

Essays in Education and Politics

by

Katherine Duch

Public Policy Studies
Duke University

Date: _____

Approved:

Seth Sanders, Supervisor

Charles Clotfelter

Nicholas Carnes

Ronald Ehrenberg

Kenneth Dodge

Dissertation submitted in partial fulfillment of the requirements
for the degree of Doctor of Philosophy in Public Policy Studies
in the Graduate School of Duke University

2013

ABSTRACT

Essays in Education and Politics

by

Katherine Duch

Public Policy Studies
Duke University

Date: _____

Approved:

Seth Sanders, Supervisor

Charles Clotfelter

Nicholas Carnes

Ronald Ehrenberg

Kenneth Dodge

An abstract of a dissertation submitted in partial fulfillment
of the requirements for the degree of Doctor of Philosophy
in Public Policy Studies in the Graduate School of Duke University

2013

Copyright by
Katherine Duch
2013

Abstract

This dissertation explores three topics relating to education and politics. The first chapter examines gender gaps in test scores among top-scoring students. Recent research by Pope and Sydnor (2010) and Nosek et al. (2008) suggests that regions where individuals hold traditional gender stereotypes may have greater gender gaps. To explore this possibility, the current study examines whether five proxy variables for traditional gender stereotypes are correlated with greater gender gaps on math, science, and reading assessments. Using data from North Carolina, the results suggest that gender gaps on 5th and 8th grade state tests are strongly associated with the proportion of voters who supported a constitutional amendment to ban same-sex marriage. A one standard deviation increase in support for the amendment is associated with a 26 percent standard deviation increase in the 5th grade gender gap and a 31 percent standard deviation increase in the 8th grade gender gap. Other proxy variables for traditional gender stereotypes, such as the proportion of voters who supported the Republican Presidential candidate in 2012 and the proportion of religious adherents who are Evangelical, are not statistically significantly related to gender gaps.

The second chapter evaluates a one-to-one laptop program implemented by a North Carolina school district. The Digital Conversion Initiative provides all 4th through 12th grade students in the Mooresville Graded School District with a laptop for

use at school and at home. Using administrative data on students in North Carolina, this paper analyzes the effect of the program on 4th through 8th grade reading and math assessments. The results suggest that the effect of the laptop program varies substantially depending on the grade in which students first received a laptop. The program had an impressive effect on test scores for students who received a laptop in 4th or 5th grades. Compared to students who were unexposed to the program, students who received a laptop in 4th grade scored 0.12 standard deviations higher on reading tests and 0.42 standard deviations higher on math tests after three years of the program. Students who received a laptop in 5th grade scored 0.22 standard deviations higher on reading tests and 0.09 standard deviations higher on math tests after three years than their unexposed peers. Meanwhile, the program had smaller effects on reading test scores and significant negative effects on math test scores for students who received a laptop in 6th or 7th grade. Compared to their unexposed peers, students who received a laptop in 6th grade scored 0.12 standard deviations lower on math tests after three years of exposure, and students who received a laptop in 7th grade scored 0.22 standard deviations lower on math tests after two years of exposure. These mixed results should serve as a warning to school districts — and to middle schools in particular — that are planning to launch one-to-one programs.

The third chapter explores whether state legislative candidates differ in their support for abortion bans and abstinence-only sex education programs based on the

gender and age of their children. Previous research by Washington (2008) suggests that legislative candidates with daughters would be less likely to support abortion bans and abstinence-only sex education programs than candidates with only sons, but the present study finds that legislative candidates with daughters feel the same about abortion and abstinence-only sex education programs as their colleagues with only sons. This finding suggests that differences in legislators' views may not explain the differences in legislators' voting identified by Washington (2008). This paper also serves as a cautionary tale to researchers who may be quick to generalize the conclusions from prior work. Although the characteristics of legislators themselves produce robust differences in voting, it appears that the characteristics of legislators' children may only matter in some circumstances.

Dedication

To Mom and Dad, with endless love and deep gratitude.

Contents

Abstract	iv
List of Figures	xi
List of Tables	xiv
Acknowledgements	xvi
1. “Allergic to algebra”? Explaining gender disparities in standardized test scores among top-scoring students	1
1.1 Conceptual framework	5
1.1.1 Do gender gaps exist?	6
1.1.2 What causes gender gaps?	8
1.1.3 Is opposition to same-sex marriage an appropriate proxy variable for traditional gender stereotypes?	27
1.1.4 What are other proxy variables for traditional gender stereotypes?	33
1.1.4.1 Support for the Republican Party	33
1.1.4.2 Adherence to Evangelism	37
1.2 Data and methodology	40
1.2.1 Data	40
1.2.2 Methodology	47
1.3 Descriptive findings	50
1.3.1 Gender gaps across test score percentiles	50
1.3.2 Gender gaps in the top 5 percent across North Carolina.....	55
1.3.3 Proxy variables for traditional gender stereotypes across North Carolina.....	62

1.4 Regression results.....	71
1.5 Robustness checks.....	82
1.5.1 Alternate specifications using large counties only.....	82
1.5.2 Alternate specifications using different calculations for the gender gap.....	85
1.5.3 Alternate specifications using gender ratios on individual subject tests.....	89
1.6 Discussion.....	94
1.7 Conclusions.....	96
2. The best school district in the country? Examining the effect of the Mooresville Graded School District’s Digital Conversion Initiative on student test scores.....	100
2.1 Conceptual framework.....	102
2.2 The Digital Conversion Initiative.....	109
2.3 Data and methodology.....	116
2.3.1 Data.....	116
2.3.2 Methodology.....	121
2.4 Descriptive findings.....	123
2.5 Regression results.....	128
2.5.1 Reading test scores.....	131
2.5.2 Math test scores.....	132
2.6 Discussion.....	134
2.7 Limitations.....	143
2.8 Conclusions.....	148
3. Does having a daughter affect state legislators’ beliefs? Evidence from a survey of state legislative candidates.....	150

3.1 Conceptual framework	152
3.2 Data and methodology	168
3.2.1. Data	168
3.2.2 Methodology	175
3.3 Descriptive findings	181
3.3.1 Support for the statement that abortion should always be illegal	181
3.3.2 Support for the statement that school sex education programs should teach abstinence-only	186
3.4 Regression results.....	190
3.5 Discussion.....	196
3.6 Conclusions	201
Appendix A.....	203
Appendix B	205
Appendix C.....	210
Appendix D.....	212
Appendix E	213
Appendix F	214
Appendix G.....	215
References	217
Biography	231

List of Figures

Figure 1: Male-female gender ratio on North Carolina math assessment	51
Figure 2: Male-female gender ratio on North Carolina science assessment.....	51
Figure 3: Female-male gender ratio on North Carolina reading assessment.....	52
Figure 4: Counties by number of students	57
Figure 5: Counties by number of students and by regions	57
Figure 6: County groups	58
Figure 7: Distribution of gender gaps in the top 5 percent by grade and area	61
Figure 8: Gender gaps in the top 5 percent in 5 th grade by area.....	61
Figure 9: Gender gaps in the top 5 percent in 8 th grade by area.....	62
Figure 10: Distribution of support for Amendment One by area	63
Figure 11: Support for Amendment One by area.....	63
Figure 12: Distribution of support for 2012 Republican presidential candidate by area..	65
Figure 13: Support for 2012 Republican presidential candidate by area	65
Figure 14: Distribution of support for 2008 Republican presidential candidate by area..	66
Figure 15: Support for 2008 Republican presidential candidate by area	66
Figure 16: Distribution of Evangelical adherents as a proportion of all religious adherents by area	68
Figure 17: Evangelical adherents as a proportion of all religious adherents by area.....	68
Figure 18: Distribution of Evangelical adherents as a proportion of non-Black Protestant religious adherents by area.....	69
Figure 19: Evangelical adherents as a proportion of non-Black Protestant religious adherents by area	69

Figure 20: Scatterplots of the gender gap in the top 5 percent as a function of support for Amendment One.....	73
Figure 21: Scatterplots of the gender gap in the top 5 percent as a function of support for the 2012 Republican Presidential candidate.....	74
Figure 22: Scatterplots of the gender gap in the top 5 percent as a function of support for the 2008 Republican presidential candidate.....	75
Figure 23: Scatterplots of the gender gap in the top 5 percent as a function of Evangelical adherents.....	76
Figure 24: Scatterplots of the gender gap in the top 5 percent as a function of non-Black Protestant Evangelical adherents.....	77
Figure 25: Proportion of all candidates who agree that abortion should always be illegal	182
Figure 26: Proportion of candidates with one child who agree that abortion should always be illegal	182
Figure 27: Proportion of candidates with two children who agree that abortion should always be illegal	183
Figure 28: Proportion of candidates with three children who agree that abortion should always be illegal	183
Figure 29: Proportion of candidates with four children who agree that abortion should always be illegal	184
Figure 30: Proportion of all candidates who agree that school sex education programs should teach abstinence only	187
Figure 31: Proportion of candidates with one child who agree that school sex education programs should teach abstinence only	187
Figure 32: Proportion of candidates with two children who agree that school sex education programs should teach abstinence only	188
Figure 33: Proportion of candidates with three children who agree that school sex education programs should teach abstinence only	188

Figure 34: Proportion of candidates with four children who agree that school sex education programs should teach abstinence only 189

List of Tables

Table 1: Estimates of gender ratios in the top 5 percent from the U.S.....	7
Table 2: Estimates of male-female variance ratios in test scores by subject, selected grades.....	10
Table 3: Observations by grade, by test, and by gender.....	42
Table 4: Gender ratios on North Carolina End of Grade tests across test score distributions.....	53
Table 5: Descriptive statistics for gender ratios	59
Table 6: Correlations between proxy variables for traditional gender stereotypes.....	72
Table 7: Estimated relationship between the average of gender ratios in the top 5 percent and proxy variables for gender stereotypes.....	79
Table 8: Estimated relationship between the average of gender ratios in the top 5 percent and proxy variables for gender stereotypes for large counties only	83
Table 9: Correlations between measures of gender gaps.....	87
Table 10: Estimated relationship between measures of gender gaps in the top 5 percent and proxy variables for gender stereotypes.....	88
Table 11: Estimated relationship between the male-female ratio in the top 5 percent on the math test and proxy variables for gender stereotypes.....	91
Table 12: Estimated relationship between the male-female ratio in the top 5 percent on the science test and proxy variables for gender stereotypes.....	92
Table 13: Estimated relationship between the female-male ratio in the top 5 percent on the reading test and proxy variables for gender stereotypes	93
Table 14: Exposure to the Digital Conversion Initiative by graduating class and year ..	112
Table 15: Observations on reading tests by exposure to the Digital Conversion Initiative and by grade	118

Table 16: Observations on math tests by exposure to the Digital Conversion Initiative and by grade	119
Table 17: Descriptive statistics	124
Table 18: Standardized means on reading tests by exposure and by grade	126
Table 19: Standardized means on math tests by exposure and by grade	127
Table 20: Estimated effect of the Digital Conversion Initiative on reading test scores...	129
Table 21: Estimated effect of the Digital Conversion Initiative on math test scores	130
Table 22: Summary statistics	170
Table 23: Estimated effects of daughters on agreement that abortion should always be illegal.....	191
Table 24: Estimated effects of daughters on agreement that school sex education programs should teach abstinence only	193
Table 25: Estimated effects of daughters on agreement that abortion should always be illegal for elected legislators only	197
Table 26: Estimated effects of daughters on agreement that school sex education programs should teach abstinence only for elected legislators only	199
Table 27: Gender ratios in the top 5 percent and variance ratios across countries.....	203
Table 28: Comparison of gender ratios on NAEP tests and NC End of Grade tests	211
Table 29: Descriptive statistics by area.....	212
Table 30: Descriptive statistics	213
Table 31: Estimated relationship between the average of gender ratios in the top 5 percent and proxy variables without controls for median household income	214
Table 32: Estimated effects of daughters on agreement with statements on abortion and abstinence only sex education.....	215

Acknowledgements

I would like to thank the members of my committee for their guidance and support on this project. I am grateful to Charles Clotfelter for providing thoughtful council throughout my doctoral program, to Nicholas Carnes for generously including my survey questions on the 2012 National Candidate Study, and to Seth Sanders for chairing my dissertation committee. I would also like to thank Kenneth Dodge, Ronald Ehrenberg, and Jacob Vigdor for their helpful comments and suggestions.

