

Cornwell, Christopher; Mustard, David B.; Van Parys, Jessica

**Working Paper**

## Non-cognitive skills and the gender disparities in test scores and teacher assessments: Evidence from primary school

IZA Discussion Papers, No. 5973

**Provided in Cooperation with:**

IZA – Institute of Labor Economics

*Suggested Citation:* Cornwell, Christopher; Mustard, David B.; Van Parys, Jessica (2011) : Non-cognitive skills and the gender disparities in test scores and teacher assessments: Evidence from primary school, IZA Discussion Papers, No. 5973, Institute for the Study of Labor (IZA), Bonn, <https://nbn-resolving.de/urn:nbn:de:101:1-201110102225>

This Version is available at:

<https://hdl.handle.net/10419/55134>

**Standard-Nutzungsbedingungen:**

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

**Terms of use:**

*Documents in EconStor may be saved and copied for your personal and scholarly purposes.*

*You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.*

*If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.*

IZA DP No. 5973

**Non-cognitive Skills and the Gender Disparities in  
Test Scores and Teacher Assessments:  
Evidence from Primary School**

Christopher M. Cornwell  
David B. Mustard  
Jessica Van Parys

September 2011

# **Non-cognitive Skills and the Gender Disparities in Test Scores and Teacher Assessments: Evidence from Primary School**

**Christopher M. Cornwell**

*University of Georgia*

**David B. Mustard**

*University of Georgia and IZA*

**Jessica Van Parys**

*Columbia University*

Discussion Paper No. 5973

September 2011

IZA

P.O. Box 7240  
53072 Bonn  
Germany

Phone: +49-228-3894-0

Fax: +49-228-3894-180

E-mail: [iza@iza.org](mailto:iza@iza.org)

Any opinions expressed here are those of the author(s) and not those of IZA. Research published in this series may include views on policy, but the institute itself takes no institutional policy positions.

The Institute for the Study of Labor (IZA) in Bonn is a local and virtual international research center and a place of communication between science, politics and business. IZA is an independent nonprofit organization supported by Deutsche Post Foundation. The center is associated with the University of Bonn and offers a stimulating research environment through its international network, workshops and conferences, data service, project support, research visits and doctoral program. IZA engages in (i) original and internationally competitive research in all fields of labor economics, (ii) development of policy concepts, and (iii) dissemination of research results and concepts to the interested public.

IZA Discussion Papers often represent preliminary work and are circulated to encourage discussion. Citation of such a paper should account for its provisional character. A revised version may be available directly from the author.

## ABSTRACT

### **Non-cognitive Skills and the Gender Disparities in Test Scores and Teacher Assessments: Evidence from Primary School**

We extend the analysis of early-emerging gender differences in academic achievement to include both (objective) test scores and (subjective) teacher assessments. Using data from the 1998-99 ECLS-K cohort, we show that the grades awarded by teachers are not aligned with test scores, with the disparities in grading exceeding those in testing outcomes and uniformly favoring girls, and that the misalignment of grades and test scores can be linked to gender differences in non-cognitive development. Girls in every racial category outperform boys on reading tests and the differences are statistically significant in every case except for black fifth-graders. Boys score at least as well on math and science tests as girls, with the strongest evidence of a gender gap appearing among whites. However, boys in all racial categories across all subject areas are not represented in grade distributions where their test scores would predict. Even those boys who perform equally as well as girls on reading, math and science tests are nevertheless graded less favorably by their teachers, but this less favorable treatment essentially vanishes when non-cognitive skills are taken into account. White boys who perform on par with white girls on these subject-area tests and exhibit the same non-cognitive skill level are graded similarly. For some specifications there is evidence of a grade “bonus” for white boys with test scores and behavior like their girl counterparts. While the evidence is a little weaker for blacks and Hispanics, the message is essentially the same.

JEL Classification: I21, I24

Keywords: gender differences, test scores, grades, educational attainment

Corresponding author:

Christopher M. Cornwell  
Department of Economics  
University of Georgia  
Athens, GA 30606  
USA  
E-mail: [cornwl@uga.edu](mailto:cornwl@uga.edu)

## 1. Introduction

“‘The challenge for all of us is how to engage these guys at a younger age,’ said Nancy Beane, a college counselor from the Westminster Schools in Atlanta.”<sup>2</sup>

This college counselor was responding to the problem of the large and still growing gender gap in college enrollment. The disparity in educational attainment between males and females has been so widely reported in recent years that the basic facts are now well known and are driving public policy debate.<sup>3</sup> As summarized in Goldin, Katz and Kuziemko (2006), the ratio of males to females graduating from a four-year college stood at 1.60 in 1960, fell to parity by 1980, and continued its decline to .74 in 2003. Thus, by 2003, there were 135 females for every 100 males who graduated from a four-year college. Not surprisingly, the gender gap in college degrees awarded is linked to differences in college attendance. In 1960, the male-female undergraduate ratio was 1.55; by 2003, it had fallen to .77. Heckman and LaFontaine (2010) show that as much as half of the current gender gap in college attendance can be linked to lower rates of high-school graduation among males, a pattern that is especially pronounced for blacks. This finding raises the question of why boys lag behind girls in high-school completion. In this paper, we push that question back to primary school, identify gender disparities as early as kindergarten and focus on the role of non-cognitive factors.

Most empirical research of the gender gap in academic achievement concentrates on disparities in post-secondary outcomes as a function of (mostly) secondary school factors.<sup>4</sup> In contrast, only a few studies (for example, Anderson 2008; Lavy and Schlosser 2011; Holmlund

---

<sup>2</sup> [“Addressing the Gender in College Aspirations”](#), *The Choice* blog, *New York Times*, 23 Oct 09.

<sup>3</sup> Kay Hymowitz (2011) *Manning Up: How the Rise of Women Has Turned Men into Boys*, is just one example.

<sup>4</sup> Some noteworthy examples are Jacob (2002), Loury (2004), Cho (2007), Goldin et al. (2006), Dynarski (2007), Frenette and Zeman (2007) and Reynolds and Burge (2007).

and Sund 2008; Husain and Millimet 2009, Fryer and Levitt 2010) examine gender differences in achievement prior to the eighth grade. These papers report gender differences in reading and math test scores as early as kindergarten. Some of the explanations offered for these differences include the gender of teacher, the ratio of boys to girls in a classroom, and whether the children attended pre-school.

Figure 1 depicts the estimated gender and race gaps in reading, math and science test scores from our ECLS-K (Early Childhood Longitudinal Study – Kindergarten) kindergarten sample, conditional on a range of personal, family and school characteristics described in section 2. Even after netting out the effects of other factors, gender differences in reading, math, and science emerge early. In addition, the gender gap in reading—which favors girls—is over 50% larger than the corresponding black and Hispanic achievement gaps. The estimated gender gap increases into the fifth grade and becomes larger in magnitude than the Hispanic gap in every subject.

Figure 2 replicates Figure 1, except the achievement measure is now a teacher’s subjective assessment of the student’s performance. The contrast with the test-score gaps is striking. The gender differences in grades emerge early in all subject areas and favor girls in every subject. Because boys out-perform girls on math and science test scores, it is surprising that girls out-perform boys on teacher grades in math and science by nearly 0.15 standard deviations. Even more surprising is that the girl-boy gap in reading grades is over 300 percent larger than the white-black reading gap and the girl-boy gaps in math and science teacher grades are about 40 percent larger than the corresponding white-black grade gaps.

This paper makes three important contributions to the research on gender differences in academic achievement. First, we extend the analysis beyond the usual emphasis on test scores to

teacher grades. This is the first paper to examine gender differences in the academic performance of primary-school students using both subjective and objective measures achievement.<sup>5</sup> While standardized tests are important, teacher-assigned grades are arguably more consequential, given the role they play in class placement, high-school graduation and college admissibility. College and university admissions generally place considerably more weight on grades because they are better predictors of college performance (Betts and Morrell 1999; Cornwell et al. 2009). We show that teachers' assessments are not aligned with test-score data, with greater gender disparities in appearing in grading than testing outcomes.

Second, we trace the misalignment of teacher grades and test scores to differences between boys and girls in their non-cognitive development, and in doing so, solve a puzzle. Unlike racial and ethnic gaps that are considerably reduced when one controls for family and school characteristics, including such control variables does little to reduce the gender gap, because there is much less difference in family and school characteristics between girls and boys than whites and blacks. We document that girls are substantially more amenable to the learning process than boys, and that this non-cognitive skill is a significant factor in teacher assessments, even after controlling for test outcomes.<sup>6</sup> Our findings are consistent with Claessens et al. (2009), who report that a range of socio-emotional skills in kindergarten affect children's standardized test scores in fifth grade.

