.05388			
		(0.02851)	(0.03257)	(0.03265)	(0.03330)			
P_READ_EVERYDAY		0.01619	0.02965	0.03517	0.00609			
		(0.01822)	(0.01905)	(0.01900)	(0.01940)			
HSORLESS		-0.18121***	-0.19262***	-0.17222**	-0.16287**	-0.14685*	-0.05474	-0.05315
		(0.05065)	(0.05832)	(0.05693)	(0.05609)	(0.05880)	(0.04963)	(0.04950)
POSTSEC		-0.10964***	-0.05830*	-0.04636	-0.04713	-0.04816	-0.03555	-0.03562
		(0.03291)	(0.02639)	(0.02613)	(0.02593)	(0.02636)	(0.02488)	(0.02477)
MASTERS		0.08323**	0.06109**	0.06760**	0.06557**	0.07219**	0.07105**	0.07020**
		(0.02923)	(0.02364)	(0.02360)	(0.02359)	(0.02350)	(0.02322)	(0.02314)
MDPHD		0.09110**	0.07021*	0.07822**	0.06303*	0.06436*	0.05636*	0.05674*
		(0.03156)	(0.02735)	(0.02659)	(0.02629)	(0.02658)	(0.02539)	(0.02539)
HSORLESSxMONTH		-0.00359						
		(0.00487)						
POSTSECxMONTH		0.00231						
		(0.00280)						
MASTERSxMONTH		-0.00467						
		(0.00262)						
MDPHDxMONTH		-0.00511*						
		(0.00261)						
HSORLESSxMONTH^2		0.00011						
		(0.00012)						
POSTSECxMONTH^2		-0.00004						
		(0.00008)						
MASTERSxMONTH^2		0.00010						
		(0.00007)						
MDPHDxMONTH^2		0.00011						
		(0.00007)						
A01to9			-0.01469	-0.01384	-0.01548	-0.00634	-0.00928	-0.00886
			(0.02890)	(0.02895)	(0.02875)	(0.02942)	(0.02146)	(0.02148)
A13to15			-0.01768	-0.01914	-0.01871	-0.02776	-0.00915	-0.00889

	(0.03426)	(0.03402)	(0.03358)	(0.03486)	(0.02292)	(0.02285)
A16to18	-0.09428**	-0.08446*	-0.08364*	-0.07927*	-0.05504*	-0.05634*
	(0.03574)	(0.03507)	(0.03531)	(0.03626)	(0.02639)	(0.02629)
A19to24	-0.07588	-0.06695	-0.05874	-0.07818	-0.04789	-0.04934
	(0.04034)	(0.03998)	(0.03909)	(0.04166)	(0.03285)	(0.03306)
A25orNever	-0.15692**	-0.09027	-0.08893	-0.08071	-0.12125**	-0.12335***
	(0.05336)	(0.05279)	(0.05313)	(0.05620)	(0.03695)	(0.03703)
A01to9xMONTH	-0.00104	-0.00081	-0.00076	-0.00013		
	(0.00265)	(0.00265)	(0.00266)	(0.00271)		
A13to15xMONTH	0.00008	0.00036	0.00044	0.00092		
	(0.00287)	(0.00288)	(0.00287)	(0.00298)		
A16to18xMONTH	-0.00118	-0.00102	-0.00093	0.00167		
	(0.00324)	(0.00325)	(0.00325)	(0.00311)		
A19to24xMONTH	0.00016	0.00079	0.00080	0.00079		
	(0.00375)	(0.00374)	(0.00373)	(0.00379)		
A25orNeverxMONTH	-0.01136*	-0.01099*	-0.01112*	-0.00905		
	(0.00457)	(0.00457)	(0.00458)	(0.00477)		
A01to9xMONTH^2	0.00000	-0.00000	-0.00000	-0.00002		
	(0.00007)	(0.00007)	(0.00007)	(0.00007)		
A13to15xMONTH^2	0.00001	0.00000	-0.00000	-0.00001		
	(0.00008)	(0.00008)	(0.00008)	(0.00008)		
A16to18xMONTH^2	0.00003	0.00002	0.00002	-0.00005		
	(0.00008)	(0.00008)	(0.00008)	(0.00008)		
A19to24xMONTH^2	-0.00001	-0.00003	-0.00003	-0.00002		
	(0.00010)	(0.00010)	(0.00010)	(0.00010)		
A25orNeverxMONTH^2	0.00028*	0.00027*	0.00027*	0.00022		
	(0.00012)	(0.00012)	(0.00012)	(0.00012)		
DISABLILITY		-0.20156***	-0.20344***	-0.17912***	-0.13980***	-0.14034***
		(0.02955)	(0.02313)	(0.02329)	(0.02206)	(0.02209)
DISABILITYxMONTH		-0.00004				
		(0.00277)				
DISABILITYxMONTH^2		-0.00001				
		(0.00007)				
RNEVER			-0.20767***	-0.13357**	-0.09966*	-0.09947*
			(0.04706)	(0.04905)	(0.04311)	(0.04281)
R1TO2			-0.01886	-0.02151	-0.02904	-0.03049
			(0.02878)	(0.02421)	(0.02258)	(0.02247)
REVERY			0.16640***	0.11570***	0.09569***	0.09588***
			(0.02822)	(0.01972)	(0.01834)	(0.01852)
RNEVERxMONTH			-0.00063			
			(0.00422)			
R1TO2xMONTH			-0.00064			
			(0.00266)			
REVERYxMONTH			-0.00213			
			(0.00227)			
RNEVERxMONTH^2			0.00011			
			(0.00012)			
R1TO2xMONTH^2			0.00004			
			(0.00007)			
REVERYxMONTH^2			-0.00001			
			(0.00006)			
EXTERN				-0.10608***	0.14863***	0.14810***
				(0.02362)	(0.02649)	(0.02654)
EXTERNxMONTH				-0.00757***	-0.00969***	-0.00969***
				(0.00197)	(0.00201)	(0.00201)
EXTERNxMONTH^2				0.00023***	0.00023***	0.00023***
				(0.00005)	(0.00005)	(0.00005)
APPROACH					0.43297***	0.43218***
					(0.02311)	(0.02321)
APPROACHxMONTH					-0.00332***	-0.00332***
					(0.00067)	(0.00067)
GIRLxMIDWEST						-0.12895*

									(0.05773)
GIRLxSOUTH									-0.07550
									(0.05609)
GIRLxWEST									-0.12548*
									(0.05705)
Random effects									
Variance: School-level	0.00012	0.000199	0.00020	0.00033	0.00034	0.00035	0.00033	0.00034	0.00034
random slope (MONTH)	(0.00001)	(0.00001)	(0.00001)	(0.00002)	(0.00002)	(0.00002)	(0.00002)	(0.00002)	(0.00002)
Variance: School-level	0.23787	0.18943	0.18152	0.21006	0.19544	0.20928	0.20160	0.20994	0.20899
random intercept	(0.01408)	(0.01291)	(0.01231)	(0.01314)	(0.01223)	(0.01308)	(0.01242)	(0.01220)	(0.01214)
School-level covariance	-0.00344	-0.00425	-0.00405	-0.00594	-0.00552	-0.00620	-0.00562	-0.00609	-0.00603
	(0.00028)	(0.00035)	(0.00034)	(0.00048)	(0.00046)	(0.00050)	(0.00045)	(0.00045)	(0.00045)
Variance: Child-level	0.00012	0.00010	0.00010	0.00009	0.00009	0.00009	0.00009	0.00009	0.00009
random slope (MONTH)	(0.00001)	(0.00001)	(0.00001)	(0.00001)	(0.00001)	(0.00001)	(0.00001)	(0.00001)	(0.00001)
Variance: Child-level	0.39901	0.32571	0.31781	0.30368	0.29786	0.29110	0.28654	0.24883	0.24842
random intercept	(0.01171)	(0.01058)	(0.01006)	(0.01408)	(0.01048)	(0.01015)	(0.01059)	(0.00970)	(0.00971)
Child-level covariance	-0.00442	-0.00347	-0.00340	-0.00325	-0.00322	-0.00311	-0.00318	-0.00290	-0.00291
	(0.00027)	(0.00025)	(0.00024)	(0.00027)	(0.00027)	(0.00027)	(0.00028)	(0.00027)	(0.00027)
Variance: Occasion-level	0.10036	0.09873	0.09870	0.09761	0.09766	0.09765	0.09668	0.09669	0.09669
residuals	(0.00226)	(0.00230)	(0.00230)	(0.00258)	(0.00259)	(0.00259)	(0.00261)	(0.00260)	(0.00260)
Num. occasions	31,070	28,280	28,200	23,480	23,410	23,390	22,280	22,270	22,270
Num. children	7,780	7,080	7,060	5,880	5,860	5,860	5,580	5,580	5,580
Num. schools	820	790	790	780	780	780	760	760	760
Log pseudolikelihood	-10,137,398	-8,697,051	-8,626,895	-6,951,344	-6,893,946	-6,856,761	-6,465,636	-6,274,634	-6,271,584
Wald χ^2	55,020***	41,650***	44,263***	31,314***	32,862***	32,405***	34,153***	33,735***	34,210***
df	5	36	51	58	61	68	62	54	57

