Noncognitive Skills and the Gender Disparities in Test Scores and Teacher Assessments

Evidence from Primary School

Christopher Cornwell
David B. Mustard
Jessica Van Parys

A B S T R A C T

Using data from the 1998–99 ECLS-K cohort, we show that the grades awarded by teachers are not aligned with test scores. Girls in every racial category outperform boys on reading tests, while boys score at least as well on math and science tests as girls. However, boys in all racial categories across all subject areas are not represented in grade distributions where their test scores would predict. Boys who perform equally as well as girls on reading, math, and science tests are graded less favorably by their teachers, but this less favorable treatment essentially vanishes when noncognitive skills are taken into account. For some specifications there is evidence of a grade "bonus" for boys with test scores and behavior like their girl counterparts.

I. Introduction

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.[1] As summarized in Goldin, Katz, and Kuziemko (2006), the ratio of males to females graduating from a four-year college stood at 1.60

1. Kay Hymowitz (2011) Manning Up: How the Rise of Women Has Turned Men into Boys, is just one example.

*Christopher Cornwell is a professor of economics at the University of Georgia. David B. Mustard is an associate professor of economics at the University of Georgia. Jessica Van Parys is an economics PhD student at Columbia University.**Previous versions of this paper were presented at the University of Georgia and the Federal Reserve Bank of New York and the 2010 SOLE/EALE, 2011 AEFP and the 2011 ESEM annual conferences. The authors thank Ian Schmutte for his helpful comments and suggestions. The data used in this article can be obtained beginning July 2013 through June 2016 from Christopher Cornwell, Department of Economics, University of Georgia, Athens GA cornwl@uga.edu.*
[Submitted September 2011, accepted February 2012]
SSN 022 166X E ISSN 1548 8004 8 2013 2 by the Board of Regents of the University of Wisconsin System

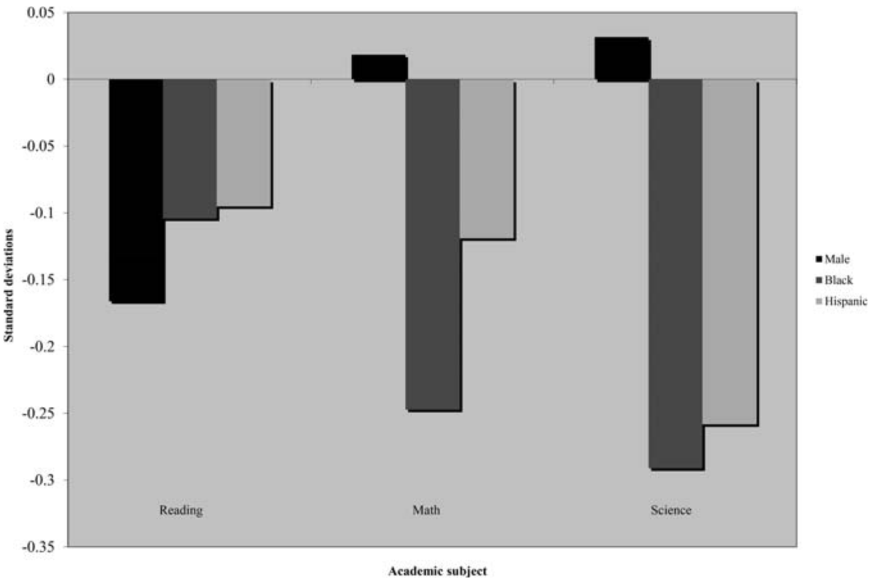

Figure 1
Gender and Race Gaps in Kindergarten Test Scores

in 1960, fell to parity by 1980, and continued its decline to 0.74 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 0.77. 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 and focus on the role of noncognitive 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.[2] In contrast, only a few studies (for example, Anderson 2008; Fryer and Levitt 2010; Holmlund and Sund 2008; Husain and Millimet 2009; Lavy and Schlosser 2011) 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 the teacher, the ratio of boys to girls in a classroom, and whether the children attended preschool.

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)

2. Some noteworthy examples are Cho (2007); Dynarski (2007); Frenette and Zeman (2007); Goldin et al. (2006); Jacob (2002); Loury (2004); and Reynolds and Burge (2007).

