Do Non-cognitive Skills Explain the Gender Gap in Educational Achievement?

Christopher M. Cornwell
cornwl@terry.uga.edu

David B. Mustard
mustard@terry.uga.edu

and

Jessica Van Parys
jnv2106@columbia.edu

Department of Economics
University of Georgia
Working Paper

Last modified: December 10, 2009

Abstract: This paper uses the National Center for Education Statistics (NCES) Early Childhood Longitudinal Study–Kindergarten Cohort 1998-99 (ECLS-K) to examine how a range of educational achievement measures vary by gender, controlling for (other) individual, family, teacher and school characteristics. The paper identifies gender differences separately for white, black, and Hispanic students. We find that girls of all races perform better on reading tests. White boys earn higher scores on math and science tests, while black and Hispanic boys earn the same scores as black and Hispanic girls on math tests, but earn higher scores on science tests. These differences appear in kindergarten and persist through the fifth grade. The gender gaps in test scores (objective assessments) are not mirrored in classroom grades (subjective assessments), which are considerably higher for girls in reading, but (statistically) similar for white girls and boys in math and science. Black and Hispanic boys earn considerably lower grades in math and science despite their test scores being the same or higher than the test scores of black and Hispanic girls. Controlling for a lagged measure of classroom behavior achieves convergence on these objective and subjective assessments. Instrumenting contemporaneous behavior scores with lagged behavior scores results in boys and girls earning the same grades in reading, but boys earning higher grades in math and science. These results suggest an important role for non-cognitive skills in explaining the gender achievement gap.

1. Introduction

Extensive research exists on the racial and ethnic disparities in educational achievement.[1] The gaps by race and ethnicity are economically and statistically significant and are generally persistent over time. Better understanding these gaps is important for the formulation of public policies to increase the educational and labor market opportunities of people in the lower-performing groups. In line with the research on inequality, this paper addresses another large and important, but less frequently analyzed education gap. Specifically, we examine the rising educational achievement of females relative to males.

The gender achievement gap was first identified in higher education, when in the late 1970s, the number of women earning Bachelor's degrees exceeded the number of men earning the same degree. Since the 1980s, the college-going rates of females increased dramatically, relative to the college-going rates of males (NCES 2009-020). For example, in 1970 42% of baccalaureate degree holders were female; in 2000, 56% were female (NCES 2005-016). Today many colleges and universities have student-bodies that are over 60% female. In addition to higher rates of college attendance, females are more likely to graduate from high school, enroll in college immediately following high school graduation, and earn a bachelor's degree within six years of enrolling in college. These gender disparities are particularly acute for black and Hispanic students (NCES 2005-016). The remarkable relative educational success of women has had implications for the gender gap in earnings (O'Neill 1985; Averett and Burton 1996; Weinberger and Kuhn 2006; Blau and Kahn 2006).

While the research on the educational gender gap is much less extensive, the vast majority of it focuses on higher education, and attempts to explain the variance in college

[1] Some examples include Jencks and Phillips (1998), Fryer and Levitt (2004), Clotfelter et al. (2009), Hanushek and Rivkin (2009), and Reardon and Galindo (2009).

attendance rates for males and females (Jacob 2002; Loury 2004; Cho 2006; Goldin et al. 2006; Dynarski 2007; Frenette and Zeman 2007; Reynolds and Burge 2007). These studies find that gender differences in college enrollment and success are largely a function of gender differences in academic preparedness prior to college and females' expectations for future work.

Though pre-college academic preparedness determines college enrollment, only a few studies examine gender differences in achievement prior to the 8th grade (Anderson 2006; Lavy and Schlosser 2007; Holmlund and Sund 2008; Husain and Millimet 2009, Fryer and Levitt 2009). These papers discover gender differences in reading and math test scores as early as kindergarten. Possible explanations for these differences include the gender of teacher, the ratio of boys to girls in a classroom, and whether the children attended pre-school.

This paper adds to the early-education gender literature in three ways. First, it analyzes both objective test scores and subjective teacher grades while the literature that examines gaps in educational achievement primarily evaluates standardized tests. This is the first paper to examine gender differences of primary-school students in both subjective and objective measures of student performance.[2] While standardized tests are important measures to consider when evaluating gender gaps, teacher-assigned grades are even more consequential. Disparities in high school grades are important to explore because admissions offices generally place more weight on this measure than on standardized test scores and because grades are better predictors of college performance (Betts and Morrell 1999; Cornwell et al. 2009). Moreover, the gender literature consistently reports that girls, on average, earn higher grades than boys. If this difference begins early in students' academic careers and if it represents a persistent trend from

[2] Burgess and Greaves (2009) use administrative data from the National Pupil Database (NPD) that combines basic individual data with assessment data to explore gaps in educational achievement by race, ethnicity, and nationality for students in England. Lavy (2008) compares blind and non-blind scores on matriculation exams of male and female high school students in Israel, and finds evidence that teachers discriminate against male students in favor of female students.

elementary school through high school, it may help to explain the differing academic trajectories of girls and boys.

Second, we are one of the few papers to integrate the rapidly growing literature on non-cognitive skills to assess the gender gap. We conclude that a large share of the gender disparities in educational success is attributable to differences between boys and girls in their non-cognitive development. We find one non-cognitive skill that is particularly significant – the expressed desire to learn in the classroom. Teachers regard girls as substantially more amenable to the learning process than boys. Our finding that this "skill" affects academic outcomes is consistent with Claessens et al. (2009), who estimate how a range of socio-emotional skills in kindergarten affect children's standardized test scores in fifth grade. We find, however, more evidence that this skill matters for describing gender differences, particularly differences in teacher-assigned grades.

Last, this paper is the most comprehensive study of gender differences of primary and secondary students in the United States. Holmlund and Sund (2008) conclude that the teacher's gender does not reduce the gender gap in student achievement among upper-secondary students in Sweden. Hussain and Millimet (2009) examine only reading and math standardized test scores of U.S. students through the third grade. In contrast, we evaluate children through fifth grade, include more standardized tests (e.g., science scores), and include teacher-assigned grades, which other papers omit entirely. Because different measures often show different results, our additional measures provide a more complete understanding of the gender gap.

This paper examines the gender achievement gap from a fresh angle, focusing on differences in educational outcomes among primary school age children. With data on children that begin in the kindergarten year, we document how the achievement gap changes as children move from kindergarten to the fifth grade. We divide students by race and ethnicity to test how relevant the gender achievement gap is for black, white, and Hispanic students. Test scores measure objective skills in reading, math, and science, while teacher grades provide a subjective comparison for the same curriculum. Test scores and teacher grades provide a more detailed analysis of the gender achievement gap. Indeed, the differences in test scores do not mirror the differences in teacher grades. We use a social rating scale (SRS) to explain this discrepancy. The SRS serves as a proxy for non-cognitive abilities, and its use in the model eliminates the difference between objective and subjective assessments.

The empirical results indicate that girls of all races score higher than boys on reading tests. White boys score higher than white girls on math and science tests. Black and Hispanic boys score the same as black and Hispanic girls on math tests, but score higher on science tests. These differences emerge in kindergarten and persist through the fifth grade. Girls, however, earn considerably higher reading grades and there is no statistically significant difference between white boys and girls in math and science grades. Black and Hispanic boys earn lower grades in math and science despite their same or relatively higher test scores. Controlling for a lagged measure of classroom behavior achieves convergence on these objective and subjective assessments; boys no longer earn lower grades than their test scores suggest and girls no longer earn higher grades than their test scores suggest. When we instrument contemporaneous classroom behavior with lagged classroom behavior there are no statistically significant differences between boys and girls in reading grades, but boys earn higher math and science grades. In other words, when our models do not account for classroom behavior, girls achieve at relatively higher levels than boys. When our models account for current classroom behavior we find the opposite result. These results suggest that differences in non-cognitive skills explain a considerable portion of the gender differences between test scores and teacher grades.

The paper proceeds in the following form: Section 2 describes the data, Section 3 discusses the variables and methodology, Section 4 explains the empirical results, and Section 5 concludes with a discussion.

2. Data and Variables

We use data from the Early Childhood Longitudinal Study – Kindergarten Cohort 1998-99 (ECLS-K), administered by the National Center for Education Statistics (NCES). In the fall of 1998, NCES randomly sampled schools (the primary sampling units) from across the United States. Within each school, all kindergarten classrooms were selected, from which children (units of observation) were randomly selected. Classrooms were required to have at least five kindergartners to qualify for the sample. NCES administered reading, math, and science tests to each child, collected information on each child's school, and submitted detailed questionnaires to each child's parents and teachers. Parents and teachers were asked to comment on their own personal characteristics and experiences, as well as on their relationship with the child.