Third, compared with previous studies of achievement differences between primary-school-aged boys and girls, we consider more testing areas (adding science) and follow students

---

<sup>5</sup> 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.

<sup>6</sup> Teachers are unaware of students' test scores when making their assessments.

longer (through fifth grade). Consequently, this paper is far more comprehensive than its antecedents.

Our analysis is based on data from the 1998-99 ECLS-K cohort administered by the National Center for Education Statistics (NCES). For the kindergarten through fifth-grade, we first present evidence on gender differences in reading, math and science test scores and their evolution as children advance through primary school. Then, we examine the relationship between the (objective) test-score differences and (subjective) teacher grades. Finally, we investigate the role of non-cognitive skills, as measured by the social rating indices contained in the ECLS-K, in explaining achievement differences.

Our findings can be summarized as follows. First, girls in every racial category outperform boys on reading tests and the differences are statistically significant in every case except for black fifth-graders. In general, boys score at least as well on math tests as girls, but the evidence for a gender gap is weaker than in reading. The strongest case exists among whites, where statistically significant performance differences emerge in kindergarten and persist through the fifth grade.

Second, given their test-score results, girls predictably receive higher reading grades than boys, but the gender disparities in grades are typically much larger. Boys occupy places in the grade distribution even lower than those in the test-score distribution. The story is similar in math and science. Despite performing as least as well as girls on math tests, and significantly better on science tests, boys are not commensurately graded by their teachers. Boys in all racial categories are not represented in the math and science grade distributions as their test scores would predict.



Third, statistically significant gender differences in teacher grades still exist even after controlling for test scores. For example, conditional on the reading test score, as much as two-thirds of the grading disparity for white children remains.

However, this inconsistency between test scores and grades is largely accounted for by non-cognitive skills. White boys who perform as well as white girls on these subject-area tests and exhibit the same attitude towards learning as white girls in the classroom are graded similarly. For some specifications there is evidence of a grade “bonus” for white boys with test scores and behavior like their girl counterparts. While the evidence is a little weaker for blacks and Hispanics, the message is essentially the same.

## **2. The ECLS-K Data**

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 drawn. 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 report 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 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 ECLS-K longitudinal file begins with approximately 17,000 observations and concludes with roughly 9,000 observations in the fifth grade. Of the 17,000, only children who passed an English language screening test were administered the reading, math, and science assessments. Our sample begins with about 9,400 observations in kindergarten and concludes with 5,800 observations in the fifth grade, and includes all observations with valid data. We restrict the sample to white, black, and Hispanic children, since those groups are our populations of interest. There are 5,841 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, NCES randomly collected grades for half of the students taking math and half of the students taking science. This resulted in 2,820 observations on students with math test scores and grades, and 2,747 observations on students with science test scores and grades.

As our analysis advances from kindergarten through fifth grade, it is important to consider how attrition might affect our results. It is true that blacks are more likely to leave the sample, but we do not make comparisons between racial or ethnic groups. Also, children who remain in the sample through fifth grade (non-movers) have, on average, higher kindergarten test scores and grades than children who do not (movers). However, boys and girls are equally likely to be movers, so the difference-in-differences in kindergarten achievement between gender and moving status are not statistically significant. As long as the reasons for moving out of the sample are unrelated to gender, our analysis should be largely unaffected, except for a decrease in precision.

NCES prepared the objective reading, math, and science assessments. Each test was divided into two parts. How well a 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. Still, higher scores indicate higher levels of academic achievement.

Academic achievement was also measured with subjective assessments. Teachers rated each student's mastery of specific skills in reading, math, and science. NCES translated these assessments into "grades" by constructing a continuous 0-4 point "Academic Rating Scale" (ARS) scale where 0 indicates no understanding of the content or skill and 4 indicates complete mastery. The ARS measures the same skills as those found on the objective reading, math, and science assessments. Significantly for us, teachers were unaware of their students' test scores when they provided their assessments for the ARS.

In addition, teachers rated their children along several dimensions of classroom behavior that reflect non-cognitive skills. For example, teachers reported how well each child was engaged in the classroom, how often the child externalized or internalized problems, how often the child lost control, and how well the child developed interpersonal skills. NCES combined the answers to such questions to create a continuous 0-3 point "Social Rating Scale" (SRS) for measuring "Approaches to Learning," "Self-Control," "Internalizing Problems," "Externalizing Problems," and "Interpersonal Skills." In this paper, we focus on the SRS for "Approaches to Learning" (ATL) as our non-cognitive-skill measure.<sup>7</sup> As with the ARS scale higher SRS scores represent higher skill levels.

---

<sup>7</sup> In our empirical work with the ECLS-K, we examined the effects of all five SRS indices. The ATL measure exhibited the greatest economic and statistical significance.

Table 1 reports descriptive statistics for the test scores and teacher grades for reading, math and science, and SRS scores for ATL, by gender. Several empirical facts are readily apparent. First, girls score higher than boys on reading tests at every grade level, while boys perform better on math and science tests. Second, girls receive higher grades on average than boys in reading, consistent with reading test scores, but receive higher grades in science and comparable grades in math, despite lower average test scores in those subjects. Third, the average ATL rating for girls is consistently about 15% greater than the average score for boys. Finally, boys generally have higher variance in test scores, teacher grades and non-cognitive skill ratings; the standard deviation of male achievement is typically greater across subjects and grade levels.

### 3. Baseline Achievement Regressions

To examine the relationship between gender and academic achievement, we estimate empirical models of the form

$$y_i = \delta_0 + \delta_1 male_i + X_i\beta + u_i \quad (1)$$

where  $y$  is either a test score and or teacher-assigned grade (defined on the ARS) for student  $i$  in reading, math, and science. We regress the achievement measures on a gender (male) indicator and a set of family, teacher, and school characteristics ( $X$ ), separately for whites, blacks and Hispanics by grade level. In each case, we incorporate the NCES sample weights in estimation and report OLS standard errors that reflect their use.<sup>8</sup> Each cross-sectional wave includes students who were assessed in the spring of that school year.

---

<sup>8</sup> In particular, we employ the jackknife procedure provided by the Stata `svy` command. Inference is unaffected if we use heteroscedasticity-robust standard errors instead.

The ECLS-K provides information on a range 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 whether the mother received WIC (Women, Infants and Children supplemental nutritional) benefits during pregnancy. The socioeconomic (SES) index is comprised of five variables: family income, the parents' highest levels of educational attainment, and the parents' occupational prestige rankings. The index is normalized to have a mean of zero and standard deviation equal to one. These family characteristics are the same as those used by Fryer and Levitt (2004) to evaluate the black-white achievement gap.

In addition, the ECLS-K supplies important information about a child's teacher and school. For teachers, their highest level of educational attainment and years of experience are reported. Teachers are categorized as having either a bachelor's degree, some additional training beyond a bachelor's degree, a master's degree, or another advanced degree such as a PhD. For schools, we are provided public/private status, location information (whether urban, suburban or rural, and whether located in the south) and the share of the student body that is a racial minority.

Tables 2 and 3 give the descriptive statistics for the family, teacher and school characteristics, as well as the gender and racial breakdown, by grade. Table 2 shows that the sample is gender-balanced in all grades. Blacks comprise 13% of kindergartners, but only 8% of fifth-graders, as they experience the greatest attrition from the ECLS-K. Hispanics make up a consistent 14-15% of the sample. In the first wave, 37% of the children's mothers received WIC benefits; 24% of the mothers were teenagers at first birth; the average age at kindergarten entry was 65.7 months; there were an average of 80 books in the home; and the average SES index stood at .08. These characteristics follow patterns from the first wave that are to be expected

with aging households and sample attrition. By the fifth grade, 6,496 of the original 10,604 children remain in the sample.

Table 3 shows that kindergarten teachers average 9 years of experience, while first-fifth grade teachers average 5-6 years more, and the typical teacher in each grade has some certification beyond a bachelor's degree, but less than a master's degree.<sup>9</sup> About 80% of the schools in the sample are public; 38% are located in urban districts and 24% in rural districts; about a third are located in the south; and just over 25% have student bodies in which a racial minority holds at least a 50% share.

#### **4. Baseline Findings**

Tables 4A-C report our baseline results by subject area, grade level and race. For each subject area and grade level, we report the estimated coefficient of the male dummy ( $\delta_1$ ) from test-score and teacher-grade regressions for whites, blacks and Hispanics. In every case, test scores and grades are normalized to have zero means and unit variances, so the estimated coefficients can be interpreted as the effects of standard deviation changes.