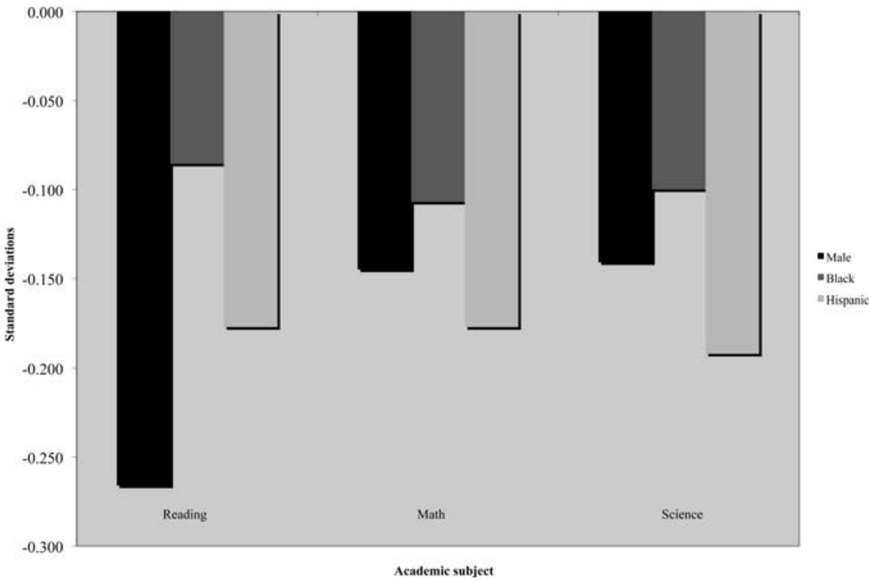

Figure 2
Gender and Race Gaps in Kindergarten Teacher Grades

kindergarten sample, conditional on a range of personal, family and school characteristics described in Section II. 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 percent 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 outperform girls on math and science test scores, it is surprising that girls outperform 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 two 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.[3] While standardized tests are important, teacher-assigned

3. 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, ethnic-

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 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 noncognitive 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 noncognitive skill is a significant factor in teacher assessments, even after controlling for test outcomes.[4]

Our analysis is based on data from the 1998–99 ECLS-K cohort administered by the National Center for Education Statistics (NCES). For 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 noncognitive 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 and science 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, the inconsistency between test scores and grades is largely accounted for by noncognitive skills. White boys who perform as well as white girls on these subject-area tests and exhibit the same attitude toward 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.

---

ity, and nationality for students in England. Lavy (2008) compares blind and nonblind 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.

4. Our findings are consistent with Claessens, Duncan and Engel (2009), who report that a range of socio-emotional skills in kindergarten affect children's standardized test scores in fifth grade.

II. 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 followup 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 children who passed an English language screening test, about 13,300 kindergartners had nonmissing data on test scores and teacher grades in reading, math, and science.

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.[5] We restrict the sample to white, black, and Hispanic children, since those groups are our populations of interest. 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.

As our analysis advances from kindergarten through fifth grade, it is important to consider how attrition might affect our results. The two primary forms of attrition are that blacks and low-performing students are more likely to leave the sample. However, this attrition does not affect our results, because 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.[6] Of course, attrition also affects the degree to which our findings generalize to the entire population of primary school children.

NCES prepared the objective reading, math, and science assessments. 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

5. In our web appendix we give an explicit accounting of the effects of the data requirements on sample selection.

6. To investigate this directly, we create a balanced panel of students with valid data in all grades for the reading analysis. We focus on reading because the manner in which math and science grades were collected in the fifth grade reduces the sample by half, as explained above. There are no qualitative differences between the results from the balanced panel and those reported in Tables 5A and 6A, and the magnitudes are strikingly similar.

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), 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 noncognitive 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 noncognitive-skill measure.[7] As with the ARS scale, higher SRS scores represent higher skill levels.

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 having lower average test scores in those subjects. Third, the average ATL rating for girls is consistently about 15 percent greater than the average score for boys. Finally, boys generally have higher variance in test scores, teacher grades and noncognitive skill ratings; the standard deviation of male achievement is typically greater across subjects and grade levels.

III. Baseline Achievement Regressions

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

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

where y is either a test score or teacher-assigned grade for student i in reading, math or 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

7. We have experimented with all five SRS indices, first giving each a turn as the measure of noncognitive skills, and then including them in the regressions all together. Individually, the ATL measure has the greatest explanatory power and behavioral significance. Not surprisingly, the indices are correlated with each other. ATL is more strongly correlated with Self-Control and Interpersonal Skills (with correlation coefficients in the 0.65–0.72 range) than Internalizing Problems (correlations around –0.40) or Externalizing Problems (correlations between –0.50 to –0.58). Compared with using ATL exclusively, including all SRS variables in the regression has little effect on the estimated gender coefficients and never adds more than 0.01 to the regression R^2 The results using all five SRS variables together are available as a web appendix.