Once children were selected for the fall 1998 sample, NCES administered follow-up assessments and questionnaires in the springs of 1999, 2000, 2002, and 2004. A "freshening" process occurred in the springs of kindergarten and the first grade, whereby a subset of "movers" were followed to their new schools. The remaining "movers" were replaced by a new sample of students from the original schools. The freshening process was discontinued after the first grade, and sample attrition set in as children moved to new schools. The data begin with 21,000 observations and conclude with approximately 9,000 observations in the fifth grade. Of the

21,000, only children who passed an English language screening test were administered the reading, math, and science assessments. This paper's analysis begins with approximately 10,600 observations in kindergarten and concludes with 6,500 observations in the fifth grade.[3] These children were selected because they had all of the requisite data to estimate the models.

This paper analyzes a range of objective and subjective assessments. NCES prepared the objective reading, math, and science assessments. Each test was divided into two parts. How well the child scored on the first portion of the assessment determined which second portion he or she would receive. Thus, scores used in this analysis are not raw scores, but rather item response theory (IRT) scores. Higher scores, however, still indicate higher levels of academic achievement.

Academic achievement was also measured with subjective assessments. Teachers were asked to rate each student's mastery of specific skills in reading, math, and science. NCES totaled these responses and constructed a continuous 0-4 point scale score for each child in each academic subject, where 0 indicates no understanding of the content or skill and 4 indicates complete mastery of the content or skill. This "Academic Rating Scale" (ARS) was designed to measure the same skills as those found on the objective reading, math, and science assessments. Moreover, teachers were unaware of their students' scores on the objective assessments when they provided answers for the ARS.

Another subjective assessment, the "Social Rating Scale" (SRS), asked teachers to rate their children along several dimensions of classroom behavior. For example, teachers reported how engaged each child was in the classroom, how often the child externalized or internalized

[3] There are 6,441 child observations in the fifth grade for all outcome variables except math and science grades. Fifth grade students had different teachers for each subject, so NCES did not ask the math and science teachers to administer grades for all of the children. Instead they randomly collected grades for half of the students taking math and half of the students taking science. This resulted in 3,148 observations on students with math grades and 3,039 observations on students with science grades.

problems, how often the child lost control, and how well the child had developed interpersonal skills. NCES combined the answers to these types of questions to create a continuous scale score for each child, ranging from 0-3, which measured the child's "Approaches to Learning," "Self-Control," "Internalizing Problems," "Externalizing Problems," and "Interpersonal Skills." Similar to the ARS scale, higher scores represent higher levels of non-cognitive skills.

This paper first utilizes the "Approaches to Learning" scores for the child in period t-1 as a control variable to explain academic achievement in period t. We next use the Approaches to Learning score in period t-1 as an instrument for Approaches to Learning scores in period t, and re-estimate our model with this instrumental variable. The Approaches to Learning Scale "measures behaviors that affect the ease with which children can benefit from the learning environment. It includes six items that rate the child's attentiveness, task persistence, eagerness to learn, learning independence, flexibility, and organization" (ECLS-K First Grade User Manual, pg. 2-14). Thus, "Approaches to Learning" scores, which we refer to as SRS scores throughout the paper, is used to proxy for a set of non-cognitive skills that differs by gender in the classroom environment. Non-cognitive skills are theoretically important, as they are correlated with students' academic and labor market outcomes (Heckman 2008).

Table 1 reports mean statistics for the test scores, teacher grades, and SRS scores. The left portion of each table contains means and standard deviations for the girls in the sample, while the right portion of each table does the same for the boys. For example, girls earn on average, 43 points on the kindergarten reading test, while boys score 40 points. Girls also earn higher teacher grades in kindergarten reading—3.55 points compared with 3.33 points for boys. There is no difference in mean test scores between boys and girls in kindergarten math and general knowledge, but boys earn higher scores on these tests in later years. Girls earn higher teacher grades in reading, math, science, and Approaches to Learning in nearly every evaluation cycle. The standard deviations in test scores, grades, and behavioral assessments are lower for girls, thus indicating boys have a wider achievement spread.

3. Methodology

3.1 Baseline Education Production Function

To understand the effect of gender on various measures of educational achievement, we estimate empirical models of the form,

$$(a) \qquad Y_{i,s,t} = \alpha + \beta * gender_i + \underline{\phi} * race_i + \underline{\varphi} * gender_i * race_i + \underline{\gamma} * family_{i,t} + \underline{\lambda} * teacher_{i,s,t} + \underline{\delta} * school_{i,t} + \varepsilon_{i,s,t}$$

where *i* indexes children, *s* indexes the subject, and *t* indexes the grade level. Each cross-sectional wave includes students who were assessed in the spring of that school year (i.e., spring kindergarten, spring first grade, spring third grade, spring fifth grade). The model incorporates sampling weights to account for the over-sampling of Pacific Islanders and attrition as children left the sample over time. The model uses the replicate sample weights provided by NCES to conduct linearized jackknife variance estimation.

Y includes test scores and teacher grades in reading, math, and science. *Race* is a vector of personal characteristics, including the child's race and ethnicity. Children are classified into one of six categories: white, black, Hispanic, Asian, Pacific Islander, Native American, or mixed racial identity. *Gender*race* is a vector of race and gender interaction variables, which allows the effect of gender to vary by race and ethnicity.

Family is a set of family characteristics, including the age of the child at kindergarten entry, the age of the mother at first birth, the number of books in the home, the socioeconomic status of the family, and if the mother received WIC benefits during pregnancy. The socioeconomic (SES) index is comprised of five variables—family income, the mother's and father's highest levels of educational attainment, and the mother and father's occupational prestige rankings. The index is normalized to have a mean of zero and standard deviation equal to one. These family characteristics are the control variables that Levitt and Fryer (2004) use to evaluate the black-white gap in educational achievement.

Teacher is a vector of teacher characteristics including the teacher's highest level of educational attainment and her years of teaching experience. Teachers were sorted into one of four categories for their highest level of educational attainment; they either had a bachelor's degree, some additional training beyond a bachelor's degree, a Master's degree, or another advanced degree such as a PhD.

School is a vector of school characteristics that includes whether the school is located in an urban, suburban, or rural district, if it is a public institution, if it is located in the south, and the percentage of the student body that qualifies as a racial or ethnic minority. Finally, ε is the variation in educational achievement that is unexplained by gender or the other control variables.

Tables 2 and 3 give descriptive statistics for these personal, family, teacher, and school characteristics. The columns of these tables represent the grade level. Unlike Table 1, where the outcome variables show differences by gender, these tables report means and standard deviations for the entire sample of students. This is because there is little variation by gender in the control variables. For example, in kindergarten 24% of mothers with boys and 24% of mothers with girls were teenagers at first birth.

Table 2 shows that the sample is approximately 50% male across each grade level. In kindergarten, the sample is 13% black, 14% Hispanic, 4% Asian, 1% Pacific Islander, 2% Native

American, and 3% of mixed racial identity. Of the racial and ethnic groups that comprise the sample, blacks leave the sample at the highest rate. By fifth grade only 8% of the sample is black. In the kindergarten year, 37% of the children's mothers received WIC benefits, 24% of the mothers were teenagers at first birth, while 13% of the mothers were greater than 30 years old at first birth. The children's average age at kindergarten entry was 65.7 months. The average number of books in the home was approximately 80 in the kindergarten year, which increased to 116 by the fifth grade. The average socioeconomic status of the children's families remained relatively constant from kindergarten to fifth grade. In kindergarten, the average family had an SES index of 0.08, and in fifth grade it is the same. Sample attrition occurs as the number of child observations drops from 10,604 in kindergarten to 6,496 in fifth grade.

The top section of Table 3 shows that the average number of years of teaching experience rises from kindergarten to fifth grade. Kindergarten teachers have taught for 9 years, while fifth grade teachers have 14.6 years of teaching experience. The average level of teacher education is consistent across grades. The average teacher has some certification beyond a bachelor's degree, but just less than a master's degree.

The lower section of Table 3 gives the means and standard deviations for characteristics describing the schools in the sample. In kindergarten approximately 80% of schools are public institutions, 38% are located in urban districts, while 24% are located in rural districts. Nearly one-third of the schools are located in the southern United States, and approximately 18% of schools have minority students comprising more than 75% of the student body.