##### *Reading*

The results for reading test scores and grades are presented in Table 4A. First consider test scores. Girls in all racial categories outperform boys on reading tests and the differences are statistically significant at the 5% level in every case but for fifth-grade blacks. Beginning in kindergarten, white boys score 0.16 standard deviations lower than white girls on reading tests, but the gap falls to 0.11 standard deviations by the fifth grade. Black and Hispanic boys also

---

<sup>9</sup> Educational attainment is defined over 5 categories; a value of "2" indicates at least one year beyond a bachelor's degree and value of "3" indicates a master's degree.

score lower than their girl counterparts on the reading tests. These disparities start at roughly the same level as whites, but in contrast to white children, the gaps grow in the years beyond kindergarten. By grade three, the gender gap in test scores rises to 0.27 standard deviations for blacks and 0.32 standard deviations for Hispanics. For Hispanics, the difference remains above 0.32 in the fifth grade.

Now turn to teacher-assigned grades. Given the test-score results, girls predictably receive higher reading grades than boys, but the gender disparities in grades are even larger. Now, in every case, the estimated male coefficients are negative and statistically significant at the 5% level. In kindergarten, white boys receive grades that are 0.25 standard deviations lower than white girls, on average, and the gap remains relatively constant through the fifth grade. Therefore, throughout primary school, white boys score lower on reading tests and receive lower grades in reading, but occupy places in the grade distribution even lower than those in the test-score distribution.

The pattern of estimated gender disparities in grades is generally similar for black children. Black boys receive 0.34 standard deviations lower reading grades than black girls in kindergarten. The gap does decrease to 0.26 standard deviations in first grade, but increases to 0.40 in third grade and remains 0.35 standard deviations in grade five. As with white boys, black boys receive substantially lower grades than their test scores might suggest.

For Hispanic boys the pattern is a little different. They earn lower grades than their girl counterparts, but the estimated grade disparities are more on par with those associated with test scores. Hispanic boys earn 0.27 standard deviations lower reading grades in kindergarten compared with Hispanic girls and this gap increases through the fifth grade to 0.43 standard

deviations. So, compared with whites and blacks, teacher assessments for Hispanics appear to be more in line with the results of the reading tests.

Finally, the observables explain more of the variance in reading test scores than grades and the relatively better fit for the test-score regressions increases with grade level. By fifth grade, observables explain 21% of the variance in whites' test scores, 34% of the variance in blacks' scores and 26% of the variance for Hispanics' scores. In contrast, the same observables produce  $R^2$ s of only 0.17, 0.19 and 0.16, respectively, in the fifth-grade teacher-grade regressions. This pattern is replicated in the math and science results, suggesting that the process teachers follow to assess achievement is generally noisier, at least from the perspective of the econometrician.

### *Math*

The results for math test scores and math grades are reported in Table 4B. In general, boys score at least as well on math tests as girls, but the evidence for a gender gap is less overwhelming than in reading. The strongest case exists among whites, where statistically significant performance differences emerge in kindergarten and persist through the fifth grade. White boys score 0.06 standard deviations higher than white girls in kindergarten and at least 0.13 standard deviations higher thereafter. In contrast, the male coefficient estimates for blacks and Hispanics are, for the most part, small in magnitude and not very precisely estimated. Only for black fifth-graders and Hispanic third-graders are they positive and statistically significant.

Despite generally performing on par with girls on math tests, and significantly better in the case of whites, boys are not commensurately graded by their teachers. White boys receive 0.12 standard deviations *lower* grades in kindergarten and the difference is statistically



significant. After kindergarten, the disparity in grading largely disappears, with the estimated male coefficient being small and statistically insignificant. But this means, like in reading, test-score performance and teacher grades are not aligned. Although white boys score higher than girls on the math tests, teachers do not differentiate between them in their grading. White boys and girls occupy essentially the same places in the grade distribution even though the boys are more likely to appear in the top half of the test-score distribution.

For black and Hispanic children, test scores and grades are also not aligned. Although math test performance is roughly the same for black and Hispanic boys and girls, the boys of both groups generally receive lower grades. With the exception of fifth-graders, the gender gaps in grades are greater for blacks – ranging from 0.11 to 0.24 standard deviations compared with 0.06 to 0.17 for Hispanics – and more precisely estimated. Thus, like their white counterparts, black and Hispanic boys are not represented in the math grade distribution as their math test scores would predict.

### *General Knowledge/Science*<sup>10</sup>

Table 4C reports our findings for science test scores and grades. In terms of test performance, the pattern for whites is basically the same as it was in math. Boys start out in kindergarten with slightly higher test scores and the difference increases in magnitude after that. At each grade level, the male coefficient estimate is statistically significant. However, the science test scores for blacks and Hispanics depart from their pattern in math. While no statistically significant performance differences show up in kindergarten or first grade, black and Hispanic boys score markedly better than their girl counterparts in third and fifth grade. For

---

<sup>10</sup> 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.

blacks, the estimated test-score differences are 0.20 and 0.36 standard deviations respectively. For Hispanics, the estimated differences are also large: 0.11 and 0.31 standard deviations.

As in math, boys' test-score performance is not reflected in the grades they receive from their teachers. In kindergarten and first grade, white boys' grades are *lower* by 0.11 and 0.06 standard deviations, even though their test scores are higher. After first grade, white boys and girls are graded similarly, but the disparity between their test performance and teacher assessment grows. From kindergarten to fifth grade, the top half of the test-score distribution for whites is increasingly populated by boys, while the grade distribution provides no corresponding evidence that boys are out-performing girls.

The disparity between test performance and grading is even sharper for black and Hispanic children. The estimated male coefficient in the teacher-grade regression is negative in every case. The magnitude of the estimated gender gap in grades ranges from essentially -0.12 to -0.26 for blacks and -0.02 to -.26 for Hispanics, but the misalignment of grades with test scores steadily increases as black and Hispanic students advance in school. By fifth grade, there is over a one-half standard deviation disparity between the estimated gender gaps in test scores and teacher grades for both blacks and Hispanics.

## **5. Grades, Test Scores and the Role of Non-cognitive Skills**

### **5.1 Connecting grades to test scores and approaches toward learning**

Now we turn specifically to the relationship between teacher grades, test scores and non-cognitive skills. To examine the link we re-estimate (1) with the subject-area grade as the achievement measure, incrementally adding the contemporaneous subject-area test score and ATL score from the previous grade level. So, the estimating equation becomes

$$grade_i = \delta_0 + \delta_1 male_i + \alpha_1 testscore_i + \alpha_2 ATL_{i,t-1} + X_i\beta + u_i \quad (2)$$

Because teachers were unaware of students' test scores when they provided their subjective assessments, the test score is exogenous. As in the baseline case, we estimate (2) separately for each race and grade level.<sup>11</sup> Tables 5A-C report these findings for reading, math and science, first reproducing the baseline results for comparison's sake (column a), then adding the subject-area test score (column b) and lagged ATL score (column c). The kindergarten case is omitted because there is no pre-kindergarten behavioral assessment.

Equation (2) embodies the proposition that students who perform equally well on subject-area tests should receive (roughly) the same subject-area assessment from the teacher. If this assertion holds in the data, controlling for the test score should eliminate the estimated gender gap in grades. If not, then the question remains regarding what accounts for the test-score/grade disparity. We explore the role of non-cognitive skills as measured by the ATL score. As evidenced in Table 1, the average ATL score for boys is roughly 15% lower than for girls and the variance in boys' scores is greater in every grade. Thus, boys are less likely to sit for long periods of time, participate or demonstrate knowledge in the classroom, or supply effort on assignments and homework. Initially, we employ the lagged ATL score to avoid the possibility of bias that might arise through feedback of the subject-area grade to the behavioral assessments. Bear in mind that the lag entails two years for grades three and five. Nevertheless, to the extent that "approaches to learning" behavior is persistent, students with higher lagged ATL scores will be assessed more favorably by their teachers.

---

<sup>11</sup> We also estimate a quantile version of (2) to examine whether the estimated gender gaps in grades varied across the grade distribution. We considered cut points at each quintile and found no statistically significant differences in the estimated gender gaps, so we report only the OLS estimates.

## *Reading*

The top panel of Table 5A shows the reading results for whites. Column (b) for each grade level reports the estimated male coefficient controlling for the reading test score. Adding the test score reduces the estimated gender gap in teacher grading by at least one-third in every case, which means that holding test performance constant, about two-thirds of the grading disparity is left unexplained. Boys who score as well as girls on the reading test still receive reading grades from their teachers that are 0.15-0.21 standard deviations lower and the differences are statistically significant. A standard deviation increase in the reading test score is associated with at least a 0.60-0.69 standard deviation increase in the grade assigned by the teacher. Finally, including the test score increases the regression  $R^2$  by a factor of at least 2.5.