Table 1
Descriptive Statistics: Academic Achievement and Noncognitive Skills

	Female		Male		
	Mean	Standard Deviation	Mean	Standard Deviation	N
Reading Scores					
Kindergarten	42.52	13.29	40.22	13.48	9,454
First	76.12	21.16	72.41	21.49	8,401
Third	124.80	22.69	120.69	24.40	5,793
Fifth	144.21	20.86	141.23	23.07	5,841
Reading grades					
Kindergarten	3.55	0.76	3.34	0.78	9,454
First	3.62	0.89	3.40	0.88	8,401
Third	3.50	0.84	3.26	0.84	5,793
Fifth	3.62	0.80	3.37	0.82	5,841
Math scores					
Kindergarten	34.12	10.60	34.52	12.21	9,454
First	59.11	15.24	61.13	17.52	8,401
Third	93.57	19.74	98.10	20.82	5,793
Fifth	114.49	19.61	118.88	19.72	2,820
Math grades					
Kindergarten	3.68	0.79	3.57	0.84	9,454
First	3.54	0.85	3.54	0.89	8,401
Third	3.13	0.70	3.14	0.73	5,793
Fifth	3.44	0.65	3.45	0.72	2,820
Science scores					
Kindergarten	27.93	7.49	28.31	7.89	9,454
First	35.33	7.05	36.18	7.13	8,401
Third	45.81	13.15	49.05	13.78	5,793
Fifth	58.39	13.86	62.09	13.10	2,747
Science grades					
Kindergarten	3.76	0.92	3.65	0.97	9,454
First	3.42	0.94	3.37	0.96	8,401
Third	3.26	0.89	3.24	0.91	5,793
Fifth	3.41	0.86	3.35	0.86	2,747
SRS score for ATL					
Kindergarten	2.30	0.62	2.00	0.68	9,454
First	2.23	0.66	1.93	0.69	8,356
Third	2.26	0.61	1.94	0.66	5,781
Fifth	2.30	0.60	1.94	0.67	5,815

Hispanics by grade level. In each case, we incorporate the NCES sample weights in estimation and report OLS standard errors that reflect their use.[8] Each cross-sectional wave includes students who were assessed in the spring of that school year.

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 includes 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.[9] For teachers, their highest level of educational attainment and years of experience are reported.[10] 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 are 15 percent of kindergartners, but only 9 percent of fifth graders, as they experience the greatest attrition from the ECLS-K. Hispanics make up a consistent 15–17 percent of the sample. In the first wave, 36 percent of the children's mothers received WIC benefits; 24 percent of the mothers were teenagers at first birth. These characteristics follow patterns from the first wave that are to be expected with aging households and sample attrition. By the fifth grade, 5,841 of the original 9,454 children remain in the sample.

Table 3 shows that kindergarten teachers average nine years of experience, while first-fifth grade teachers average five-six years more, and the typical teacher in each grade has some certification beyond a bachelor's degree, but less than a master's degree.[11] About 80 percent of the schools in the sample are public; 37 percent are located in urban districts and 23 percent in rural districts; about a third are located in the south; and 25 percent have student bodies in which a racial minority holds at least a 50 percent share.

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

9. A potentially important teacher characteristic that is not accounted for is gender. First, the information is suppressed in ECLS-K for first grade and beyond. Second, in kindergarten, where in principle it could be included, there is virtually no variation—more than 98 percent of kindergarten teachers are women.

10. Arguably, a better way to specify the model would be with teacher fixed effects. However, after the first grade, the median number of students per teacher is one; in the first grade it is only two. For kindergarteners, there are three students per teacher at the median and nine at the 90th percentile, so we did experiment with teacher fixed effects in the kindergarten regressions. The impact on the male coefficient estimates was small, but generally in the direction of greater gender disparities in teacher grades favoring girls.

11. Educational attainment is defined over five categories; a value of "2" indicates at least one year beyond a bachelor's degree and value of "3" indicates a master's degree.

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

	K	First	Third	Fifth
Personal characteristics				
Male	0.51	0.50	0.50	0.49
	(0.50)	(0.50)	(0.50)	(0.50)
Black	0.15	0.13	0.10	0.09
	(0.35)	(0.34)	(0.30)	(0.29)
Hispanic	0.15	0.16	0.15	0.17
	(0.36)	(0.36)	(0.36)	(0.37)
Family characteristics				
WIC benefits	0.36	0.33	0.30	0.30
	(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	0.12	0.13	0.14	0.14
	(0.33)	(0.34)	(0.35)	(0.35)
Age at *K* entry	65.77	65.89	65.89	65.82
	(4.13)	(4.14)	(4.18)	(4.18)
Number of books in the home	81.7	112.36	135.98	117.91
	(60.27)	(147.95)	(189.80)	(177.96)
SES index	0.08	0.08	0.09	0.08
	(0.76)	(0.78)	(0.76)	(0.78)
Observations	9,454	8,401	5,793	5,841

IV. 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 (δ_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. The normalization uses the full sample at each grade level; for example, N=9454 in kindergarten.

A. 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 percent 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

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

	K	First	Third	Fifth
Teacher characteristics				
Teacher experience	9.10	14.89	15.34	14.63
	(7.66)	(10.09)	(10.08)	(10.29)
Teacher education	2.10	2.13	2.20	2.23
	(0.91)	(0.93)	(0.92)	(0.93)
School characteristics				
Public school	0.80	0.79	0.77	0.78
	(0.40)	(0.41)	(0.42)	(0.41)
Urban school	0.37	0.36	0.34	0.35
	(0.48)	(0.48)	(0.47)	(0.48)
Rural school	0.23	0.23	0.26	0.26
	(0.42)	(0.42)	(0.44)	(0.44)
Southern school	0.35	0.34	0.32	0.30
	(0.48)	(0.47)	(0.47)	(0.46)
Percent minority < 10	0.39	0.39	0.42	0.40
	(0.49)	(0.49)	(0.49)	(0.49)
Percent minority 10–25	0.20	0.20	0.20	0.20
	(0.40)	(0.40)	(0.40)	(0.40)
Percent minority 25–50	0.16	0.16	0.15	0.17
	(0.37)	(0.36)	(0.36)	(0.38)
Percent minority 50–75	0.09	0.09	0.08	0.07
	(0.29)	(0.29)	(0.27)	(0.26)
Percent minority >75	0.16	0.16	0.14	0.16
	(0.37)	(0.37)	(0.35)	(0.37)
Observations	9,454	8,401	5,793	5,841

the fifth grade. Black and Hispanic boys also 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.