Note: All sample sizes have been rounded to the nearest ten in compliance with the ECLS-K:2011 restricted-use data policies.
All random effect variances and covariances are statistically significant with p<0.001.
For fixed effect coefficients: *** *p<0.001,* ***p<0.01,* **p<0.05.*

Appendix F. Visualization of Fitted Models

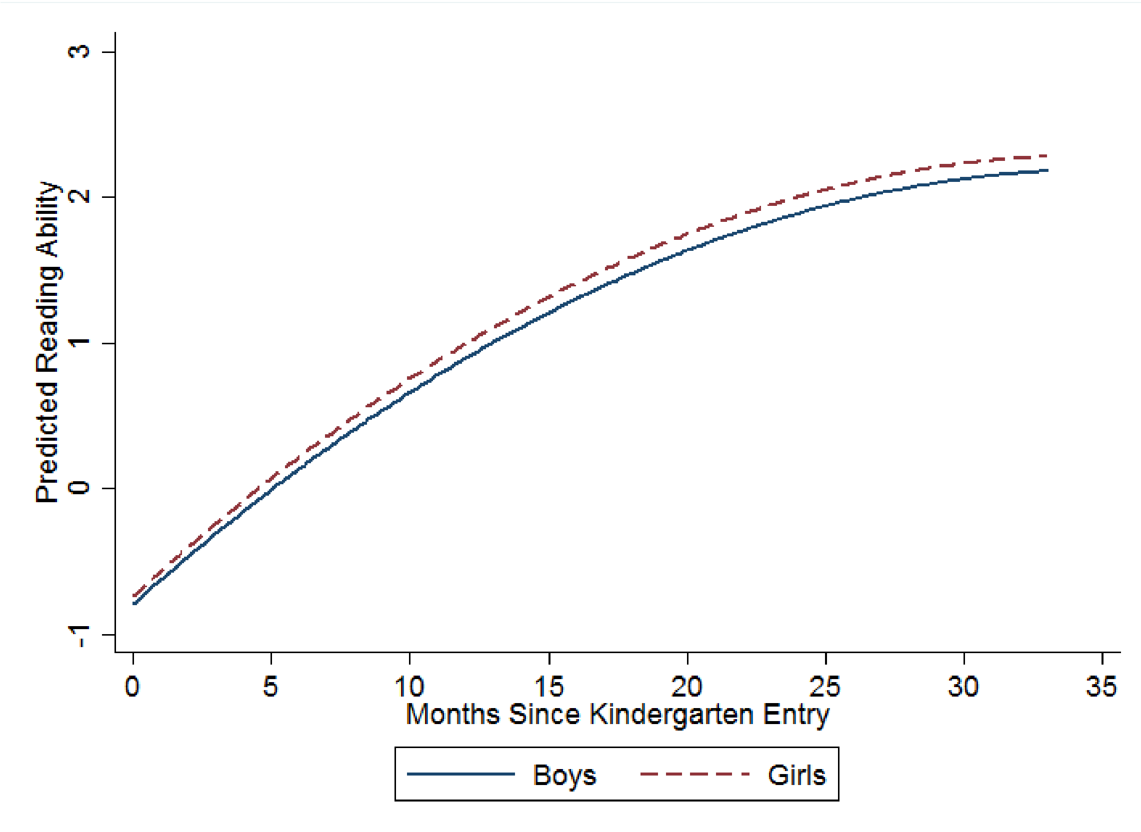

Figure 2. Predicted mean reading growth trajectory from Kindergarten through 2nd grade for girls and boys (Model A)

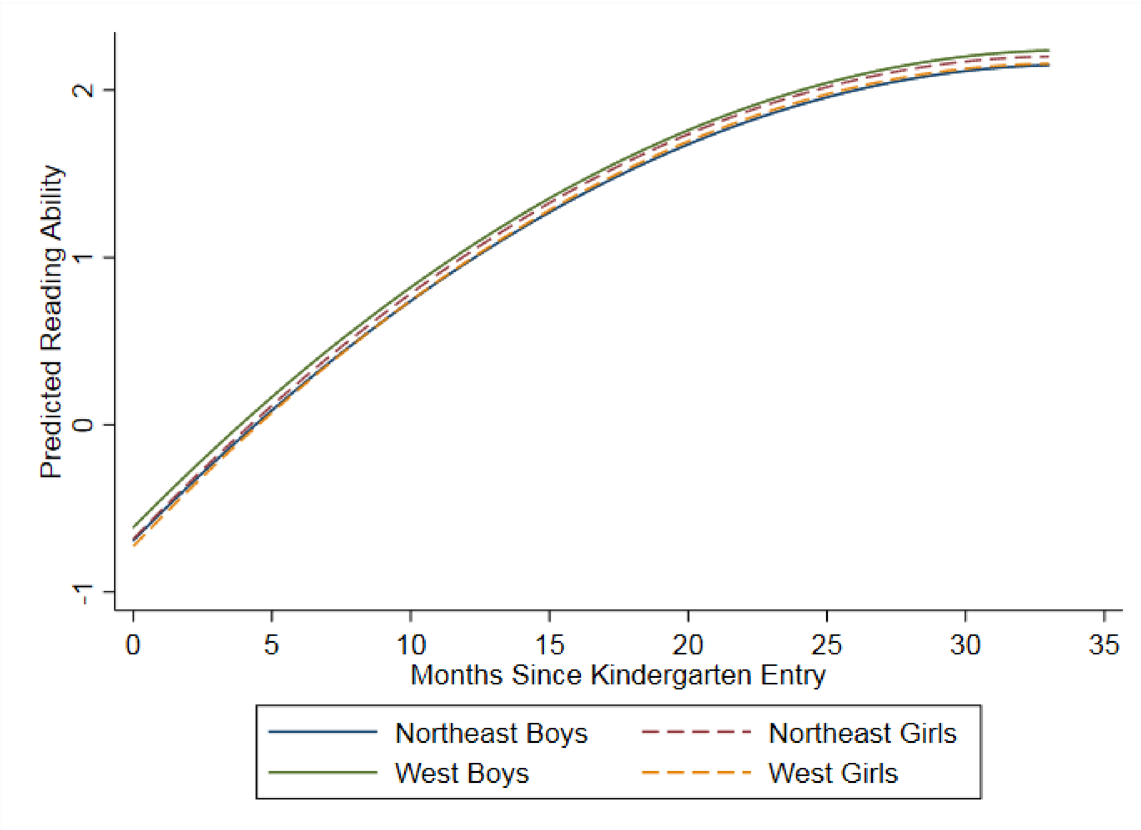

Figure 3. Predicted mean reading growth trajectory from Kindergarten through 2nd grade for girls and boys in the Northeast versus West with all other categorical variables set at the reference categories (white students of average SES, primary home language is English, did not repeat Kindergarten, attended public schools, had a parent who expected them to earn a Bachelor's degree, began speaking at 10-12 months of age, read independently 3-5 times per week, did not have a disability) and continuous variables set at their predicted population means (Model I)