Column (c) introduces the lagged ATL score. Controlling for non-cognitive skills, as measured by the ATL index constructed one to two years earlier, almost eliminates the estimated gender gap in reading grades. The male coefficient estimate is less than 0.09 standard deviations in every case. Thus, white boys who perform on par with white girls on the reading test and have the same “approaches to learning” are graded similarly. A standard deviation increase in the lagged ATL score is associated with a 0.19-0.26 standard deviation rise in reading grades. The lagged ATL score also explains a portion of the reading test-score effect, from 10% in kindergarten (0.68 to 0.62) to more than 15% in fifth grade (0.60 to 0.49).

Qualitatively, the results for blacks and Hispanics follow the same basic pattern. Introducing the reading test score reduces the magnitude of the estimated male coefficient, though not to zero, and improves the regression’s fit substantially. Compared with whites, the effects of a standard deviation increase in reading test scores are higher for blacks (except in fifth grade) and lower for Hispanics (except in first grade). Adding the lagged ATL score further

reduces the male coefficient estimate and accounts for some (though less) of the test-score effect given in column (b).

However, there are important quantitative differences between the white and non-white students. First, even when non-cognitive skills are held constant, there remains a statistically significant gender difference in reading grades in the third and fifth grade for blacks and in the fifth grade for Hispanics. Black boys still receive a 0.16 and 0.22 standard deviation lower reading grade than black girls in the third and fifth grade, respectively; for Hispanic boys the fifth-grade difference is 0.14 standard deviations. Second, the effect of a standard deviation increase in the lagged ATL score on reading grades is generally smaller for blacks and Hispanics. The  $ATL_{t-1}$  coefficient estimate ranges between 0.13 to 0.16 for blacks and 0.15 to 0.19 for Hispanics, but 0.19 to 0.26 for whites.

### *Math*

Table 5B reports the findings for math. In contrast to reading, controlling for the test score in the math-grade regressions amplifies the disparity favoring girls. In all but two cases, the estimated male coefficient is negative and larger in magnitude. For whites the grading disparity starts at 0.06 standard deviations in the first grade and rises to 0.15 standard deviations by grade five. For blacks and Hispanics the gender gap in grades is less precisely estimated but typically greater in magnitude, reaching as high as 0.25 standard deviations for third-graders.

However, adding the lagged ATL score generally eliminates the increases in the estimated gender gaps produced by the introduction of the test score. For whites, there are now no statistically significant differences between boys and girls in their math grades. The same is true for blacks and Hispanics, except in the third grade where there is still evidence that teachers

grade girls more generously. Remarkably, in these instances, even those black and Hispanic boys who score as well as and approach learning as maturely as their girl counterparts receive distinctly different assessments from their teachers.

Except in grade three, the math test-score effect is greater for both blacks and Hispanics than whites. The range of  $ATL_{t-1}$  coefficient estimates is roughly the same for each racial group, but the influence of non-cognitive skills on math grades diminishes in importance for whites relative to blacks and Hispanics as children advance through school.

### *Science*

Finally we turn to the science results in Table 5C. As in math, holding the test score constant increases the disparity in grades favoring girls. Compared with the math findings, the results for science are somewhat stronger. For whites, the gender difference in science grades peaks at 0.14 standard deviations in fifth grade. For blacks, the estimated male coefficient rises above 0.30 standard deviations in both the third and fifth grades, while for Hispanics it jumps from 0.17 to 0.37 standard deviations between the third and fifth grades.

Again, including the non-cognitive skills measure largely erases the gender gap in teacher grades. For whites, the male coefficient estimate is now less than 0.01 standard deviations with standard errors more than twice as large in every grade. Although the estimated male coefficients remain negative (except for Hispanic first-graders) and larger in magnitude for blacks and Hispanics, they are not statistically significant (except for black third-graders).

Across racial groups, the test-score coefficient estimates are smaller for science than reading and math, while the  $ATL_{t-1}$  coefficient estimates are similar in magnitude. Also, the

observables explain less of the variation in science grades than they do for reading and math grades.

## **5.2 Refining the connection with a contemporaneous measure of non-cognitive skills**

Explaining the gender gap in teacher grades using a measure of non-cognitive skills that is one to two years old is obviously problematic. It would be preferable to relate the grades assigned by teachers to a contemporaneous measure. However, as we noted earlier, the contemporaneous ATL score may not be strictly exogenous; there could be feedback from the subject-area grade to the behavioral assessments. Our solution is to instrument the contemporaneous ATL score with its lag. To the degree that attitudes to learning are correlated across grade levels, the instrumented contemporaneous score should reflect behavioral patterns that persist as children advance through school.

The lagged ATL score is a strong instrument, entering each regression with an estimated coefficient typically above 0.37 and a standard error that is no more than one-fifth as large. Also, the estimated male-indicator coefficient is negative and statistically significant in every case, indicating boys receive lower behavioral assessments, conditional on contemporaneous test scores and past behavior scores. The estimated gender disparity ranges from 0.11 to 0.37 standard deviations and is generally larger for blacks and Hispanics.

Tables 6A-C present the Instrumental Variable (IV) results for reading, math and science. Two broad patterns stand out. First, the effect of behavior on grades is sharply higher when we use the instrumented contemporaneous ATL score. The estimated “attitude toward learning” effect is roughly two to three times greater in Table 6A-C than Table 5A-C. For whites, it now

dominates the subject-area test-score effect in every subject. The same is generally true for blacks in reading and math and for Hispanics in reading and science.

Second, there is now no statistically significant evidence of a gender gap in grading favoring girls. In reading, the grading disparity for whites and Hispanics has actually reversed. For these groups, the male coefficient estimate is positive in every grade, albeit generally less than 0.10 standard deviations. For black fifth graders the estimated male coefficient is still negative and about 0.14, but its t-ratio is only 1.3. There is also evidence of a gender gap reversal in math and science. White boys now receive significantly higher grades in math and science at every grade level. Tables 6B and C indicate that white boys are assigned math grades that are 0.12-0.23 standard deviations higher and science grades that are 0.15-0.21 standard deviations higher, holding test scores and behavior constant. White boys who perform as well as white girls on these subject-area tests and exhibit the same attitude towards learning as white girls in the classroom are rewarded with a kind of grade “bonus”. While the evidence is a little weaker for Hispanics, the message is essentially the same. For blacks, on the other hand, the story is more mixed, with generally imprecisely estimated male coefficients ranging from minus 0.14 (in grade-three math) to positive 0.39 (in grade-five science).

Why are boys graded more favorably than their girl counterparts in test-score performance and classroom behavior? One potential explanation is that teachers—who, in primary school, are overwhelmingly female—develop assumptions about typical boy and girl classroom behavior. Girls may be expected to possess a better “attitude toward learning”. The gender differences in ATL scores depicted in Table 2 provide support for such expectations. Then, boys who act “out of character” by displaying the same non-cognitive skills as girls with



similar ability may receive special recognition. They may be, in essence, compensated for exceeding expectations.

## **6. Conclusion**

This paper extends the analysis of early-emerging gender differences in academic achievement to include both (objective) test scores and (subjective) teacher assessments. Using data from the 1998-99 ECLS-K cohort, we show that the grades awarded by teachers are not aligned with test scores, with the disparities in grading exceeding those in testing outcomes and uniformly favoring girls. In addition, the misalignment of grades and test scores can be linked to differences between boys and girls in their non-cognitive development.

First, we establish the existence of empirically meaningful gender gaps in reading, math and science test scores. Girls in every racial category outperform boys on reading tests and the differences are statistically significant in every case except for black fifth-graders. While boys score at least as well on math and science tests as girls, the evidence varies by racial category and grade level. The strongest case exists among whites, where statistically significant performance differences emerge in kindergarten and persist through the fifth grade.

Not surprisingly, we also find gender gaps in teacher-assigned grades. However, they are generally not aligned with the test-score gaps and always favor girls. Girls predictably receive higher reading grades than boys, but the gender disparities in grades are typically much larger. Despite performing as well as girls on math tests, and significantly better on science tests, boys are not commensurately graded by their teachers. Boys in all racial categories across all subject areas are represented in grade distributions below where their test scores would predict.

We also find that sizeable estimated gender differences in grades still remain after controlling for test scores and many personal, family, and school characteristics. Boys who perform equally to girls on reading, math and science tests are nevertheless graded less favorably by their teachers. However, this less favorable treatment essentially vanishes when non-cognitive skills are taken into account. White boys who perform on par with white girls on these subject-area tests and exhibit the same non-cognitive skill level are graded similarly. For some specifications we report evidence of a grade “bonus” for white boys with test scores and behavior like their girl counterparts. While the evidence is a little weaker for blacks and Hispanics, the message is essentially the same.