Next we 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 percent 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

Table 4a
Estimated Gender Gap in Reading Test Scores and Grades, by Race and Ethnicity

	Kindergarten		First Grade		Third Grade		Fifth Grade	
	Test	Grade	Test	Grade	Test	Grade	Test	Grade
I. Whites								
Male	–0.162***	–0.250***	–0.151***	–0.247***	–0.146***	–0.272***	–0.113*	–0.278***
	(0.023)	(0.028)	(0.028)	(0.028)	(0.032)	(0.035)	(0.049)	(0.050)
R^2	0.136	0.136	0.132	0.118	0.186	0.144	0.208	0.168
N	6,638	6,638	5,983	5,983	4,338	4,338	4,327	4,327
II. Blacks								
Male	–0.149*	–0.337***	–0.199***	–0.257***	–0.268**	–0.402***	–0.165	–0.350**
	(0.061)	(0.061)	(0.060)	(0.072)	(0.083)	(0.092)	(0.110)	(0.121)
R^2	0.175	0.177	0.180	0.149	0.232	0.143	0.340	0.190
N	1,387	1,387	1,094	1,094	576	576	538	538
III. Hispanics								
Male	–0.199***	–0.268***	–0.217***	–0.201**	–0.320***	–0.407***	–0.344***	–0.427***
	(0.049)	(0.065)	(0.057)	(0.064)	(0.074)	(0.080)	(0.088)	(0.076)
R^2	0.228	0.187	0.178	0.124	0.295	0.163	0.264	0.160
N	1,429	1,429	1,324	1,324	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 percent levels, respectively.

children. 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. So, compared with whites and blacks, teacher assessments for Hispanics are more in line with the results of the reading tests.

Finally, the estimated gender effects in both the test-score and grades regressions are robust to variations in the control set. Incrementally adding the family, teacher, and school characteristics produces essentially the same male coefficient estimates, albeit with increasingly smaller standard errors. This robustness is evident across racial groups, grades, and subject areas. At the same time, 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 percent of the variance in whites' test scores, 34 percent of the variance in blacks' scores, and 26 percent of the variance for Hispanics' scores. In contrast, the same observables produce R^2s 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.

B. 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 than 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.

Table 4b
Estimated Gender Gap in Math Test Scores and Grades, by Race and Ethnicity

	Kindergarten		First Grade		Third Grade		Fifth Grade	
	Test	Grade	Test	Grade	Test	Grade	Test	Grade
I. Whites								
Male	0.063*	–0.121***	0.125***	0.000	0.246***	0.067	0.207**	–0.026
	(0.028)	(0.029)	(0.027)	(0.028)	(0.032)	(0.038)	(0.072)	(0.070)
R^2	0.189	0.115	0.151	0.115	0.194	0.116	0.2	0.118
N	6,638	6,638	5,983	5,983	4,338	4,338	2,113	2,113
II. Blacks								
Male	–0.047	–0.235***	–0.061	–0.114	0.018	–0.244**	0.361**	0.029
	(0.049)	(0.065)	(0.052)	(0.072)	(0.086)	(0.092)	(0.124)	(0.151)
R^2	0.185	0.140	0.153	0.115	0.222	0.104	0.415	0.301
N	1,387	1,387	1,094	1094	576	576	245	245
III. Hispanics								
Male	–0.080	–0.158*	0.086	–0.044	0.163*	–0.172*	–0.042	–0.168
	(0.052)	(0.069)	(0.056)	(0.060)	(0.071)	(0.078)	(0.110)	(0.122)
R^2	0.267	0.135	0.180	0.133	0.216	0.141	0.229	0.077
N	1,429	1,429	1,324	1,324	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 percent levels, respectively.

C. General Knowledge/Science[12]

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.

As in math, boys' test-score performances are 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 increasing populated by boys, while the grade distribution provides no corresponding evidence that boys are outperforming 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, and 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.

V. Grades, Test Scores and the Role of Noncognitive Skills

A. Connecting grades to test scores and approaches toward learning

Now we turn specifically to the relationship between teacher grades, test scores and noncognitive skills. To examine the link we reestimate Equation 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.[13] So, the estimating equation becomes

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

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

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

13. Proceeding in this way does not treat cognitive and noncognitive skills in a parallel fashion in the sense that lagged test scores are omitted from the specifications. This is a potential problem if lagged test scores explain some of the gender gap in grades. However, including lagged test scores either as a covariate or using it as an instrument (like we do with the lagged ATL score in the next section) has no impact on the estimated gender gap in teacher grades. The results of this exercise are available as a web appendix.