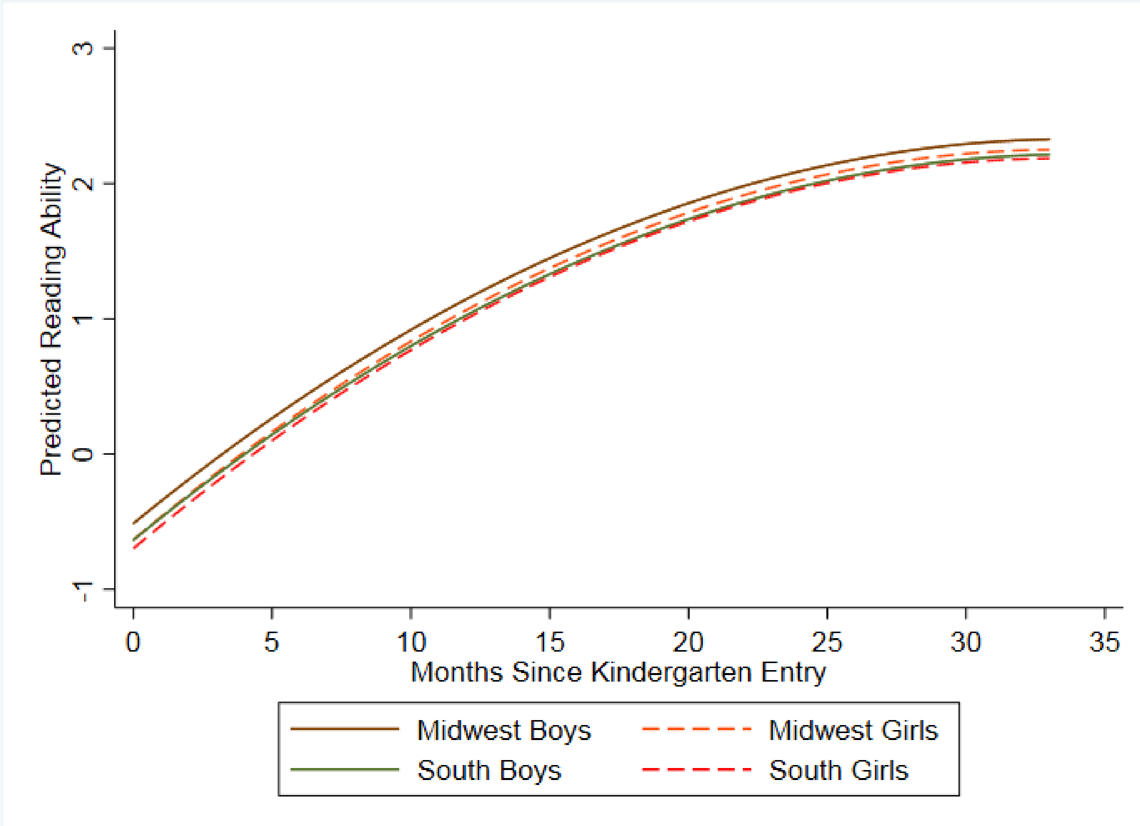

Figure 4. Predicted mean reading growth trajectory from Kindergarten through 2nd grade for girls and boys in the South versus Northwest with all other categorical variables set at the reference categories (white students of average SES, primary home language is English, did not repeat Kindergarten, attended public schools, had a parent who expected them to earn a Bachelor's degree, began speaking at 10-12 months of age, read independently 3-5 times per week, did not have a disability) and continuous variables set at their predicted population means (Model I)

Appendix G. Regression Diagnostics and Sensitivity Analyses

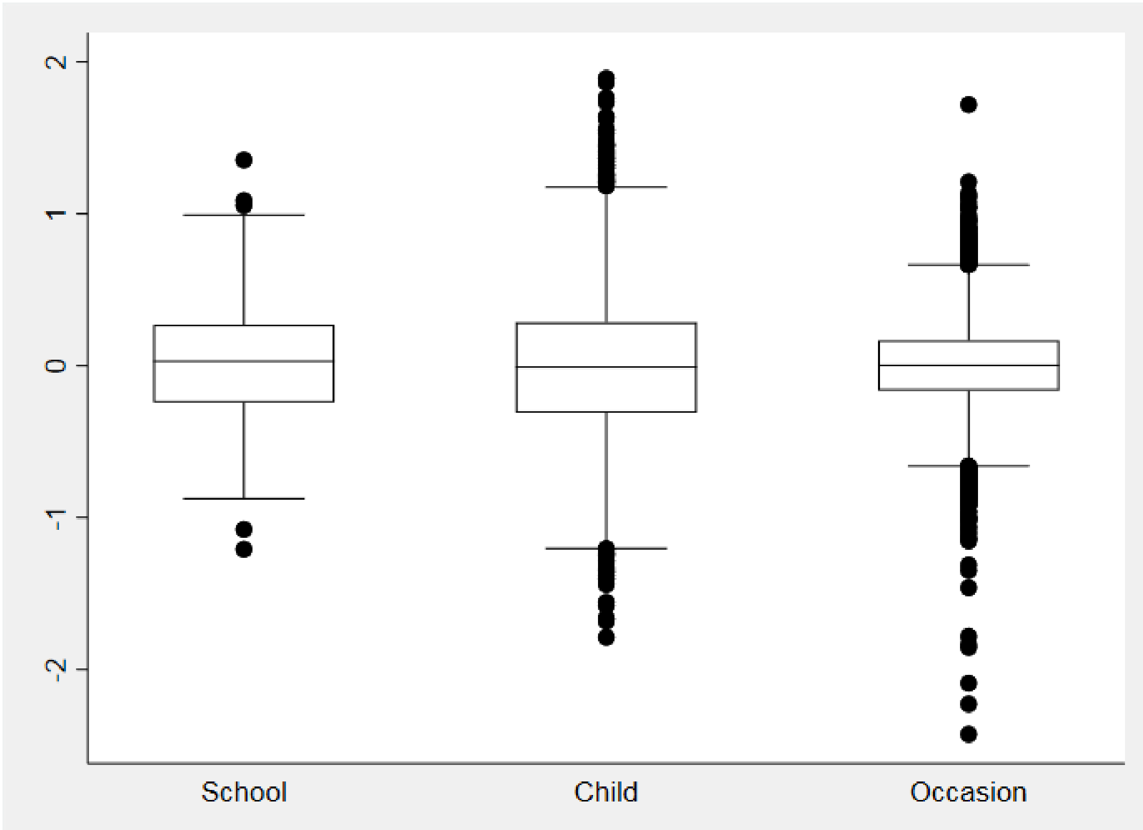

Figure 5. Box plots of empirical Bayes predictions for random intercepts at the school level and child level as well as residuals at the occasion level from the final model (Model I)