Our paper shines a light on the teacher’s role in assessing academic achievement. If, as the data suggest, young girls display a more developed “attitude toward learning” and teachers (consciously or subconsciously) reward these attitudes by giving girls higher marks than warranted by their test scores, the seeds of a gender gap in educational attainment may be sown at an early age, because teachers’ grades strongly influence grade-level placement, high-school graduation and college admission prospects. Consequently, our results may spur further educational innovation in the early grades, such as developing ways to improve boys’ non-cognitive skills, creating alternative methods of instruction that may communicate more effectively to boys who have different non-cognitive skill sets, and experimenting with single-gender instruction.

## References

- Anderson, Michael. 2008. "Multiple Inference and Gender Differences in the Effects of Early Intervention: A Reevaluation of the Abecedarian, Perry Preschool, and Early Training Projects." *Journal of the American Statistical Association*, Vol. 103, No. 484: 1481-1495.
- 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.
- Burgess, Simon and Ellen Greaves. 2009. "Test scores, Subjective Assessments and Stereotyping of Ethnic Minorities." Working Paper.
- 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.

- 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. 2010. "An Empirical Analysis of the Gender Gap in Mathematics." *American Economic Journal: Applied Economics*, Vol. 2, No. 2: 210-240.
- 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.
- Heckman, James and Paul A. LaFontaine. 2010. "The American High School Graduation Rate: Trends and Levels." *The Review of Economics and Statistics*, Vol. 92, No. 2: 244-262.
- 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.
- Lavy, Victor. 2008. "Do Gender Stereotypes Reduce Girls' or Boys' Human Capital Outcomes? Evidence from a Natural Experiment." *Journal of Public Economics*, Vol. 92, No. 10: 2083-2105.
- Lavy, Victor and Anal Schlosser. 2011. "Mechanisms and Impacts of Gender Peer Effects at School." *American Economic Journal: Applied Economics*, Vol. 3, No. 2: 1-33.
- Loury, Linda Datcher. 2004. "Siblings and Gender Differences in African-American College

Attendance.” *Economics of Education Review*, Vol. 23, No. 3: 213–219.

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.

**Table 1. Descriptive Statistics  
Academic Achievement and Non-Cognitive Skills**

	Female		Male		
	Mean	Std. Dev.	Mean	Std. Dev.	N
<b>Reading Scores</b>					
Kindergarten	42.52	13.29	40.22	13.48	9454
First	76.12	21.16	72.41	21.49	8401
Third	124.80	22.69	120.69	24.4	5793
Fifth	144.21	20.86	141.23	23.07	5841
<b>Reading Grades</b>					
Kindergarten	3.55	0.76	3.34	0.78	9454
First	3.62	0.89	3.40	0.88	8401
Third	3.50	0.84	3.26	0.84	5793
Fifth	3.62	0.80	3.37	0.82	5841
<b>Math Scores</b>					
Kindergarten	34.12	10.60	34.52	12.21	9454
First	59.11	15.24	61.13	17.52	8401
Third	93.57	19.74	98.1	20.82	5793
Fifth	114.49	19.61	118.88	19.72	2820
<b>Math Grades</b>					
Kindergarten	3.68	0.79	3.57	0.84	9454
First	3.54	0.85	3.54	0.89	8401
Third	3.13	0.70	3.14	0.73	5793
Fifth	3.44	0.65	3.45	0.72	2820
<b>Science Scores</b>					
Kindergarten	27.93	7.49	28.31	7.89	9454
First	35.33	7.05	36.18	7.13	8401
Third	45.81	13.15	49.05	13.78	5793
Fifth	58.39	13.86	62.09	13.10	2747
<b>Science Grades</b>					
Kindergarten	3.76	0.92	3.65	0.97	9454
First	3.42	0.94	3.37	0.96	8401
Third	3.26	0.89	3.24	0.91	5793
Fifth	3.41	0.86	3.35	0.86	2747
<b>SRS Score for ATL</b>					
Kindergarten	2.30	0.62	2.00	0.68	9454
First	2.23	0.66	1.93	0.69	8356
Third	2.26	0.61	1.94	0.66	5781
Fifth	2.30	0.60	1.94	0.67	5815

**Table 2. Descriptive Statistics  
Gender, Race and Family Characteristics  
(Standard deviations in parentheses)**

<b>Personal Characteristics</b>	<b>K</b>	<b>First</b>	<b>Third</b>	<b>Fifth</b>
Male	0.51 (0.50)	0.50 (0.50)	0.50 (0.50)	0.49 (0.50)
Black	0.15 (0.35)	0.13 (0.34)	0.10 (0.30)	0.09 (0.29)
Hispanic	0.15 (0.36)	0.16 (0.36)	0.15 (0.36)	0.17 (0.37)
<b>Family Characteristics</b>	<b>K</b>	<b>First</b>	<b>Third</b>	<b>Third</b>
WIC Benefits	0.36 (0.48)	0.33 (0.47)	0.30 (0.46)	0.30 (0.46)
Teenage Mom	0.24 (0.42)	0.22 (0.41)	0.19 (0.39)	0.19 (0.39)
Mom > 30 years old	0.12 (0.33)	0.13 (0.34)	0.14 (0.35)	0.14 (0.35)
Age at K entry	65.77 (4.13)	65.89 (4.14)	65.89 (4.18)	65.82 (4.18)
# Books in the home	81.7 (60.27)	112.36 (147.95)	135.98 (189.80)	117.91 (177.96)
SES Index	0.08 (0.76)	0.08 (0.78)	0.09 (0.76)	0.08 (0.78)
Observations	9454	8401	5793	5841

**Table 3. Descriptive Statistics  
Teacher and School Characteristics  
(Standard deviations in parentheses)**

<b>Teacher Characteristics</b>	<b>K</b>	<b>First</b>	<b>Third</b>	<b>Fifth</b>
Teacher experience	9.10 (7.66)	14.89 (10.09)	15.34 (10.08)	14.63 (10.29)
Teacher education	2.10 (0.91)	2.13 (0.93)	2.20 (0.92)	2.23 (0.93)
<b>School Characteristics</b>	<b>K</b>	<b>First</b>	<b>Third</b>	<b>Fifth</b>
Public school	0.80 (0.40)	0.79 (0.41)	0.77 (0.42)	0.78 (0.41)
Urban school	0.37 (0.48)	0.36 (0.48)	0.34 (0.47)	0.35 (0.48)
Rural school	0.23 (0.42)	0.23 (0.42)	0.26 (0.44)	0.26 (0.44)
Southern school	0.35 (0.48)	0.34 (0.47)	0.32 (0.47)	0.30 (0.46)
% Minority < 10	0.39 (0.49)	0.39 (0.49)	0.42 (0.49)	0.40 (0.49)
% Minority 10-25	0.20 (0.40)	0.20 (0.40)	0.20 (0.40)	0.20 (0.40)
% Minority 25-50	0.16 (0.37)	0.16 (0.36)	0.15 (0.36)	0.17 (0.38)
% Minority 50-75	0.09 (0.29)	0.09 (0.29)	0.08 (0.27)	0.07 (0.26)
% Minority >75	0.16 (0.37)	0.16 (0.37)	0.14 (0.35)	0.16 (0.37)
Observations	9454	8401	5793	5841



**Table 4A. Estimated Gender Gap in Reading Test Scores and Grades, by Race and Ethnicity**

	Kindergarten		First Grade		Third Grade		Fifth Grade	
<b>I. Whites</b>								
	<b>Test</b>	<b>Grade</b>	<b>Test</b>	<b>Grade</b>	<b>Test</b>	<b>Grade</b>	<b>Test</b>	<b>Grade</b>
Male	-0.162*** (0.0296)	-0.250*** (0.0281)	-0.151*** (0.0280)	-0.247*** (0.0281)	-0.146*** (0.0322)	-0.272*** (0.0351)	-0.113* (0.0487)	-0.278*** (0.0501)
R <sup>2</sup>	0.136	0.136	0.132	0.118	0.186	0.144	0.208	0.168
N	6638	6638	5983	5983	4338	4338	4327	4327
	Kindergarten		First Grade		Third Grade		Fifth Grade	
<b>II. Blacks</b>								
	<b>Test</b>	<b>Grade</b>	<b>Test</b>	<b>Grade</b>	<b>Test</b>	<b>Grade</b>	<b>Test</b>	<b>Grade</b>
Male	-0.149* (0.0612)	-0.337*** (0.0610)	-0.199*** (0.0595)	-0.257*** (0.0723)	-0.268** (0.0833)	-0.402*** (0.0921)	-0.165 (0.110)	-0.350** (0.121)
R <sup>2</sup>	0.175	0.177	0.180	0.149	0.232	0.143	0.340	0.190
N	1387	1387	1094	1094	576	576	538	538
	Kindergarten		First Grade		Third Grade		Fifth Grade	
<b>III. Hispanics</b>								
	<b>Test</b>	<b>Grade</b>	<b>Test</b>	<b>Grade</b>	<b>Test</b>	<b>Grade</b>	<b>Test</b>	<b>Grade</b>
Male	-0.199*** (0.0486)	-0.268*** (0.0645)	-0.217*** (0.0566)	-0.201** (0.0638)	-0.320*** (0.0744)	-0.407*** (0.0795)	-0.344*** (0.0882)	-0.427*** (0.0757)
R <sup>2</sup>	0.228	0.187	0.178	0.124	0.295	0.163	0.264	0.160
N	1429	1429	1324	1324	879	879	976	976

Notes: Test scores and grades are normalized to have mean=0 and variance=1. All regressions control for family, teacher, and school characteristics. Standard errors are in parentheses; \*\*\*, \*\*, and \* indicate statistical significance at 0.1%, 1%, and 5% levels, respectively.

