STUDENT GENDER AND TEACHER GENDER:

WHAT IS THE IMPACT ON HIGH STAKES TEST SCORES?

Dr. John M. Krieg
MS-9074
Western Washington University
Bellingham, WA 98225
John.Krieg@wwu.edu
Phone: 360-650-7405
Fax: 360-650-4844

Abstract: A large literature establishes that boys and girls are treated differently in the classroom. Research suggests that this treatment depends upon the gender of the teacher. Using a large data set that observes a matched teacher/student sample over multiple years, this paper explores the impact of teacher and student gender differences on standardized test scores. Three notable findings are found: 1) conditional upon their test scores at the end of third grade, boys perform worse and gain less on math, reading, and writing during the 4th grade; 2) regardless of gender, students of male teachers perform worse than students of female teachers and; 3) there is no significant differential impact of male teachers on boys versus girls—both do equally poorly relative to students of female teachers. These findings cast doubt on the argument that teachers instruct students differentially based upon student gender.

Abstract: A large literature establishes that boys and girls are treated differently in the classroom. Research suggests that this treatment depends upon the gender of the teacher. Using a large data set that observes a matched teacher/student sample over multiple years, this paper explores the impact of teacher and student gender differences on standardized test scores. Three notable findings are found: 1) conditional upon their test scores at the end of third grade, boys perform worse and gain less on math, reading, and writing during the 4^{th} grade; 2) regardless of gender, students of male teachers perform worse than students of female teachers and; 3) there is no significant differential impact of male teachers on boys versus girls—both do equally poorly relative to students of female teachers. These findings cast doubt on the argument that teachers instruct students differentially based upon student gender.

A large literature examines the effect of teacher and student gender on teacher-student interactions, yet little research investigates if these interactions impact student outcomes as measured by standardized tests. Because of the high-stakes nature of standardized tests under The No Child Left Behind Act (NCLBA), it is imperative that researchers better understand the impact of teacher-student interactions on standardized test performance. For instance, many researchers argue that teacher gender differentially impacts the teacher's relationship with male and female students.[1] If the quality of student-teacher relationship impacts test performance, then teacher gender may place one gender-group of students at a disadvantage when taking standardized tests.

Researchers have found that teachers interact differently with students of similar gender than they do with students of opposite gender.[2] This includes evidence suggesting disciplinary procedures and proclivity to discipline vary by both student and teacher gender. Likewise, a teacher's perception of student characteristics and abilities appear to systematically vary by gender. Other studies find male students benefit at the expense of female students in the amount and quality of interaction received from teachers of both genders. What has yet to be determined is how these differences in discipline, perceptions of student ability, and interactions between student and teacher influence student outcomes as measured by standardized exams.

This paper explores this issue by following a large subset of Washington 3rd graders over a two year period that concludes with students completing the Washington Assessment of Student Learning (WASL). The WASL is the standardized test the state of Washington has chosen to employ to comply with the NCLBA. Combining these test results with specific teacher information provides a comprehensive data set that allows one to test the impact of

[1] See for instance Meece (1987), Hopf and Hatzichristou (1999) and Rodriguez (2002).

student and teacher gender on standardized test results. After controlling for measurable student and teacher characteristics, this paper demonstrates three interesting findings. First, like a considerable amount of previous research suggests, boys score considerably worse on the math, reading, and writing sections of the WASL after controlling for test scores given during previous academic years. Secondly, on average, students of male teachers score worse on the WASL than do students of female teachers. Finally, although students of either gender score worse on the WASL when instructed by a male teacher, there is no differential impact of male teachers on the WASL scores of boys compared to girls. This evidence suggests that although disciplinary procedures, perceptions of gender differences, and interactions with students may differ between teachers by gender, these differences do not result in differential test scores between boys and girls.

This paper proceeds as follows: in the next section, previous research documenting the gender differences in the classroom is summarized. The second section describes the WASL exam and documents regression results that investigate the question of gender differences. The final section presents discussion and conclusions.

I. Literature Review

The amount and type of attention students receive from teachers has long been a topic of interest to researchers. Numerous studies examine gender differences and the patterns of these interactions (Lockheed & Harris, 1984; Sadker, Sadker & Bauchner, 1984; Massey & Christensen, 1990; Rodriguez, 2002; Einarsson & Granström, 2002) with most documenting greater amounts of teacher attention directed toward boys rather than girls. Research that delves carefully into the reasons under which this "overattention" to boys occurs suggests a

---

[2] See for instance Etaugh and Hughes (1975), and McCandless, Bush and Carden (1976).

host of potential causes. For instance, if society stresses the success of males above that of females, then teachers may unconsciously promote male students by paying greater attention to them.

While a large body of research focuses on the gender of students, less research explores the impacts of a teacher's gender on students (Hopf & Hatzichristou, 1999). Evidence suggests that male teachers tend to be more authoritative whereas female teachers tend to be more supportive and expressive (Meece, 1987). A survey of 20 teachers indicates that male teachers are likely to select a more aggressive disciplinary approach toward boys while teachers of either gender tended to ignore boys' disruptive behavior than that of girls when the behavior was not aggressive (Rodriguez, 2002).

Teacher gender is also systematically related to class environment. A number of studies suggest that male teachers provide a more positive atmosphere for boys (Etaugh & Hughes, 1975; McCandless, Bush & Carden, 1976); however, relative to male teachers, Stake and Katz (1982) suggest that female teachers tend to provide a more positive classroom atmosphere overall. After observing 40 class sessions, Einarsson and Granström (2002) find that male teachers increase the attention paid to girls as pupils age while female teachers consistently give more attention to boys.