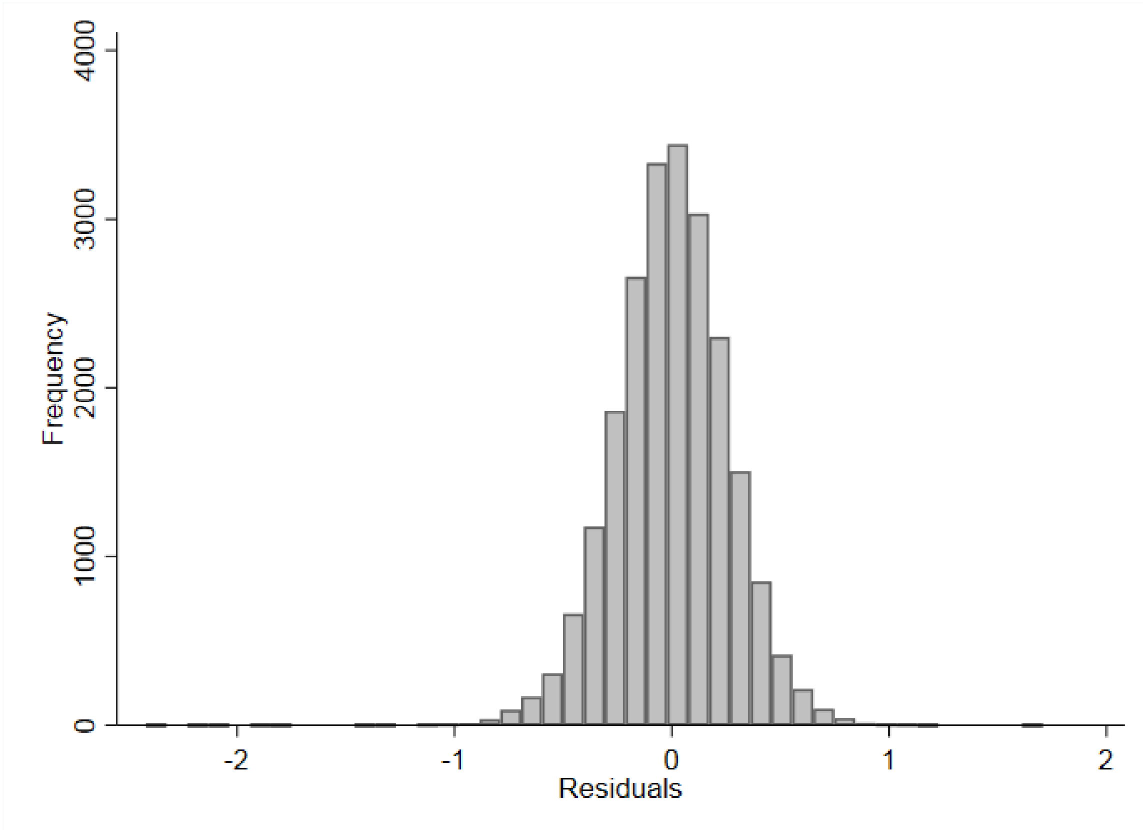

Figure 6. Histogram of Level-1 residuals from the final model (Model I)

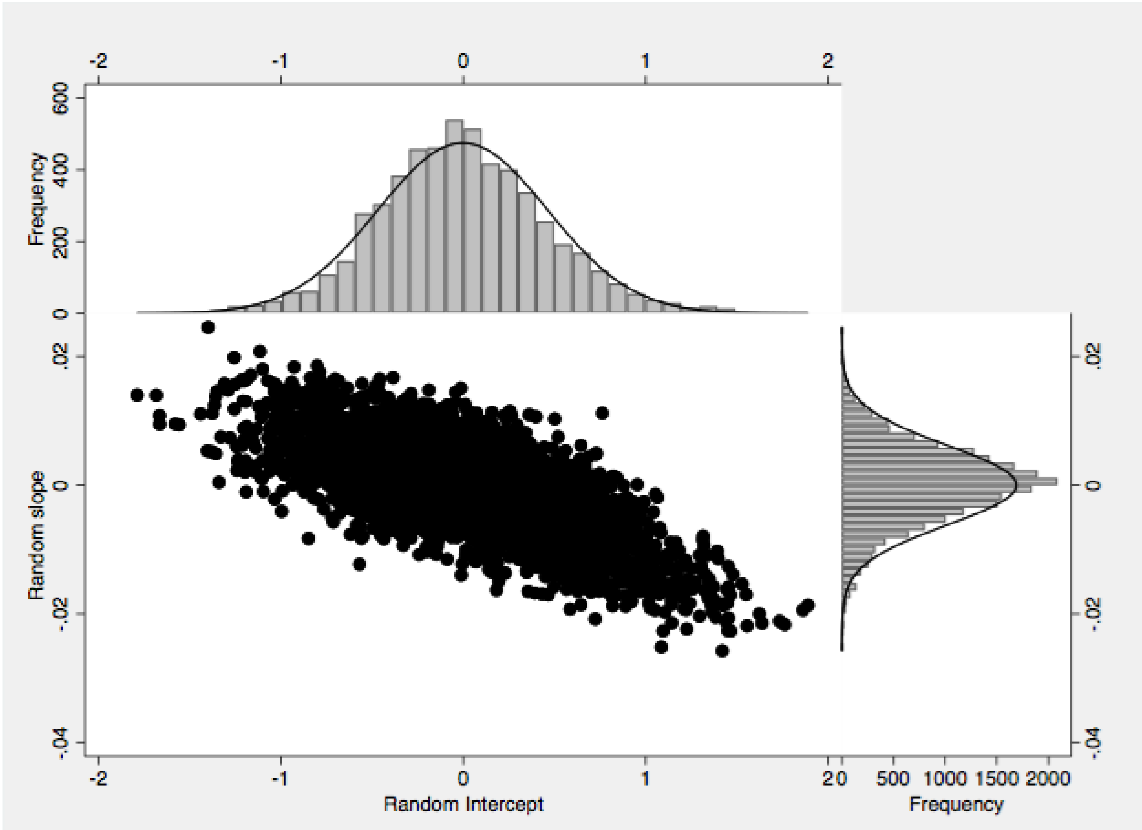

Figure 7. Bivariate and univariate (with Normal curve overlaid) distributions of empirical Bayes predictions for random intercepts and random slopes at the child level from the final model (Model I)

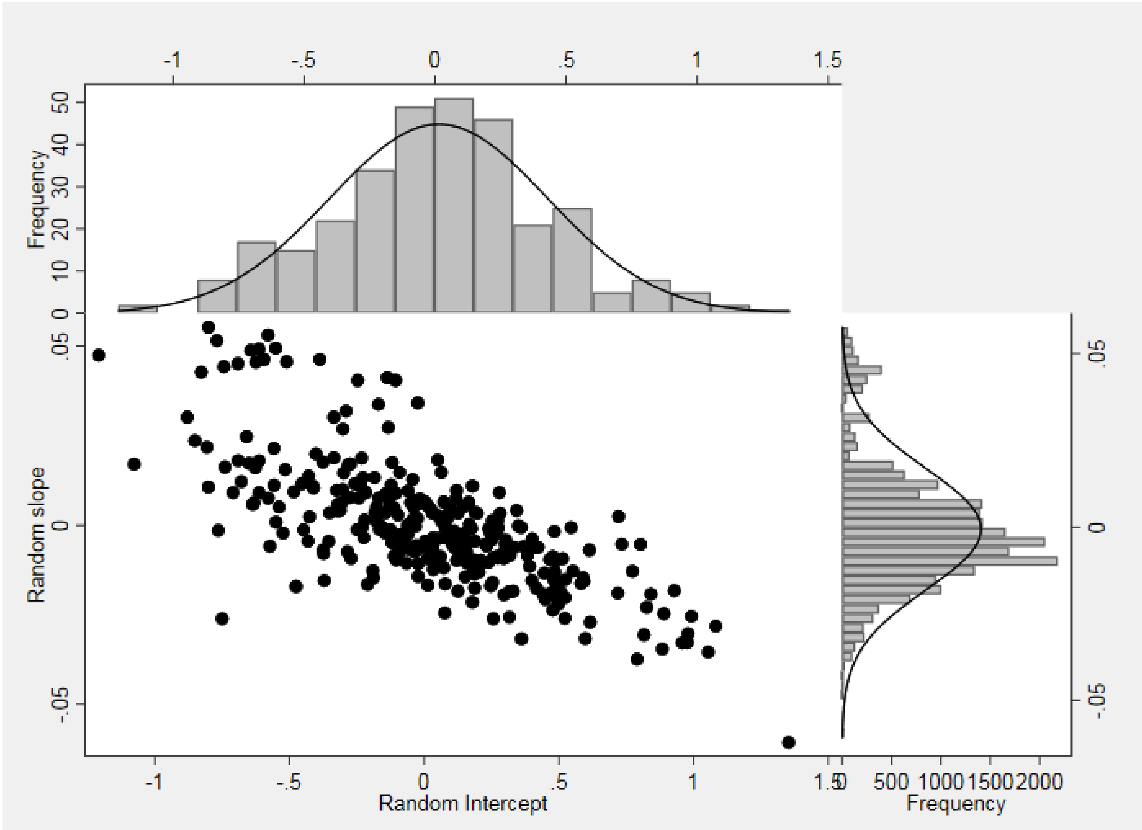

Figure 8. Bivariate and univariate (with Normal curve overlaid) distributions of empirical Bayes predictions for random intercepts and random slopes at the school level from the final model (Model I)