I would like to extend a special thank you to Marie Hull, Kristoph Kleiner, and Brian Clark. Your friendship has been the cornerstone of my degree. Thank you for being with me from math camp to Sunday dinners. It is impossible to imagine the past four years without you.

1. “Allergic to algebra”? Explaining gender disparities in standardized test scores among top-scoring students

Implicit and explicit messages about gender norms abound. In the United States, three national apparel retailers advertised clothing with clear messages about gender stereotypes during 2011. Teen apparel retailer Forever 21 marketed a shirt to young women that read, “Allergic to Algebra.” Department store J.C. Penney advertised a shirt to pre-teen girls that proclaimed, “I’m too pretty to do homework so my brother has to do it for me.” And, ensuring that messages about gender stereotypes start early, children’s clothing merchant Gymboree designed onesies that boasted “Smart like Dad” for newborn boys and “Pretty like Mommy” for newborn girls. The messages are unambiguous: boys are expected to be smart but girls are not, especially when it comes to algebra.

Amidst these messages, women have made significant progress towards gender equality over the past fifty years. In 1960, women earned 35 percent of bachelor’s degrees, 32 percent of master’s degrees, and only 11 percent of doctoral degrees (U.S. Department of Education 2012, Table 310). In 2010, women earned 62 percent of bachelor’s degrees, 57 percent of master’s degrees, and 52 percent of doctoral degrees (U.S. Department of Education 2012, Table 310). Women have earned the majority of bachelor’s degrees since 1982, the majority of master’s degrees since 1987, and the majority of doctoral degrees since 2006 (U.S. Department of Education 2012, Table 310).

However, progress towards gender equality in traditionally male fields such as science, technology, engineering, and mathematics (collectively known as STEM fields) has been more muted. Women earned 50 percent of bachelor's degrees, 46 percent of master's degrees, and 40 percent of doctoral degrees in science and engineering fields in 2008 (National Science Foundation 2011, Table 3). Women are further underrepresented among science and engineering faculty at colleges and universities, comprising 28 percent of tenured or tenure-track faculty and 19 percent of full professors (National Science Foundation 2008, Table 5). Women fare better but remain underrepresented among faculty with recent science and engineering doctorates, representing 42 percent of tenured or tenure-track faculty and 28 percent of full professors (National Science Foundation 2008, Figure 1).¹ Women also fare better in psychology, social sciences, and life sciences departments, representing 32-46 percent of tenure or tenure-track faculty and 23-33 percent of full professors. In contrast, women comprise less than 20 percent of tenured or tenure-track faculty and less than 10 percent of full professors in engineering, mathematics, and physical science faculty (National Science Foundation 2008, Table 5).

Researchers continue to debate the causes of these remaining disparities. Former Harvard President Lawrence Summers made national news for suggesting that gender

¹ Examining faculty who recently earned their doctorate helps to account for the fact that degree statistics are "flow" variables whereas employment statistics are "stock" variables. Tenured or tenure-track faculty were considered "recent" doctorates if they earned their Ph.D. within seven years of the survey. Full professors were considered "recent" doctorates if they earned their Ph.D. between 11 and 15 years prior to the survey (National Science Foundation 2008, Figure 1).

differences in innate ability may contribute to the underrepresentation of women among science and engineering faculty. Recent research documenting changes in gender gaps over time (Monastersky 2005; Wai et al. 2010) and differences in gender gaps across countries (Feingold 1994; Guiso et al. 2008; Kane and Mertz 2010; Nosek et al. 2008) present an alternative hypothesis: that differences in socialization explain at least some of the gender gap at the top of the distribution. For example, one study reports that high school females with more feminine names are less likely to take advanced math and science courses than their biological sisters with less feminine names (Figlio 2008). Another study observes that countries where individuals have a greater implicit association between men and science also have greater gender gaps on international science and math assessments (Nosek et al. 2008).

Based on this emerging area of research, the current study examines the relationship between gender gaps and traditional gender stereotypes. Judd and Park (1993) summarize the consensus among psychologists that stereotypes are “generalizations about the shared attributes of a group of people” (page 109). The current paper extends this description to define traditional gender stereotypes as “generalizations about the shared attributes of a group of people on the basis of traditional sex roles.” Statements such as “men are better scientists than women” and “women care more about raising a family than pursuing a career” reflect traditional gender stereotypes. For brevity, this paper uses the term “gender stereotypes” to refer

to “traditional gender stereotypes” or stereotypes on the basis of traditional sex roles.² This paper also uses the term “greater gender stereotypes” to refer to stronger traditional gender associations, such as stronger associations between men and science or women and families.

The current study hypothesizes that greater gender gaps at the top of the distribution are associated with greater traditional gender stereotypes. This study further hypothesizes that the proportion of voters who oppose same-sex marriage can be used as a proxy variable to measure the strength of traditional gender stereotypes in a region. This paper tests these hypotheses by evaluating whether gender gaps among top-scoring students on standardized assessments are associated with the proportion of voters who supported a constitutional amendment to ban same-sex marriage in North Carolina. The results show that counties with greater gender gaps among students scoring in the top 5 percent also had higher support for the amendment; a one standard deviation increase in the proportion of voters who supported the amendment is associated with a 26 percent standard deviation increase in the 5th grade gender gap and a 31 percent standard deviation increase in the 8th grade gender gap.³ Other possible proxy variables for traditional gender stereotypes — such as the proportion of voters

² This study does not examine other gender stereotypes, such as stereotypes that women are less competitive, more empathetic, or worse drivers than men.

³ Gender gaps are defined as the arithmetic average of the male-female ratio in the top 5 percent on the math assessment, the male-female ratio in the top 5 percent on the science assessment, and the female-male ratio in the top 5 percent on the reading assessment.

who supported the Republican presidential candidate in 2012 and the proportion of religious adherents who are Evangelical — are not statistically significantly related to gender gaps in test scores. In short, the results are consistent with the conclusion that greater gender gaps exist in areas with greater traditional gender stereotypes.

These findings have two important implications. First, the results cast doubt on the hypothesis that differences in ability fully explain gender gaps. Instead, they support the alternative hypothesis that differences in socialization contribute to gender gaps. Second, the results have practical implications for educators aiming to reduce gender gaps. By better understanding gender gaps, we can design more effective programs and policies to mitigate them.

1.1 Conceptual framework

This paper presents two arguments: (1) that greater gender gaps among top-scoring students are associated with greater traditional gender stereotypes, and (2) that the proportion of voters who oppose same-sex marriage can be used as a proxy variable to measure traditional gender stereotypes. Sections A and B focus on the first claim. Section A presents estimates of gender gaps among top-scoring students, and Section B details the evidence on the possible causes for these gaps. Section C focuses on the second claim: that the proportion of voters who oppose same-sex marriage can be used as a proxy variable to measure traditional gender stereotypes. Section D describes other potential proxy variables to capture the strength of traditional gender stereotypes and

the limitations of these proxy variables as compared to the proportion of voters who oppose same-sex marriage.

1.1.1 Do gender gaps exist?

There is no shortage of research on the existence and purported causes of gender gaps in various educational and career outcomes (see Ceci, Williams, and Barnett 2009 and Spelke 2005 for recent reviews). Whereas studies conclude that males and females are equally represented at the mean on intelligence and achievement tests (Ceci, Williams, and Barnett 2009; Hyde 2005; Hyde and Linn 2006; Hedges and Nowell 1995), studies also conclude that males are overrepresented among top-scoring students on a variety of tests (Arden and Plomin 2006; Deary et al. 2003; Hedges and Nowell 1995; Hyde et al. 2008; Leahey and Guo 2001; Lohman and Lakin 2009; Pope and Sydnor 2010; Strand et al. 2006). This pattern has been documented on general intelligence tests (Arden and Plomin 2006; Deary et al. 2003) and on math and science achievement tests (Hedges and Nowell 1995; Hyde et al. 2008; Leahey and Guo 2001).

Male-female ratios at the top of the distribution on math and science tests tend to range between one and two, depending on the test, grade, and percentile. Table 1 presents recent estimates of gender ratios in the top 5 percent from the United States. For 11th grade white students, Hyde et al. (2008) report male-female ratios of 1.45 in the top 5 percent and 2.06 in the top 1 percent on a Minnesota state math assessment. Machin and Pekkarinen (2008) report a male-female ratio of 1.72 in the top 5 percent on

Table 1: Estimates of gender ratios in the top 5 percent from the U.S.

	M-F ratio in top 5% in math	M-F ratio in top 5% in science	F-M ratio in top 5% in reading
Hyde et al. (2008)	1.45		
Machin and Pekkarinen (2008)	1.72		1.67
Lohman and Lakin (2009)	1.55 *		1.09 *
Pope and Sydnor (2010)	1.40	1.87	2.31
Duch	1.24	1.85	1.15

Notes: Math gender ratios represent the male-female ratio in the top 5 percent on math assessments. Science gender ratios represent the male-female ratio in the top 5 percent on the science assessment. Reading gender ratios represent the female-male ratio in the top 5 percent on the reading assessment. * indicates that values are calculated from the top stanine, roughly corresponding to the top 8 percent (Lohman and Lakin 2009, Table 4). Ratios in Hyde et al. (2008) are computed using 11th grade test scores from Minnesota state assessments. Ratios in Machin and Pekkarinen (2008) are computed using test scores from 15 year olds from the Program for International Student Assessment. Ratios in Lohman and Lakin (2009) are computed using 3rd-11th grade test scores from the Cognitive Abilities Test. Ratios in Pope and Sydnor (2008) are computed using 8th grade test scores from the National Assessment of Educational Progress (NAEP). Ratios in Duch are computed using 8th grade test scores from North Carolina End of Grade assessments.

Sources: Hyde et al. (2008); Machin and Pekkarinen (2008); Lohman and Lakin (2009); Pope and Sydnor (2010); North Carolina Education Research Data Center, End of Grade files.

the math portion of the Program for International Student Assessment (PISA) and a female-male ratio of 1.67 in the top 5 percent on the reading portion. Pope and Sydnor (2010) report male-female ratios of 1.40 and 1.87 for 8th grade white students scoring in the top 5 percent on the math and science sections, respectively, of the National Assessment of Educational Progress (NAEP).

Some studies document male overrepresentation at the top of the distribution as early as kindergarten (Penner and Paret 2008) and 3rd grade (Leahey and Guo 2001; Lohman and Lakin 2009; Swiatek et al. 2000), while Arden and Plomin (2006) reports

that male overrepresentation does not emerge until 5th grade. However, studies consistently find that gender gaps exist by middle school and persist into high school (Benbow 1992; Deary et al. 2003; Hyde et al. 2008; Leahey and Guo 2001; Lohman and Lakin 2009; Pope and Sydnor 2010; Reis and Park 2001; Strand et al. 2006).

While males tend to outnumber females at the top of the distribution on math and science tests, females tend to outnumber males at the top on reading tests (Lohman and Lakin 2009; Pope and Sydnor 2010; Strand et al. 2006). Pope and Sydnor (2010) report a female-male of 1.67 among 8th grade students scoring in the top 5 percent on the NAEP reading assessment, indicating that 70 percent of top-scoring students are female. Machin and Pekkarinen (2008) document a female-male ratio of 1.67 among 15 year old students scoring in the top 5 percent of the PISA reading test, implying that 63 percent of top-scoring students are female.

1.1.2 What causes gender gaps?

The purported causes for these gender gaps at the top of the distribution are fiercely debated. Studies tend to argue that gender gaps at the top of the distribution exist for one of two reasons: (1) innate differences in aptitude or (2) differences in socialization. The first hypothesis received national attention in 2005 when the President of Harvard University, Lawrence Summers, argued that “different availability of aptitude at the high end” may contribute to the underrepresentation of women in tenured positions at colleges and universities (Summers 2005). Specifically, Summers

suggested that males have greater variance in intrinsic math and science aptitude than females. This difference produces more males at the very top (and at the very bottom) of the distribution. First proposed in 1894, this greater male variability hypothesis encompasses both a testable conjecture (that males exhibit greater variance in scores on intelligence and achievement tests than females) and an underlying biological theory to explain the result (that males have higher variance in intrinsic aptitude than females) (Feingold 1992).