Previous research also suggests that differences in teacher’s perceptions of student abilities and characteristics are related to teacher gender. Parker-Price and Claxton (1996) surveyed teachers regarding their perceptions of student abilities. They learned that male teachers are more likely to believe that boys are superior visual learners while girls are more helpful in the classroom. On the other hand, female teachers do not demonstrate these differences in belief but do tend to think that boys are better with quantitative skills.

While it is clear that teachers treat and perceive boys and girls differently, it is less clear how this differential treatment impacts student performance on standardized exams. Of course, a large literature establishes differences on standardized exams by gender of student, but no research connects test results to teacher gender and its interaction with student gender. If, as the previously mentioned studies suggest, male teachers treat students differently than female teachers, then one would expect teacher gender to influence student outcomes on standardized exams. Further, if male teachers treat boys differently than girls, then one would also expect standardized test score differences between boys and girls to vary systematically by teacher gender. Although teachers may overtly treat students differently by gender, overt treatment need not be the sole vehicle for generating gender-based test score differences. If, as Parker-Price and Claxton suggest, boys learn better through visual experiences, then it would be natural for a male teacher, who also learned better through these experiences, to revert to visual teaching leading to better performance by the boys in his class. The next section tests the impact and interaction of teacher gender and student gender on student test performance.

II. Estimation Procedure and Results

The NCLBA was signed into law by President Bush on January 8, 2002 and its provisions will be phased in over a period of several years. The law places important conditions on the use of federal Title I funds targeted to aid students in high poverty schools. States are required to assess the performance of schools and to reward schools that perform well while prescribing corrective action for schools that fail to meet benchmarks set by law. No specific assessment instruments are prescribed, but these assessment methods must test performance of all public school students within the state in at least two core areas:

reading/language arts and mathematics. The results of these tests must be stated in terms of proficiency levels of students rather than percentile scores.

The WASL is the state of Washington's diagnostic tool intended to identify faltering schools under NCLBA. The WASL is a mixed open-ended, short answer, and multiple choice exam covering four distinct areas of learning: reading, writing, listening and mathematics.[3] The intent of the WASL is to measure the application of basic skills to real-world situations with a large number of comprehension, application, and analysis questions as categorized by Bloom's Taxonomy. The WASL is administered in grades 4, 7, and 10 and, under current state legislation, students need to pass the WASL in order to receive a high school diploma. For each section of the WASL the state chooses a minimum score required for passing that section. In the 2002-2003 academic year 34.4% of 4th graders, 27.2% of all 7th graders, and 33.5% of all 10th graders met all four WASL standards.[4] This work measures student performance on the WASL in two ways: by creating a binary variable equal to one if the student passes all four sections of the WASL and zero otherwise; and by measuring each student's score on the individual reading, writing, listening and mathematics sections of the WASL. In order to make comparisons with other standardized tests easier, the raw scores on

---

[3] The listening section of the WASL has long been criticized for not testing material in direct relation to standard academic goals and for having a very large variance in outcomes. After the 2002-2003 year examined in this work, Washington decided to phase the listening section out to be replaced by a science section. Currently, the science section is undergoing state-wide tests of validity.

[4] Taylor (2000) reported on the psychometric properties of the WASL administered to fourth graders in 1999. The validity of the WASL was examined using the correlations among WASL strand scores (correlations ranged between .42 and .75) and an explanatory factor analysis. The latter analysis revealed found factor load ratings in writing, mathematics, and reading ranging from .62 to .79. Internal consistency reliability coefficients for the 1999 grade 4 WASL were satisfactory for listening (α=.61), reading (α=.86), math (α=.88), and writing (α=.81). Inter-rater reliability coefficients for the open-ended and writing questions were high with correlations ranging from .97 to .98. Finally, Brickell and Lyon (2003, p. 3) find that the WASL has made annual improvements in reliability scores to the point that it has "reached traditionally reported levels found for standardized achievement tests." Further, after analyzing multiple years of WASL results, Brickell and Lyon conclude that the Standard Error of Measurement (SEM) have been consistent across years although the level of the SEM is large enough to caution against using the WASL solely to make decisions regarding a student's measured academic progress.

each the WASL's individual sections have been normalized so that the mean of the observations are zero with standard deviation of one.[5]

The data set employed by this paper examines the 49,415 4th graders who took the WASL exam at the end of the 2002-2003 academic year. A majority of these 4th graders took the Iowa Test of Basic Skills (ITBS) in the previous year. The Iowa tests are annual standardized exams intended to identify a student's developmental level and to measure annual academic growth. Measuring a student's 3rd grade ITBS results against their 4th grade WASL results allows one to estimate the gains made in the student's 4th grade year. A further benefit to merging WASL and ITBS results is that students taking the ITBS exam provide a wealth of personal and demographic information that is likely correlated with WASL test performance. This information is incorporated in the empirical strategy used in this paper.

The sample of 49,415 students represent 2,519 different classrooms distributed over 965 school buildings in 251 school districts. This accounts for 49.6% of all Washington 4th graders in 85% of buildings that offer 4th grade in 84.7% of all Washington districts. One method of measuring the impact of teacher and student gender on test performance is to examine descriptive statistics which are provided in Table 1. Of these students, 51.3% are male, 4.7% are black. 6.8% are Asian and 11.3% are Hispanic. Of special interest to this paper is the relatively small numbers of male students that are taught by male teachers; even though over half of students are male, only 10.2% of student-teacher combinations are both male. Of course, this is because male teachers are relatively rare at the 4th grade level.[6] Table

---

[5] Raw scores for each section of the WASL varied by section. For the sample of students used in this paper, the raw math scores ranged from 3 to 55 (mean 35.2), for reading 1 to 40 (mean 31.1), for listening 0 to 10 (mean 8.8) and for writing 0 to 12 (mean 8.3).