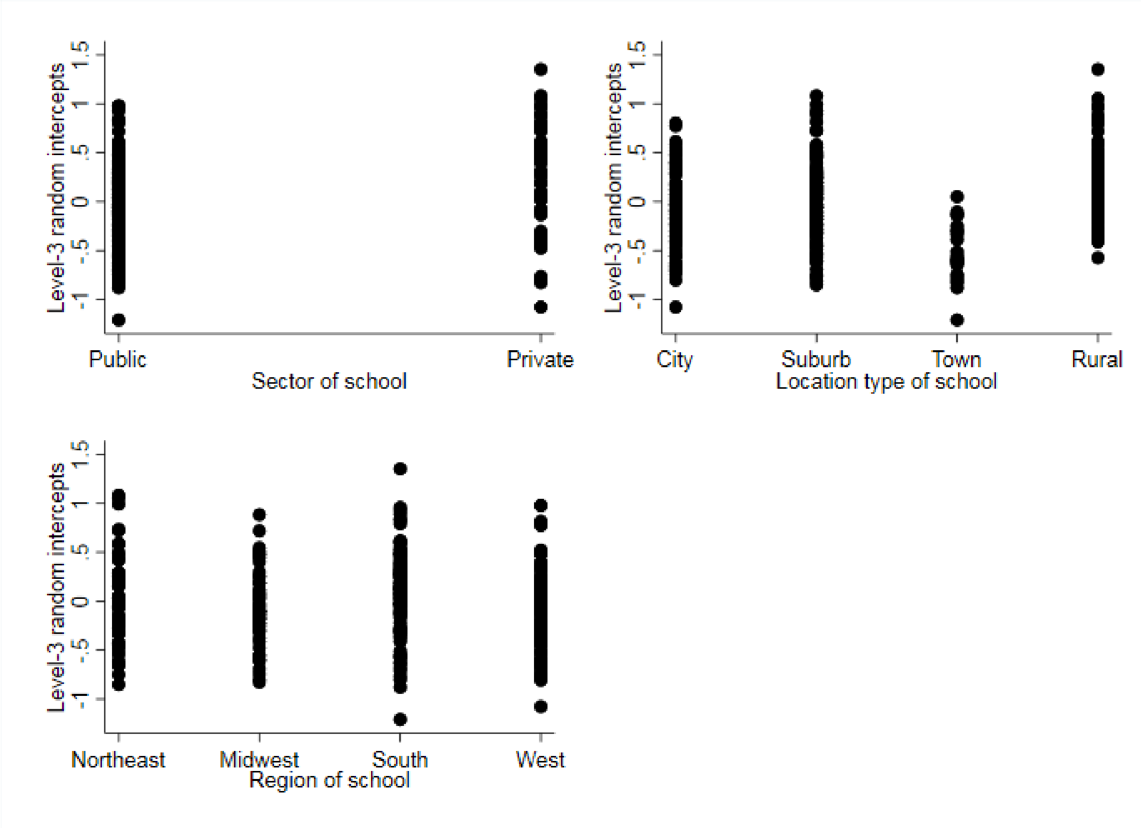

Figure 9. Empirical Bayes predictions for random intercepts at the school level from the final model (Model I) plotted on each of the school-level predictors (location type, sector, and region) to assess for homoscedasticity at level-3.

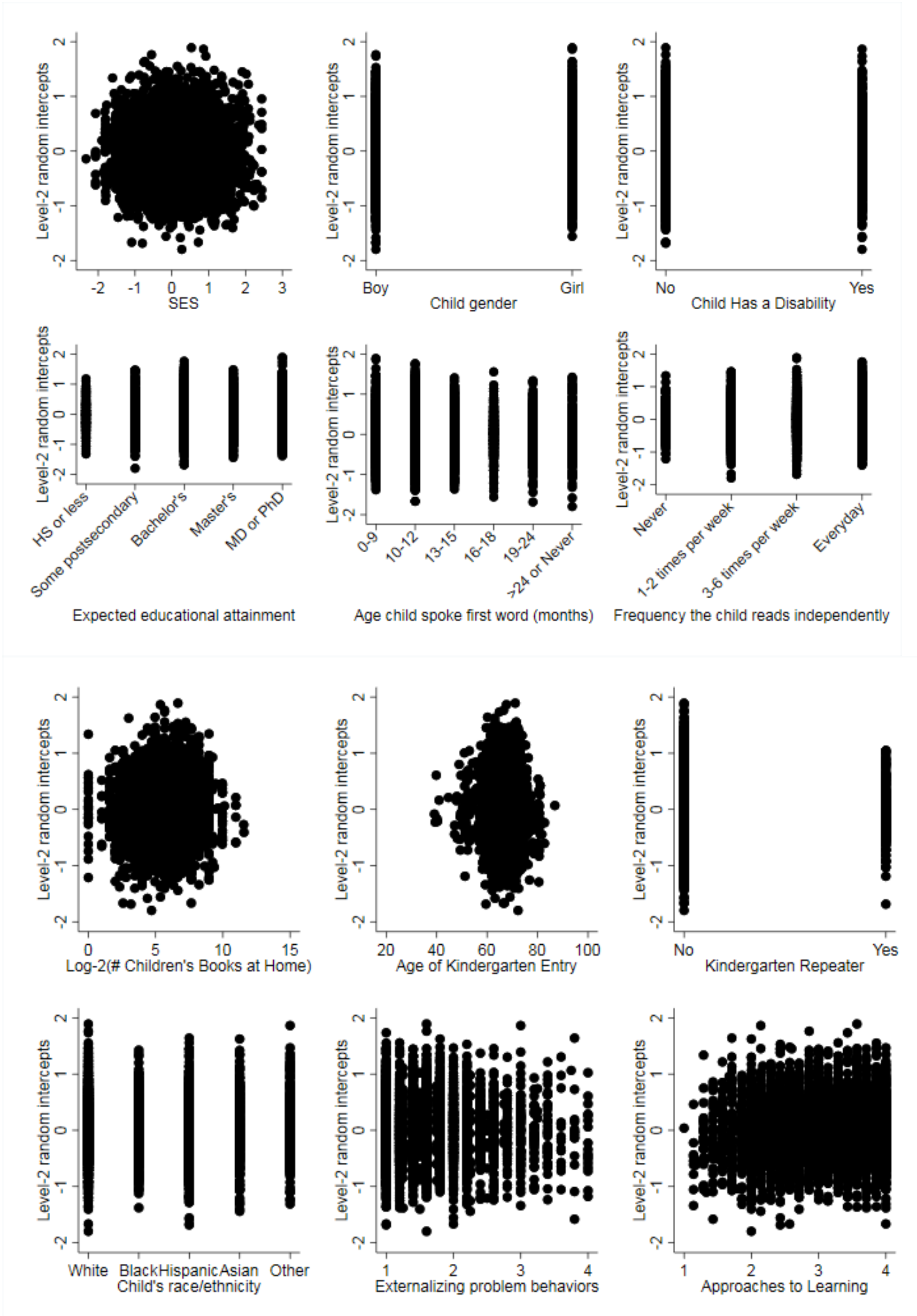

Figure 10. Empirical Bayes predictions for random intercepts at the child level from the final model (Model I) plotted on all of the child-level predictors included in the model (SES, gender, disability status, expected educational attainment, age of first word, frequency the child reads outside of school, number of books in the child's home, age of Kindergarten entry, whether or not the child repeated Kindergarten, race/ethnicity, externalizing behaviors and approaches to learning) to assess for homoscedasticity at level-2.