3.2 Teacher Grade Production Function with Lagged Non-cognitive Skills

To estimate whether a specific non-cognitive skill – classroom participation and eagerness to learn – affects a child's grade assignment, we estimate the following least squares equations:

(b) $$Y_{i,s,t} = \alpha + \beta{*}gender_i + \underline{\theta}{*}test\_score_{i,s,t} + \underline{\phi}{*}race_i + \underline{\varphi}{*}gender_i{*}race_i + \underline{\eta}{*}X_{i,s,t} + \varepsilon_{I,s,t}$$

(c) $$Y_{i,s,t} = \alpha + \beta{*}gender_i + \rho{*}learn_{i,t-1} + \underline{\theta}{*}test\_score_{i,s,t} + \underline{\phi}{*}race_i + \underline{\varphi}{*}gender_i{*}race_i + \underline{\eta}{*}X_{i,s,t} + \varepsilon_{i,s,t}$$

where $X_{i,s,t}$ contains the same family, teacher, and school characteristics from equation (a) and $Y_{i,s,t}$ is a vector containing teacher grades in reading, math, and science for the first, third, and fifth grade years. Model (b) tests whether there is a gender gap in assessments – between test scores and teacher grades. If no systematic gender difference exists between these two outcome variables, then the coefficient estimate on gender should be close to zero and lack statistical significance.

Model (c) tests whether the gender gap in assessments – between test scores and teacher grades – disappears when we incorporate a lagged proxy for non-cognitive ability[4]. In particular, specification (c) tests the hypothesis that a student's gender does not *cause* him (her) to earn lower (higher) teacher grades than his (her) test scores would suggest. The teacher grade production function likely contains a form of omitted variable bias in which an unobserved characteristic is correlated with gender and with the teacher's assessment of the child's cognitive ability. We hypothesize the omitted variable is the child's classroom behavior. Potential behavioral impediments to classroom learning may include boys' relative inability to sit still for

[4] We also estimate this model using quantile regressions at the 0.2, 0.4, 0.6, and 0.8 cut-points. We did this to test whether the differences by gender in test scores and teacher grades are larger at different points in the teacher grade distribution. The results (omitted here) show that there are not statistically significant differences at different points in the grade distribution, and thus, we report only the least-squares analysis at the mean.

long periods of time, their relative lack of participation or demonstration of knowledge in the classroom, or their relative supply of effort on assignments and homework. We use a lagged SRS score to ensure that different teachers assign the grade and the behavioral assessment. This removes the potential for endogeneity due to reverse causality. To the extent that behavior is persistent, the SRS score will have an independently positive and statistically significant effect on the teacher-assigned grade. To the extent that it is correlated with gender, it will drive the coefficient estimate on gender in the teacher grade equation to zero.

3.3 Instrumental Variables Estimation and Contemporaneous Non-cognitive Skills

In two out of three cases, the lagged SRS score was given two years prior to the time the teacher grade was produced. For example, we estimate the fifth grade year's reading grade as a function of the fifth grade test score, gender, other control variables, and the *third* grade SRS score. Consequently, controlling for the lagged SRS score only compares children who had the same behavioral score two years ago; it does not guarantee that the children behave the same way today. To understand the effect of contemporaneous SRS score on teacher grades, we instrument the contemporaneous SRS score with the lagged SRS score and estimate the following model by instrumental variables:

$$(d)\quad Y_{i,s,t} = \alpha + \beta * gender_i + \rho * learn\_hat_{i,t} + \underline{\theta} * test\_score_{i,s,t} + \underline{\phi} * race_i + \underline{\varphi} * gender_i * race_i + \underline{\eta} * X_{i,s,t} + \varepsilon_{i,s,t}$$

where we assume (1) that lagged SRS score is correlated with the contemporaneous SRS score and (2) that the lagged SRS score only affects the contemporaneous teacher grade through its effect on the contemporaneous SRS score. *Learn_hat* are the predicted values from the first stage regression of contemporaneous SRS on lagged SRS, contemporaneous test score, gender, and the other control variables in the model. Table 4 reports results from this first stage regression. The instrument, lagged SRS score, has an independently positive and statistically significant effect on the contemporaneous SRS score. Thus, classroom behavior is, to some extent, persistent, which ensures that the lagged SRS score is not a weak instrument. The coefficient estimate on male is also negative and statistically significant in the first stage regression. This says, for example, that given the same classroom behavior in the third grade and the same test score in the fifth grade, boys have lower behavioral scores in fifth grade. This finding is explained in further detail in the IV estimation results section.

4. Results

There are three sections of results. Section 4.1 describes and analyzes the gender differences in test scores and teacher grades for reading, math, and science. Section 4.2 examines why differences in test scores and teacher grades exist and tests the hypothesis that grades are a function of cognitive and non-cognitive abilities. Section 4.3 discusses (1) the gender differences in teacher grades after accounting for contemporaneous SRS scores and test scores, and (2) the relative importance of SRS scores and test scores in the model of teacher grades. Each regression model controls for personal, family, teacher, and school characteristics; however, only the coefficient estimates describing the gender gaps by race and ethnicity are reported in the first section. The second and third sections add the coefficient estimates on test scores and SRS scores for the teacher grade equations.

Because the paper explores how the gender achievement gap changes according to the subjective or objective nature of the outcome variable, Table 4 shows how the (Boy – Girl) gap differs by race, outcome variable, and grade level. Table 5 illustrates what happens to the gender gaps in teacher grades once the model controls for test scores and classroom behavior. Table 7 recreates columns (III) of Table 5 using IV estimation. The results are reported in standard deviations, where the test scores and teacher grades are normalized to have a mean equal to zero and a variance equal to one. Standard errors on the coefficient estimates are reported in parentheses. Asterisks (*) and crosses (+) indicate statistical significance above the 5% and 10% confidence levels, respectively.

4.1 Gender Differences in Test Scores and Teacher Grades

Table 4 reports results for the gender achievement gap in test scores and teacher grades for children in kindergarten, first grade, third grade, and fifth grade. The top third of the table gives results for reading, the middle third for math, and the final third for science. The columns denote whether the gap is in a test score or a teacher grade, and for which grade level the students represent. The rows compare the gender gaps by race, that is, white boy: white girl, black boy: black girl, Hispanic boy: Hispanic girl. The gaps are measured (boy – girl), so a negative sign indicates that, on average, girls earn the higher score. The third column in each sequence denotes the difference between teacher grades and test scores. This is the difference of two differences – a gender difference in test scores and a gender difference in teacher grades. The $grade_{GAP}$ = ($grade_{BOY}$ – $grade_{GIRL}$) and test $score_{GAP}$ = (test $score_{BOY}$ – test $score_{GIRL}$). The difference = $grade_{GAP}$ – test $score_{GAP}$ = ($grade_{BOY}$ – $grade_{GIRL}$) – (test $score_{BOY}$ – test $score_{GIRL}$). As the difference limits to zero, the gender gaps in test scores and teacher grades are equal. As the difference grows negative, girls achieve an advantage. This means either (1) girls earn higher grades than their test scores would suggest, (2) boys earn lower grades than their test scores would suggest, or (3) a combination of (1) and (2).

Reading

The results for reading test scores and grades are presented in Table 4. Beginning in kindergarten, white boys score 0.16 standard deviations lower than white girls on reading tests. This gap decreases to 0.13 standard deviations in the first grade, and then remains relatively constant in the higher grade-levels. Black and Hispanic boys, similarly, score lower than black and Hispanic girls on the reading tests. These gaps in kindergarten are 0.16 and 0.17 standard deviations, respectively. In contrast to the gender reading gap for white children, black and Hispanic boys score lower than their female peers in the years beyond kindergarten. Hispanic boys score 0.36 standard deviations lower than Hispanic girls in the fifth grade. Black boys score 0.26 standard deviations lower than black girls in fifth grade.

Each gender gap estimate for reading test scores is precisely estimated above the 5% level. The results are also consistent over time; girls begin kindergarten with a slight reading advantage and maintain that advantage. White boys do not fall further behind white girls in reading test scores from kindergarten to the fifth grade, but the gender gaps for black and Hispanic children widen over the years.

Contrasting the results for reading test scores with those for reading grades gives a more detailed analysis of the gender gap. Results for reading grades are also located in Table 4. In kindergarten, white boys earn 0.26 standard deviations lower reading grades than white girls. This gap is precisely estimated and remains relatively constant through the fifth grade. Similar to the results for reading test scores, white girls begin with an advantage in reading grades and maintain that advantage over the years, but these gaps are twice the size as those for reading test scores. The differences between reading grades and test scores for white children are 0.09 to 0.13

standard deviations and represent statistically significant differences. Thus, girls earn considerably higher classroom grades than boys, but have only marginally higher test scores. Such discrepancy between test scores and teacher grades is a principal result for the paper, and a testable explanation is provided in the section 4.2.