[6] In fact, only 28.9% of all certified employees in Washington (including principals and district administrators) are men. At the elementary level, only 12.2% of all certified employees are men. Although the percent of male teachers represented in the sample are small, it is important to note that the size of the overall sample is larger than any mentioned in the literature review that measure gender effects. Thus, roughly 10% male teachers

1 also presents direct comparisons of differences-in-means test results. Compared to male students, girls score significantly better than boys on the reading and writing components of the WASL and slightly worse on the listening component. Girls are also more likely to use a computer for school work, are more likely to read often for fun, and are more likely to come from a home in which English is never spoken, while boys are more likely to be held back at least one grade in the past.

Table 1 also provides a set of comparisons between those students who share the same gender with their teachers and those who do not. Interestingly, students of the same gender as their teacher score better on reading and writing and were overall more likely to pass the WASL exam than students of opposite gender than their teachers. While this may indicate that students benefit from being instructed by teachers of similar gender, it is important to remember that these descriptive statistics do not control for other factors that might influence student test scores. The remainder of this work uses regression analysis to determine the conditional impact of teacher and student gender on test scores.

This paper measures the impact of teacher gender on students by estimating variants of the following equation:

(1) $$\text{WASL Score Result}_i = B_0 + B_1\text{Student Male}_i + B_2\text{Teacher Male}_i + B_3\text{Same Gender}_i + \Gamma\mathbf{X}_i + \varepsilon_i$$

where the B's measure the marginal impact of the variables on the WASL score, **X** is a matrix of control variables, Γ are the coefficients corresponding to the control variables, ε represents a random error term, and i indexes individual students. The three variables of interest to this paper, Student Male, Teacher Male, and Same Gender are zero-one binary variables. In the case of Student and Teacher Male, these variables equal one if the observation is a male and

represent about 250 male teachers (and rougly 5,000 students) which is a large number of male teachers compared to Einersson and Granström who examine 40 total teachers.

zero otherwise. The variable Same Gender is equal to 1 if both the student and their teacher are of the same gender and 0 otherwise.

Equation (1) presents the opportunity to test the impact of gender on student performance. A negative estimate of B_1 indicates that on average, girls score better on the WASL than boys. This is likely to be the case if girls develop academically faster than boys. B_2, the coefficient on Teacher Male, represents the impact on a student's WASL score if their teacher is a man. If students respond better to male teachers than female teachers, then the estimate of B_2 will be positive. On the other hand, if students respond to the more positive attitudes of female teachers, as suggested by Stake and Katz, then B_2 would be negative. Finally, the coefficient B_3 determines the impact of sharing the same gender with their teacher on a student's test scores. A positive estimated value of B_3 indicates that boys [girls] perform better on the WASL exam when taught by male [female] teachers. On the other hand, an estimated negative value of B_3 indicates that students perform better on the WASL if they are of opposite gender than their teacher. One might expect B_3 to be positive if male teachers focus more on male students (as suggested by Etaugh and Hughes).

The control variables in equation (1) include both individual student and teacher measures. Basic student demographic measures are controlled for such as race, migrant status, and the frequency with which English is spoken in the student's home. Other student measures include the length of time a student has been enrolled in both their current school and the school district, if they changed schools in the middle of their fourth grade year, if the student has computer access at home and if computers are used for homework. Further student measures contain the frequency that students read books for fun, the frequency they watch television, and if they have ever been held back a grade in school. Individual teacher characteristics included in Equation (1) are the teacher's race, their level of college degree (bachelors, masters, or doctorate), the number of academic credits and in-service credits earned since beginning employment as a teacher, and years of experience and its square in the teaching profession. Measures of the length of the school day and the school year are controlled for as are binary variables indicating if the teacher is in their first year of teaching, or new to either their building or district. In sum, 55 control variables are included in Equation (1) in addition to the variables that measure the impact of student and teacher gender.

The results of estimating equation (1) and four variants are presented in Table 2.[7] All regressions contain the control variables listed above as well as standard errors corrected for groupwise heteroskedasticity as suggested by Wooldridge (2002).[8] In order to make later comparisons, Panel A of Table 2 presents results of baseline regressions that contain the control variables and only the Student Male variable. Later panels will introduce the other measures of teacher and student gender. The purpose of this piecewise introduction of gender variables is not to determine correct functional form of the regression, but rather to follow the impact of the introduction of the additional gender variables on the ones already included. If the already included coefficients on the gender variables remain constant with addition of new variables, one can conclude that no significant relationship exists between the variables in question.

From the regression presented in panel A, it is clear that male fourth graders perform differently than do female fourth graders on the WASL exam. On average, boys score slightly better than girls on the math and listening sections of the exam but considerably

[7] In order to save space, the entire regression results, including the coefficients on the 55 control variables have been suppressed. The author will gladly share the full set of results upon request.
[8] Groupwise heteroskedasticity, which is caused in this application by students being nested within buildings which are nested within districts, typically causes one to underestimate the standard errors of the coefficients. Wooldridge's procedure provides consistent standard errors.

worse on the reading and writing portions of the exam. Specifically, boys on average score .37 standard deviations and almost .17 standard deviations worse on the writing and reading sections on the exam than girls. To get a feel for the importance of these numbers, the non-reported estimated coefficient on a dummy variable indicating if the student is black is -.52.[9] In other words, on average white students score just over one-half of a standard deviation better than their black counterparts. Thus, the deficit of .37 standard deviations that boys face when compared to girls on the writing section of the test is about 70% of the size of the deficit black students face compared to whites.