Table 4c
Estimated Gender Gap in Science Test Scores and Grades, by Race and Ethnicity

	Kindergarten		First Grade		Third Grade		Fifth Grade	
	Test	Grade	Test	Grade	Test	Grade	Test	Grade
I. Whites								
Male	0.055*	–0.107***	0.103***	–0.064*	0.258***	–0.003	0.287***	0.003
	(0.023)	(0.028)	(0.022)	(0.029)	(0.030)	(0.036)	(0.055)	(0.074)
R^2	0.250	0.110	0.187	0.102	0.175	0.105	0.232	0.125
N	6,638	6,638	5,983	5,983	4,338	4,338	2,021	2,021
II. Blacks								
Male	–0.024	–0.257***	0.091	–0.122	0.198**	–0.220*	0.364**	–0.161
	(0.052)	(0.064)	(0.059)	(0.072)	(0.067)	(0.090)	(0.136)	(0.147)
R^2	0.293	0.121	0.236	0.121	0.298	0.104	0.472	0.200
N	1,387	1,387	1,094	1,094	576	576	265	265
III. Hispanics								
Male	–0.002	–0.175**	0.068	–0.018	0.114	–0.122	0.306**	–0.258*
	(0.055)	(0.063)	(0.056)	(0.063)	(0.066)	(0.076)	(0.111)	(0.111)
R^2	0.354	0.142	0.307	0.147	0.314	0.176	0.311	0.16
N	1,429	1,429	1,324	1,324	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 percent levels, respectively.

Equation 2 separately for each race and grade level.[14] 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 prekindergarten 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 noncognitive skills as measured by the ATL score. As evidenced in Table 1, the average ATL score for boys is roughly 15 percent 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.

1. 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 0.14–0.21 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 noncognitive 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 lagged "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 percent in kindergarten (0.68 to 0.62) to more than 15 percent 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 coef-

14. We also estimated a quantile version of Equation 2 to examine whether the estimated gender gaps in grades vary across the grade distribution. We considered cut points at each quintile and find no statistically significant differences in the estimated gender gaps, so we report only the OLS estimates.

Table 5a
Estimated Gender Gap in Reading Grades, Controlling for Test Scores and Noncognitive Skills

	First Grade			Third Grade			Fifth Grade		
	(a)	(b)	(c)	(a)	(b)	(c)	(a)	(b)	(c)
I. Whites									
Male	–0.247***	–0.144***	–0.074***	–0.272***	–0.172***	–0.089***	–0.278***	–0.210***	–0.090*
	(0.028)	(0.021)	(0.020)	(0.035)	(0.027)	(0.027)	(0.050)	(0.042)	(0.042)
Test score$_t$		0.684***	0.620***		0.687***	0.593***		0.603***	0.493***
		(0.012)	(0.013)		(0.016)	(0.017)		(0.026)	(0.029)
ATL score$_{t-1}$			0.191***			0.234***			0.260***
			(0.012)			(0.016)			(0.027)
R^2	0.118	0.516	0.544	0.144	0.476	0.516	0.168	0.423	0.471
N	5,983	5,983	5,983	4,338	4,338	4,338	4,327	4,327	4,327
II. Blacks									
Male	–0.257***	–0.086	–0.030	–0.402***	–0.214**	–0.161*	–0.350**	–0.253*	–0.220*
	(0.072)	(0.051)	(0.051)	(0.092)	(0.072)	(0.073)	(0.121)	(0.102)	(0.106)
Test score$_t$		0.863***	0.801***		0.704***	0.631***		0.588***	0.544***
		(0.028)	(0.031)		(0.042)	(0.044)		(0.056)	(0.059)
ATL score$_{t-1}$			0.146***			0.160***			0.131*
			(0.025)			(0.040)			(0.052)
R^2	0.149	0.589	0.604	0.143	0.481	0.500	0.190	0.435	0.448
N	1,094	1,094	1,094	576	576	576	538	538	538

III. Hispanics									
Male	–0.201**	–0.036	0.004	–0.407***	–0.221**	–0.136	–0.427***	–0.247***	–0.141*
	(0.064)	(0.047)	(0.047)	(0.080)	(0.070)	(0.071)	(0.076)	(0.067)	(0.067)
Test score$_t$		0.762***	0.701***		0.580***	0.497***		0.523***	0.454***
		(0.027)	(0.030)		(0.044)	(0.046)		(0.038)	(0.043)
ATL score$_{t-1}$			0.152***			0.193***			0.193***
			(0.028)			(0.035)			(0.038)
R^2	0.124	0.487	0.503	0.163	0.392	0.419	0.160	0.398	0.428
N	1,324	1,324	1,324	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 percent levels, respectively.

ficient, 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 nonwhite students. First, even when noncognitive skills are held constant, there remains a statistically significant gender difference in reading grades for black third and fifth graders and Hispanic fifth graders that is at least 50 percent larger than the estimated difference for whites in these grades. Second, the effect of a standard deviation increase in the lagged ATL score on reading grades is generally smaller for blacks and Hispanics.