The gender gaps in reading grades are also quite large for black and Hispanic children. Black boys earn 0.35 standard deviations lower reading grades than black girls in kindergarten. This gap temporarily decreases to 0.24 standard deviations in first grade, but increases to 0.40 in third grade. Similar to the results for white children, black boys earn lower reading grades than they do reading test scores. These differences are on the magnitude of 0.07 (first grade) to 0.19 (kindergarten) and with the exception of first grade, represent statistically significant differences.

Reading grades for Hispanic boys are also lower than the grades for Hispanic girls, but unlike the results for white and black children, these gaps mirror the reading gaps in test scores, or at least do not represent statistically significant differences. Hispanic boys earn 0.22 standard deviations lower reading grades in kindergarten compared with Hispanic girls. This gap increases through the fifth grade, where Hispanic boys earn 0.38 standard deviations lower grades than Hispanic girls. The differences between reading grades and test scores, however, lack statistical significance for Hispanic children. Reading grades for Hispanic children are highly correlated with their test scores, while reading grades for black and white children are not.

The models of reading test scores fit the data better than the models for reading grades, and this difference increases over time. In kindergarten, the model of reading grades explains 16% of the variance in those grades, while the model for test scores explains 18%. By fifth grade, the model of test scores explains 31% of the variance in test scores, while the model for reading grades only explains 17% of the variance in those grades. This suggests that the variance in test scores is more easily explained by observable characteristics, whereas explaining the variance in teacher grades requires additional information on unobservable characteristics.

Math

The results for math test scores and math grades are reported in middle of Table 4. Beginning in kindergarten, white boys earn 0.06 standard deviations higher math test scores than white girls. The gap increases to 0.14 standard deviations in first grade, 0.28 in third grade, and 0.19 in fifth grade. All of these results are significant at the 5% level. With the exception of the third grade year, Hispanic boys do not earn higher math test scores than Hispanic girls. The same is true for black boys, who do not earn higher math test scores than black girls. Thus, the gender gap in math test scores is primarily a phenomenon among white children.

Contrasting the results for math test scores with those results for math grades tells a very different story. White boys earn 0.13 standard deviations *lower* grades in kindergarten, and there is no statistically significant difference in math grades between white boys and girls in the first and fifth grades. White boys earn 0.09 standard deviations higher math grades in the third grade, but this is markedly lower than their 0.28 standard deviations higher test scores. The differences between the math grades and test scores of white children are consistently negative, statistically significant, and range from 0.15 to 0.18 in favor of girls.

This result holds for black and Hispanic children as well. Although black and Hispanic boys and girls earn approximately the same scores on math tests, the boys' math grades are significantly lower. Black boys earn grades that are 0.11 (first grade) to 0.28 (fifth grade) standard deviations lower than black girls. The black gender gap in math grades is not always statistically significant (first and fifth grades), but the difference between the math grades and test scores is significant in every year except first grade. Indeed, this difference reaches its zenith in fifth grade at 0.46 standard deviations in favor of black girls. The gender gaps in teacher grades usually lack statistical significance for Hispanic children, but the difference between grades and test scores is large and usually statistically significant (except kindergarten). Like that of black students, this difference reaches 0.21 standard deviations in fifth grade. These results demonstrate that although boys score higher or the same as girls on math tests, teachers rate boys as less likely to excel in mathematics in the classroom.

General Knowledge/Science[5]

The bottom section of Table 4 shows the results for science test scores and grades. Similar to the pattern for math grades and test scores, white boys score higher than white girls on science tests, but earn lower teacher grades in the kindergarten and first grade years. In the third and fifth grade years, white boys earn the same teacher grades as girls, despite their considerably higher test scores. In third grade specifically, boys earn 0.28 standard deviations higher science test scores, but essentially the same teacher grades.

The results are similar for black and Hispanic boys. In third grade, black and Hispanic boys earn 0.23 and 0.13 standard deviations higher science test scores, respectively, than their female peers. Yet their grades are 0.22 and 0.04 standard deviations *lower* than their female peers. As in reading and math, the differences between teacher grades and test scores in science are large and statistically significant. The differences are also consistently negative, and therefore in favor of girls.

[5] In the kindergarten and first grade years, these are "general knowledge" test scores and grades. General knowledge questions cover a combination of social science and natural science subject matter. In the third and fifth grade years, these test scores and grades reflect science curriculum only.

4.2 Subjective vs. Objective Assessments and the Role of Classroom Behavior

Table 4 shows that there are statistically significant differences between test scores and teacher grades by gender. Girls earn higher grades relative to their test scores, while boys earn lower grades relative to their test scores. Models of teacher grades fit the data only half as well as models of test scores. For example, the fifth grade science model explains approximately 39% of the variance in test scores, but only 14% of the variance in teacher grades. This leads us to believe that classroom behavior is an unobserved variable that will explain more of the variance in teacher grades. We hypothesize that including classroom behavior will eliminate the gender gaps between test scores and teacher grades and increase the teacher grade model's goodness of fit. Table 5 investigates this hypothesis.

Column (I) in Table 5 denotes the baseline model of teacher grades, conditional on gender and the other observed control variables. Column (II) adds each child's contemporaneous test score. If there are no gender differences between test scores and teacher grades, the coefficient estimate on gender in the second column should be approximately zero. The results from Table 4 suggest the gender coefficient should be negative, and indeed, Table 5's results corroborate these findings. Column (III) of Table 5 then adds a control for classroom behavior in period t-1. Using the "Approaches to Learning" variable from the SRS assessment in period t-1 guarantees that different teachers assessed the child's behavior and academic achievement. We are less concerned with interpreting the coefficient estimate on behavior, as we are on how the gender coefficient estimate changes in this framework. We hold that gender differences in teacher grades are not necessarily causal, and more likely, represent a form of omitted variable bias. By teasing out lagged classroom behavior we expect the coefficient estimates on gender to converge to zero, and therefore, to eliminate the gaps by gender in test scores and teacher grades.

Reading

The top half of Table 5 gives the results for reading. Three time periods are shown – first, third, and fifth grade – because the data does not contain a pre-kindergarten assessment of behavior. The third column in each sequence shows the gender achievement gap in teacher grades when the estimation procedure conditions on the test score in the current period and behavior score from the previous period. For example, the gender gap estimate in reading grades for first grade students accounts for those students' behavior in the kindergarten year.

The results for white children demonstrate that controlling for behavior substantially reduces the assessment gap by gender in reading. Whereas boys earned 0.24 standard deviations lower reading grades in first grade, controlling for test scores reduces the gap to 0.16, but controlling for classroom behavior further reduces the gap to 0.09 standard deviations. This result occurs in the third and fifth grade years as well. Conditional on test scores, white boys earn approximately 0.17 standard deviations lower reading grades; however, conditioning on test scores and lagged classroom behavior cuts this estimate in half to 0.08 standard deviations.

The gender gaps for black and Hispanic children exhibit the same trend. Controlling for first grade behavior and third grade test scores, the gender gap in third grade reading is reduced from -0.23 to -0.17 standard deviations for black boys and from -0.18 to -0.10 standard deviations for Hispanic boys. This reduction occurs consistently from first to fifth grade.

The results for reading further show that the lagged SRS score has a positive and statistically significant effect on the teacher grade equation. The test scores and SRS scores are normalized to have a mean equal to zero and variance equal to one. With each standard deviation increase in the child's reading test score, the model predicts the child will earn 0.51-0.65

standard deviations higher reading grades. Similarly, with each standard deviation increase in lagged SRS score, the child will earn 0.17-0.21 standard deviations higher reading grades. Not surprisingly, the contemporaneous test score is more highly correlated with teacher grades than the lagged SRS score.

Controlling for lagged SRS increases the model's goodness of fit. Conditional on test scores, the R^2 measure for teacher grades is now approximately 0.44-0.52, depending on the grade level. Adding the lagged proxy for classroom behavior increases the model's goodness of fit by approximately three percentage points. This suggests that teachers implicitly or explicitly account for classroom behavior in their evaluations of students' academic progress.

Math and Science

Table 4 shows that boys earn the same or higher test scores in math and science, and yet those gender differences do not readily appear in teacher grades. Table 5 shows that controlling for test scores, white boys earn 0.07 standard deviations lower math grades, while black and Hispanic boys earn more than 0.22 standard deviations lower math grades in the third grade and beyond. Conditioning on previous behavior nearly eliminates this discrepancy, particularly for white children. Despite controlling for lagged behavior and test scores black boys continue to earn lower math grades than black girls; however, the differences are not as large had we excluded behavior from the model. Most of the gender gap in math assessments is eliminated for Hispanic boys once we control for behavior.