The evidence tends to support the conjecture but to refute the biological theory. Table 2 illustrates that males exhibit greater variance than females across grades on a variety of intelligence and achievement tests (Arden and Plomin 2006; Deary et al. 2003; Feingold 1992; Hedges and Nowell 1995; Hyde et al. 2008; Lohman and Lakin 2009; Machin and Pekkarinen 2008; Strand et al. 2006).⁴ Males typically display 10 to 30 percent greater variance on math and other quantitative tests than females (Arden and Plomin 2006; Deary et al. 2003; Hedges and Nowell 1995; Hyde et al. 2008; Strand et al. 2006). Some studies document an even larger variance: Lohman and Lakin (2009) calculate 56 percent greater male variance on the quantitative section of the Cognitive Abilities Test among 11th grade students who completed the test in 2000. The gender difference on reading and verbal tests is less clear. Some studies suggest 5 to 15 percent

⁴ Appendix A also provides male-female variance ratios across countries on the reading and math portions of the Program for International Student Assessment (PISA) and on the math portion of the Third International Mathematics and Sciences Study (TIMSS).

Table 2: Estimates of male-female variance ratios in test scores by subject, selected grades

	Age or grade	Math or quantitative	Reading or verbal	General intelligence
<i>2nd grade</i>				
Arden and Plomin (2006)	7 years			1.13
Hyde et al. (2008)	2nd grade	1.11		
<i>3rd grade</i>				
Hyde et al. (2008)	3rd grade	1.11		
Lohman and Lakin (2010)	3rd grade	1.25	1.05	
<i>4th grade</i>				
Arden and Plomin (2006)	9 years			1.06
Hyde et al. (2008)	4th grade	1.11		
Lohman and Lakin (2010)	4th grade	1.25	1.14	
<i>5th grade</i>				
Arden and Plomin (2006)	10 years			1.20
Hyde et al. (2008)	5th grade	1.14		
Lohman and Lakin (2009)	5th grade	1.27	1.09	
<i>6th grade</i>				
Deary et al. (2003)	11 years			1.12
Hyde et al. (2008)	6th grade	1.14		
Lohman and Lakin (2009)	6th grade	1.31	1.15	
Strand et al. (2006)	6th grade	1.18	1.09	
<i>7th-8th grade</i>				
Hyde et al. (2008)	7th grade	1.16		
Hyde et al. (2008)	8th grade	1.21		
Lohman and Lakin (2009)	7th and 8th grades	1.39	1.15	
<i>9th-11th grade</i>				
Hyde et al. (2008)	9th grade	1.14		
Hyde et al. (2008)	10th grade	1.18		
Hyde et al. (2008)	11th grade	1.17		
Lohman and Lakin (2010)	9th and 10th grades	1.34	1.08	
Lohman and Lakin (2010)	11th grade	1.56	1.08	

Notes: Variance ratios are calculated by dividing the variance of test scores among males by the variance of test scores among females. Variance ratios greater than one represent greater male variance. Ratios in Arden and Plomin (2006) are computed using scores from several intelligence tests administered in England and Wales between 2004 and 2006. Ratios in Deary et al. (2003) are computed using scores from the Scottish Mental Survey administered in 1932. Ratios in Hyde et al.

(2008) are computed using scores from U.S. state assessments administered in 10 states. Ratios in Lohman and Lakin (2010) are computed using scores from the Cognitive Abilities Test administered to U.S. students in 2000. Ratios in Strand et al. (2006) are computed using scores from the Cognitive Abilities Test administered to U.K. students between 2001 and 2003.

Sources: Arden and Plomin (2006); Deary et al. (2003); Hyde et al. (2008); Lohman and Lakin (2010); Strand et al. (2006).

greater male variance (Hedges and Nowell 1995; Lohman and Lakin 2009; Strand et al. 2006), while other studies report equal variances (Feingold 1992; Maccoby and Jacklin 1974).

However, the finding that males exhibit greater variance on aptitude tests does not imply that males have higher variance in intrinsic aptitude. A growing body of research accepts the results cited above but refutes the biological theory that underlies the greater male variability hypothesis. One criticism of the greater male variability hypothesis is that, although the studies cited above provide evidence of greater male variability on aptitude tests, they do not necessarily provide evidence of greater male variability in intrinsic ability. For example, Spelke (2005) cautions against using the results of aptitude tests to infer innate differences in ability. As evidence, she cites longitudinal research conducted using the Study of Mathematically Precocious Youth (SMPY). During the 1970s, Benbow and Stanley (1980) conducted a series of talent searches for high-ability boys and girls. Students were eligible to participate in the study if they were enrolled in 7th grade (or if they were enrolled in a later grade but were the same age as 7th grade students) and if they scored in the top 3 percent on a

standardized math achievement test. Benbow and Stanley (1980) document that more than eight times as many boys as girls scored above 600 on the math portion of the Scholastic Aptitude Test (SAT-M) among 7th grade students tested in 1976, 1978, and 1979 (Table 1, page 1263). In research conducted on these students twenty years later, Benbow et al. (2000) find that men and women in the study from these cohorts majored in mathematics and pursued master's degrees in mathematics at nearly the same rate (Table1, page 475). Spelke (2005) notes: "If one gauges students' talent at mathematics by their successful mastery of the demanding material required of college mathematics majors, one will conclude that men and women have equal aptitude for mathematics" (page 956).

Another criticism of the greater male variability hypothesis is that, if males are overrepresented at the top of the distribution due to differences in intrinsic aptitude, then gender gaps should remain relatively constant over time and across countries. In contrast, gender gaps among top-scoring students have narrowed substantially in recent years (Monastersky 2005; Wai et al. 2010). In one of Benbow and Stanley's talent searches from the early 1980s, the authors report a male-female ratio of 13.0 among students who scored above 700 on the SAT-M before the age of 13 (Benbow and Stanley 1983). By 2005, Stanley estimated that this ratio had declined dramatically to roughly 2.8 (Monastersky 2005). Wai et al. (2010) document similar declines from the Duke University Talent Identification Program 7th Grade Talent Search. Students were eligible

for the study if they were enrolled in 7th grade and if they had scored in the top 5 percent on a standardized test. The authors report a steep decline in the male-female ratio among students scoring above 700 on the SAT-M, from 13.5 in the early 1980s to 3.8 in the late 2000s.

Recent research also documents that gender ratios and variance ratios differ across countries (Feingold 1994; Guiso et al. 2008; Kane and Mertz 2012; Machin and Pekkarinen 2008; Nosek et al. 2008) and across states (Pope and Sydnor 2010). Appendix A documents differences across countries reported by Machin and Pekkarinen (2008) using test scores from the Program for International Student Assessment (PISA) and by Kane and Mertz (2012) using test scores from the Third International Mathematics and Sciences Study (TIMSS). Machin and Pekkarinen (2008) examine scores from the PISA assessments administered to 15 year old students in 2003. Appendix A illustrates that male-female ratios in the top 5 percent of the math assessment range from 0.91 in Indonesia to 2.55 in Korea (Machin and Pekkarinen (2008), Table S1 of the supporting online material; see also Guiso et al. (2008), Figure S2A in the supporting online material).⁵ More girls than boys score in the top 5 percent of the math test in both Indonesia and Thailand. Pope and Sydnor (2010) examine gender ratios across states from the NAEP assessment. They report a strikingly similar range in male-female ratios

⁵ Machin and Pekkarinen (2008) calculate a larger male-female ratio in Lichtenstein, but this estimate is likely attributable to a small sample size. They observe only 639 observations for this country (the smallest sample size in their study), and the standard error for the male-female ratio is larger than the point estimate.

across states on the NAEP assessment as Machin and Pekkarinen (2008) report across countries on the PISA assessment. Male-female ratios in the top 5 percent of the math test range from 0.81 in Hawaii to 2.07 in Kentucky (page 100). Pope and Sydnor (2008) also report wide ranges on the science and reading tests; male-female ratios range from 1.30 to 3.47 on the science test, and female-male ratios range from 1.75 to 4.47 on the reading test. These studies illustrate the wide dispersion in gender ratios among top-scoring students, both internationally and domestically.

By illustrating that gender differences vary over time and differ across cultures, these papers support the hypothesis that differences in socialization contribute to gender gaps. But if differences in socialization contribute to gender gaps, then what are the characteristics of environments with smaller gender gaps? To answer this question, scholars have examined “macro-level” and “micro-level” social factors. Research on macro-level social considerations examines factors that differ across large geographic regions, such as how differences in gender equality correspond to differences in gender gaps across countries or states. Studies of micro-level social considerations examine how factors from the immediate social environment — such as teachers or testing environments — are related to gender gaps within the same geographic region.⁶

⁶ Of course, individuals in these studies of micro-level social factors are also subject to macro-level social considerations. Researchers often assume that, by selecting participants from the same macro-level environment, the subjects in their sample will be equally affected by macro-level factors. The effect of macro-level social considerations would then “cancel out,” allowing researchers to identify micro-level

To determine the effect of macro-level social factors, studies have explored differences in gender gaps across countries (Baker and Jones 1993; Charles and Bradley 2006; Guiso et al. 2008; Hyde and Mertz 2009; Kane and Mertz 2012; Nosek et al. 2008; Penner 2008) and across states (Pope and Sydnor 2010). The majority of these studies conclude that gender gaps among top students are associated with measures of gender equality (such as women's labor force participation) or measures of social attitudes about women (such as the proportion of individuals who agree that it is necessary for a woman "to have children in order to be fulfilled").⁷ As Penner (2008) explains, this cross-country variation is useful for identifying "how macrolevel societal factors influence gender differences" (page S140).

Both Guiso et al. (2008) and Hyde and Mertz (2009) identify statistically significant correlations between gender gaps at the top 5 percent of the PISA and the World Economic Forum's Gender Gap Index (GGI). The GGI "synthesizes the position of women in any given country by taking into account economic opportunities, economic participation, educational attainment, political achievements, and health and well-being," with larger values indicating greater equality (Guiso et al. 2008 supporting

social concerns. If this assumption does not hold, then the micro-level effects identified by researchers may reflect both micro- and macro-level social considerations.

⁷ The only exception is Charles and Bradley (2006). This paper finds that the proportion of female graduates from computer science programs is unrelated to economic development, women's participation in the labor market, women's educational attainment, or women's representation in high-status professional occupations (page 194).

online material, page 3). Guiso et al. (2008) find that a one standard deviation increase in the GGI is associated with a 0.09 increase in the female-male ratio at the top 5 percent of the math test and a 0.25 increase in the female-male ratio at the top 5 percent of the reading test (author's calculation from Tables S2 and S3 in the supporting online material). This study also examines other measures of gender equality, including women's labor force participation and women's political participation. The authors find a statistically significant relationship between each measure of gender equality and gender ratios at the 95th and 99th percentiles (Table S3 in the supporting online material). A one standard deviation increase in women's labor force participation or women's political participation is associated with a 0.08-0.09 increase in the female-male ratio at the top 5 percent of the math test and a 0.18-0.22 increase in the female-male ratio at the top 5 percent of the reading test (author's calculation from Tables S2 and S3 in the supporting online material). In short, females performed better relative to males on both math and reading tests in countries with greater gender equality.

Penner (2008) examines whether measures of gender equality or social attitudes are correlated with gender gaps on the Third International Mathematics and Sciences Study (TIMSS). He examines measures of gender equality in broad domains such as educational attainment and labor force equality.⁸ He also captures differences in social

⁸ He aggregates 3-5 variables to calculate each of these measures (Penner 2008a, page 159). Educational attainment includes the male-female ratio in secondary enrollment, the male-female ratio in years of education, and public educational expenditure. Labor force equality includes the percent of women in the

attitudes by examining measures of women's status (such as the female-male ratio in mean age at marriage) and measures of "domesticity" (such as the proportion of individuals who agree that it is necessary for a woman "to have children in order to be fulfilled").⁹

Penner finds that females are significantly more likely to score in the top 5 percent of the math test in countries with greater labor force equality (Table 5, page S157). A one standard deviation change in labor force equality is associated with nearly a 40 percent greater probability that a female will score in the top 5 percent (author's calculation). Penner (2008a) further examines whether the proportion of women in various occupations is related to the likelihood that females score in the top 10 percent of the TIMSS math assessment (Table A.5, page 158). He finds that females are less likely to receive a top score in countries with greater occupational segregation. He also finds that females are more likely to score highly in countries where females represent a larger

labor force, the percent of the labor force that is female, occupational segregation (measured as the percentage of women who would have to change occupations to achieve equal representation across occupations), the proportion of managers who are female, and the proportion of individuals who agree that "when jobs are scarce, men should have more right to a job than women". He compiles these variables from various sources including the World Bank, the United Nations, the World Values Survey, the International Labour Organization, and the International Social Survey Programme (Penner 2008, page S138).