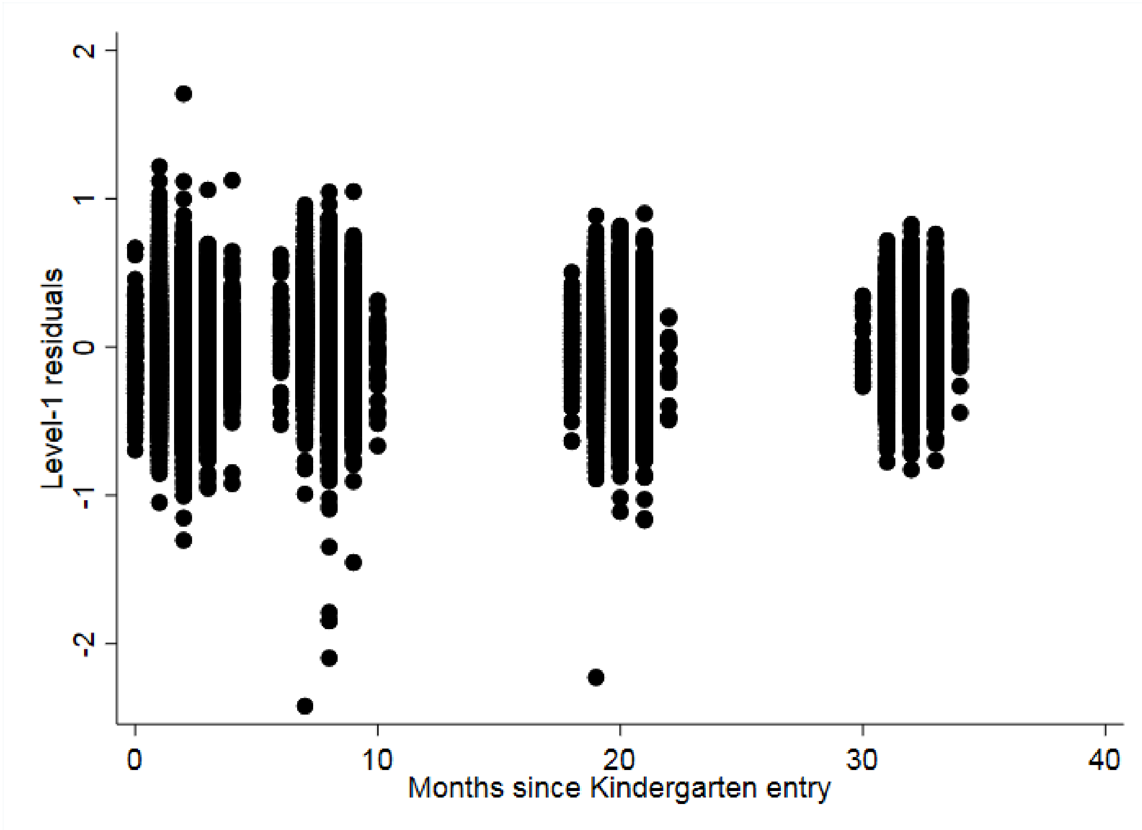

Figure 11. Residuals at the occasion level from the final model (Model I) plotted on the only occasion-level predictor, months since Kindergarten entry, to assess for homoscedasticity at level-1.

Table 8.

Taxonomy of Selected Hierarchical Linear Models from Sensitivity Analyses Exploring Whether Results of Analysis Exploring Gender Differences in Reading Development from Kindergarten Through Second Grade Hold When Data are Unweighted and/or Additional Time Points are Added. Unstandardized estimated regression coefficients and (robust standard errors) are presented below. The first three models are the same as Model A, H, and I from Table 7 except that the data were not weighted (although the analysis was restricted to the same subsample of participants). The last three models are the same as Models A, H, and I from Table 7 except that data from all six time points was used and therefore the W6CF6P_2T0 sampling weight was applied.

	(1) Model A-nw b/se	(2) Model H-nw b/se	(3) Model I-nw b/se	(4) Model A-6 b/se	(5) Model H-6 b/se	(6) Model I-6 b/se
Fixed Effects						
Intercept	-0.76188***	-3.86977***	-3.89419***	-0.80805***	-4.00324***	-4.04758***
	(0.01872)	(0.15831)	(0.15898)	(0.03336)	(0.34035)	(0.33849)
MONTH	0.16775***	0.23398***	0.23398***	0.14980***	0.21128***	0.21128***
	(0.00098)	(0.00510)	(0.00510)	(0.00256)	(0.01021)	(0.01021)
MONTH2	-0.00237***	-0.00284***	-0.00284***	-0.00184***	-0.00215***	-0.00215***
	(0.00003)	(0.00007)	(0.00007)	(0.00007)	(0.00017)	(0.00017)
GIRL	0.07042***	-0.06929***	-0.01548	0.09759**	-0.01069	0.10378
	(0.01836)	(0.01926)	(0.03552)	(0.03266)	(0.03785)	(0.06593)
GIRLxMONTH	0.00624***	0.00576***	0.00576***	0.00215	-0.00080	-0.00081
	(0.00135)	(0.00161)	(0.00161)	(0.00299)	(0.00376)	(0.00376)
GIRLxMONTH2	-0.00015***	-0.00012**	-0.00012**	-0.00003	0.00003	0.00003
	(0.00004)	(0.00004)	(0.00004)	(0.00008)	(0.00010)	(0.00010)
BLACK		-0.05211	-0.05203		-0.09085	-0.08896
		(0.03690)	(0.03689)		(0.08731)	(0.08656)
HISPANIC		-0.07886**	-0.07698*		-0.13652*	-0.13130*
		(0.03018)	(0.03018)		(0.06000)	(0.05959)
ASIAN		0.29567***	0.29545***		0.24518**	0.25064**
		(0.04403)	(0.04402)		(0.09200)	(0.09173)
OTHER		0.05337	0.05335		-0.07079	-0.06824
		(0.03777)	(0.03776)		(0.07264)	(0.07186)
BLACKxMONTH		-0.00233*	-0.00234*		0.00070	0.00070
		(0.00104)	(0.00104)		(0.00248)	(0.00248)
HISPANICxMONTH		0.00112	0.00112		0.00430*	0.00430*
		(0.00087)	(0.00087)		(0.00169)	(0.00169)
ASIANxMONTH		-0.00687***	-0.00687***		-0.00403	-0.00403
		(0.00128)	(0.00128)		(0.00278)	(0.00278)
OTHERxMONTH		-0.00106	-0.00106		0.00141	0.00141
		(0.00111)	(0.00111)		(0.00168)	(0.00168)
SES		0.26579***	0.26643***		0.23429***	0.23357***
		(0.01458)	(0.01459)		(0.03145)	(0.03160)
SESxMONTH		-0.00791***	-0.00791***		-0.00329	-0.00329

	(0.00103)	(0.00103)	(0.00263)	(0.00263)
SESxMONTH^2	0.00016***	0.00016***	0.00002	0.00002
	(0.00003)	(0.00003)	(0.00007)	(0.00007)
SUBURB	0.02411	0.02353	-0.04717	-0.04961
	(0.03248)	(0.03249)	(0.06406)	(0.06432)
TOWN	-0.04796	-0.04849	-0.29570**	-0.29672**
	(0.05076)	(0.05079)	(0.10530)	(0.10508)
RURAL	0.02654	0.02636	-0.10486	-0.10171
	(0.03593)	(0.03595)	(0.07928)	(0.07921)
SUBURBxMONTH	-0.00080	-0.00080	0.00067	0.00067
	(0.00090)	(0.00090)	(0.00180)	(0.00180)
TOWNxMONTH	0.00131	0.00131	-0.00022	-0.00022
	(0.00140)	(0.00140)	(0.00256)	(0.00256)
RURALxMONTH	-0.00090	-0.00090	-0.00108	-0.00108
	(0.00099)	(0.00099)	(0.00221)	(0.00221)
MIDWEST	0.05354	0.09771**	-0.03224	0.04858
	(0.02897)	(0.03586)	(0.07350)	(0.09032)
SOUTH	0.00259	0.01925	-0.02911	0.02820
	(0.02696)	(0.03312)	(0.06484)	(0.07566)
WEST	-0.00159	0.04328	-0.01411	0.06362
	(0.02907)	(0.03611)	(0.06308)	(0.07417)
NON_ENG	-0.19408***	-0.19439***	-0.33542***	-0.33953***
	(0.03441)	(0.03440)	(0.06806)	(0.06805)
NON_ENGxMONTH	0.00313**	0.00314**	0.00486*	0.00486*
	(0.00098)	(0.00098)	(0.00194)	(0.00194)
KREPEAT	0.51604***	0.51565***	0.63173***	0.62641***
	(0.04892)	(0.04891)	(0.11247)	(0.11224)
KREPEATxMONTH	-0.04435***	-0.04435***	-0.04790***	-0.04790***
	(0.00388)	(0.00388)	(0.00926)	(0.00926)
KREPEATxMONTH^2	0.00078***	0.00078***	0.00079**	0.00079**
	(0.00011)	(0.00011)	(0.00025)	(0.00025)
AGEENT	0.02237***	0.02238***	0.02675***	0.02652***
	(0.00211)	(0.00211)	(0.00445)	(0.00449)
AGEENTxMONTH	-0.00054***	-0.00054***	-0.00062***	-0.00062***
	(0.00006)	(0.00006)	(0.00013)	(0.00013)
PRIVATE	0.00621	0.00613	0.16109	0.16068
	(0.03951)	(0.03953)	(0.09602)	(0.09573)
PRIVATExMONTH	-0.00417	-0.00416	-0.00487	-0.00487
	(0.00243)	(0.00243)	(0.00796)	(0.00796)
PRIVATExMONTH^2	0.00014*	0.00014*	0.00007	0.00007
	(0.00006)	(0.00006)	(0.00023)	(0.00023)
Log_2NUMBOOKS	0.04901***	0.04891***	0.03444**	0.03424**
	(0.00574)	(0.00574)	(0.01105)	(0.01104)
HSORLESS	-0.10141**	-0.09994**	0.11105	0.11758
	(0.03567)	(0.03566)	(0.07369)	(0.07325)
POSTSEC	-0.04211	-0.04191	-0.00697	-0.00487
	(0.02167)	(0.02165)	(0.04497)	(0.04482)