In science the gender gap in assessments is also large for white, black, and Hispanic children. White boys earn, on average, -0.11 to -0.17 standard deviations lower science grades, conditional on their test scores. Incorporating a lagged SRS score eliminates this difference.

Given science test scores and previous classroom behavior, there is no gender difference in teacher grades for white students. For black and Hispanic students, the gender gaps in science assessments are larger than those for white students. Accounting for lagged behavior eliminates this difference in most years (first and fifth grade), and substantially reduces it in the remaining years (third grade).

Like the model for reading grades, the lagged SRS scores have a positive and statistically significant effect on teacher grades in math and science. For each standard deviation increase in lagged SRS score, a student receives 0.17-0.25 standard deviations higher grades. Contemporaneous test scores also enter significantly. In science a one standard deviation increase in the child's test score leads to a 0.36-0.41 standard deviation increase in science grades. In math a one standard deviation increase in the child's test score leads to a 0.47-0.57 increase in the child's grade. The model's goodness of fit also increases quite significantly. Where the model previously explained anywhere from 12-15% of the variance in math and science grades, with test scores and lagged SRS scores it now explains 38-46% of the variance in math and 27-33% of the variance in science. Lagged SRS scores increase the model's goodness of fit by approximately three percentage points.

4.3 Instrumental Variables Estimation and Contemporaneous Classroom Behavior

Table 7 reports gender gaps in teacher grades conditional on the test score and an instrumented contemporaneous SRS score. This Table replicates Column (III) from Table 5, only with an instrumented contemporaneous SRS score rather than a lagged SRS score. Table 6 reports results from the first stage regression where the contemporaneous SRS score is regressed on the lagged SRS score, the contemporaneous test score, gender, and the model's remaining control variables. For each grade level and subject, the lagged SRS score has an independently positive and statistically significant effect on the current SRS score. This implies that some of children's behavior is persistent, and hence, the lagged SRS score is not a weak instrument for the current SRS score.

The top third of Table 7 gives results for reading grades. The gender gaps in teacher grades for white students entirely disappear in the first and third grade years. In fifth grade, white boys earn *higher* teacher grades by 0.12 standard deviations. Similarly the gender gap for black students is eliminated in first and third grade. Black boys earn 0.15 standard deviations lower reading grades in fifth grade, but this estimate lacks statistical significance and is lower than the estimate found in Column (III) of Table 5 where the model includes lagged SRS scores (-0.19 standard deviations). Hispanic boys earn slightly higher reading grades in all years, yet the estimates lack statistical significance beyond the first grade year. These results show that controlling for the student's reading test score and current classroom behavior eliminates the gender gap in reading grades, or in some cases, boys receive an advantage.

Both the test score and the instrumented SRS scores positively and precisely explain the variance in reading grades. In fact beyond the first grade year, current SRS scores are marginally more important than the reading test score in explaining the variance in teacher grades. In the fifth grade, for each standard deviation increase in the reading test, students are expected to earn 0.46 standard deviations higher grades, while for each standard deviation increase in the instrumented SRS scores students are expected to earn 0.51 standard deviations higher reading grades.

From Columns (III) of Table 5, we did not find differences between white boys and girls in their math and science grades. We also found smaller differences in these grades between black and Hispanic boys and girls when we used a lagged SRS score. In Table 7, however, with the instrumented SRS scores, white boys earn 0.15-0.22 standard deviations higher grades in math and 0.14-0.18 standard deviations higher grades in science. For Hispanic children there is no difference in math grades in the first and third grade years, then Hispanic boys earn higher grades in the fifth grade year. Hispanic boys also earn higher grades in science in all years. Black boys earn higher math (science) grades in the first (fifth) grade, but still lower grades in third (first) and fifth (third) grades. Though black boys continue to earn somewhat lower grades than black girls, these differences are smaller than those produced with the lagged SRS score in Table 5.

Contemporaneous test scores and SRS scores are equally significant determinants of math grades. Each standard deviation increase in the math test score (SRS score) leads to an increase of 0.33-0.50 (0.41-0.55) standard deviations in the math grade. In science the SRS score enters more significantly than the test score in the teacher grade equation. For each standard deviation increase in the SRS score (test score), we expect students to earn 0.48-0.59 (0.25-0.34) standard deviations higher science grades.

Why do boys earn higher grades when we use contemporaneous behavior scores? Table 5 shows that there is little difference between boys' and girls' grades when we use a lagged behavioral score. However, with the exception of first grade, the lagged score was recorded two years prior to the test scores and grades – in the third grade equation the SRS scores come from the first grade year, in the fifth grade equation they come from the third grade year. Although we compare differences in teacher grades for boys and girls who behaved the same way two years ago, this does not imply those children behave the same today. Indeed, the first-stage regressions reported in Table 6 show that conditional on the lagged behavioral score and today's test score boys are substantially more likely to receive lower behavioral scores today. Thus, when we condition on the lagged behavioral score in the teacher grade equation, we compare boys and girls who are somewhat similar today, because their behavioral scores were the same two years ago, but they are not entirely the same, because the boys' behavioral scores are likely lower today. The negative and statistically significant coefficient estimates for boys in Table 6 show that boys exhibit a consistent downward trend in their behavioral scores relative to girls. Thus, the results in Table 5 imply that boys and girls receive the same grades when boys have contemporaneous SRS scores that are lower than the girls' contemporaneous scores, as long as the differences in these scores is not considerably large. The results from Table 7 show that when boys and girls receive the same SRS scores today, boys earn higher grades. This may result from teachers' expectations for how groups of students should behave in the classroom. For example, if teachers expect boys to participate less frequently in class discussion, but there is a boy who participates at the same frequency as the average girl, he may be rewarded for such behavior. In sum, boys do not have to earn the same behavioral scores as girls to receive the same grades, but if they do, we expect them to be compensated for their behavior with higher grades.

5. Conclusion

The results of this paper confirm that a gender gap in educational achievement exists among children ages 5-12. The nature of the gap is consistent over time. Girls score higher than boys on reading tests. White boys score higher than white girls on math and science tests. Black and Hispanic boys score higher than black and Hispanic girls on science tests, but there are not statistically significant gender differences for these children on math tests. The gender gap is notably pronounced for outcomes of relatively subjective measurement, where teachers evaluate the students in the form of grades. Girls earn higher teacher grades than their test scores suggest, while boys earn lower grades than their test scores suggest. Since the objective and subjective assessments are intended to measure the same skills, these gender differences are particularly noteworthy. They also corroborate consistent findings that girls tend to earn higher grades than boys.

With lagged behavior in the model of teacher grades, the gender gaps in grades approach zero for white and Hispanic children, and they are considerably reduced for black children. Without accounting for lagged classroom behavior, boys earn significantly lower reading grades (0.16 standard deviations lower); with a lagged behavioral score their grades nearly equal the grades of girls. Similarly boys' math and science grades are lower without accounting for lagged behavior, yet statistically similar when the model includes lagged behavior.

When we instrument contemporaneous behavior scores with the lagged behavioral scores, the gender differences in reading grades are entirely eliminated for students of all races and ethnicities. White boys earn higher math and science grades than white girls, black boys earn the slightly lower math grades and higher science grades than black girls, and Hispanic boys earn the same math grades and higher science grades than Hispanic girls.

Cumulatively, these results suggest that classroom behavior matters. According to teachers, girls are better able to communicate a desire to learn. Teachers consciously or subconsciously reward these attitudes by giving girls higher marks for cognitive ability. When boys receive the same lagged behavioral score as girls, gender differences in teacher grades disappear. When boys receive the same contemporaneous behavioral score, boys earn *higher* teacher grades.

This paper demonstrates that the presence and magnitude of gender differences depend critically on the type of assessment used to evaluate academic aptitude. Classroom behavior is also decisively important for understanding gender differences in teacher grades. Teachers' assessments have long-term consequences for students. Teachers often have the ability to influence the child's class placement in the following school year; for example, teachers suggest grade retention or that the child be placed in remedial, on-level, or advanced classes at the next grade level. Teachers' assessments will also have implications for gender differences in the rates of high school graduation and college enrollment. Grades are the most important factor for deciding whether a student may graduate from high school. Moreover, many post-secondary schools emphasize the importance of high school grades and curriculum difficulty (e.g., Advanced Placement courses) for admission and financial aid. Thus, these early differences in teacher assessments may set boys and girls on different trajectories for educational attainment.

References

Anderson, Michael. 2006. "Multiple Inference and Gender Differences in the Effects of Preschool: A Reevaluation of the Abecedarian, Perry Preschool, and Early Training Projects." Working Paper.