2. *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 blacks and Hispanics the gender gap in grades is less precisely estimated but typically greater in magnitude. 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. 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 noncognitive skills on math grades diminishes in importance for whites relative to blacks and Hispanics as children advance through school.

3. *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. Again, including the noncognitive 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.

Table 5b
Estimated Gender Gap in Math Grades, Controlling for Test Scores and Noncognitive Skills

	First Grade			Third Grade			Fifth Grade		
	(a)	(b)	(c)	(a)	(b)	(c)	(a)	(b)	(c)
I. Whites									
Male	0.001	–0.062*	0.045	0.067	–0.082*	0.023	–0.026	–0.149**	–0.029
	(0.028)	(0.025)	(0.025)	(0.038)	(0.034)	(0.036)	(0.070)	(0.058)	(0.053)
Test score$_t$		0.499***	0.420***		0.608***	0.525***		0.592***	0.518***
		(0.014)	(0.014)		(0.020)	(0.022)		(0.039)	(0.041)
ATL score$_{t-1}$			0.233***			0.204***			0.172***
			(0.014)			(0.021)			(0.033)
R^2	0.115	0.326	0.367	0.116	0.366	0.395	0.118	0.361	0.382
N	5,983	5,983	5,983	4,338	4,338	4,338	2,113	2,113	2,113
II. Blacks									
Male	–0.114	–0.069	0.024	–0.244**	–0.254**	–0.220**	0.029	–0.187	–0.084
	(0.072)	(0.061)	(0.059)	(0.092)	(0.081)	(0.084)	(0.151)	(0.134)	(0.129)
Test score$_t$		0.743***	0.638***		0.538***	0.500***		0.597***	0.487***
		(0.042)	(0.044)		(0.049)	(0.056)		(0.065)	(0.069)
ATL score$_{t-1}$			0.213***			0.073			0.237***
			(0.033)			(0.053)			(0.067)
R^2	0.115	0.391	0.425	0.104	0.309	0.313	0.301	0.537	0.574
N	1,094	1,094	1,094	576	576	576	245	245	245

(*continued*)

Table 5b (*continued*)

	First Grade			Third Grade			Fifth Grade		
	(a)	(b)	(c)	(a)	(b)	(c)	(a)	(b)	(c)
III. Hispanics									
Male	–0.044	–0.094	–0.048	–0.172*	–0.252***	–0.169*	–0.168	–0.136	0.021
	(0.060)	(0.051)	(0.053)	(0.078)	(0.072)	(0.071)	(0.122)	(0.081)	(0.088)
Test score$_t$		0.587***	0.542***		0.495***	0.439***		0.742***	0.663***
		(0.038)	(0.038)		(0.040)	(0.047)		(0.053)	(0.053)
ATL score$_{t-1}$			0.120***			0.128**			0.212***
			(0.029)			(0.043)			(0.045)
R^2	0.133	0.341	0.351	0.141	0.322	0.335	0.077	0.492	0.523
N	1,324	1,324	1,324	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 percent levels, respectively.

Table 5c
Estimated Gender Gap in Science Grades, Controlling for Test Scores and Noncognitive Skills

	First Grade			Third Grade			Fifth Grade		
	(a)	(b)	(c)	(a)	(b)	(c)	(a)	(b)	(c)
I. Whites									
Male	–0.064*	–0.109***	0.010	–0.003	–0.119***	0.001	0.003	–0.138	0.009
	(0.029)	(0.027)	(0.027)	(0.036)	(0.035)	(0.035)	(0.074)	(0.072)	(0.070)
Test score$_t$		0.440***	0.350***		0.451***	0.383***		0.491***	0.393***
		(0.020)	(0.020)		(0.020)	(0.021)		(0.063)	(0.064)
ATL score$_{t-1}$			0.264***			0.248***			0.271***
			(0.016)			(0.020)			(0.045)
R^2	0.102	0.200	0.254	0.105	0.240	0.288	0.125	0.254	0.306
N	5,983	5,983	5,983	4,338	4,338	4,338	2,021	2,021	2,021
II. Blacks									
Male	–0.122	–0.166*	–0.052	–0.220*	–0.333***	–0.250**	–0.161	–0.326**	–0.201
	(0.072)	(0.065)	(0.065)	(0.090)	(0.083)	(0.086)	(0.147)	(0.125)	(0.131)
Test score$_t$		0.477***	0.401***		0.571***	0.497***		0.452***	0.376***
		(0.038)	(0.038)		(0.059)	(0.060)		(0.078)	(0.084)
ATL score$_{t-1}$			0.230***			0.151***			0.242***
			(0.032)			(0.044)			(0.061)
R^2	0.121	0.269	0.313	0.104	0.260	0.278	0.200	0.334	0.387
N	1,094	1,094	1,094	576	576	576	265	265	265

(*continued*)

Table 5c (*continued*)