⁹ Status includes the percent of parliamentary seats filled by women, women's mean age at marriage, and the female-male ratio in mean age at marriage. Domesticity includes the fertility rate, the number of hours that women report spending on housework, and the proportion of individuals who agree that it is necessary for a woman "to have children in order to be fulfilled."

portion of managers, researchers, and teachers, but that they are less likely to score highly in countries with a higher proportion of women in government.

Looking beyond labor market outcomes, Penner (2008) also finds that females are more likely to score in the top 5 percent in countries with greater gender equality in educational attainment, in countries where women have higher status, and in countries where individuals have a weaker association between women and home production. However, these differences are all insignificant, suggesting that gender gaps are more attributable to differences in labor market opportunities than to differences in educational opportunities or to cultural attitudes.

Rather than examine broad measures of gender equality or social attitudes, Nosek et al. (2008) examine whether the strength of the association between males and science across countries is related to gender gaps on the TIMSS science assessment. To measure the strength of the association between males and science, they use results from the Implicit Association Test (IAT), an online test where participants categorize words as quickly as possible. In one part of the test, respondents are asked to categorize words associated with males (such as *boy* and *grandpa*) with words associated with science (such as *chemistry* and *geology*) and to categorize words associated with females (such as *girl* and *grandma*) with words associated with liberal arts (such as *history* and *music*). In another part of the test, respondents are asked to categorize words according to the reverse associations, grouping males with liberal arts and females with science. By

calculating the difference in the time required to complete the task under these two conditions, the researchers measure the strength of participants' reflexive association between males and science. They find that the average score on the IAT by country is statistically significantly associated with the gender gap on the TIMSS science and math assessments. A one standard deviation increase in the average IAT score is associated with a 0.56 standard deviation increase in the science gender gap (page 10594) and a 0.63 standard deviation increase in the math gender gap (page 10595).

While these studies all examine macro-level social factors across countries, Pope and Sydnor (2010) examine macro-level social factors across U.S. states. Using NAEP test scores, the authors create a "stereotype adherence index" in each state by averaging the male-female ratio in the top 5 percent on the math test, the male-female ratio in the top 5 percent on the science test, and the female-male ratio in the top 5 percent on the reading test. They find that this stereotype adherence index is correlated with a measure of cultural attitudes among adults from the General Social Survey and a measure of gender stereotypes among children from the NAEP assessment. A one standard deviation increase in the proportion of adults who agree that it is "much better for everyone involved if the man is the achiever outside the home and the woman takes care of the home and family" is associated with a 0.65 standard deviation increase in the stereotype adherence index (author's calculation from pages 104-105). They also find a significant relationship between the stereotype adherence index and the proportion of

children who do not disagree that “math is for boys.” A one percentage point increase in the proportion of boys who do not disagree with this statement is associated with a 0.16 standard deviation increase in the stereotype adherence index (author’s calculation from page 105).

The studies cited previously provide evidence that gender gaps are associated with macro-level social factors, such as gender equality, labor market equality, and cultural attitudes. Studies have also explored the relationship between gender differences and micro-level social factors, such as parents, teachers, testing environments, and the tests themselves (Dar-Nimrod and Heine 2006; Davies et al. 2002; Dee 2005; Figlio 2008; Inzlicht and Ben-Zeev 2000; Murphy, Steele, and Gross 2007; Nosek et al. 2002; Spencer, Steele, and Quinn 1999; Steele and Aronson 1995; Smyth et al. 2009; Walton and Cohen 2003).

This branch of research has expanded considerably since Steele and Aronson (1995) first identified the phenomenon known as stereotype threat. Steele and Aronson (1995) define stereotype threat as a “social-psychological predicament” that occurs when individuals are reminded about negative stereotypes about their race, gender, behavior, beliefs, or other personal characteristics and subsequently conform to that stereotype (page 797). The authors first identified this threat among black college students who performed worse relative to white students on a verbal test when they were informed that the test reflected intellectual ability, but who performed just as well as white

students when they were informed that the test was simply a task unrelated to intellectual ability (page 799). Subsequent tests illustrated that simply asking students to report their race in the demographic questions prior to the test increased racial gaps on the assessment (page 806).

Researchers have also examined stereotype threat among women. Spencer, Steele, and Quinn (1999) find that women and men performed the same on a difficult math test when students were informed that the test had not revealed gender differences previously. However, when students were informed that the test *had* revealed gender differences previously, women performed significantly worse than men and significantly worse than women who were administered the same test but were told that the test had not revealed gender differences. Further research determined that this marked decline in females' test scores was more attributable increased anxiety (measured by responses to statements such as "I feel nervous") than to increased concern for others' perceptions ("People will look down on me if I do not do well on this test") or increased self-doubt ("I am uncertain that I have the mathematical knowledge to do well on this test") (pages 18-19).

Dar-Nimrod and Heine (2006) find evidence of a more nuanced version of stereotype threat. The authors recruited female subjects to take two math tests and one reading test modeled after the Graduate Record Exam (GRE). After completing the first math test, subjects read a passage as part of the reading comprehension test. The

researchers distributed four different passages: (1) a passage arguing that men and women perform equally well on math tests, (2) a passage that discussed gender (“the role of the female body in the arts”) without mentioning math performance, (3) a passage arguing that men perform better than women on math tests due differences in genetics, and (4) a passage arguing that men perform better than women on math tests due to differences in experiences (specifically that “teachers biased their expectations during early school formative years”) (page 3 of supporting online material). The subjects answered questions about the passage and then completed the second math test.

Consistent with prior research on stereotype threat, Dar-Nimrod and Heine (2006) find that women who read passage (2) performed significantly worse on the second math test than women who read passage (1), conditional on their scores from the first math test. Even though passage (2) did not mention stereotypes about women’s math abilities, it emphasized participants’ gender, reminded them of negative stereotypes associated with their gender, and lowered their math test scores. Additionally, the researchers find that women who read passage (3) performed significantly worse on the second math test than women who read passage (4), conditional on their scores from the first math test. The authors note that women appear to instinctively associate negative stereotypes about gender with genetic factors, given the similarity in performance between women who read passage (2) and women who

read passage (3). However, stereotype threat is alleviated by associating negative stereotypes about gender with experiential factors rather than with genetic factors. Their paper suggests that remarks that attribute gender gaps to differences in ability — such as those by former Harvard president Lawrence Summers — may activate stereotype threat. Dar-Nimrod and Heine (2006) explain: “What President Summers perhaps intended to be a provocative call for more empirical research on biological bases of achievement may inadvertently exacerbate the gender gap in science through stereotype threat” (page 435).

Inzlicht and Ben-Zeev (2000) and Murphy, Steele, and Gross (2007) illustrate that stereotype threat also emerges from unbalanced gender ratios. Inzlicht and Ben-Zeev (2000) document that, among 3-person groups, females score highest on a difficult math test when randomly assigned to a room with two other females and lowest when randomly assigned to a room with two males. Murphy, Steele, and Gross (2007) find that female STEM majors who watched a video from a conference with an unbalanced gender ratio were less interested in participating in the conference than female majors who watched a video from a conference with a balanced gender ratio.

These studies on stereotype threat illustrate that micro-level situational factors, such as changes to the tests or testing conditions, can amplify or mitigate gender differences in performance. But do other situational factors, unrelated to the tests or the testing conditions, affect gender gaps? To explore this question, Davies et al. (2002)

conducted an experiment where they exposed college men and women to six (real) television commercials. While four of the commercials did not depict any humans, two of the commercials either reinforced gender stereotypes or contradicted gender stereotypes.¹⁰ After watching the commercials, participants completed a lexical-decision task where they were instructed “to identify as quickly as possible whether letter strings flashed on a computer screen are words or nonwords” and completed a math test composed of 12 questions from the GRE math subject assessment. Both men and women who watched the commercials that reinforced gender stereotypes were faster to identify words related to female stereotypes than participants who watched the commercials contradicting gender stereotypes (page 1620). Women who watched the commercials that reinforced gender stereotypes also scored lower on the math assessment than women who watched the commercials that countered gender stereotypes. The study further revealed that faster lexical-decision task completion largely explained lower math scores, hinting that the activation of gender stereotypes caused the decline in math performance among women.

Researchers have also examined whether micro-level factors such as teachers affect gender gaps. Dee (2005) examines how teachers’ perceptions of students’

¹⁰ One commercial that reinforced gender stereotypes “portrayed a young woman who was so excited about being a consumer of a new acne product that she bounced on her bed with joy, whereas the other portrayed a woman ‘drooling’ with anticipation to try a new brownie mix” (page 1619). The first commercial that contradicted gender stereotypes “portrayed an attractive woman impressing a man with her knowledge of automotive engineering and the second revealed a woman speaking intelligently about health care concerns” (page 1619).

behavior differs depending on whether the teacher shares the same gender or racial identity as the student. Using data from the National Education Longitudinal Study of 1988 (NELS: 88), Dee finds that teachers who did not share a student's gender were 37 percent more likely to rate the student as "frequently disruptive," 20 percent more likely to rate the student as "consistently inattentive," and 15 percent more likely to report that the student "rarely completed homework" than teachers who were the same gender (Table 2, page 161). Dee also finds that the discrepancies in rating students as "frequently disruptive" or "consistently inattentive" were greater when teachers were male and students were female than when teachers were female and students were male. Whereas male and female teachers were equally likely to rate male students as "consistently inattentive," male teachers were 33 percent more likely to report that female students were inattentive than female teachers. The reverse was true for the discrepancies in reporting that students "rarely completed homework": male and female teachers were equally likely to report that female students rarely completed their homework, whereas female teachers were 26 percent more likely to report that male students rarely completed their homework than male teachers.

Although Dee (2005) focuses on differences in perceived student behavior, the results suggest that the interaction between teacher gender and student gender may affect actual student performance as well, especially if teachers devote less time to students whom they consider "consistently inattentive" or "frequently disruptive."

Research on college freshmen has supported the hypothesis that the interaction between professor gender and student gender has a significant effect on students' performance. Carrell, Page, and West (2010) report that women who are randomly assigned to female professors score 0.10 standard deviations higher in first-year math and science courses than women who are randomly assigned to male professors (page 1122). They also find that this effect is even larger for women with the highest SAT-M scores and that, "at the top of the distribution, having a female professor completely closes the gender gap" (page 1123).

A final study of note is Figlio (2008). Figlio examines the "femininity" of names by identifying 1,700 linguistic attributes of names and uses these attributes to predict the likelihood that a child with that name is female.¹¹ He finds that high school females with more feminine names are less likely to take advanced math and science courses than their biological sisters with less feminine names. A high-achieving girl whose name is one standard deviation greater than average in femininity is 2.6 percentage points less likely to take calculus and 3.6 percentage points less likely to take physics relative to her sister whose name is average in femininity (Table 8, page 27). Using these estimates, Figlio predicts that a high-achieving girl named Meagan would be more than twice as

¹¹ Among popular names for girls in the early 2000s, his model identifies Kayla, Isabella, and Anna as the three most feminine names and Alexis, Madison, and Taylor as the three least feminine names (Table 2, page 20). He notes that "almost all girls' names — regardless of name femininity — tend to be overwhelmingly given to girls" and that, even among the names identified as the least feminine, "nine out of ten recipients of those names are themselves girls" (page 6).

likely as her equally high-achieving sister named Haylie to take calculus by the end of high school (Table 9, page 28). Although he cannot identify a mechanism to explain his findings, Figlio hypothesizes that “sisters with different types of names may be treated differently by their parents or their teachers, they may independently develop a self-concept based on their name, or they may select different peers (or be treated differently by their friends) because of their names” (page 18). The findings from Dee (2005) and from Carrell, Page, and West (2010) suggest that the interaction between student gender and teacher gender may contribute to Figlio’s results. Just as teachers perceive the behavior of students differently based on their own gender, perhaps math and science teachers perceive female students differently based on the femininity of their names.

The evidence cited above casts doubt on the hypothesis that differences in ability fully explain gender gaps and supports the alternative hypothesis that differences in socialization at least contribute to the gaps. Recent research also illustrates that both macro-level and micro-level social factors may intensify or attenuate gender gaps. The present study focuses on one macro-level social factor — the strength of traditional gender stereotypes in a region — and examines its relationship to gender gaps among top-scoring students.

1.1.3 Is opposition to same-sex marriage an appropriate proxy variable for traditional gender stereotypes?

The studies reviewed previously reveal explicit and implicit associations between men and science, men and math, women and reading, men and careers, and

women and families. In addition to these gender stereotypes, I stipulate that Americans may hold another gender stereotype: that men marry women and that women marry men.