Averett, Susan L. and Mark L. Burton. 1996. "College Attendance and the College Wage Premium: Differences by Gender." *Economics of Education Review,* Vol. 15, No. 1: 37-49.

Betts, Julian and Darlene Morell. 1999. "The Determinants of Undergraduate Grade Point Average: The Relative Importance of Family Background. High School Resources, and Peer Group Effects." *Journal of Human Resources,* Vol. 34, No. 2: 268-93.

Blau, Francine and Lawrence M. Kahn. 2006. "The U.S. Gender Pay Gap in the 1990s: Slowing Convergence." IZA Working Paper #2176.

Burgess, Simon and Ellen Greaves. 2009. "Test scores, Subjective Assessments and Stereotyping of Ethnic Minorities." Working Paper.

Clotfelter, Charlie T., Helen F. Ladd, and Jacob L. Vigdor. 2009. "The Academic Achievement Gap in Grades 3 to 8." *Review of Economics and Statistics*, Vol. 91, No. 2: 398-419.

Cho, Donghun. 2007. "The Role of High School Performance in Explaining Women's Rising College Enrollment." *Economics of Education Review*, Vol. 26, No. 4: 450-462.

Claessens, Amy, Greg Duncan, and Mimi Engel. 2009. "Kindergarten skills and fifth-grade achievement: Evidence from ECLS-K." *Economics of Education Review*, Vol. 28, No. 4: 415-427.

Cornwell, Christopher M., David B. Mustard, and Jessica Van Parys. 2009. "How Does the New SAT Predict Academic Achievement in College?" University of Georgia Working Paper.

Dynarski, Susan. 2007. "Cradle to College: The Puzzle of Gender Differences in Educational

Outcomes." Working Paper.

Frenette, Marc and Klarka Zeman. 2007. "Why Are Most University Students Women? Evidence Based on Academic Performance, Study Habits and Parental Influences." *Business and Labour Market Analysis*, No. 303: 5-25.

Freeman, Catherine. 2004. "Trends in Educational Equity of Girls and Women: 2004." National Center for Education Statistics, Statistical Analysis Report (NCES 2005-016): 1-109.

Fryer, Roland G. and Steven D. Levitt. 2004. "Understanding the Black-White Test Score Gap in the First Two Years of School." *Review of Economics and Statistics*, Vol. 86, No. 2: 447-464.

Fryer, Roland G. and Steven D. Levitt. 2009. "An Empirical Analysis of the Gender Gap in Mathematics." NBER Working Paper No. 15430.

Goldin, Claudia, Lawrence Katz, and Ilyana Kuziemko. 2006. "The Homecoming of American College Women: The Reversal of the College Gender Gap." *Journal of Economic Perspectives*, Vol. 20, No. 4: 133-156.

Hanushek, Eric A. and Steven G. Rivkin. 2009. "Harming the Best: How Schools Affect the Black-White Achievement Gap." *Journal of Policy Analysis and Management*, Vol. 28, No. 3: 366-393.

Heckman, James. 2008. "Schools, Skills, and Synapses." NBER Working Paper No. 14064.

Holmlund, Helena and Krister Sund. 2008. "Is the Gender Gap in School Performance Affected by the Sex of the Teacher?" *Labour Economics*, Vol. 15, No. 1: 37-53.

Husain, Muna and Daniel L. Millimet. 2009. "The Mythical 'Boy Crisis'?" *Economics of Education Review*, Vol. 28, No. 1: 38-48.

Jacob, Brian. 2002. "Where the Boys aren't: Non-cognitive Skills, Returns to School and the Gender Gap in Higher Education." *Economics of Education Review*, Vol. 21, No. 6: 589-598.

Jencks, Christopher and Meredith Phillips. 1998. *The Black-White Test Score Gap*. Brookings Institution Press, Washington, D.C.

Lavy, Victor. 2008. "Do Gender Stereotypes Reduce Girls' or Boys' Human Capital Outcomes? Evidence from a Natural Experiment." *Forthcoming in Journal of Public Economics.*

Lavy, Victor and Anal Schlosser. 2007. "Mechanisms and Impacts of Gender Peer Effects at School." NBER Working Paper No. 13292.

Loury, Linda Datcher. 2004. "Siblings and Gender Differences in African-American College Attendance." *Economics of Education Review*, Vol. 23, No. 3: 213–219.

O'Neill, June. 1985. "The Trend in the Male-Female Wage Gap in the US," *Journal of Labor Economics*, Vol. 3, No. 2: S91-S116.

Reardon, Sean F. and Claudia Galindo. 2009. "The Hispanic-White Achievement Gap in Math and Reading in the Elementary Grades." *American Educational Research Journal*, Vol. 46, No. 3: 853-891.

Reynolds, John R. and Stephanie Woodham Burge. 2007. "Educational Expectations and the Rise in Women's Post-secondary Attainments." *Social Science Research*, Vol. 37, No. 2: 485-499.

U.S. Department of Education, National Center for Education Statistics. 2009. Digest of Education Statistics, (NCES 2009-020), Chapter 3.

U.S. Department of Education, National Center for Education Statistics. 2002. User Manual for the ECLS-K First Grade Public-Use Data Files and Electronic Codebook (NCES 2002-

135).

Weinberger, Catherine and Peter Kuhn. 2006. "The Narrowing of the U.S. Gender Earnings Gap: 1959-1999: A Cohort-Based Analysis," Institute for the Study of Labor Discussion Paper No. 2007.

Table 1 Descriptive Statistics

Academic Achievement

	Female		Male	
Reading Scores	**Mean**	**Std. Dev.**	**Mean**	**Std. Dev.**
Kindergarten	42.71	13.68	40.30	13.69
First	76.36	21.40	72.58	21.69
Third	124.52	22.63	120.29	24.47
Fifth	144.17	20.81	141.02	23.06
Reading Grades				
Kindergarten	3.55	0.76	3.33	0.78
First	3.62	0.89	3.40	0.88
Third	3.51	0.84	3.27	0.85
Fifth	3.62	0.80	3.37	0.82
Math Scores				
Kindergarten	34.09	10.65	34.50	12.39
First	59.03	15.25	60.97	17.53
Third	93.47	19.72	97.89	21.01
Fifth	114.16	20.11	118.74	20.06
Math Grades				
Kindergarten	3.67	0.79	3.56	0.84
First	3.54	0.85	3.53	0.89
Third	3.14	0.71	3.15	0.73
Fifth	3.45	0.65	3.47	0.72
Science Scores				
Kindergarten	27.66	7.51	27.92	7.98
First	35.14	7.12	35.91	7.22
Third	45.64	13.15	48.77	13.82
Fifth	58.28	13.72	61.65	13.20
Science Grades				
Kindergarten	3.75	0.92	3.62	0.97
First	3.41	0.94	3.35	0.96
Third	3.26	0.89	3.25	0.92
Fifth	3.41	0.85	3.36	0.87
SRS Score				
Kindergarten	2.30	0.62	2.00	0.68
First	2.23	0.66	1.94	0.69
Third	2.27	0.61	1.95	0.66
Fifth	2.30	0.59	1.95	0.67

Table 2 Descriptive Statistics

Personal and Family Characteristics

Personal Characteristics	**K**	**First**	**Third**	**Fifth**
Male	0.51	0.50	0.50	0.50
	(0.50)	(0.50)	(0.50)	(0.50)
Black	0.13	0.12	0.09	0.08
	(0.34)	(0.32)	(0.29)	(0.28)
Hispanic	0.14	0.14	0.14	0.15
	(0.34)	(0.35)	(0.35)	(0.36)
Asian	0.04	0.04	0.04	0.05
	(0.21)	(0.21)	(0.20)	(0.22)
Pacific Islander	0.01	0.01	0.01	0.01
	(0.11)	(0.11)	(0.11)	(0.11)
Native American	0.02	0.01	0.01	0.01
	(0.13)	(0.12)	(0.11)	(0.12)
Mixed race	0.03	0.03	0.02	0.03
	(0.17)	(0.16)	(0.15)	(0.16)
Family Characteristics	**K**	**First**	**Third**	**Third**
WIC Benefits	0.37	0.34	0.30	0.31
	(0.48)	(0.47)	(0.46)	(0.46)
Teenage Mom	0.24	0.22	0.19	0.19
	(0.42)	(0.41)	(0.39)	(0.39)
Mom > 30 years old	0.13	0.13	0.14	0.15
	(0.33)	(0.34)	(0.35)	(0.35)
Age at K entry	65.70	65.79	65.79	65.73
	(4.13)	(4.15)	(4.19)	(4.17)
# Books in the home	79.44	110.12	132.66	116.08
	(60.18)	(152.56)	(185.52)	(178.60)
SES Index	0.08	0.09	0.09	0.08
	(0.77)	(0.78)	(0.77)	(0.79)
Observations	10604	9299	6362	6496

Notes: Standard errors are in parentheses.