MASTERS		0.06836***	0.06725***		-0.02609	-0.02709
		(0.01915)	(0.01914)		(0.04064)	(0.04078)
MDPHD		0.05296**	0.05316**		0.06517	0.06862
		(0.01984)	(0.01984)		(0.04361)	(0.04376)
A01to9		-0.02681	-0.02689		0.00534	0.00271
		(0.01806)	(0.01805)		(0.03800)	(0.03777)
A13to15		-0.01279	-0.01250		0.03097	0.03101
		(0.02022)	(0.02021)		(0.04046)	(0.04019)
A16to18		-0.04758*	-0.04881*		-0.10087*	-0.10122*
		(0.02330)	(0.02330)		(0.04797)	(0.04748)
A19to24		-0.06628*	-0.06736*		-0.04114	-0.04246
		(0.02674)	(0.02673)		(0.06248)	(0.06212)
A25orNever		-0.13230***	-0.13364***		-0.21163**	-0.21385**
		(0.03121)	(0.03120)		(0.06691)	(0.06631)
DISABLILITY		-0.14491***	-0.14477***		-0.09398*	-0.09334*
		(0.01745)	(0.01744)		(0.04114)	(0.04110)
RNEVER		-0.12274***	-0.12174***		-0.06006	-0.06576
		(0.03457)	(0.03455)		(0.08611)	(0.08562)
R1TO2		-0.03338	-0.03400		-0.10617**	-0.11021**
		(0.01807)	(0.01807)		(0.04058)	(0.03948)
REVERY		0.08825***	0.08841***		0.10589**	0.10586**
		(0.01593)	(0.01592)		(0.03471)	(0.03438)
EXTERN		0.11950***	0.11924***		0.16270***	0.16362***
		(0.01951)	(0.01951)		(0.03779)	(0.03782)
EXTERNxMONTH		-0.01038***	-0.01037***		-0.00828**	-0.00828**
		(0.00143)	(0.00143)		(0.00306)	(0.00306)
EXTERNxMONTH^2		0.00026***	0.00026***		0.00020*	0.00020*
		(0.00004)	(0.00004)		(0.00009)	(0.00009)
APPROACH		0.41054***	0.41014***		0.42529***	0.42632***
		(0.01694)	(0.01694)		(0.03976)	(0.03995)
APPROACHxMONTH		-0.00357***	-0.00357***		-0.00307**	-0.00307**
		(0.00049)	(0.00049)		(0.00114)	(0.00114)
GIRLxMIDWEST			-0.08882*			-0.16661
			(0.04262)			(0.09094)
GIRLxSOUTH			-0.03306			-0.11588
			(0.03939)			(0.07383)
GIRLxWEST			-0.09060*			-0.15945
			(0.04314)			(0.08196)
Random effects						
Variance: School-level random slope (MONTH)	0.00005	0.00004	0.00004	0.00012	0.00015	0.00004
	(0.00001)	(0.00001)	(0.00001)	(0.00001)	(0.00002)	(0.00001)
Variance: School-level random intercept	0.14063	0.06118	0.06131	0.24213	0.17671	0.05925
	(0.01034)	(0.00629)	(0.00629)	(0.02667)	(0.02257)	(0.00525)
School-level covariance	-0.00186	-0.00132	-0.00132	-0.00323	-0.00330	-0.00114
	(0.00018)	(0.00015)	(0.00015)	(0.00048)	(0.00052)	(0.00012)
Variance: Child-level random slope (MONTH)	0.00017	0.00015	0.00015	0.00010	0.00005	0.00009
	(0.00001)	(0.00001)	(0.00001)	(0.00002)	(0.00003)	(0.00001)
Variance: Child-level random intercept	0.47802	0.31021	0.30989	0.39021	0.23289	0.30575
	(0.00929)	(0.00772)	(0.00771)	(0.02147)	(0.02103)	(0.00686)
Child-level covariance	-0.00584	-0.00415	-0.00415	-0.00336	-0.00134	-0.00274

	(0.00020)	(0.00019)	(0.00019)	(0.00059)	(0.00073)	(0.00018)
Variance: Occasion-level residuals	0.09874	0.09542	0.09542	0.15940	0.17815	0.15680
	(0.00112)	(0.00128)	(0.00128)	(0.01964)	(0.02757)	(0.0155)
Num. occasions	31,070	22,270	22,270	18,910	12,840	12,840
Num. children	7,780	5,580	5,580	3,160	2,150	2,150
Num. schools	820	760	760	280	270	270
Log (pseudo)likelihood	-21,615	-13,938	-13,935	-17,443,371	-12,065,156	-12,060,772
Wald χ^2	84,949***	78,495***	78,509***	18,726***	21,737***	22,133***
df	5	54	57	5	54	57

Note: All sample sizes have been rounded to the nearest ten in compliance with the ECLS-K:2011 restricted-use data policies.
All random effect variances and covariances are statistically significant with p<0.001.
For fixed effect coefficients: *** *p<0.001,* ***p<0.01,* **p<0.05.*