For many years, this stereotype was reinforced by the fact that no same-sex couple requested a marriage license. The issue of same-sex marriage in the United States first arose in 1970, when Richard Baker and James McConnell were denied a marriage license at the Hennepin County District Court in Minnesota.¹² The men took their case to the courts, invoking the U.S. Supreme Court's *Loving v. Virginia* decision that struck down state statutes prohibiting interracial marriages. However, the Minnesota Supreme Court determined that the men were not entitled to a marriage license because, "in commonsense and in a constitutional sense, there is a clear distinction between a marital restriction *based merely upon race* and one *based upon the fundamental difference in sex*" (291 Minn. 310; emphasis added).

Even after same-sex couples began requesting marriage licenses, the stereotype that individuals marry members of the opposite sex has persisted because every state prohibited same-sex marriage until 2003, when the Massachusetts Supreme Court ruled that the state could not "deny the protections, benefits, and obligations conferred by civil marriage to two individuals of the same sex who wish to marry" (440 Mass. 2309).

¹² See Appendix B for more details on the history of same-sex marriage. The main points are summarized here.

Other state supreme courts followed suit, including New Jersey in 2006, Connecticut in 2008, and Iowa in 2009. Perhaps the most notable court ruling on same-sex marriage, however, occurred in 2013 when the U.S. Supreme Court declared the Defense of Marriage Act unconstitutional under the due process clause of the U.S. Constitution. Enacted in 1996, the Defense of Marriage Act (DOMA) restricted the definition of marriage to “a legal union between one man and one woman as husband and wife” (110 Stat. 2419, page 1), thereby prohibiting same-sex marriages from being legally recognized by the federal government. In the *United States v. Windsor* decision repealing DOMA, Justice Kennedy explains that “DOMA’s principal effect is to identify and make unequal a subset of state-sanctioned marriages” (570 U.S.). States remain free to withhold marriage licenses to same-sex couples if they choose, but the federal government is required to recognize marriages between gay and lesbian couples. Currently, 13 states and the District of Columbia allow same-sex marriage, while 35 states prohibit same-sex marriage through constitutional or statutory provisions (National Conference of State Legislatures).

This paper argues that the proportion of voters who oppose same-sex marriage today can be used as a proxy variable to measure the strength of traditional gender stereotypes in a region. The basis for this claim is the correlation found in previous research between traditional views (including views of gender roles) and opposition to same-sex marriage. A number of studies find that voters who oppose gay rights have

more traditional views (Brewer 2003; Campbell and Larson 2007; Duckitt et al. 2010; Gaines and Garand 2010; McVeigh and Diaz 2009; Wilcox et al. 2007). Using survey responses from the American National Election Studies (ANES), Brewer (2003) and Gaines and Garand (2010) measure “moral traditionalism” by combining responses to four statements: “(1) the world is always changing and we should adjust our view of moral behavior to those changes; (2) the newer lifestyles are contributing to the breakdown of our society; (3) we should be more tolerant of people who choose to live according to their own moral standards, even if they are very different from our own; and (4) this country would have fewer problems if there were more emphasis on traditional family ties” (Gaines and Garand 2010 online appendix, page 1). Brewer (2003) examines the 1992, 1996, and 2000 ANES surveys and finds that moral traditionalism “consistently produced opposition to gay rights” (page 1216). He measures gay rights by combining answers to two questions: “Do you favor or oppose laws to protect homosexuals against job discrimination?” and “Do you think homosexuals should be allowed to serve in the United States Armed Forces, or don’t you think so?” (page 1212). While Brewer (2003) does not examine opposition to same-sex marriage specifically, Gaines and Garand (2010) examine responses to the 2004 ANES survey which asked the following question: “Should same-sex couples be allowed to marry, or do you think they should not be allowed to marry?” (page 554). They find that support for “moral traditionalism” is negatively and statistically significantly

related to support for same-sex marriage (Table 1, page 559). They estimate that, while 75 percent of individuals with the lowest support for moral traditionalism favor same-sex marriage, only 6 percent of individuals with the highest support for moral traditionalism favor same-sex marriage (page 559).

Duckitt et al. (2010) also examine the strength of the relationship between “traditionalist” social attitudes and gay rights. The authors measure traditionalist social views by surveying individuals on their agreement with 10-12 statements such as “It is important that we preserve our traditional values and moral standards” and “The radical and sinful new ways of living and behaving of many young people may one day destroy our society” (page 712). They find a large, statistically significant negative relationship between traditionalist views and respondents’ agreement with the statements “Gay and lesbian marriage should be permitted” and “I support gay and lesbian rights” (page 696).¹³

The prior studies indicate that opposition to gay rights is correlated with support for more traditionalist views in general. Further, some of these studies find that voters who oppose same-sex marriage have more traditional views specifically regarding the role of women. Using the ANES, Brint and Abrutyn (2010) and Gaines and Garand (2010) measure support for equal gender roles using a seven-point scale that ranges from

¹³ It should be noted that their findings are based on “samples of convenience” that include only 67 participants in the U.S. (pages 693-694). This finding is robust to another small sample of 106 participants in Israel.

“men and women should have equal roles” to “women’s place is in the home” (Brint and Abrutyn 2010: Table 2, page 338). Brint and Abrutyn (2010) find that support with male-dominant gender roles is negatively associated with support for gay rights (Table 5, page 342). Similarly, Gaines and Garand (2010) report that support for equal gender roles is positively associated with support for same-sex marriage (Table 1, page 559). They hypothesize that this finding “reflects individuals’ gender positioning in the culture wars, with gender-role traditionalists on one side of the same-sex marriage issue and gender-role egalitarians on the other side” (page 561).

Gaines and Garand (2010) also explore whether voters who support same-sex marriage are more supportive of women’s rights. They base their measure of women’s rights on responses to statements such as “women demanding equality seek special favors” and “women complaining about harassment cause problems” (online appendix, page 2). However, they do not find a statistically significant relationship between support for same-sex marriage and support for women’s rights. The authors argue that “the concepts of gender and sexual orientation are intertwined,” but the relationship appears nuanced; views on same-sex marriage are related to views of gender roles but unrelated to views of women’s rights.

The findings that opposition to same-sex marriage is associated with more traditional views — both traditional views in general and traditional views regarding the role of women specifically — suggest that the proportion of voters who oppose

same-sex marriage can be used as a proxy variable to measure the strength of traditional gender stereotypes. Although the studies above do not describe their findings in terms of stereotypes, statements such as “this country would have fewer problems if there were more emphasis on traditional family ties” (Brewer 2003; Gaines and Garand 2010) and “the ‘old-fashioned ways’ and ‘old-fashioned values’ still show the best way to live” (Duckitt et al. 2010) conjure images with implicit traditional gender stereotypes, such as a male breadwinner and a female housewife. Hence, it seems reasonable to hypothesize that, because opposition to same-sex marriage is associated with more traditional views, opposition to same-sex marriage may be associated with greater traditional gender stereotypes.

1.1.4 What are other proxy variables for traditional gender stereotypes?

The close association between same-sex marriage and traditional views supports using the proportion of voters who oppose same-sex marriage as a proxy variable for greater traditional gender stereotypes. However, there are two variables that are closely correlated with opposition to same-sex marriage that may also reflect greater traditional gender stereotypes: support for the Republican Party and adherence to Evangelicalism.

1.1.4.1 Support for the Republican Party

Many studies illustrate that support for same-sex marriage is correlated with partisan identification and political ideology. Both Brewer (2003) and Gaines and Garand (2010) use the seven-point scales from the American National Election Studies

(ANES) to measure partisan identification (ranging from strong Democrat to strong Republican) and political ideology (ranging from strong liberal to strong conservative). Brewer (2003) reports that more liberal ideology or stronger Democrat identification are associated with greater support for gay rights (Table 2, page 1215). Gaines and Garand (2010) reach a parallel conclusion: individuals with more conservative beliefs or with a stronger Republican identification are less likely to support same-sex marriage (Table 1, page 559). However, Brewer (2003) finds that only partisan identification is statistically significant, while Gaines and Garand (2010) report that only political ideology is statistically significant.

McVeigh and Diaz (2009) also examine the relationship between support for Republican candidates and support for same-sex marriage. They explore county-level variation across 2,231 counties from 28 states where same-sex marriage initiatives appeared on the ballot between 2000 and 2008. They find that counties where a greater proportion of voters supported George W. Bush in 2000 also had a greater proportion of voters support ballot measures to ban same-sex marriage. They estimate that a one percentage point increase in support for the Republican presidential candidate is associated with roughly a 0.30 percentage point increase in support for same-sex marriage bans (Table 2, page 901).

The studies above illustrate that opposition to same-sex marriage is correlated with support for the Republican Party. However, opposition to same-sex marriage is

useful in the present study because it is plausibly associated with greater traditional gender stereotypes. Support for the Republican Party will only be useful in the present study if it is also plausibly associated with greater traditional gender stereotypes. On the one hand, there is some evidence that suggests an association between support for the Republican Party and greater traditional gender stereotypes. Knuckey (2005) examines the relationship between Republican Party identification and moral traditionalism. He measures “moral traditionalism” using responses to the four survey questions from the American National Election Studies (ANES) discussed previously.¹⁴ He finds a statistically significant positive relationship between Republican Party identification and moral traditionalism (Table 1, page 657). He further reports that the most morally traditional individuals were 18 percentage points more likely to vote for Republican U.S. Congressional candidates than the most morally progressive individuals between 1994 and 2000 (page 659). Brint and Abrutyn (2010) also use the ANES and find a statistically significant positive relationship between Republican Party identification and moral traditionalism (Table 5, page 343). Even more compelling, the authors report a statistically significant positive relationship between Republican Party identification and male-dominant gender roles (Table 5, page 343).

¹⁴ For convenience, the survey statements are as follows: “(1) the world is always changing and we should adjust our view of moral behavior to those changes; (2) the newer lifestyles are contributing to the breakdown of our society; (3) we should be more tolerant of people who choose to live according to their own moral standards, even if they are very different from our own; and (4) this country would have fewer problems if there were more emphasis on traditional family ties” (Gaines and Garand 2010 online appendix, page 1).

Although there is some evidence that suggests a potential association between support for the Republican Party and greater traditional gender stereotypes, there are reasons to believe that support for the Republican Party will be a weaker proxy variable of greater traditional gender stereotypes than opposition to same-sex marriage. It is plausible that some individuals affiliate with the Republican Party even if they support same-sex marriage. Individuals who are fiscally conservative but socially liberal may affiliate with the Republican Party if, for example, their preference for lower taxes outweighs their preference for legalizing same-sex marriage. Further, Duckitt et al. (2010) do not find a statistically significant relationship between support for the Republican Party and traditionalism. These authors measure traditionalism by surveying individuals on their agreement with 10-12 statements such as “It is important that we preserve our traditional values and moral standards” and “The country will flourish if young people stop experimenting with drugs, alcohol, and sex, and pay more attention to family values” (pages 711-712). They combine these survey responses to develop an aggregate measure of traditionalism that “expresses the value and motivational goal of maintaining traditional lifestyles, norms, and morality” (page 691). They find a positive association between support for the Republican Party and traditionalism, but the relationship is not statistically significant (page 703).

1.1.4.2 Adherence to Evangelism

There is also evidence that support for same-sex marriage is correlated with Evangelical religious beliefs. Gaines and Garand (2010) find negative, statistically significant relationship between supporting same-sex marriage and reporting an Evangelical Protestant religious denomination. They estimate that “84 percent of Evangelical Protestants are predicted to oppose same-sex marriage, compared to only 55 percent of non-Evangelicals” (page 560). Brint and Abrutyn (2010) also report a negative, statistically significant relationship between supporting gay rights and reporting an Evangelical Protestant religious denomination. In their county-level analysis, McVeigh and Diaz (2009) consider whether counties with a greater proportion of Evangelical Protestants had greater support for same-sex marriage bans. They find that a one percentage point increase the proportion of religious adherents who are Evangelical Protestant is associated with a small but statistically significant 0.03 percentage point increase in support for same-sex marriage bans (Table 2, page 901).