Panel B presents a similar regression as presented in Panel A but also includes the variable Teacher Male. Two themes are notable in these regressions. First, on average, students of male teachers perform worse on all sections of the WASL than do students of female teachers. The magnitude of this "Male Gender" deficit is fairly small; on each of the four parts of the WASL the difference is less than one-tenth of a standard deviation. Secondly, the estimates of the impact of Student Male in Panel B did not significantly change relative to those estimates in Panel A. Statistically, this indicates that the correlation between Student Male and Teacher Male is near zero. Thus, it appears as if there is little systematic placing of students in teachers classrooms based upon either the student's or teacher's gender.

Panel C introduces the teacher-student gender interaction term. The patterns indicated by Panel C follow those of previous panels: on average male students do slightly better on the mathematics and listening portions of the test and considerably worse on the reading and writing sections while students of male teachers are at a small disadvantage relative to students of female teachers. Included in Panel C are estimated coefficients corresponding to

[9] That black's score worse than whites is a standard finding of test literature. As a matter of fact, Levine and Eubanks (1990) find that black's, on average, score one-half of a standard deviation worse than whites on standardized reading tests—this is exactly the same magnitude as found in this paper.

the variable Same Gender. The purpose of including this variable is to test if students benefit by having teachers of the same gender. If, as previous research suggests, teachers treat students of similar gender differently, then one would expect statistically significant coefficients on this variable. The positive coefficients on Same Gender in Panel C indicate that students of the same gender as their teachers benefit in a small, statistically significant way on only the math and reading sections of the WASL. On average, students of the same gender as their teachers score .026 standard deviations higher on the math and .019 standard deviations higher on the reading tests than students of opposite gender than their teachers. Although these estimates are statistically different than zero, these estimates are relatively unimportant when compared to coefficients on other variables. For instance, the estimated (and unreported) impact of being held back one grade is that test scores fall by .349 standard deviations. Likewise, the black-white test gap of .520 standard deviations, changing schools in the middle of the year (a fall of .233 standard deviations), and simply being from a home with a computer (positive .250 standard deviations) all dwarf the impact of sharing gender with one's teacher. As a matter of fact, the expected impact on student test scores for another year of teacher experience is .013 standard deviations. Thus, the expected benefit of having a teacher of the same gender as a student amounts to about the same benefit of having a teacher with two additional years of teaching experience, holding all else constant.

One concern with the results of panels A, B, and C is that students' standardized test scores are likely to be results of cumulative education occurring in previous grades. If true, then it would not be surprising to see that the gender of the student's fourth grade teacher has little impact on test scores because these scores account for cumulative impacts of previous educational experiences. A further concern with panels A, B, and C is the relatively small amount of WASL variance explained by the included 55 variables and measured by the adjusted R^2.[10] One way to address both concerns is to use students' performance on standardized tests given in the third grade as explanatory variables in the WASL regressions. If standardized tests measure accumulated learning over past grades and are also highly predictive of future performance, then including test scores from the third grade to equation (1) will generate regression results that control for this accumulation upon entering the fourth grade and provide a greater explanation of test score variation. Specifically, the following regression is estimated:

(2) $\text{WASL Score Result}_i = B_0 + B_1\text{Student Male}_i + B_2\text{Teacher Male}_i + B_3\text{Same Gender}_i + B_4 3^{rd}\text{ Grade Test Score}_i + \Gamma\mathbf{X}_i + \varepsilon_i$

To summarize, the difference between equation (2) and (1) is that by including the 3rd grade test score, the coefficients of interest in equation (2) measure the impact of student and teacher gender on fourth grade test scores holding students' ability in the 3rd grade constant. Put another way, equation (2) measures the value added to test scores over only the 4th grade year.

The State of Washington administers the WASL in the 4th grade year. The measure of third grade test scores employed in equation (2) is student's performance on the individual components of the ITBS. Specifically, the student's score on the math section of the ITBS is matched with the math WASL, the ITBS listening score with the WASL listening score, and the ITBS reading score with the WASL reading score. As the ITBS does not offer a writing test for third graders, the ITBS vocabulary score was matched with the fourth grade WASL writing score.[11] The ITBS is administered near the end of the third grade year, thus it is likely

[10] One reviewer noted the rather low adjusted R^2 of the regressions in panels A through C. This is fairly typical of regressions that examine standardized test scores which are notoriously difficult to predict. Further, the statistically significant F-tests measuring the overall fit of the model, as well as the large number of statistically significant regression coefficients (including those that measure the impact of gender) indicate that despite the low adjusted R^2, we are still able to draw valid statistical inferences from these models.

[11] Although it is likely that the open-ended writing WASL actually tests higher order cognitive skills than the vocabulary based ITBS, there is no 3rd grade test that more closely matches the WASL writing test. This likely

to be an appropriate control variable for student's abilities at the beginning of the next academic year. Because many students who took the WASL in the fourth grade were either unable to take the ITBS in the third grade or were not tracked by the state between the two years, including ITBS scores in equation 2 decreases the sample size of the regression from 49,415 to 39,124 observations.[12]

Results of estimating equation (2) are presented in panel D of Table 2. Given the large increase in adjusted R^2, the performance on the third grade ITBS test is a significant and important predictor of fourth grade WASL results. A student who scores one standard deviation above average on the math ITBS is expected to score .685 standard deviations above average on the math WASL. The importance of the ITBS on the other WASL subjects are equally impressive; coefficients of .615 on the reading, .436 on writing, and .345 on listening are all statistically significant and meaningful coefficients.