Table 3 Descriptive Statistics

Teacher and School Characteristics

Teacher Characteristics	**K**	**First**	**Third**	**Fifth**
Teacher experience	9.06	14.85	15.11	14.62
	(7.61)	(10.09)	(10.07)	(10.22)
Teacher education	2.11	2.13	2.20	2.24
	(0.90)	(0.93)	(0.92)	(0.92)
School Characteristics	**K**	**First**	**Third**	**Fifth**
Public school	0.80	0.79	0.78	0.78
	(0.40)	(0.41)	(0.42)	(0.41)
Urban school	0.38	0.37	0.34	0.35
	(0.48)	(0.48)	(0.47)	(0.48)
Rural school	0.24	0.23	0.26	0.26
	(0.43)	(0.42)	(0.44)	(0.44)
Southern school	0.34	0.36	0.30	0.29
	(0.47)	(0.48)	(0.46)	(0.45)
% Minority < 10	0.36	0.36	0.40	0.37
	(0.48)	(0.48)	(0.49)	(0.48)
% Minority 10-25	0.20	0.20	0.20	0.19
	(0.40)	(0.40)	(0.40)	(0.40)
% Minority 25-50	0.16	0.16	0.16	0.17
	(0.37)	(0.37)	(0.36)	(0.38)
% Minority 50-75	0.10	0.10	0.09	0.08
	(0.29)	(0.29)	(0.28)	(0.27)
% Minority >75	0.18	0.18	0.16	0.18
	(0.39)	(0.39)	(0.37)	(0.39)
Observations	10604	9494	6658	6496

Notes: Standard errors are in parentheses.

Table 4 Results
Gender Gap in Test Scores and Teacher Grades by Race and Ethnicity

	Kinder. (Spring)			**First Grade (Spring)**			**Third Grade (Spring)**			**Fifth Grade (Spring)**		
I. Reading												
	Test	**Grade**	**Diff.**	**Test**	**Grade**	**Diff.**	**Test**	**Grade**	**Diff.**	**Test**	**Grade**	**Diff.**
White	-0.164*	-0.255*	-0.091*	-0.129*	-0.247*	-0.118*	-0.130*	-0.258*	-0.128*	-0.161*	-0.285*	-0.124*
	(0.025)	(0.024)		(0.033)	(0.032)		(0.038)	(0.040)		(0.057)	(0.057)	
Black	-0.155*	-0.349*	-0.194*	-0.169*	-0.239*	-0.070	-0.245*	-0.402*	-0.157+	-0.260*	-0.373*	-0.113*
	(0.046)	(0.053)		(0.068)	(0.080)		(0.081)	(0.091)		(0.120)	(0.131)	
Hispanic	-0.173*	-0.220*	-0.047	-0.233*	-0.208*	0.025	-0.252*	-0.333*	-0.081	-0.358*	-0.375*	-0.017
	(0.043)	(0.054)		(0.064)	(0.072)	-	(0.083)	(0.090)		(0.096)	(0.077)	
R^2	0.18	0.16		0.19	0.14		0.29	0.17		0.31	0.17	
N	10604	10604		9299	9299		6362	6362		6496	6496	

	Kinder. (Spring)			**First Grade (Spring)**			**Third Grade (Spring)**			**Fifth Grade (Spring)**		
II. Math												
	Test	**Grade**	**Diff.**	**Test**	**Grade**	**Diff.**	**Test**	**Grade**	**Diff.**	**Test**	**Grade**	**Diff.**
White	0.057*	-0.125*	-0.182*	0.143*	-0.002	-0.145*	0.276*	0.094*	-0.182*	0.192*	0.022	-0.170*
	(0.024)	(0.025)		(0.031)	(0.032)		(0.037)	(0.042)		(0.062)	(0.063)	
Black	-0.057	-0.278*	-0.221*	-0.076	-0.113	-0.037	0.033	-0.243*	-0.276*	0.178	-0.281	-0.459*
	(0.041)	(0.055)		(0.058)	(0.080)		(0.086)	(0.091)		(0.137)	(0.186)	
Hispanic	-0.028	-0.088	-0.060	0.094	-0.061	-0.155*	0.200*	-0.125	-0.325*	0.022	-0.183	-0.205*
	(0.044)	(0.055)		(0.065)	(0.067)		(0.077)	(0.089)		(0.105)	(0.114)	
R^2	0.25	0.15		0.23	0.13		0.29	0.14		0.29	0.15	
N	10604	10604		9299	9299		6362	6362		6441	3148	

Notes:
(a) Difference = Grade - Test Score; White = White male - white female; Black = Black male - black female; Hispanic = Hispanic male - Hispanic female
(b) Coefficient estimates are reported in standard deviations. I.e., test scores and grades are normalized to have mean=0 and variance=1.
(c) (*) and (+) indicate statistical significance above the 5% and 10% confidence levels, respectively.
(d) All regressions control for family, teacher, and school characteristics.
(e) Standard errors are in parentheses.
(f) N is the entire sample size, including white, black, and Hispanic students.

Table 4 Continued

Gender Gap in Test Scores and Teacher Grades by Race and Ethnicity

	Kinder. (Spring)			**First Grade (Spring)**			**Third Grade (Spring)**			**Fifth Grade (Spring)**		
III. Science												
	Test	**Grade**	**Diff.**	**Test**	**Grade**	**Diff.**	**Test**	**Grade**	**Diff.**	**Test**	**Grade**	**Diff.**
White	0.059*	-0.106*	-0.165*	0.111*	-0.063+	-0.174*	0.282*	0.006	-0.276*	0.248*	-0.026	-0.274*
	(0.021)	(0.024)		(0.025)	(0.033)		(0.035)	(0.042)		(0.053)	(0.074)	
Black	-0.020	-0.273*	-0.253*	0.046	-0.132	-0.178*	0.225*	-0.223*	-0.448*	0.560*	0.070	-0.490*
	(0.044)	(0.054)		(0.068)	(0.082)		(0.066)	(0.091)		(0.131)	(0.150)	
Hispanic	0.052	-0.109*	-0.161*	0.052	-0.031	-0.083	0.131*	-0.041	-0.172*	0.290*	-0.207+	-0.497*
	(0.045)	(0.052)		(0.065)	(0.070)		(0.069)	(0.086)		(0.103)	(0.116)	
R^2	0.41	0.14		0.36	0.12		0.37	0.14		0.39	0.14	
N	10604	10604		9299	9299		6362	6362		6441	3039	

Notes:

(a) Difference = Grade - Test Score; White = White male - white female; Black = Black male - black female; Hispanic = Hispanic male - Hispanic female

(b) Coefficient estimates are reported in standard deviations. I.e., test scores and grades are normalized to have mean=0 and variance=1.

(c) (*) and (+) indicate statistical significance above the 5% and 10% confidence levels, respectively.

(d) All regressions control for family, teacher, and school characteristics.

(e) Standard errors are in parentheses.

(f) N is the entire sample size, including white, black, and Hispanic students.

Table 5: Gender Gap in Teacher Grades
(I) Baseline, (II) Conditional on Test Scores, and (III) Conditional on Test Scores and the Lagged SRS Score

	First Grade (Spring)			Third Grade (Spring)			Fifth Grade (Spring)		
I. Reading	**(I)**	**(II)**	**(III)**	**(I)**	**(II)**	**(III)**	**(I)**	**(II)**	**(III)**
White	-0.247*	-0.161*	-0.089*	-0.258*	-0.171*	-0.087*	-0.285*	-0.185*	-0.067
	(0.032)	(0.023)	(0.023)	(0.040)	(0.032)	(0.030)	(0.057)	(0.045)	(0.046)
Black	-0.239*	-0.091+	-0.039	-0.402*	-0.227*	-0.166*	-0.373*	-0.207*	-0.191+
	(0.080)	(0.054)	(0.055)	(0.091)	(0.071)	(0.069)	(0.131)	(0.100)	(0.106)
Hispanic	-0.208*	-0.031	0.003	-0.333*	-0.182*	-0.103	-0.375*	-0.197*	-0.115+
	(0.072)	(0.053)	(0.053)	(0.090)	(0.077)	(0.078)	(0.077)	(0.068)	(0.067)
Test Score$_t$		0.715*	0.654*		0.663*	0.574*		0.600*	0.511*
		(0.011)	(0.012)		(0.018)	(0.018)		(0.021)	(0.024)
SRS Score$_{t-1}$			0.171*			0.216*			0.218*
			(0.011)			(0.015)			(0.024)
R^2	0.14	0.52	0.55	0.17	0.47	0.50	0.17	0.44	0.48
N	9299	9299	9299	6362	6362	6362	6496	6496	6496