Again, adherence to Evangelicalism will only be useful in the present study if it is plausibly associated with greater traditional gender stereotypes, not just associated with same-sex marriage. Using the oft-studied ANES, Brint and Abrutyn (2010) report a positive correlation between Evangelical affiliation and moral traditionalism. Weisberg (2005) examines the relationship between Evangelical affiliation and moral beliefs using survey responses to The Ohio Political Study (TOPS). This study, conducted by Ohio

State University in 2002, asked the following questions: “This country would have many fewer problems if there were more emphasis on traditional family ties,” “The newer lifestyles are contributing to the breakdown of society,” and “Do you think it is more important for government to encourage traditional family values, or encourage tolerance of nontraditional families?” (page 664). He finds a statistically significant positive relationship between moral traditionalism and Evangelical Protestants, but he does not find a statistically significant relationship between moral traditionalism and any other religious denomination (Table 2, page 654). Using a sample of respondents in New Zealand, Duckitt et al. (2010) find a large, statistically significant correlation between traditionalist views and religious fundamentalism, where fundamentalism is measured by respondents’ answers to statements such as, “To lead the best, most meaningful life, one must belong to the one, true religion” (page 695).

Although the prior studies suggest that adherence to Evangelicalism is associated with traditionalism, there are reasons to believe that adherence to Evangelicalism will be a weaker proxy variable of greater traditional gender stereotypes than opposition to same-sex marriage. Campbell and Larson (2007) argue that many religious denominations have split into two factions on social issues: the traditionalists and the progressives. They explain: “people with these [various religious denominations] who hold traditional beliefs have found that they have more in common with other traditionalists, even those of other faiths, than members of their own religion

who hold what are often called progressive beliefs” (page 132). If some Evangelical Protestants are less traditional and exhibit weaker traditional gender stereotypes, then Evangelical adherence will be less representative of traditional gender stereotypes than support for same-sex marriage.

To summarize, Section 1.1 has presented evidence of the following: that gender gaps exist among top-scoring students, that the gaps are at least partly attributable to differences in socialization, and that opposition to same-sex marriage is associated with more traditional views on the role of women. I hypothesize that the proportion of voters who oppose same-sex marriage can be used as a proxy variable to measure the strength of traditional gender stereotypes in a region and that the strength of traditional gender stereotypes in a region is associated with greater gender gaps among top-scoring students on state assessments. Hence, the following analyses explore whether the proportion of voters who oppose same-sex marriage is associated with greater gender gaps among top-scoring students on state assessments. For comparison, the following analyses also explore whether the proportion of voters who supported Republican presidential candidates and the proportion of religious adherents who are Evangelical are also associated with greater gender gaps on state assessments.

1.2 Data and methodology

1.2.1 Data

North Carolina requires students in 3rd through 8th grade to take annual standardized assessments in math and reading, jointly referred to as End of Grade tests. Students in 5th and 8th grades are also required to take a standardized assessment in science. Each year, the North Carolina Department of Public Instruction compiles a database containing the test scores and demographic characteristics of all students enrolled in a public school in the state. This database is provided to the North Carolina Education Research Data Center at Duke University, where data center administrators remove identifying student characteristics (e.g., names, birthdates, and Social Security Numbers) and replace them with unique student codes that remain with students throughout their years in the state public school system. This restricted database is then shared with researchers.

This paper analyzes math, science, and reading test scores for 5th and 8th grade students between 2008 and 2011.¹⁵ I normalize test scores by grade, test, and year. Then, I restrict the sample to white students because gender gaps vary by race (American Association of University Women 2008; Brandon et al. 1987; Schratz 1978) and because gender gaps are more prevalent for white students than for students of other ethnicities (American Association of University Women 2008; Friedman 1989;

¹⁵ The math and reading tests have been administered to 3rd through 8th grade students since 1993, but the science test has only been administered to 5th and 8th grade students since 2008.

Hyde et al. 2008). Previous studies have also restricted their analyses to white students (Arden and Plomin 2006; Pope and Sydnor 2010), so this condition is consistent with the literature.¹⁶

The sample contains 235,482 students in 5th grade and 232,481 students in 8th grade.¹⁷ Table 3 provides further details on the number of students by grade, by test, and by gender. Males comprise 51.3 percent of 5th grade students and 51.4 percent of 8th grade students. Males also comprise between 50.8 percent and 51.0 percent of test takers on any given assessment in either grade. Hence, males are more likely to be missing test scores than females, but the difference is no more than half of one percentage point.

After mapping school districts to counties, I count the number of boys and the number of girls in each county who score in the top 5 percent on the math assessment. I repeat this process for the science and reading assessments. I use these numbers to calculate the male-female ratio in math, the male-female ratio in science, and the female-male ratio in reading for each county. I calculate the arithmetic average of these three ratios and refer to this ratio as the “gender gap” in the county. I present results using this definition of gender gaps in Sections 1.3 and 1.4. In the robustness checks described

¹⁶ The results presented below are robust to the inclusion of all students in the state.

¹⁷ The sample omits 192 observations for students whose school district is classified as the North Carolina Department of Health and Human Services or the North Carolina Department of Juvenile Justice, 22 observations for students whose school district is missing, and 3 observations for students whose gender is missing.

Table 3: Observations by grade, by test, and by gender

	All students	Students with math scores	Students with science scores	Students with reading scores
Panel A. 5th grade				
Total	235,482	228,951	228,368	227,910
Male	120,782 51.3%	116,588 50.9%	116,175 50.9%	115,767 50.8%
Female	114,700 48.7%	112,363 49.1%	112,193 49.1%	112,143 49.2%
Panel B. 8th grade				
Total	232,481	226,017	225,597	225,724
Male	119,542 51.4%	115,341 51.0%	115,050 51.0%	115,064 51.0%
Female	112,939 48.6%	110,676 49.0%	110,547 49.0%	110,660 49.0%

Notes: The sample includes white students enrolled in 5th or 8th grade in a public school in North Carolina between 2008 and 2011. The sample omits 192 observations from students whose school districts are identified as the North Carolina Department of Juvenile Justice or the North Carolina Department of Health and Human Services, 22 observations from students whose school districts are missing, and 3 observations from students whose genders are missing.

Source: North Carolina Education Research Data Center, End of Grade files.

in Section 1.5, I evaluate the results using two alternative measures of gender gaps: (1) computing a weighted average with the male-female ratio in math weighted at 25 percent, the male-female ratio in science weighted at 25 percent, and the female-male ratio in reading weighted at 50 percent, and (2) calculating the geometric mean of the three ratios.

As explained in Section 1.1, I hypothesize that opposition to same-sex marriage can be used as a proxy variable for traditional gender stereotypes in a region. To capture views on same-sex marriage, I use voting records from a ballot measure that appeared during the May 2012 primary in North Carolina. The measure asked voters whether they supported amending the North Carolina Constitution to include the following section: "Marriage between one man and one woman is the only domestic legal union that shall be valid or recognized in this State. This section does not prohibit a private party from entering into contracts with another private party; nor does this section prohibit courts from adjudicating the rights of private parties pursuant to such contracts." Known as Amendment One, 61 percent of over two million voters in the state supported the measure. I calculate the share of voters in each county who supported Amendment One using publicly-available data from the North Carolina Board of Elections.

I also hypothesize that support for the Republican Party and adherence to Evangelicalism may reflect the strength of traditional gender stereotypes in a region. To capture support for the Republican Party, I calculate the share of voters in each county who supported the Republican presidential candidate during the 2012 and 2008 general elections, again using publicly-available data from the North Carolina Board of Elections. North Carolina voters were more likely to support the Republican presidential candidates in both of these elections than voters nationwide. Across the

U.S., 47 percent of voters supported Mitt Romney in 2012, and 46 percent of voters supported John McCain in 2008 (Federal Election Commission 2012; Federal Election Commission 2008). In North Carolina, 50 percent of 4.5 million voters supported Mitt Romney, and 49 percent of 4.3 million voters supported John McCain.

To capture adherence to Evangelicalism, I use data from the U.S. Religion Census, a survey conducted by the Association of Statisticians of American Religious Bodies (ASARB). The survey asked religious groups to estimate the number of adherents in each county across the U.S. The census defined adherents as members of the congregation “with full membership status,” as well as individuals who regularly attend services but who are not considered members, such as “those not confirmed” and “those not eligible for Communion” (Association of Religion Data Archives 2010). The Association of Statisticians of American Religious Bodies followed Steensland et al. (2000) to classify 146 of the 236 religious groups as Evangelical Protestant, including the American Baptist Association, Southern Baptist Convention, Southern Methodist Church, Lutheran Church-Missouri Synod, Wisconsin Evangelical Lutheran Synod, Churches of God, and Assemblies of God (McVeigh and Diaz 2009; Steensland et al. 2000).¹⁸ I calculate the share of all religious adherents in each county and the share of

¹⁸ The Association of Statisticians of American Religious Bodies consulted other resources — such as the Encyclopedia of American Religion and the Handbook of Denominations in the United States — to classify denominations not examined by Steensland et al. (2000). A full list of denominations classified as Evangelical Protestant by the Association of Statisticians of American Religious Bodies can be found at the following website: <http://www.thearda.com/rcms2010/evangelical.asp>

non-Black Protestant religious adherents in each county who are affiliated with an Evangelical Protestant religious denomination.

The census recorded over 150 million adherents nationally (49 percent of the country's population) and over 4.5 million adherents in North Carolina (48 percent of the state's population). Although one-third of adherents nationally are Evangelical, over half of adherents in North Carolina are affiliated with an Evangelical denomination. The largest religious denomination in the state — the Southern Baptist Convention, an Evangelical Protestant denomination — has over 1.5 million adherents. The next largest religious denomination — the United Methodist Church, a Mainline Protestant denomination — has 660,000 adherents, fewer than half the number of Southern Baptists. Other Evangelical denominations reporting large numbers of adherents in North Carolina include the Church of God (Cleveland, Tennessee denomination) with 90,000 adherents, the International Pentecostal Holiness Church with 57,000 adherents, and the Assemblies of God with 56,000 adherents (Association of Religion Data Archives 2010, State Membership Report for North Carolina).

Together, the North Carolina Board of Elections and the U.S. Religious Census provide five proxy variables for traditional gender stereotypes: the proportion of voters who supported Amendment One, the proportion of voters who supported Mitt Romney in 2012, the proportion of voters who supported John McCain in 2008, the proportion of religious adherents who are Evangelical, and the proportion of non-Black Protestant

religious adherents who are Evangelical. The first three proxy variables should have little measurement error, as these variables are collected from the North Carolina Board of Elections. The last two proxy variables, however, may have more substantial errors in measurement. The Association of Statisticians of American Religious Bodies (ASARB) asked each religious group to estimate the number of adherents in each county. Some religious groups provided ASARB with yearbooks listing the members of each congregation, allowing the ASARB to calculate the number of (listed) members in each county. Other religious groups instructed the ASARB to estimate the number of members “according to some formula” or provided the ASARB with a list of members in each county without specifying how the groups determined those numbers (Association of Religion Data Archives 2010). As a result, the ASARB reports that 31 counties have more religious adherents than residents. The Association of Religion Data Archives explains: “Reasons for the discrepancy will differ from county to county, but the most plausible would include U.S. Census undercount, church membership overcount, and county of residence differing from county of congregational membership” (Association of Religion Data Archives 2010). I attempt to minimize these errors in measurement by calculating Evangelical adherents as a proportion of all adherents or as a proportion of non-Black Protestant adherents, rather than calculating Evangelical adherents as a proportion of the population. If Evangelical religious groups estimated membership data with the same degree of error as other religious groups, then the errors in

measurement will cancel out. If Evangelical religious groups were more likely to overestimate or underestimate membership data than other religious groups, then the calculated proportions will be biased upwards or downwards.¹⁹

This paper estimates the relationship between gender gaps among top-scoring students and these five proxy variables for traditional gender stereotypes by exploring variation across counties in North Carolina. I restrict the analysis to North Carolina because the North Carolina Education Research Data Center provides student-level data only for students in North Carolina public schools. Further, I focus the analysis on counties because the North Carolina Board of Elections and the U.S. Religious Census release data by county.