It is interesting to note how including measures of ITBS scores influences the estimated coefficients on Student Male, Teacher Male, and the same gender measures. After controlling for ability at the end of the third grade, compared to girls, boys do worse on the math, reading, and writing components of the WASL and better only on the listening. Other than the math result, this is identical to the previous regressions. If one views the results in panel D as the value added to a student's performance by the 4^{th} grade, then the negative math coefficient estimated for boys indicates that they grow during the fourth grade slower than girls. Given the previous positive coefficients on the math variables, this suggests that boys started with more mathematical aptitude than girls but girls close the gap over time.

---

accounts for the large differences in coefficients found in Tables 2 and 3 between the writing test and the reading and math tests.

[12] The reason for the decrease in the sample size is that the state of Washington does not match ITBS and WASL scores for students who switch buildings or districts between their ITBS and WASL years. Thus, the regressions in panels D, E, and Table 3 consist only of students who were able to be followed for both their 3^{rd} and 4^{th} grade years.

Another reoccurring result from panel D is that students of male teachers are less likely to score well on the math, reading, and writing sections of the WASL. While these patterns are similar to those demonstrated in the previous regressions, significant changes occur in the coefficients on Same Gender. The same gender coefficients are statistically no different than zero for reading, writing, and listening results. Sharing the same gender as the teacher only gives a very small impact (.013 of a standard deviation) for math results and this impact is statistically significant only at the 10% level. These results suggest that if favoritism or benefit exists between teachers and students of similar gender, then its impact on standardized test scores is so small that its impact is hardly important.

Another potential concern regarding the analysis performed so far is that each student is nested within schools which, in turn, are nested within districts. In order to control for the variation in student test scores caused by building and district-level impacts, building and district fixed effects were added to equation (2). The following equation was estimated:

$$(3)\qquad \text{WASL Score Result}_{isd} = B_0 + B_1\text{Student Male}_{isd} + B_2\text{Teacher Male}_{isd} + B_3\text{Same Gender}_{isd} + B_4 3^{rd}\text{ Grade Test Score}_{isd} + \Gamma\mathbf{X}_{isd} + v_s + w_d + \varepsilon_i$$

In this model, s indexes school buildings, d indexes schools districts, v and w are school and district fixed effects that vary for each building/district combination in Washington. If gains to the WASL are related to the gender variables and to individual building or district policies, then equation (3) will control for this relationship leaving the coefficients on B_1, B_2, and B_3 the unbiased estimates of the effect of gender on the WASL.

Panel E of Table 2 reports the coefficients on the gender variables in the presence of building and district fixed effects. The estimated coefficients changed very little from those estimated in panel D indicating the relative unimportance of building and district fixed effects upon the gender coefficients.[13] Further, the inclusion of building and district fixed effects actually reduced the adjusted R^2's in all four regressions. This is because after controlling for a student's ability through the past ITBS test (which is likely a function of building and district effects), there is little extra variation in the WASL test correlated to these fixed effects.[14]

Rather than investigating the individual sections of the WASL exam, an experiment that conforms more closely with the spirit of high stakes tests is to inquire about the impact of gender on the ability to pass these tests. Each year the WASL is given, and as mandated by NCLBA, a state mandated minimum score on each of the sections is required to demonstrate proficiency. A student must meet this score on each of the four sections in order to pass the WASL. Rather than using OLS to estimate the impact of gender on individual test scores, I propose to estimate the following fixed-effects logit model that predicts if students pass the exam:

$$(4) \quad \text{WASL Pass}_{isd} = f(B_0 + B_1\text{Student Male}_{isd} + B_2\text{Teacher Male}_{isd} + B_3\text{Same Gender}_{isd} + B_4 3^{rd}\text{ Grade Test Score}_{isd} + \Gamma\mathbf{X}_i + v_s + w_d + \varepsilon_i)$$

In equation (4), WASL Pass is a binary variable taking on a value of 1 if the student passed the WASL and 0 if the student failed the WASL and f represents the standard fixed effects

[13] As pointed out by one reviewer, a possible reason why the interactive gender term does not statistically impact student test scores is because the interactive gender term is correlated with the other explanatory variables. In order to check this, variance inflation factors were computed for each model estimated in panel E of Table 2. The highest VIF estimated on the Student/Teacher gender interaction term was 1.58—a finding much less than the general rule of thumb of 10 that indicates the presence of severe collinearity.

[14] In order to check for robustness of results, the residuals from panel E were examined for normality and for outliers that may be driving the results. First, any observation resulting in a residual more than 3 standard deviations from 0 were eliminated and the models were re-estimated. The math results, for example, had only 25 observations that fit in this category. After eliminating those, no significant differences in the results presented were found. Again, using the math regressions as an example, skewness (.38) and kurtosis (4.10) statistics demonstrate that the residuals were near-normal. This is a similar finding for the other three models.

logit function.[15] The results of estimating equation (4) are presented in the first column of Table 3.

The estimates of the logit model follow closely those of the estimates of equation (4). Holding other independent variables constant, male students are 8.3% less likely to pass the WASL compared to girls. Likewise, students of male teachers are 3.4% less likely to pass the WASL compared to students of female teachers. Finally, the estimated probability of passing the WASL if students share the gender of their teacher decreases by a statistically insignificant .4%. This provides further evidence that students do not benefit simply because they share the same gender as their teacher.

Table 3 presents a subset of independent variables included, but not reported, in the other regressions. The coefficients on these variables are not surprising; students from homes that speak a language other than English are 13% less likely to pass the WASL while students who have been held back a grade in the past are 6% less likely to pass. Likewise, students of teachers with greater years of experience are more likely to pass the WASL although, as indicated by the negative coefficient on squared experience, the impact of an additional year of teacher experience on the probability of passing the WASL diminishes as teacher experience grows.