**Table 4B. Estimated Gender Gap in Math Test Scores and Grades, by Race and Ethnicity**

	Kindergarten		First Grade		Third Grade		Fifth Grade	
<b>I. Whites</b>								
	<b>Test</b>	<b>Grade</b>	<b>Test</b>	<b>Grade</b>	<b>Test</b>	<b>Grade</b>	<b>Test</b>	<b>Grade</b>
Male	0.0627*	-0.121***	0.125***	0.000456	0.246***	0.0674	0.207**	-0.0259
	(0.0280)	(0.0289)	(0.0269)	(0.0283)	(0.0317)	(0.0375)	(0.0722)	(0.0698)
R <sup>2</sup>	0.189	0.115	0.151	0.115	0.194	0.116	0.2	0.118
N	6638	6638	5983	5983	4338	4338	2113	2113
	Kindergarten		First Grade		Third Grade		Fifth Grade	
<b>II. Blacks</b>								
	<b>Test</b>	<b>Grade</b>	<b>Test</b>	<b>Grade</b>	<b>Test</b>	<b>Grade</b>	<b>Test</b>	<b>Grade</b>
Male	-0.0469	-0.235***	-0.0609	-0.114	0.0182	-0.244**	0.361**	0.0288
	(0.0492)	(0.0647)	(0.0518)	(0.0723)	(0.0863)	(0.0915)	(0.124)	(0.151)
R <sup>2</sup>	0.185	0.140	0.153	0.115	0.222	0.104	0.415	0.301
N	1387	1387	1094	1094	576	576	245	245
	Kindergarten		First Grade		Third Grade		Fifth Grade	
<b>III. Hispanics</b>								
	<b>Test</b>	<b>Grade</b>	<b>Test</b>	<b>Grade</b>	<b>Test</b>	<b>Grade</b>	<b>Test</b>	<b>Grade</b>
Male	-0.0795	-0.158*	0.0855	-0.0438	0.163*	-0.172*	-0.0419	-0.168
	(0.0524)	(0.0690)	(0.0561)	(0.0601)	(0.0711)	(0.0784)	(0.110)	(0.122)
R <sup>2</sup>	0.267	0.135	0.180	0.133	0.216	0.141	0.229	0.077
N	1429	1429	1324	1324	879	879	462	462

Notes: Test scores and grades are normalized to have mean=0 and variance=1. All regressions control for family, teacher, and school characteristics. Standard errors are in parentheses; \*\*\*, \*\*, and \* indicate statistical significance at 0.1%, 1%, and 5% levels, respectively.

**Table 4C. Estimated Gender Gap in Science Test Scores and Grades, by Race and Ethnicity**

	Kindergarten		First Grade		Third Grade		Fifth Grade	
<b>I. Whites</b>								
	<b>Test</b>	<b>Grade</b>	<b>Test</b>	<b>Grade</b>	<b>Test</b>	<b>Grade</b>	<b>Test</b>	<b>Grade</b>
Male	0.0547*	-0.107***	0.103***	-0.0641*	0.258***	-0.00255	0.287***	0.00313
	(0.0233)	(0.0284)	(0.0216)	(0.0288)	(0.0304)	(0.0364)	(0.0554)	(0.0744)
R <sup>2</sup>	0.250	0.110	0.187	0.102	0.175	0.105	0.232	0.125
N	6638	6638	5983	5983	4338	4338	2021	2021
	Kindergarten		First Grade		Third Grade		Fifth Grade	
<b>II. Blacks</b>								
	<b>Test</b>	<b>Grade</b>	<b>Test</b>	<b>Grade</b>	<b>Test</b>	<b>Grade</b>	<b>Test</b>	<b>Grade</b>
Male	-0.0242	-0.257***	0.0908	-0.122	0.198**	-0.220*	0.364**	-0.161
	(0.0516)	(0.0640)	(0.0588)	(0.0721)	(0.0665)	(0.0897)	(0.136)	(0.147)
R <sup>2</sup>	0.293	0.121	0.236	0.121	0.298	0.104	0.472	0.200
N	1387	1387	1094	1094	576	576	265	265
	Kindergarten		First Grade		Third Grade		Fifth Grade	
<b>III. Hispanics</b>								
	<b>Test</b>	<b>Grade</b>	<b>Test</b>	<b>Grade</b>	<b>Test</b>	<b>Grade</b>	<b>Test</b>	<b>Grade</b>
Male	-0.00213	-0.175**	0.0679	-0.0178	0.114	-0.122	0.306**	-0.258*
	(0.0548)	(0.0629)	(0.0562)	(0.0626)	(0.0657)	(0.0755)	(0.111)	(0.111)
R <sup>2</sup>	0.354	0.142	0.307	0.147	0.314	0.176	0.311	0.16
N	1429	1429	1324	1324	879	879	461	461

Notes: Test scores and grades are normalized to have mean=0 and variance=1. All regressions control for family, teacher, and school characteristics. Standard errors are in parentheses; \*\*\*, \*\*, and \* indicate statistical significance at 0.1%, 1%, and 5% levels, respectively.

**Table 5A. Estimated Gender Gap in Reading Grades, Controlling for Test Scores and Non-Cognitive Skills**

	First Grade			Third Grade			Fifth Grade		
	(a)	(b)	(c)	(a)	(b)	(c)	(a)	(b)	(c)
<b>I. Whites</b>									
Male	-0.247*** (0.0281)	-0.144*** (0.0205)	-0.0741*** (0.0202)	-0.272*** (0.0351)	-0.172*** (0.0272)	-0.0889*** (0.0268)	-0.278*** (0.0501)	-0.210*** (0.0421)	-0.0902* (0.0416)
Test Score <sub>t</sub>		0.684*** (0.0120)	0.620*** (0.0125)		0.687*** (0.0161)	0.593*** (0.0167)		0.603*** (0.0262)	0.493*** (0.0287)
ATL Score <sub>t-1</sub>			0.191*** (0.0117)			0.234*** (0.0157)			0.260*** (0.0268)
R <sup>2</sup>	0.118	0.516	0.544	0.144	0.476	0.516	0.168	0.423	0.471
N	5983	5983	5983	4338	4338	4338	4327	4327	4327
<b>II. Blacks</b>									
Male	-0.257*** (0.0723)	-0.0857 (0.0506)	-0.0303 (0.0507)	-0.402*** (0.0921)	-0.214** (0.0723)	-0.161* (0.0726)	-0.350** (0.121)	-0.253* (0.102)	-0.220* (0.106)
Test Score <sub>t</sub>		0.863*** (0.0284)	0.801*** (0.0312)		0.704*** (0.0417)	0.631*** (0.0437)		0.588*** (0.0555)	0.544*** (0.0587)
ATL Score <sub>t-1</sub>			0.146*** (0.0252)			0.160*** (0.0400)			0.131* (0.0517)
R <sup>2</sup>	0.149	0.589	0.604	0.143	0.481	0.500	0.190	0.435	0.448
N	1094	1094	1094	576	576	576	538	538	538
<b>III. Hispanics</b>									
Male	-0.201** (0.0638)	-0.0362 (0.0466)	0.00421 (0.0468)	-0.407*** (0.0795)	-0.221** (0.0696)	-0.136 (0.0712)	-0.427*** (0.0757)	-0.247*** (0.0672)	-0.141* (0.0670)
Test Score <sub>t</sub>		0.762*** (0.0268)	0.701*** (0.0295)		0.580*** (0.0440)	0.497*** (0.0461)		0.523*** (0.0376)	0.454*** (0.0428)
ATL Score <sub>t-1</sub>			0.152*** (0.0275)			0.193*** (0.0346)			0.193*** (0.0382)
R <sup>2</sup>	0.124	0.487	0.503	0.163	0.392	0.419	0.160	0.398	0.428
N	1324	1324	1324	879	879	879	976	976	976

Notes: Test scores and grades are normalized to have mean=0 and variance=1. All regressions control for family, teacher, and school characteristics. Standard errors are in parentheses; \*\*\*, \*\*, and \* indicate statistical significance at 0.1%, 1%, and 5% levels, respectively.