1.2.2 Methodology

Section 1.1 presented two arguments. The first argument is that the strength of traditional gender stereotypes in a region is associated with greater gender gaps among top-scoring students. To test this hypothesis, I would like to estimate the following regression:

$$GenderGap_{ig} = \beta_0 + \beta_1 GenderStereotype_i + \gamma X_{ig} + \varepsilon_{ig} \quad (1)$$

¹⁹ The results presented below are robust to using Evangelical adherents as a proportion of the population instead of Evangelical adherents as a proportion of religious adherents.

where $GenderGap_{ig}$ would be some average of the male-female ratio in math, the male-female ratio in science, and the female-male ratio in reading among students in the top 5 percent in area i and grade g ; $GenderStereotype_i$ would be a measure of the strength of traditional gender stereotypes in area i ; X_{ig} would be a matrix of controls for area i and grade g ; and ε_{ig} would be an error term. Then, the coefficient β_1 would represent the relationship between traditional gender stereotypes and gender gaps among top-scoring students. The problem with this approach is that there are few direct measures of the strength of traditional gender stereotypes in a region. Nosek et al. (2008) use implicit associations between males and science to calculate the strength of traditional gender stereotypes, but their study relies on individuals who self-select into their research sample by completing their online Implicit Association Test (IAT). Pope and Sydnor (2010) use explicit responses to the General Social Survey and calculate the proportion of adults who agreed that it is “much better for everyone involved if the man is the achiever outside the home and the woman takes care of the home and family” (page 104), but there are far too few respondents to explore variation across counties.²⁰

To address this problem, Section 1.1 presented a second argument: that the proportion of voters who oppose same-sex marriage can be used as a proxy variable to measure the strength of traditional gender stereotypes. The studies cited in Section 1.1.3

²⁰ Even Pope and Sydnor (2010) have to restrict their analysis to 37 states, indicating that 13 states have fewer than 100 respondents to this question between 1972 and 2006.

illustrate that opposition to same-sex marriage is associated with traditional views and suggest that opposition to same-sex marriage may also be associated with traditional gender stereotypes. Therefore, I estimate the following ordinary least squares (OLS) regression:

$$GenderGap_{ig} = \beta_0 + \beta_1 AmendmentSupport_i + \gamma X_{ig} + \varepsilon_{ig} \quad (2)$$

where $GenderGap_{ig}$ is the arithmetic average of the male-female ratio in math, the male-female ratio in science, and the female-male ratio in reading among students in the top 5 percent in area i and grade g ; $AmendmentSupport_i$ is the proportion of voters who supported Amendment One in area i ; X_{ig} is a matrix of controls for area i and grade g ; and ε_{ig} is an error term. Here, the coefficient β_1 represents the relationship between opposition to same-sex marriage and gender gaps among top-scoring students. I estimate separate regressions for each grade g .

Opposition to same-sex marriage is also strongly correlated with support for the Republican Party and adherence to Evangelical religious denominations. In alternate specifications, I replace the variable $AmendmentSupport_i$ with the variable $RepublicanSupport_i$, which represents the proportion of voters in area i who supported the Republican presidential candidate in 2012 or in 2008. I also replace the variable $AmendmentSupport_i$ with the variable $EvangelicalAdherence_i$, which represents the

proportion of all religious adherents or the proportion of non-Black Protestant religious adherents who are Evangelical in area i .

The matrix of controls includes the number of students in grade g and the median household income in area i . I include the median household income as a control because it is a significant predictor of gender gaps across states (Pope and Sydnor 2010). Similar research has also identified gross domestic product (GDP) and gross national income (GNI) as significant predictors of gender gaps across countries (Guiso et al. 2008; Nosek et al. 2008; Penner 2008).

1.3 Descriptive findings

1.3.1 Gender gaps across test score percentiles

Figures 1, 2, and 3 depict the gender ratios for each test at various points of the test score distribution, and Table 4 provides detailed descriptive statistics on gender ratios and gender gaps. The three figures illustrate that the gender ratios are always greater than one. Collectively, students always adhere to the gender norm of more boys than girls scoring in a top percentile group on math and science tests and more girls than boys scoring in a top percentile group on reading tests. The figures also illustrate that, for both grades and for all points in the distribution, gender ratios are the largest for the science test. The male-female ratio on the 5th grade science test ranges from 1.14 in the top half of the distribution to 2.40 in the top one percent of the distribution. In other words, nearly 12,000 more boys than girls score in the top half of the distribution

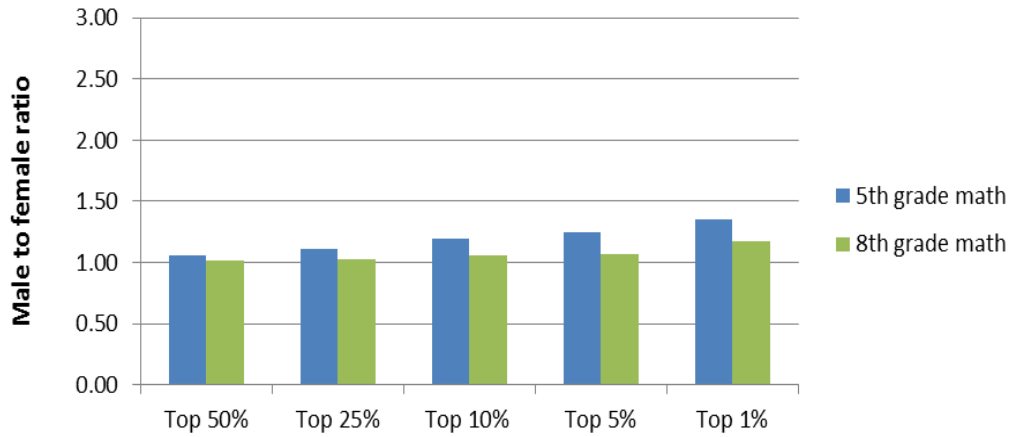


Figure 1: Male-female gender ratio on North Carolina math assessment

Notes: The sample includes white students enrolled in 5th or 8th grade in a public school in North Carolina between 2008 and 2011. The sample omits 192 observations from students whose school districts are identified as the North Carolina Department of Juvenile Justice or the North Carolina Department of Health and Human Services, 22 observations from students whose school districts are missing, and 3 observations from students whose genders are missing. *Source:* North Carolina Education Research Data Center, End of Grade files.

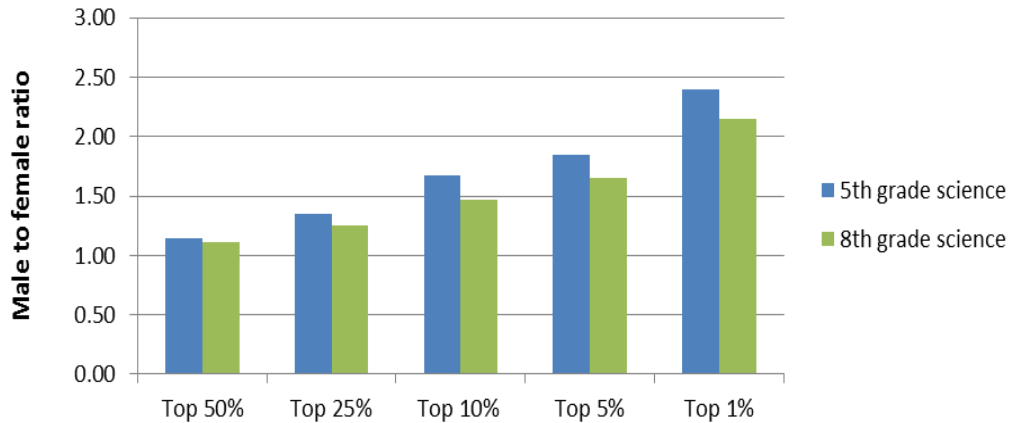


Figure 2: Male-female gender ratio on North Carolina science assessment

Note: See Figure 1 for a description of the sample. *Source:* North Carolina Education Research Data Center, End of Grade files.

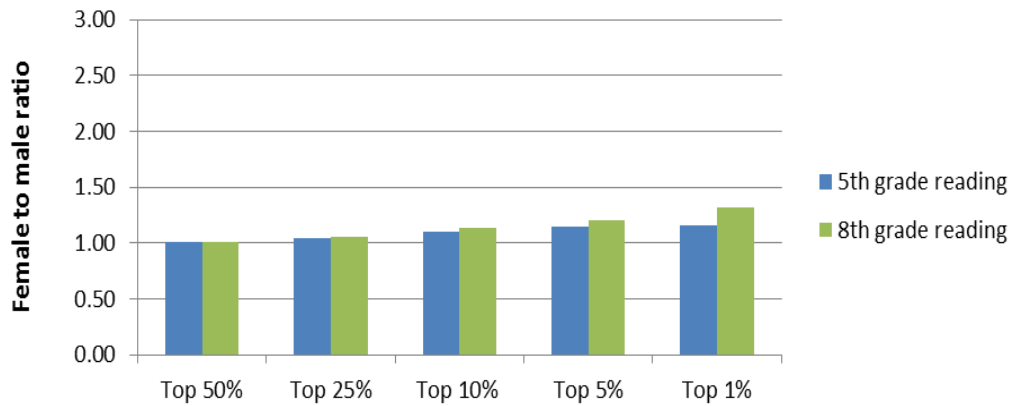


Figure 3: Female-male gender ratio on North Carolina reading assessment

Note: See Figure 1 for a description of the sample. *Source:* North Carolina Education Research Data Center, End of Grade files.

on the 5th grade science test, and over 70 percent of students who score in the top one percent on the 5th grade science test are boys. The statistics are similar for 8th grade students. Table 4 also illustrates that gender ratios and measures of gender gaps increase monotonically as the test score percentile increases. The arithmetic average of gender ratios increases from 1.07 in the top half of the distribution to 1.64 in the top one percent of the distribution for 5th grade students. The numbers are smaller but similar for 8th grade students and for alternate measures of gender gaps. In short, the gender gaps are much larger at the top of the test score distribution than at the middle of the test score distribution.

Table 4: Gender ratios on North Carolina End of Grade tests across test score distributions

		Top 50%		Top 25%		Top 10%		Top 5%		Top 1%	
Panel A. 5th grade											
<i>Number of students by test</i>		Male	Female	Male	Female	Male	Female	Male	Female	Male	Female
Math		85,411	80,572	48,753	43,781	22,110	18,492	10,559	8,501	2,338	1,733
Science		94,805	82,974	59,950	44,552	25,941	15,510	14,527	7,869	3,091	1,287
Reading		84,254	85,346	46,465	48,787	17,792	19,678	8,647	9,901	1,914	2,212
<i>Gender ratios by test</i>											
Math (M-F)		1.06		1.11		1.20		1.24		1.35	
Science (M-F)		1.14		1.35		1.67		1.85		2.40	
Reading (F-M)		1.01		1.05		1.11		1.15		1.16	
<i>Combined gender ratios</i>											
ES	Average of ratios, equal weighting	1.07		1.17		1.32		1.41		1.64	
	Average of ratios, 25% math, 25% science, 50% reading	1.06		1.14		1.27		1.34		1.52	
	Geometric mean	1.07		1.16		1.30		1.38		1.55	
Panel B. 8th grade											
<i>Number of students by test</i>		Male	Female	Male	Female	Male	Female	Male	Female	Male	Female
Math		81,080	79,741	45,570	44,586	19,194	18,065	9,588	8,960	1,999	1,703
Science		89,718	80,827	53,579	42,621	23,613	16,060	12,612	7,616	2,737	1,272
Reading		85,276	85,684	46,806	49,383	18,759	21,290	9,333	11,243	1,657	2,179
<i>Gender ratios by test</i>											
Math (M-F)		1.02		1.02		1.06		1.07		1.17	
Science (M-F)		1.11		1.26		1.47		1.66		2.15	
Reading (F-M)		1.00		1.06		1.13		1.20		1.32	

<i>Combined gender ratios</i>					
Average of ratios, equal weighting	1.04	1.11	1.22	1.31	1.55
Average of ratios, 25% math, 25% science, 50% reading	1.03	1.10	1.20	1.28	1.49
Geometric mean	1.04	1.11	1.21	1.29	1.49

Note: See Table 1 for a description of the sample.

Source: North Carolina Education Research Data Center, End of Grade files.

1.3.2 Gender gaps in the top 5 percent across North Carolina

To develop a more detailed picture of gender gaps, I explore how gender ratios at the top 5 percent differ across the state's 100 counties. Although the sample contains 17,783 observations from 5th grade students in Wake County, it contains only 84 observations from 5th grade students in Hyde County. Wake County had 4,878 students score in the top 5 percent on at least one 5th grade assessment over this time period, but Hyde County had only 20 students score in the top 5 percent on at least one 5th grade assessment. Between 2008 and 2011, five counties had no female students score in the top 5 percent on the 5th grade math test, one county had no female students score in the top 5 percent on the 5th grade science test, and another county had no male students score in the top 5 percent on the 5th grade reading test. However, every county had at least one male student score in the top 5 percent on the 5th grade math and science tests, and every county had at least one female student score in the top 5 percent on the 5th grade reading test.