The analysis presented so far may suffer from bias caused by an important omitted variable. Consider a set of parents who have a high level of concern for their child's education. Because of this concern, these parents are likely to spend additional time and resources promoting their child's education and hence are likely to have children that pass the WASL with greater frequency. If these parents also believe their students benefit from having teachers of the same gender, then these parents will lobby the school administration for their

[15] Chapter 19 of Greene (2000) gives details on the fixed effects logit specification used here.

students to share the same gender as their teacher. Thus, the coefficients on Same Gender may proxy for the fact that students with caring parents do better because of their unobserved background rather than any impact of sharing gender with their teacher. If this is the case, then the coefficients reported previously will be biased in a positive direction.

In order to account for this type of omitted variable, all schools were eliminated from the sample that employed fourth grade teachers of different genders. Examining students whose parents are unable to choose between teachers by gender eliminates the possibility that the variable Same Gender proxies for parental sorting of students into classrooms based upon teacher and student gender. After eliminating from the sample students attending schools with fourth grade teachers of different gender, 20,075 observations remain. Using these remaining observations, Equation (4) is re-estimated with results reported in the second column of Table 3.

Very little substantive differences exist when comparing the results based upon the partial sample with the results from the complete sample. As expected, the coefficient on Same Gender moves in a negative direction lending some support for the hypothesis that concerned parents may place their students into classrooms based partially upon the gender of the teacher. The estimated impact of sharing the teacher's gender is that students are 3.1% more likely to fail than students who do not. More importantly though is the fact that after correcting for this potential, the coefficient on Student Male remains small in magnitude and statistically no different than zero. This supports the earlier findings that students do not perform better on standardized exams because they share the same gender as their teachers.

Interestingly, when gender choice is eliminated from the choice set of parents, the impact of teacher gender on passing the WASL grows. In the complete sample, a student of a male teacher was expected to fail the WASL with 2.7% more likelihood than a student of a female teacher. When the sample consists only of buildings that have either all male or all female teachers, students of male teachers are expected to fail the WASL 6.9% of the time. This suggests that some type of systematic sorting occurs in buildings with both male and female teachers. For instance, it may be that male teachers are given the more advanced students and female teachers assigned the less advanced. Thus, in the prior regressions that analyzed the entire sample, the impact of male teachers was estimated to be smaller than it really is because they are dealing with better students than female teachers. In the later case, when better students cannot be sorted into classrooms by gender, the estimated impact is much larger because male teachers would be teaching a more representative sample of students.

III. Discussion and Conclusion

Earlier work on gender in the classroom suggests that teachers treat students of their gender differently compared to students of opposite gender. Some of these differences include disciplinary interactions, perceptions of student characteristics, and the amount of attention devoted to students. While not directly testing for differential treatment within classrooms, this paper is asks if differential outcomes on high stakes tests depend upon student and teacher genders.

Previous research suggests male teachers discipline boys differently than girls, provide a more positive atmosphere for boys, and have different perceptions of boys ability relative to girls. If true, one might expect boys to perform differently on standardized exams when in a male teacher's classroom than in a female teacher's class. Using a large matched sample of Washington 4th graders and their teachers, the most reliable estimates this paper finds no statistically significant impact of the interaction between student and teacher gender. In other words, no evidence is found that students and teachers sharing the same gender impacts student performance on standardized tests suggesting that the impact of the differential treatment found by other authors either is insignificant to academic progress or results in changes not measured by high stakes testing.

Although no evidence is found to support the hypothesis that the interaction of student and teacher gender impacts test scores, a number of findings indicate teacher and student genders are correlated with test outcomes. For instance, regardless of their gender, students of male teachers are 2.7% less likely to pass the WASL than students of female teachers. This may be a function of differences in education philosophies by male teachers compared to female teachers. If male teachers are viewed by students as being more strict, less caring, or more aloof then it would not be a surprise that all students respond less well to male teachers than female teachers. Perhaps this finding is related to Hopf and Hatzichristou's finding that female teachers tend to be more supportive towards all of their students and this support is a needed component in education. Of course it may also be related to the argument proposed by Etaugh and Hughes, as well as McCandless, Bush & Carden, that male teachers provide a positive atmosphere for boys. If boys are relatively less needy of such amenities, then perhaps in providing a good atmosphere for boys, male teachers reduce their overall effectiveness resulting in poorer performing students overall.

A second finding suggests that male teachers may actually cause their students to perform more poorly than the 2.7% decline in pass rates indicates. After eliminating all buildings that employ both male and female fourth grade teachers, this paper estimates that male teachers have students that fail the WASL with 6.9% greater frequency than female teachers. Eliminating all buildings with choice in the gender of fourth grade teachers reduces the possibility that the impact of teacher gender on student performance is biased by the nonrandom sorting of high-ability students into a male teacher's classroom. Since the impact of male teachers on students increased in this scenario, it is possible that parents or principals place high ability students with male fourth grade teachers leading to the lower estimated failure rates in the complete sample model. Another possibility is that schools who hire only male fourth grade teachers share some unmeasured characteristic that causes all students to fail the WASL more frequently and this impact is being attributed to male teachers.

Regardless of teacher gender, this work also finds that boys tend to perform less well than girls. As a matter of fact, boys are expected to pass the WASL 8.6% less often than girls, even after controlling for past performance on standardized exams and other individual characteristics. This finding is not surprising given the fact that much research argues that boys in the fourth grade are less academically developed than girls. In conclusion, while this paper does not address if students are treated differentially by teachers of similar gender, it does suggest that if some type of gender bias occurs, it has little impact student's standardized test scores.