	First Grade			Third Grade			Fifth Grade		
	(a)	(b)	(c)	(a)	(b)	(c)	(a)	(b)	(c)
III. Hispanics									
Male	–0.018	–0.045	0.016	–0.122	–0.173**	–0.078	–0.258*	–0.366***	–0.205
	(0.063)	(0.058)	(0.058)	(0.076)	(0.066)	(0.070)	(0.111)	(0.107)	(0.105)
Test score$_t$		0.407***	0.374***		0.451***	0.399***		0.353***	0.304***
		(0.033)	(0.033)		(0.048)	(0.048)		(0.056)	(0.055)
ATL score$_{t-1}$			0.167***			0.154***			0.223***
			(0.033)			(0.039)			(0.060)
R^2	0.147	0.256	0.277	0.176	0.300	0.322	0.160	0.254	0.296
N	1,324	1,324	1,324	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 percent levels, respectively.

4. Grading disparities and teacher characteristics

While the lagged ATL score accounts for most, if not all, of the overall gender disparity in grades, a natural question to ask is whether the estimated disparity varies by teacher characteristics. On this point, a potentially important characteristic is teacher gender, but as we explained earlier, the ECLS-K supplies this information only for kindergarten teachers, and this group is 98 percent female. Of the observable characteristics — experience and education — the literature emphasizes the role of the former over the latter in teacher performance (Hanushek and Rivkin 2010). It is certainly reasonable to suppose that experience improves teachers' assessments of students in a manner that reduces the gender gap in grading. So we replicated the results in Tables 5a–c, allowing the (student) gender effect to vary with teacher experience. We specified the interaction first using the continuous measure of experience, and then a binary measure distinguishing "experienced" teachers (more than two years) from the inexperienced (less than two years). Either way, the estimated coefficient of the interaction of the male dummy and experience varies in sign from case to case, but is typically small and statistically insignificant. Thus, we find no evidence that the effect of gender on grades depends systematically on teacher experience.

B. Refining the connection with a contemporaneous measure of noncognitive skills

Explaining the gender gap in teacher grades using a measure of noncognitive 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 first-stage regressions indicate that the lagged ATL score is a strong instrument. Its estimated coefficient is typically above 0.37 and five times larger than the standard error. Also, the male coefficient estimate is negative and statistically significant in every first-stage regression, 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 math and science 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,

Table 6a
Estimated Gender Gap in Reading Grades, Controlling for Test Scores and Noncognitive Skills – IV

	First	Third	Fifth
I. Whites			
Male	0.048*	0.042	0.085
	(0.020)	(0.028)	(0.046)
Test $score_t$	0.493***	0.463***	0.415***
	(0.016)	(0.020)	(0.032)
ATL $score_t$	0.500***	0.525***	0.593***
	(0.029)	(0.032)	(0.053)
R^2	0.624	0.574	0.521
N	5,973	4,329	4,309
II. Blacks			
Male	0.014	–0.009	–0.141
	(0.048)	(0.083)	(0.107)
Test $score_t$	0.641***	0.533***	0.480***
	(0.044)	(0.055)	(0.075)
ATL $score_t$	0.399***	0.428***	0.328*
	(0.062)	(0.102)	(0.129)
R^2	0.676	0.549	0.538
N	1,092	574	536
III. Hispanics			
Male	0.141**	0.090	0.039
	(0.048)	(0.076)	(0.082)
Test $Score_t$	0.568***	0.333***	0.356***
	(0.042)	(0.055)	(0.060)
ATL $Score_t$	0.442***	0.611***	0.508***
	(0.071)	(0.095)	(0.109)
R^2	0.593	0.521	0.465
N	1,321	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 percent levels.

albeit generally less than 0.10 standard deviations. For black third and fifth graders the estimated male coefficient is still negative, but it is not statistically significant. 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

Table 6b
Estimated Gender Gap in Math Grades, Controlling for Test Scores and Noncognitive Skills – IV

	First	Third	Fifth
I. Whites			
Male	0.232***	0.176***	0.123
	(0.028)	(0.041)	(0.065)
Test score$_t$	0.277***	0.422***	0.463***
	(0.019)	(0.027)	(0.044)
ATL score$_t$	0.605***	0.453***	0.401***
	(0.037)	(0.041)	(0.076)
R^2	0.431	0.445	0.398
N	5,973	4,329	2,105
II. Blacks			
Male	0.113*	–0.140	–0.040
	(0.056)	(0.110)	(0.140)
Test score$_t$	0.432***	0.465***	0.324*
	(0.055)	(0.070)	(0.104)
ATL score$_t$	0.559***	0.189	0.596**
	(0.083)	(0.139)	(0.179)
R^2	0.522	0.348	0.466
N	1,092	574	245
III. Hispanics			
Male	0.088	0.037	0.192
	(0.062)	(0.095)	(0.116)
Test score$_t$	0.444***	0.332***	0.575***
	(0.043)	(0.061)	(0.072)
ATL score$_t$	0.339***	0.410***	0.473***
	(0.074)	(0.115)	(0.113)
R^2	0.465	0.443	0.513
N	1,321	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 percent levels.

these subject-area tests and exhibit the same attitude toward 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.