**Table 5B. Estimated Gender Gap in Math Grades, Controlling for Test Scores and Non-Cognitive Skills**

	First Grade			Third Grade			Fifth Grade		
<b>I. Whites</b>	<b>(a)</b>	<b>(b)</b>	<b>(c)</b>	<b>(a)</b>	<b>(b)</b>	<b>(c)</b>	<b>(a)</b>	<b>(b)</b>	<b>(c)</b>
Male	0.000456 (0.0283)	-0.0621* (0.0245)	0.0452 (0.0245)	0.0674 (0.0375)	-0.0823* (0.0335)	0.0226 (0.0361)	-0.0259 (0.0698)	-0.149** (0.0576)	-0.0290 (0.0533)
Test Score <sub>t</sub>		0.499*** (0.0143)	0.420*** (0.0144)		0.608*** (0.0199)	0.525*** (0.0220)		0.592*** (0.0392)	0.518*** (0.0406)
ATL Score <sub>t-1</sub>			0.233*** (0.0144)			0.204*** (0.0207)			0.172*** (0.0332)
R <sup>2</sup>	0.115	0.326	0.367	0.116	0.366	0.395	0.118	0.361	0.382
N	5983	5983	5983	4338	4338	4338	2113	2113	2113
<b>II. Blacks</b>	<b>(a)</b>	<b>(b)</b>	<b>(c)</b>	<b>(a)</b>	<b>(b)</b>	<b>(c)</b>	<b>(a)</b>	<b>(b)</b>	<b>(c)</b>
Male	-0.114 (0.0723)	-0.0687 (0.0614)	0.0237 (0.0593)	-0.244** (0.0915)	-0.254** (0.0814)	-0.220** (0.0837)	0.0288 (0.151)	-0.187 (0.134)	-0.0839 (0.129)
Test Score <sub>t</sub>		0.743*** (0.0421)	0.638*** (0.0439)		0.538*** (0.0488)	0.500*** (0.0555)		0.597*** (0.0646)	0.487*** (0.0685)
ATL Score <sub>t-1</sub>			0.213*** (0.0326)			0.0727 (0.0527)			0.237*** (0.0672)
R <sup>2</sup>	0.115	0.391	0.425	0.104	0.309	0.313	0.301	0.537	0.574
N	1094	1094	1094	576	576	576	245	245	245
<b>III. Hispanics</b>	<b>(a)</b>	<b>(b)</b>	<b>(c)</b>	<b>(a)</b>	<b>(b)</b>	<b>(c)</b>	<b>(a)</b>	<b>(b)</b>	<b>(c)</b>
Male	-0.0438 (0.0601)	-0.0940 (0.0509)	-0.0479 (0.0532)	-0.172* (0.0784)	-0.252*** (0.0720)	-0.169* (0.0710)	-0.168 (0.122)	-0.136 (0.0813)	0.0206 (0.0877)
Test Score <sub>t</sub>		0.587*** (0.0382)	0.542*** (0.0381)		0.495*** (0.0402)	0.439*** (0.0465)		0.742*** (0.0525)	0.663*** (0.0526)
ATL Score <sub>t-1</sub>			0.120*** (0.0285)			0.128** (0.0426)			0.212*** (0.0448)
R <sup>2</sup>	0.133	0.341	0.351	0.141	0.322	0.335	0.077	0.492	0.523
N	1324	1324	1324	879	879	879	462	462	462

Notes: Test scores and grades are normalized to have mean=0 and variance=1. All regressions control for family, teacher, and school characteristics. Standard errors are in parentheses; \*\*\*, \*\*, and \* indicate statistical significance at 0.1%, 1%, and 5% levels, respectively.

**Table 5C. Estimated Gender Gap in Science Grades, Controlling for Test Scores and Non-Cognitive Skills**

	First Grade			Third Grade			Fifth Grade		
<b>I. Whites</b>	<b>(a)</b>	<b>(b)</b>	<b>(c)</b>	<b>(a)</b>	<b>(b)</b>	<b>(c)</b>	<b>(a)</b>	<b>(b)</b>	<b>(c)</b>
Male	-0.0641*	-0.109***	0.0103	-0.00255	-0.119***	0.00122	0.00313	-0.138	0.00865
	(0.0288)	(0.0269)	(0.0272)	(0.0364)	(0.0350)	(0.0353)	(0.0744)	(0.0721)	(0.0696)
Test Score <sub>t</sub>		0.440***	0.350***		0.451***	0.383***		0.491***	0.393***
		(0.0203)	(0.0200)		(0.0203)	(0.0206)		(0.0625)	(0.0637)
ATL Score <sub>t-1</sub>			0.264***			0.248***			0.271***
			(0.0156)			(0.0196)			(0.0451)
R <sup>2</sup>	0.102	0.200	0.254	0.105	0.240	0.288	0.125	0.254	0.306
N	5983	5983	5983	4338	4338	4338	2021	2021	2021
<b>II. Blacks</b>	<b>(a)</b>	<b>(b)</b>	<b>(c)</b>	<b>(a)</b>	<b>(b)</b>	<b>(c)</b>	<b>(a)</b>	<b>(b)</b>	<b>(c)</b>
Male	-0.122	-0.166*	-0.0518	-0.220*	-0.333***	-0.250**	-0.161	-0.326**	-0.201
	(0.0721)	(0.0651)	(0.0651)	(0.0897)	(0.0829)	(0.0859)	(0.147)	(0.125)	(0.131)
Test Score <sub>t</sub>		0.477***	0.401***		0.571***	0.497***		0.452***	0.376***
		(0.0383)	(0.0384)		(0.0587)	(0.0602)		(0.0778)	(0.0843)
ATL Score <sub>t-1</sub>			0.230***			0.151***			0.242***
			(0.0318)			(0.0437)			(0.0609)
R <sup>2</sup>	0.121	0.269	0.313	0.104	0.260	0.278	0.200	0.334	0.387
N	1094	1094	1094	576	576	576	265	265	265
<b>III. Hispanics</b>	<b>(a)</b>	<b>(b)</b>	<b>(c)</b>	<b>(a)</b>	<b>(b)</b>	<b>(c)</b>	<b>(a)</b>	<b>(b)</b>	<b>(c)</b>
Male	-0.0178	-0.0454	0.0158	-0.122	-0.173**	-0.0781	-0.258*	-0.366***	-0.205
	(0.0626)	(0.0578)	(0.0578)	(0.0755)	(0.0664)	(0.0700)	(0.111)	(0.107)	(0.105)
Test Score <sub>t</sub>		0.407***	0.374***		0.451***	0.399***		0.353***	0.304***
		(0.0328)	(0.0326)		(0.0479)	(0.0483)		(0.0556)	(0.0550)
ATL Score <sub>t-1</sub>			0.167***			0.154***			0.223***
			(0.0325)			(0.0387)			(0.0602)
R <sup>2</sup>	0.147	0.256	0.277	0.176	0.300	0.322	0.160	0.254	0.296
N	1324	1324	1324	879	879	879	461	461	461

Notes: Test scores and grades are normalized to have mean=0 and variance=1. All regressions control for family, teacher, and school characteristics. Standard errors are in parentheses; \*\*\*, \*\*, and \* indicate statistical significance at 0.1%, 1%, and 5% levels, respectively.

**Table 6A. Estimated Gender Gap in Reading Grades,  
Controlling for Test Scores and Non-Cognitive Skills – IV**

<b>I. Whites</b>	<b>First</b>	<b>Third</b>	<b>Fifth</b>
Male	0.0475* (0.0202)	0.0419 (0.0275)	0.0847 (0.0456)
Test Score <sub>t</sub>	0.493*** (0.0156)	0.463*** (0.0195)	0.415*** (0.0323)
ATL Score <sub>t</sub>	0.500*** (0.0287)	0.525*** (0.0317)	0.593*** (0.0527)
R <sup>2</sup>	0.624	0.574	0.521
N	5973	4329	4309
<b>II. Blacks</b>	<b>First</b>	<b>Third</b>	<b>Fifth</b>
Male	0.0136 (0.0479)	-0.00897 (0.0827)	-0.141 (0.107)
Test Score <sub>t</sub>	0.641*** (0.0443)	0.533*** (0.0549)	0.480*** (0.0747)
ATL Score <sub>t</sub>	0.399*** (0.0617)	0.428*** (0.102)	0.328* (0.129)
R <sup>2</sup>	0.676	0.549	0.538
N	1092	574	536
<b>III. Hispanics</b>	<b>First</b>	<b>Third</b>	<b>Fifth</b>
Male	0.141** (0.0479)	0.0899 (0.0760)	0.0392 (0.0821)
Test Score <sub>t</sub>	0.568*** (0.0424)	0.333*** (0.0554)	0.356*** (0.0597)
ATL Score <sub>t</sub>	0.442*** (0.0712)	0.611*** (0.0953)	0.508*** (0.109)
R <sup>2</sup>	0.593	0.521	0.465
N	1321	878	970

Notes: Test scores and grades are normalized to have mean=0 and variance=1. All regressions control for family, teacher, and school characteristics. Standard errors are in parentheses; \*\*\*, \*\*, and \* indicate statistical significance at 0.1%, 1%, and 5% levels, respectively.