Bibliography

Brickell, J., & Lyon, D. (2003). Reliability, Validity, and Related Issues Pertaining to the WASL. Washington Education Association Research Report: Olympia, WA.

Einarsson, C., & Granström, K. (2002). Gender-biased Interaction in the Classroom: The Influence of Gender and Age in the Relationship Between Teacher and Pupil. *Scandinavian Journal of Educational Research*, 46, pp. 117 – 127.

Etaugh, C., & Hughes, V. (1975). Teacher's Evaluation of Sex-Typed Behavior in Children: The Role of Teacher Sex and School Setting. *Developmental Psychology*, 11, pp. 394-395.

Greene, W. H. (2000). *Econometric Analysis*. Upper Saddle River, NJ: Prentice Hall Publishers.

Hopf, D. & Hatzichristou, C. (1999). Teacher Gender-Related Influences in Greek Schools. *British Journal of Educational Psychology*, 69, pp. 1 – 18.

Levine, D. & Eubanks, E. (1990). Achievement Disparities Between Minority and Nonminority Students in Suburban Schools. *Journal of Negro Education*, 59, pp. 186 – 194.

Lockheed, M., & Harris, A. (1984). A Study of Sex Equity in Classroom Interaction. Final Reports #1 and #2, Educational Testing Service, Princeton, New Jersey.

Massey, D., & Christensen, C. (1990). Student Teacher Attitudes to Sex Role Sterotyping: Some Australian Data. *Educational Studies*, 16, pp. 95-107.

McCandless, B., Bush, C., & Carden, A. (1976). Reinforcing Contingencies for Sex-Role Behaviors in Preschool Children. *Contemporary Educational Psychology*, 1, pp. 241-246.

Meece, J.L., (1987). The Influence of School Experiences on the Development of Gender Schemata. In L.S. Liben & M.L. Signorella (eds.), *Children's Gender Schemata*: Jossey-Bass, San Francisco, pp. 57 – 73.

Parker-Price, S., and Claxton, A. (1996). Teacher's Perceptions of Gender Differences in Students. Paper presented at the Annual Convention of the National Association of School Psychologists.

Rodriguez, N., (2002). "Gender Differences in Disciplinary Approaches," ERIC Document SP041019.

Sadker, M., Sadker, D., & Bauchner, J., "Teacher Reactions to Classroom Responses of Male and Female Students," Washington, DC, National Institute of Education, ERIC Document ED245839, 1984.

Stake, J. & Katz, J. (1982). Teacher-Pupil Relationships in the Elementary School Classroom: Teacher-Gender and Pupil Gender Differences. *American Educational Research Journal*, 19, pp. 465-471.

Taylor, C.S., (2000). Washington Assessment of Student Learning, Grade 4, 1999, Technical Report. Washington Office of the Superintendent of Public Instruction: Olympia, WA.

Wooldridge, J. (2002). *Econometric Analysis of Cross Section and Panel Data*. Cambridge, MA: MIT Press.

Table 1: Selected Descriptive Statistics of Students

	Full Sample	Male Students		Female Students	Same Teacher/Student Gender		Different Teacher/Student Gender
WASL Math Score	0 (1)	-.005 (.999)	=	.006 (1.000)	.009 (1.001)	=	-.008 (.998)
WASL Reading Score	0 (1)	-.105 (.981)	<	.110 (1.008)	.069 (1.008)	>	-.068 (.986)
WASL Listening Score	0 (1)	.030 (.965)	>	-.032 (1.034)	-.014 (1.012)	<	.014 (.986)
WASL Writing Score	0 (1)	-.201 (1.004)	<	.212 (.949)	.125 (.984)	>	-.122 (1.00)
Student Passes the Complete WASL	.307 (.461)	.271 (.444)	<	.346 (.475)	.331 (.470)	>	.285 (.451)
Student is Male	.513 (.499)	1 (0)		0 (0)	.207 (.405)	<	.812 (.390)
Student is Male with Male Teacher	.102 (.303)						
Student is Male with Female Teacher	.411 (.491)						
Student is Female with Male Teacher	.094 (.292)						
Student is Female with Female Teacher	.392 (.488)						
Black	.047 (.212)	.047 (.212)	=	.047 (.213)	.049 (.217)	>	.045 (.209)
Asian	.068 (.251)	.069 (.253)	=	.067 (.250)	.066 (.249)	=	.069 (.254)
Hispanic	.113 (.316)	.110 (.313)	=	.116 (.320)	.113 (.317)	=	.111 (.314)
White	.725 (.446)	.727 (.445)	=	.722 (.447)	.724 (.446)	=	.727 (.445)
American Indian	.029 (.171)	.028 (.166)	=	.031 (.174)	.030 (.172)	=	.028 (.167)
Computer at Home	.648 (.477)	.646 (.478)	=	.650 (.476)	.652 (.476)	=	.645 (.478)
Computer is Used for School Work	.257 (.437)	.241 (.427)	<	.275 (.446)	.268 (.443)	>	.248 (.431)
Student Held Back a Grade in the Past	.063 (.244)	.071 (.258)	>	.055 (.228)	.058 (.235)	<	.068 (.253)
Student Reads Often for Fun	.495 (.499)	.448 (.497)	<	.544 (.498)	.525 (.499)	>	.465 (.498)
English Never Spoken at Home	.082 (.274)	.078 (.268)	<	.086 (.281)	.083 (.276)	=	.079 (.271)
N	49,415	25,373		24,042	24,436		24,979

< = > indicate statistical difference at the 5% level.