Appendix H. Figures Exploring the Interplay of Sociocultural Factors

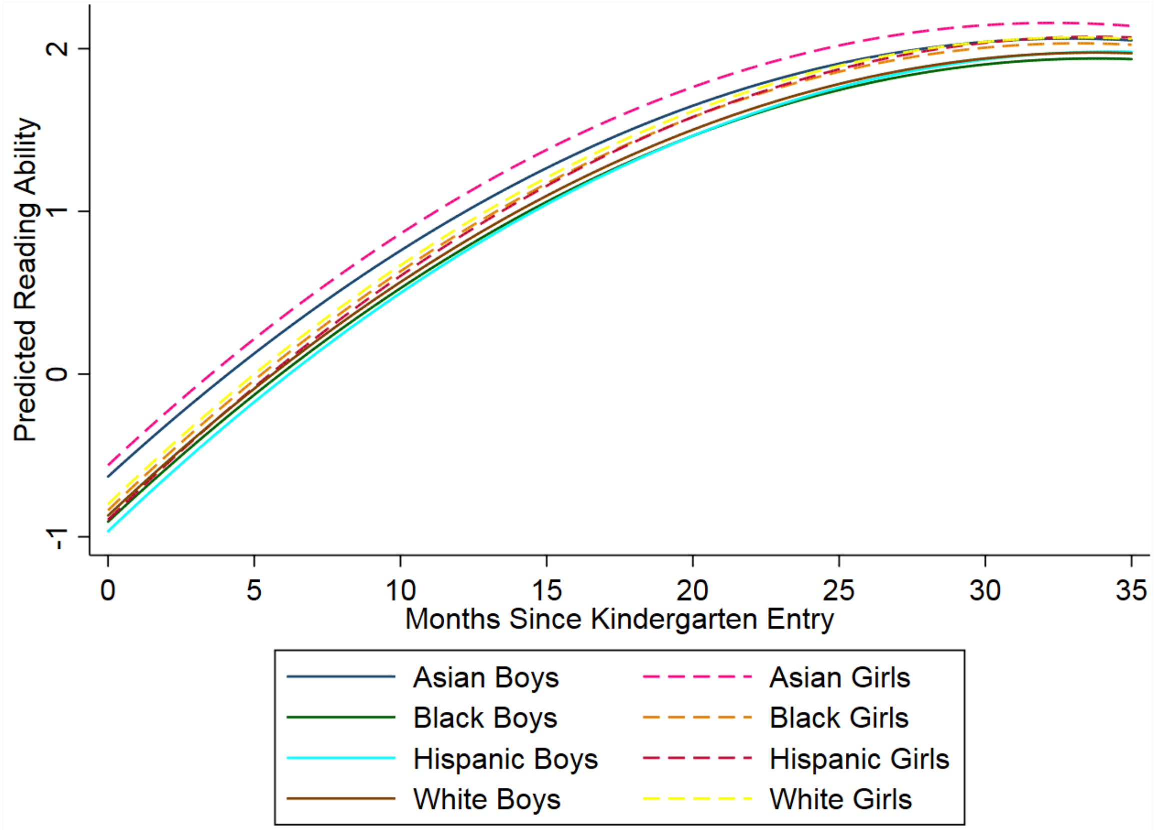

Figure 12. Predicted mean reading growth trajectories from Kindergarten through 2nd grade for girls and boys of different races/ethnicities with categorical variables set at their reference groups (students of average SES, primary home language is English, did not repeat Kindergarten, attended public schools) and continuous variables set at their predicted population means (Model B)

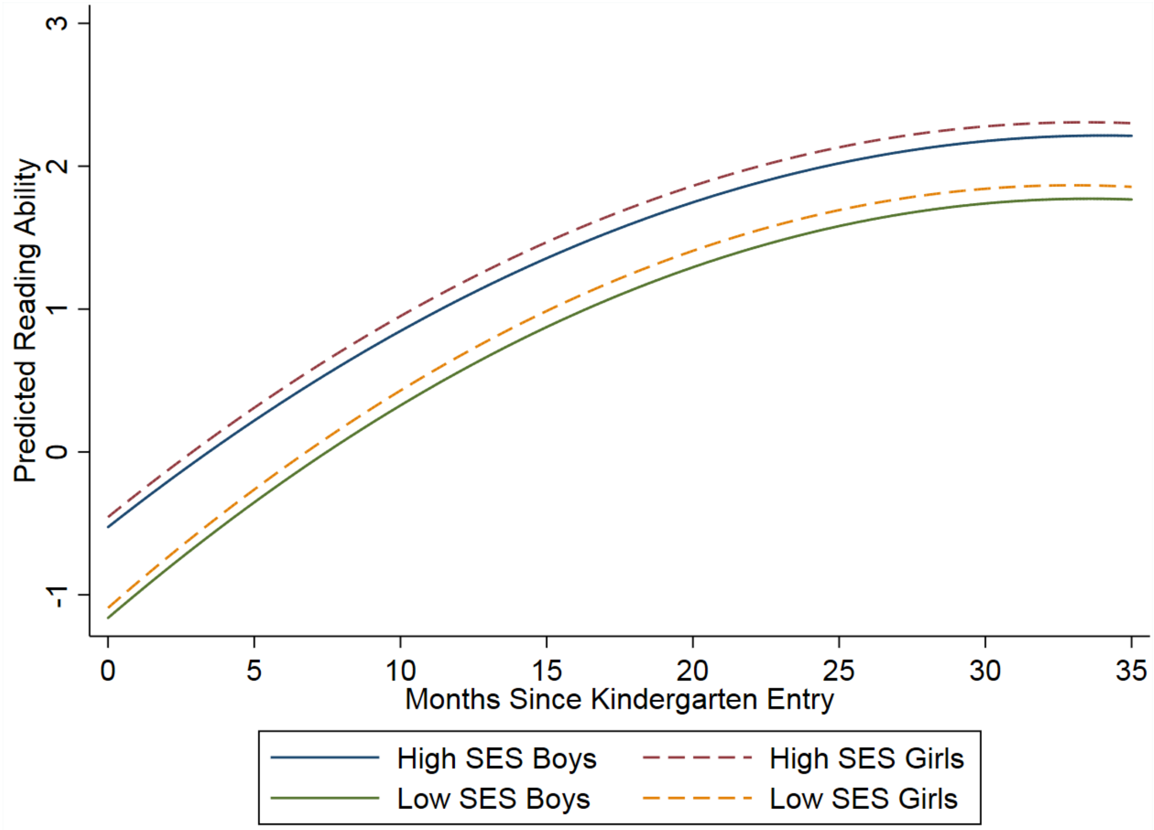

Figure 13. Predicted mean reading growth trajectories from Kindergarten through 2nd grade for girls and boys from high (90th percentile) vs. low (10th percentile) SES with categorical variables set at their reference groups (students of average SES, primary home language is English, did not repeat Kindergarten, attended public schools) and continuous variables set at their predicted population means (Model B)

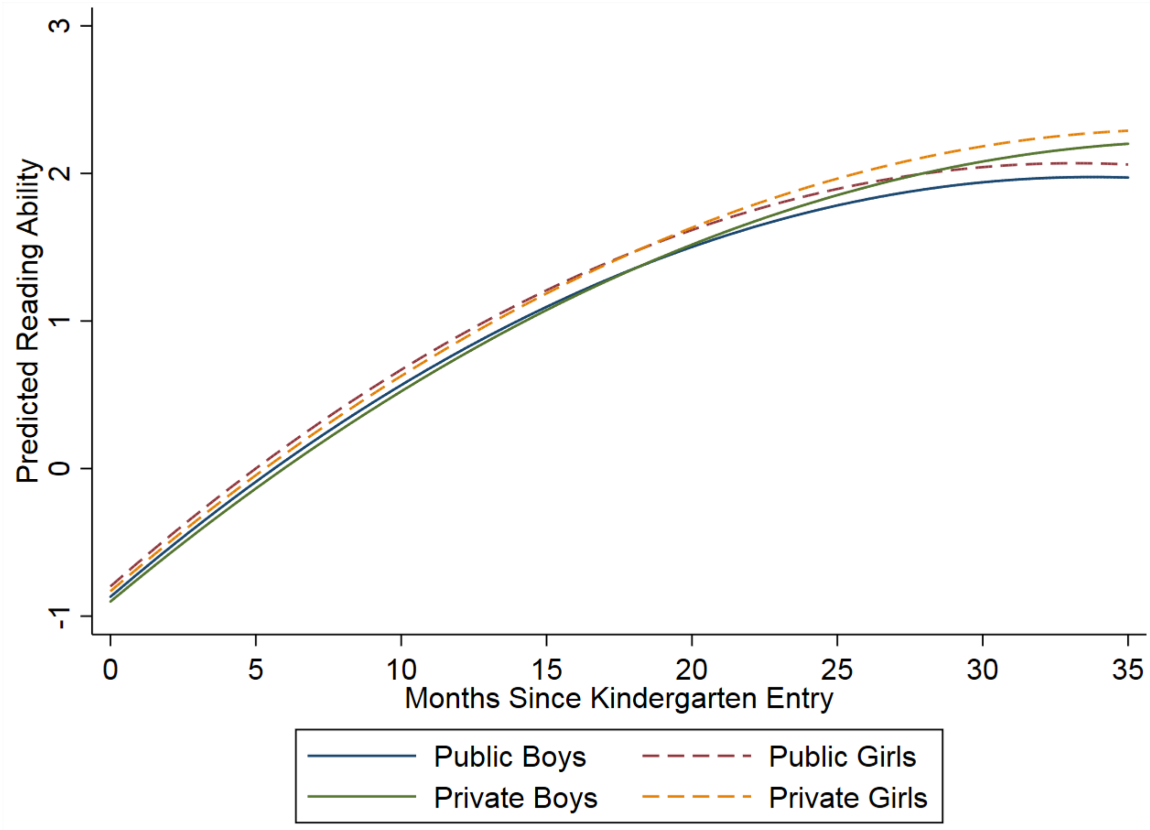

Figure 14. Predicted mean reading growth trajectories from Kindergarten through 2nd grade for girls and boys who attended public versus private schools with categorical variables set at their reference groups (students of average SES, primary home language is English, did not repeat Kindergarten, attended public schools) and continuous variables set at their predicted population means (Model B)