	First Grade (Spring)			Third Grade (Spring)			Fifth Grade (Spring)		
II. Math	**(I)**	**(II)**	**(III)**	**(I)**	**(II)**	**(III)**	**(I)**	**(II)**	**(III)**
White	-0.002	-0.072*	0.035	0.094*	-0.072+	0.033	0.022	-0.094+	0.006
	(0.032)	(0.028)	(0.028)	(0.042)	(0.038)	(0.040)	(0.063)	(0.052)	(0.048)
Black	-0.113	-0.056	0.040	-0.243*	-0.260*	-0.223*	-0.281	-0.407*	-0.250+
	(0.080)	(0.067)	(0.064)	(0.091)	(0.083)	(0.082)	(0.186)	(0.149)	(0.140)
Hispanic	-0.061	-0.115*	-0.075	-0.125	-0.229*	-0.151+	-0.183	-0.213*	-0.031
	(0.067)	(0.060)	(0.062)	(0.089)	(0.078)	(0.080)	(0.114)	(0.079)	(0.086)
Test Score$_t$		0.545*	0.471*		0.574*	0.502*		0.641*	0.566*
		(0.015)	(0.015)		(0.018)	(0.020)		(0.028)	(0.029)
SRS Score$_{t-1}$			0.210*			0.169*			0.179*
			(0.013)			(0.0188)			(0.024)
R^2	0.14	0.35	0.38	0.14	0.36	0.38	0.15	0.44	0.46
N	9299	9299	9299	6362	6362	6362	3148	3148	3148

Table 5 Continued: Gender Gap in Teacher Grades
(I) Baseline, (II) Conditional on Test Scores, and (III) Conditional on Test Scores and the Lagged SRS Score

	First Grade (Spring)			Third Grade (Spring)			Fifth Grade (Spring)		
III. Science	**(I)**	**(II)**	**(III)**	**(I)**	**(II)**	**(III)**	**(I)**	**(II)**	**(III)**
White	-0.063+	-0.110*	0.010	0.006	-0.117*	-0.000	-0.026	-0.165*	-0.032
	(0.033)	(0.031)	(0.032)	(0.042)	(0.040)	(0.041)	(0.074)	(0.069)	(0.069)
Black	-0.132	-0.154*	-0.042	-0.223*	-0.352*	-0.250*	0.070	-0.209	-0.109
	(0.082)	(0.074)	(0.074)	(0.091)	(0.084)	(0.084)	(0.150)	(0.133)	(0.147)
Hispanic	-0.031	-0.052	0.002	-0.041	-0.106	-0.006	-0.207+	-0.312*	-0.158
	(0.070)	(0.064)	(0.065)	(0.086)	(0.073)	(0.078)	(0.116)	(0.108)	(0.101)
Test Score$_t$		0.426*	0.359*		0.457*	0.390*		0.492*	0.405*
		(0.018)	(0.018)		(0.020)	(0.020)		(0.036)	(0.038)
SRS Score$_{t-1}$			0.241*			0.220*			0.254*
			(0.015)			(0.018)			(0.030)
R^2	0.12	0.23	0.27	0.14	0.27	0.30	0.14	0.28	0.33
N	9299	9299	9299	6362	6362	6362	3039	3039	3039

Notes:
(a) Standard errors are in parentheses.
(b) All regressions control for family, teacher, and school characteristics.
(c) (*) and (+) indicate statistical significance above the 5% and 10% confidence levels, respectively.
(d) Coefficient estimates on Male, Test Scores, and SRS Scores are reported in standard deviations.
(e) Model (I) explains the variance in teacher grades.
(f) Model (II) explains the variance in teacher grades, conditional on the student's test score.
(g) Model (II) explains the variance in teacher grades, conditional on the student's test score and the previous period's SRS score.

Table 6: First Stage Regressions
Lagged SRS Score as an Instrument for Contemporaneous SRS Score

I. Reading	SRS – First Grade	SRS – Third Grade	SRS – Fifth Grade
SRS_{t-1}	0.374*	0.412*	0.425*
	(0.014)	(0.016)	(0.025)
Test $Score_t$	0.270*	0.248*	0.096*
	(0.015)	(0.018)	(0.029)
Male	-0.248*	-0.262*	-0.348*
	(0.028)	(0.033)	(0.048)
R^2	0.37	0.41	0.38
N	9282	6350	6465

II. Math	SRS – First Grade	SRS – Third Grade	SRS – Fifth Grade
SRS_{t-1}	0.383*	0.414*	0.405*
	(0.014)	(0.017)	(0.025)
Test $Score_t$	0.253*	0.228*	0.160*
	(0.015)	(0.018)	(0.027)
Male	-0.316*	-0.357*	-0.374*
	(0.028)	(0.034)	(0.048)
R^2	0.36	0.40	0.39
N	9298	6350	6410

III. Science	SRS – First Grade	SRS – Third Grade	SRS – Fifth Grade
SRS_{t-1}	0.425*	0.458*	0.439*
	(0.014)	(0.017)	(0.028)
Test $Score_t$	0.189*	0.170*	0.111*
	(0.017)	(0.018)	(0.033)
Male	-0.282*	-0.324*	-0.383*
	(0.028)	(0.035)	(0.057)
R^2	0.34	0.39	0.38
N	9282	6350	6410

Notes:
(a) Standard errors are in parentheses.
(b) All regressions control for family, teacher, and school characteristics.
(c) (*) and (+) indicate statistical significance above the 5% and 10% confidence levels, respectively.
(d) Coefficient estimates on Male, Test Scores, and SRS Scores are reported in standard deviations.
(e) SRS_{t-1} is the behavior score from the previous period's behavioral assessment

Table 7: Instrumental Variables Estimation
Gender Differences in Teacher Grades, Conditional on Contemporaneous Test and SRS Scores

I. Reading	**First Grade**	**Third Grade**	**Fifth Grade**
White	0.037+	0.038	0.117*
	(0.023)	(0.030)	(0.052)
Black	0.002	-0.002	-0.146
	(0.052)	(0.079)	(0.118)
Hispanic	0.132*	0.096	0.104
	(0.055)	(0.076)	(0.076)
Test $Score_t$	0.531*	0.444*	0.462*
	(0.016)	(0.020)	(0.028)
SRS $Score_t$	0.456*	0.523*	0.512*
	(0.027)	(0.032)	(0.057)
R^2	0.63	0.57	0.55
N	9282	6350	6465

II. Math	**First Grade**	**Third Grade**	**Fifth Grade**
White	0.222*	0.180*	0.150*
	(0.033)	(0.045)	(0.059)
Black	0.118*	-0.146	-0.183
	(0.059)	(0.103)	(0.146)
Hispanic	0.048	0.043	0.181
	(0.074)	(0.100)	(0.118)
Test $Score_t$	0.333*	0.410*	0.500*
	(0.018)	(0.025)	(0.031)
SRS $Score_t$	0.548*	0.409*	0.434*
	(0.034)	(0.042)	(0.057)
R^2	0.47	0.44	0.47
N	9282	6350	3135

Notes:
(a) Standard errors are in parentheses.
(b) All regressions control for family, teacher, and school characteristics.
(c) (*) and (+) indicate statistical significance above the 5% and 10% confidence levels, respectively.
(d) Coefficient estimates on Male, Test Scores, and SRS Scores are reported in standard deviations.

Table 7 Continued: Instrumental Variables Estimation
Gender Differences in Teacher Grades, Conditional on Contemporaneous Test and SRS Scores

III. Science	**First Grade**	**Third Grade**	**Fifth Grade**
White	0.188*	0.143*	0.178*
	(0.037)	(0.042)	(0.073)
Black	0.042	-0.066	0.479+
	(0.078)	(0.105)	(0.278)
Hispanic	0.148*	0.196*	0.206
	(0.072)	(0.091)	(0.146)
Test $Score_t$	0.252*	0.308*	0.340*
	(0.020)	(0.021)	(0.040)
SRS $Score_t$	0.568*	0.480*	0.592*
	(0.035)	(0.037)	(0.070)
R^2	0.35	0.36	0.33
N	9282	6350	3021

Notes:
(a) Standard errors are in parentheses.
(b) All regressions control for family, teacher, and school characteristics.
(c) (*) and (+) indicate statistical significance above the 5% and 10% confidence levels, respectively.
(d) Coefficient estimates on Male, Test Scores, and SRS Scores are reported in standard deviations.