Table 2: Marginal Impact of Gender Characteristics on WASL Test Scores

	Variables	Math	Reading	Writing	Listening
A	Student is Male	.034*** (.008)	-.169*** (.008)	-.370*** (.000)	.083*** (.008)
	Adjusted R^2	.183	.175	.196	.089
	F-test†	91.39 (.000)	107.50	133.09 (.000)	44.34 (.000)
	F-test‡	18.58 (.000)	406.61 (.000)	2031.65 (.000)	94.29 (.000)
	N	49,415	49,415	49,415	49,415
B	Student is Male	.035*** (.008)	-.169*** (.008)	-.369*** (.008)	.084*** (.008)
	Teacher is Male	-.065*** (.019)	-.066*** (.016)	-.089*** (.018)	-.025* (.016)
	Adjusted R^2	.184	.175	.198	.091
	F-test†	89.18 (.000)	105.13 (.000)	130.83 (.000)	43.88 (.000)
	F-test‡	14.76 (.000)	212.35 (.000)	1025.78 (.000)	48.24 (.000)
	N	49,415	49,415	49,415	49,415
C	Student is Male	.051*** (.010)	-.157*** (.009)	-.366*** (.010)	.093*** (.011)
	Teacher is Male	-.066*** (.019)	-.066*** (.016)	-.089*** (.018)	-.025* (.015)
	Student and Teacher are Same Gender	.026*** (.010)	.019** (.009)	.004 (.010)	.017 (.011)
	Adjusted R^2	.183	.175	.198	.091
	F-test†	88.63 (.000)	104.54 (.000)	129.59 (.000)	43.00 (.000)
	F-test‡	11.94 (.000)	141.75 (.000)	683.98 (.000)	32.43 (.000)
	N	49,415	49,415	49,415	49,415
D	ITBS Score	.685*** (.005)	.615*** (.004)	.436*** (.005)	.345*** (.005)
	Student is Male	-.072*** (.008)	-.134*** (.008)	-.358*** (.009)	.114*** (.011)
	Teacher is Male	-.043** (.017)	-.060*** (.014)	-.073*** (.017)	-.010 (.013)
	Student and Teacher are Same Gender	.013* (.008)	.005 (.008)	-.0001 (.009)	-.0006 (.011)
	Adjusted R^2	.558	.461	.345	.186
	F-test†	385.62 (.000)	365.46 (.000)	226.06 (.000)	90.62 (.000)
	F-test‡	54.12 (.000)	125.69 (.000)	690.97 (.000)	57.07 (.000)
	N	38,537	39,006	39,110	39,124
E	ITBS Score	.706*** (.003)	.625*** (.004)	.437*** (.000)	.346*** (.005)
	Student is Male	-.076*** (.007)	-.135*** (.009)	-.363 (.009)	.114*** (.011)
	Teacher is Male	-.046*** (.009)	-.045*** (.011)	-.040*** (.012)	.0003 (.014)
	Student and Teacher are Same Gender	.013 (.008)	.006 (.008)	-.002 (.009)	.001 (.010)
	Adjusted R^2	.553	.455	.333	.184
	F-test†	722.78 (.000)	444.10 (.000)	249.31 (.000)	101.95 (.000)
	F-test‡	65.20 (.000)	128.21 (.000)	706.91 (.000)	54.61 (.000)
	N	38,515	38,398	39,088	39,102
	Building Fixed Effects?	Yes	Yes	Yes	Yes
	District Fixed Effects?	Yes	Yes	Yes	Yes

*** (**) [*] Indicate statistical significance at the 1%, (5%), [10%] level of significance.
Standard errors have been corrected for heteroskedasticity within classrooms.
F-test† tests significance of overall regression.
F-test‡ tests joint significance of the variables of interest in each regression (Student is Male, Teacher is Male, and Student and Teacher are Same Gender)
All regressions contain the 55 control measures described in the text.

Table 3: Marginal Impact of Selected Characteristics on Passing the WASL and Components

Variables	Marginal Impact Full Sample	Marginal Impact Schools with No Gender Choice
ITBS Total	.352*** (.005)	.375*** (.006)
Student is Male	-.086*** (.006)	-.113*** (.017)
Teacher is Male	-.027*** (.007)	-.069*** (.020)
Student and Teacher are Same Gender	-.004 (.006)	-.031 (.021)
Student was Held back at Least One Grade	-.061*** (.011)	-.060*** (.015)
English Never Spoken at Student's Home	-.132** (.059)	-.132* (.074)
Student is Asian	.080 (.050)	.166* (.094)
Student is Hispanic	-.013 (.009)	.016 (.079)
Student is Black	-.097** (.048)	-.016 (.087)
Teacher Holds a Bachelors Degree	-.005 (.104)	.051* (.028)
Teacher Holds a Masters Degree	.013 (.097)	.051* (.027)
Teacher's Years of Experience	.004*** (.001)	.001 (.001)
Teacher's Years of Experience Squared	-.0001*** (.00003)	-.00003 (.00004)
Teacher is in First Year of Teaching	-.024** (.012)	-.040** (.016)
Teacher is New to the Building	-.012 (.029)	-.024 (.035)
Teacher is New to the District	-.029 (.040)	.029 (.032)
N	38,230	20,075
Building Fixed Effects	Yes	Yes
District Fixed Effects	Yes	Yes
% of Passing Students Correctly Predicted	46.3%	48.8%
% of Failing Students Correctly Predicted	91.7%	90.6%

*** (**) [*] Indicate statistical significance at the 1%, (5%), [10%] level of significance.
All marginal impacts are evaluated at the sample means of the variables.
Standard errors have been corrected for heteroskedasticity within classrooms.