Why are boys graded more favorably than girls when they have the same test scores and classroom behavior? One potential explanation is that teachers — who, in primary school, are overwhelmingly female — develop assump-

Table 6c
Estimated Gender Gap in Science Grades, Controlling for Test Scores and Noncognitive Skills—IV

	First	Third	Fifth
I. Whites			
Male	0.188***	0.152***	0.218**
	(0.030)	(0.037)	(0.075)
Test Score_t	0.221***	0.311***	0.334***
	(0.023)	(0.021)	(0.068)
ATL Score_t	0.626***	0.500***	0.602***
	(0.036)	(0.036)	(0.079)
R^2	0.329	0.351	0.335
N	5,973	4,329	2,011
II. Blacks			
Male	0.048	–0.089	0.387
	(0.066)	(0.104)	(0.286)
Test score_t	0.268***	0.434***	0.261*
	(0.045)	(0.070)	(0.122)
ATL score_t	0.542***	0.363**	0.785**
	(0.075)	(0.111)	(0.258)
R^2	0.411	0.331	0.215
N	1,092	574	263
III. Hispanics			
Male	0.166**	0.124	0.098
	(0.063)	(0.089)	(0.165)
Test score_t	0.300***	0.292***	0.236**
	(0.036)	(0.053)	(0.072)
ATL score_t	0.411***	0.436***	0.664***
	(0.075)	(0.100)	(0.196)
R^2	0.371	0.411	0.118
N	1,321	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 percent levels.

tions 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 support such expectations. Then, boys who act "out of character" by displaying the same noncognitive skills as girls with similar ability may receive special recognition. They may be, in essence, compensated for exceeding expectations.

VI. Conclusion

This paper extends the analysis of early-emerging gender differences in academic achievement to examine both (objective) test scores and (subjective) teacher assessments and connect the two. Using data from the 1998–99 ECLS-K cohort, we first 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. Boys in all racial categories (white, black, and Hispanic) across all subject areas (reading, math, and science) are not represented in grade distributions where their test scores would predict. We then trace the misalignment of grades and test scores to differences between boys and girls in their noncognitive development.

Boys who perform equally as well as girls on subject-area tests are graded less favorably by their teachers, but this less favorable treatment essentially vanishes when noncognitive skills are taken into account. For some specifications there is evidence of a grade "bonus" for boys with test scores and behavior like their girl counterparts.

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 at the early grade levels, such as developing ways to improve boys' noncognitive skills, creating alternative methods of instruction to communicate more effectively to boys who have different noncognitive 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* 103(484):1481–95.

Betts, Julian R., 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* 34(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* 26(4):450–62.

Claessens, Amy, Greg Duncan, and Mimi Engel. 2009. "Kindergarten skills and fifth grade achievement: Evidence from ECLS-K." *Economics of Education Review* 28(4):415–27.

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* (303):5–25.
Fryer, Roland G. Jr., and Steven D. Levitt. 2004. "Understanding the Black-White Test Score Gap in the First Two Years of School." *Review of Economics and Statistics* 86(2):447–64.
———. 2010. "An Empirical Analysis of the Gender Gap in Mathematics." *American Economic Journal: Applied Economics* 2(2):210–40.
Goldin, Claudia, Lawrence F. Katz, and Ilyana Kuziemko. 2006. "The Homecoming of American College Women: The Reversal of the College Gender Gap." *Journal of Economic Perspectives* 20(4):133–56.
Hanushek, Eric, and Steven Rivkin. 2010. "Generalizations about Using Value-Added Measures of Teacher Quality." *American Economic Review*: Papers and Proceedings 100(May): 267–271.
Heckman, James J., and Paul A. LaFontaine. 2010. "The American High School Graduation Rate: Trends and Levels." *The Review of Economics and Statistics* 92(2):244–62.
Holmlund, Helena, and Krister Sund. 2008. "Is the Gender Gap in School Performance Affected by the Sex of the Teacher?" *Labour Economics* 15(1):37–53.
Husain, Muna, and Daniel L. Millimet. 2009. "The Mythical 'Boy Crisis'?" *Economics of Education Review* 28(1):38–48.
Jacob, Brian A. 2002. "Where the Boys Aren't: Noncognitive Skills, Returns to School and the Gender Gap in Higher Education." *Economics of Education Review* 21(6):589–98.
Lavy, Victor. 2008. "Do Gender Stereotypes Reduce Girls' or Boys' Human Capital Outcomes? Evidence from a Natural Experiment." *Journal of Public Economics* 92(10): 2083–105.
Lavy, Victor, and Analia Schlosser. 2011. "Mechanisms and Impacts of Gender Peer Effects at School." *American Economic Journal: Applied Economics* 3(2):1–33.
Loury, Linda Datcher. 2004. "Siblings and Gender Differences in African-American College Attendance." *Economics of Education Review* 23(3):213–19.
Reynolds, John R., and Stephanie Woodham Burge. 2007. "Educational Expectations and the Rise in Women's Post-Secondary Attainments." *Social Science Research* 37(2):485–99.