**Table 6B. Estimated Gender Gap in Math Grades,  
Controlling for Test Scores and Non-Cognitive Skills – IV**

<b>I. Whites</b>	<b>First</b>	<b>Third</b>	<b>Fifth</b>
Male	0.232*** (0.0277)	0.176*** (0.0406)	0.123 (0.0645)
Test Score <sub>t</sub>	0.277*** (0.0191)	0.422*** (0.0265)	0.463*** (0.0438)
ATL Score <sub>t</sub>	0.605*** (0.0365)	0.453*** (0.0414)	0.401*** (0.0763)
R <sup>2</sup>	0.431	0.445	0.398
N	5973	4329	2105
<b>II. Blacks</b>	<b>First</b>	<b>Third</b>	<b>Fifth</b>
Male	0.113* (0.0560)	-0.140 (0.110)	-0.0400 (0.140)
Test Score <sub>t</sub>	0.432*** (0.0546)	0.465*** (0.0701)	0.324* (0.104)
ATL Score <sub>t</sub>	0.559*** (0.0825)	0.189 (0.139)	0.596** (0.179)
R <sup>2</sup>	0.522	0.348	0.466
N	1092	574	245
<b>III. Hispanics</b>	<b>First</b>	<b>Third</b>	<b>Fifth</b>
Male	0.0875 (0.0618)	0.0368 (0.0948)	0.192 (0.116)
Test Score <sub>t</sub>	0.444*** (0.0432)	0.332*** (0.0614)	0.575*** (0.0718)
ATL Score <sub>t</sub>	0.339*** (0.0739)	0.410*** (0.115)	0.473*** (0.113)
R <sup>2</sup>	0.465	0.443	0.513
N	1321	878	461

Notes: Test scores and grades are normalized to have mean=0 and variance=1. All regressions control for family, teacher, and school characteristics. Standard errors are in parentheses; \*\*\*, \*\*, and \* indicate statistical significance at 0.1%, 1%, and 5% levels, respectively.

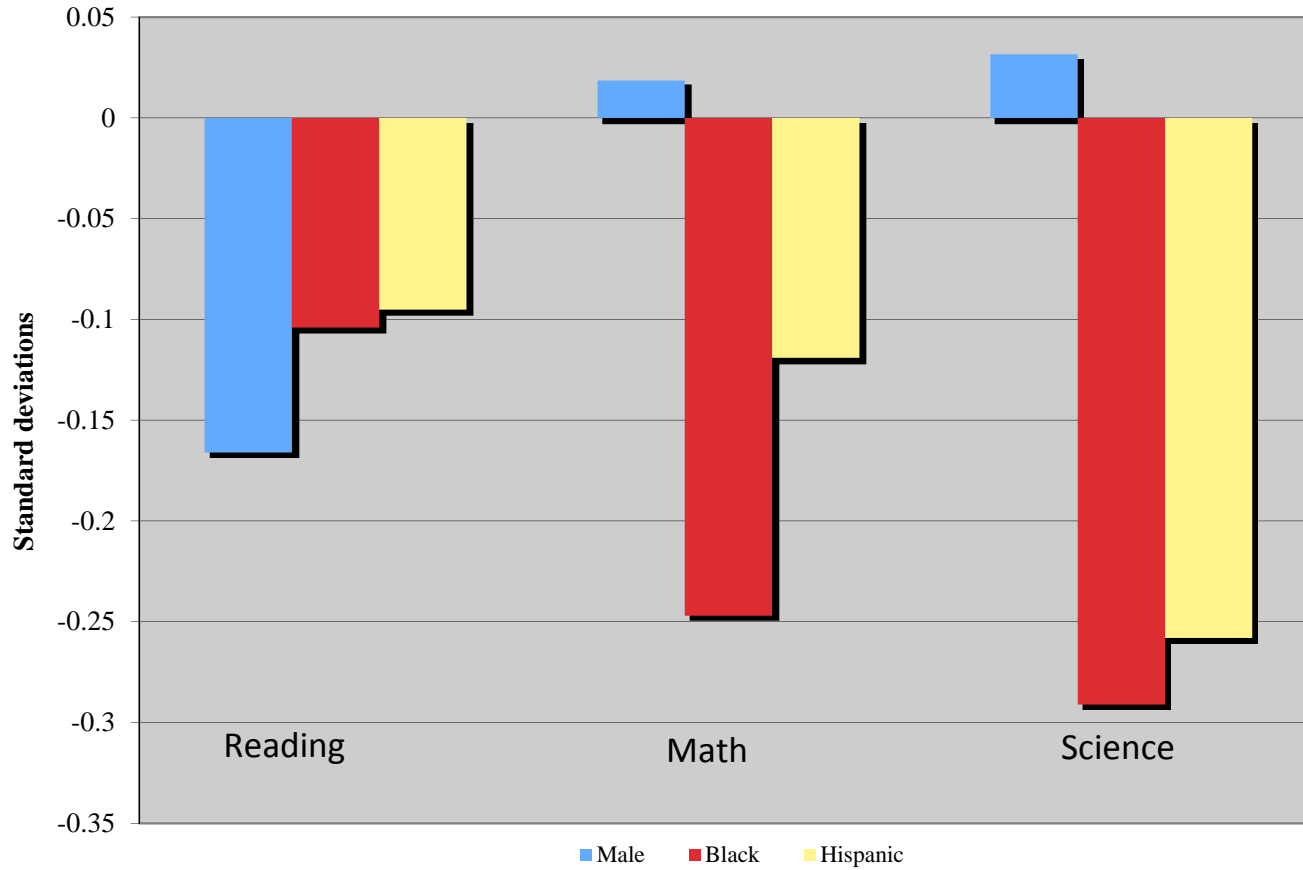


**Table 6C. Estimated Gender Gap in Science Grades,  
Controlling for Test Scores and Non-Cognitive Skills – IV**

<b>I. Whites</b>	<b>First</b>	<b>Third</b>	<b>Fifth</b>
Male	0.188*** (0.0303)	0.152*** (0.0367)	0.218** (0.0748)
Test Score <sub>t</sub>	0.221*** (0.0230)	0.311*** (0.0209)	0.334*** (0.0675)
ATL Score <sub>t</sub>	0.626*** (0.0356)	0.500*** (0.0357)	0.602*** (0.0788)
R <sup>2</sup>	0.329	0.351	0.335
N	5973	4329	2011
<b>II. Blacks</b>	<b>First</b>	<b>Third</b>	<b>Fifth</b>
Male	0.0483 (0.0657)	-0.0893 (0.104)	0.387 (0.286)
Test Score <sub>t</sub>	0.268*** (0.0449)	0.434*** (0.0697)	0.261* (0.122)
ATL Score <sub>t</sub>	0.542*** (0.0745)	0.363** (0.111)	0.785** (0.258)
R <sup>2</sup>	0.411	0.331	0.215
N	1092	574	263
<b>III. Hispanics</b>	<b>First</b>	<b>Third</b>	<b>Fifth</b>
Male	0.166** (0.0630)	0.124 (0.0894)	0.0982 (0.165)
Test Score <sub>t</sub>	0.300*** (0.0357)	0.292*** (0.0529)	0.236** (0.0717)
ATL Score <sub>t</sub>	0.411*** (0.0749)	0.436*** (0.0997)	0.664*** (0.196)
R <sup>2</sup>	0.371	0.411	0.118
N	1321	878	456

Notes: Test scores and grades are normalized to have mean=0 and variance=1. All regressions control for family, teacher, and school characteristics. Standard errors are in parentheses; \*\*\*, \*\*, and \* indicate statistical significance at 0.1%, 1%, and 5% levels, respectively.

**Figure 1. Gender and Race Gaps in Kindergarten Test Scores**



**Figure 2. Gender and Race Gaps in Kindergarten Teacher Grades**