Because some counties lacked a single female student in the top 5 percent on the math and science test while other counties lacked a single male student in the top 5 percent on the reading test, it would be impossible to compute the male-female ratio in math and science or the female- gender ratio in reading for these counties. In an effort to observe at least 100 students in the top 5 percent in each area, I group counties with fewer than 2,000 students into "county groups." First, I identify the 61 small counties

with fewer than 2,000 students over this period, illustrated in Figure 4. Then, I divide these 61 counties across three regions — Mountain, Piedmont, and Coastal — using the boundaries specified by the North Carolina Department of Public Instruction (DPI), illustrated in Figure 5. I further divide each region horizontally into a northern and a southern section, illustrated in Figure 6. This process yields six groups of counties: North Mountain, South Mountain, North Piedmont, South Piedmont, North Coastal, and South Coastal. The 61 small counties are distributed roughly evenly across the six county groups: 11 counties in the North Mountain group, 8 counties in the South Mountain group, 9 counties in the North Piedmont group, 11 counties in the South Piedmont group, 14 counties in the North Coastal group, and 8 counties in the South Coastal group. Hence, I create 45 different areas across the state: 39 large counties and six county groups.

Table 5 provides descriptive statistics of gender ratios for the full sample, for the 39 large counties, and for the six county groups.²¹ On average, the county groups are larger than the individual counties. The 39 large counties have an average of 4,741 students in 5th grade, and the six county groups have an average of 8,433 students in 5th grade. While the numbers of students differ, the gender ratios among top-scoring students are similar. For example, the average of the three subject ratios in 5th grade is exactly the same for large counties and for county groups. The average of the three

²¹ Appendix D presents descriptive statistics for each county group separately.

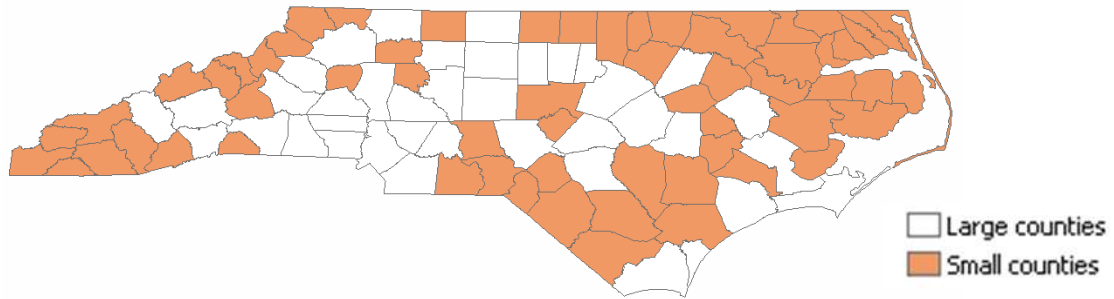


Figure 4: Counties by number of students

Notes: Small counties are defined as counties with fewer than 2,000 students in 5th grade or fewer than 2,000 students in 8th grade between 2008 and 2011. Large counties are defined as counties with more than 2,000 students in 5th grade and more than 2,000 students in 8th grade between 2008 and 2011. See Figure 1 for a description of the sample. *Source:* North Carolina Education Research Data Center, End of Grade files.

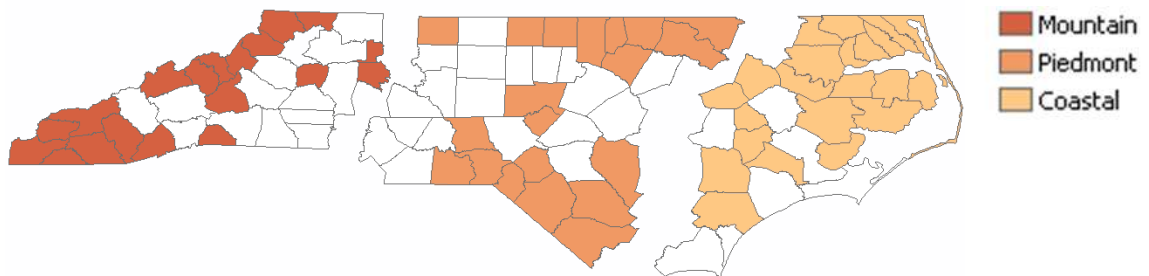


Figure 5: Counties by number of students and by regions

Notes: These regions are determined by the North Carolina Department of Public Instruction. See Figure 1 for a description of the sample. *Sources:* North Carolina Education Research Data Center, End of Grade files; North Carolina Department of Public Instruction.

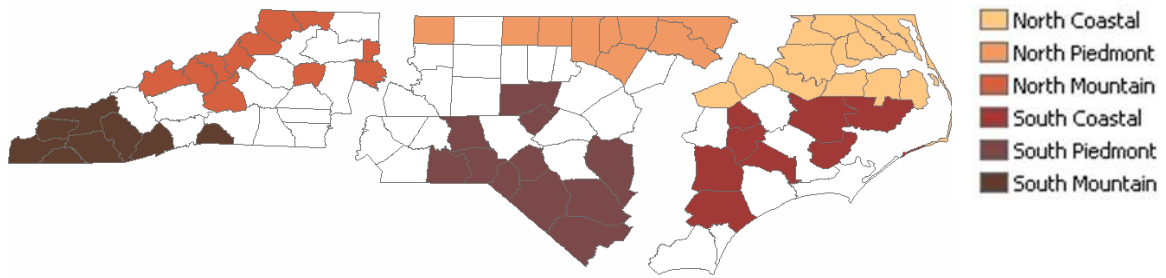


Figure 6: County groups

Notes: County groups are determined by dividing each of the regions from the prior figure into a northern and a southern area. See Figure 1 for a description of the sample. *Sources:* North Carolina Department of Public Instruction; North Carolina Education Research Data Center, End of Grade files.

subject ratios in 8th grade is only slightly larger (4 percent) in county groups than in large counties. While Table 5 does not suggest that the county groups are biasing the results, I restrict the sample to large counties in alternate specifications presented in Section 1.5.

Now that I have identified the 45 areas that will serve as my units of observation, I explore how gender gaps differ across them.²² Figure 7 illustrates the distribution of gender gaps in the top 5 percent. Most areas exhibit gender gaps between 1.3 and 1.5, indicating that roughly 55 to 60 percent of students who score in the top 5 percent in these areas adhere to prescribed gender norms. The gender gap is also greater in 5th grade than in 8th grade. At the one extreme, six areas have a gender gap greater than 1.6 among 5th grade students, but only one area has a gender gap greater than 1.6 among 8th grade students. At the other extreme, only two areas have a gender gap less

²² Appendix E provides full descriptive statistics for all independent and dependent variables.

Table 5: Descriptive statistics for gender ratios

	Number of observations	Number of students	Math, top 5%			Science, top 5%			Reading, top 5%			Average of ratios, top 5%
			(a) Males	(b) Females	(c) M-F	(a) Males	(b) Females	(c) M-F	(a) Males	(b) Females	(c) F-M	
Panel A. 5th grade												
Total	45	5,233	235	189	1.26	323	175	1.95	192	220	1.15	1.45
Large counties	39	4,741	231	185	1.27	303	166	1.95	188	214	1.14	1.45
County groups	6	8,433	258	214	1.21	450	231	1.95	217	257	1.20	1.45
Panel B. 8th grade												
Total	45	5,166	213	199	1.06	280	169	1.73	207	250	1.23	1.34
Large counties	39	4,665	211	197	1.06	276	168	1.73	201	240	1.22	1.33
County groups	6	8,424	227	210	1.11	307	175	1.76	250	314	1.26	1.38

Notes: See Table 1 for a description of the sample. Figures in the "total" rows reflect the average across the 45 areas. Figures in the "large counties" rows reflect the average across the 39 large counties. Figures in the "county groups" rows reflect the average across the 6 groups of small counties.

Source: North Carolina Education Research Data Center, End of Grade files.

than 1.3 among 5th grade students, but 15 areas have a gender gap of less than 1.3 among 8th grade students.

Figures 8 and 9 illustrate the spatial distribution of gender gaps in the top 5 percent. The average gender ratio is the greatest in Harnett County among 5th grade students and in Carteret County among 8th grade students. Harnett County has an average gender ratio of 1.94, implying that two-thirds of 5th grade students in the top 5 percent are the “expected” gender according to gender norms. In contrast, the gender gap in the top 5 percent is only 1.23 among 5th grade students in Orange County and 1.18 among 8th grade students in Surry County.

Although the average of gender ratios is greater than one in every area, gender ratios on the math and reading assessments are sometimes less than one. The lowest gender ratio on a math test is 0.76 among 8th grade students in Haywood County, indicating that over 55 percent of students who score in the top 5 percent on this test are girls. The lowest gender ratio on a reading test is 0.6 among 5th grade students in Stanly County, revealing that over 60 percent of students who score in the top 5 percent on this test are boys. The gender ratios are less than one in three counties for the 5th grade math test, 17 areas for the 8th grade math test, six counties for the 8th grade math test, and one county for the 8th grade reading test. A t-test that 17 areas would display a gender ratio less than one on the 8th grade math test by chance is rejected at the 0.051 level. For every other test and grade, a t-test of the hypothesis that the distribution would arise by

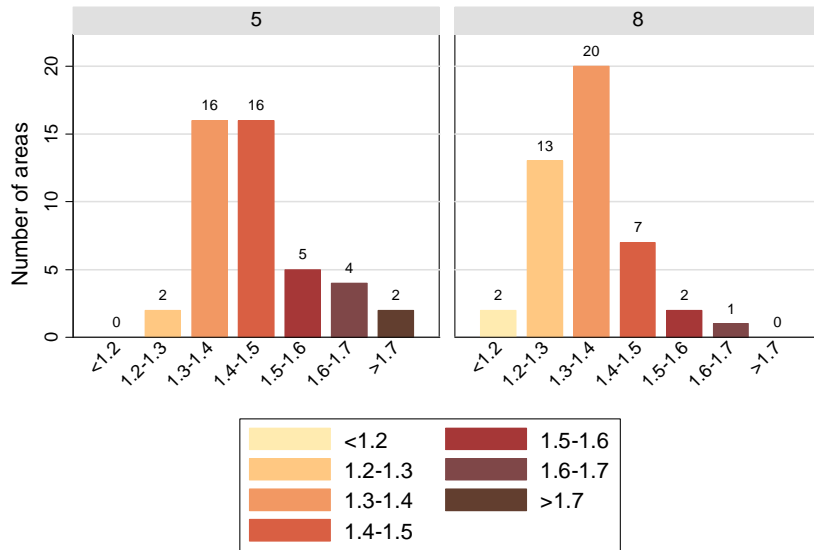


Figure 7: Distribution of gender gaps in the top 5 percent by grade and area

Notes: The sample includes 39 large counties and 6 county groups. The gender gap in the top 5 percent is defined as the average of the male-female ratio in the top 5 percent on the math test, the male-female ratio in the top 5 percent on the science test, and the female-male ratio in the top 5 percent on the reading test. Each observation represents the gender gap in a large county or a county group. *Source:* North Carolina Education Research Data Center, End of Grade files.

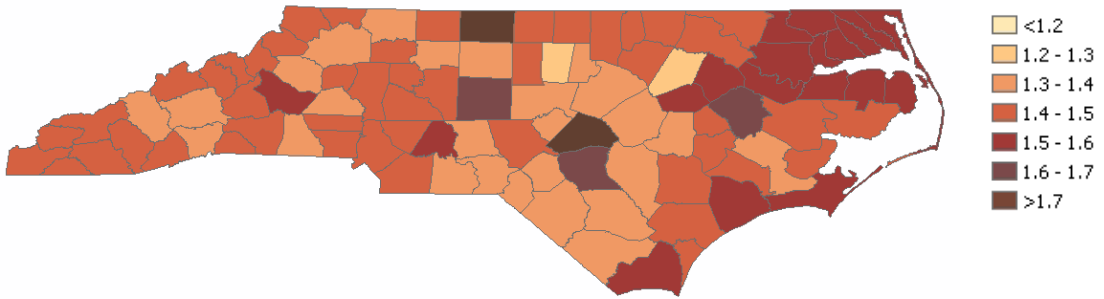


Figure 8: Gender gaps in the top 5 percent in 5th grade by area

Note: See Figure 7 for a definition of gender gaps in the top 5 percent. *Source:* North Carolina Education Research Data Center, End of Grade files.

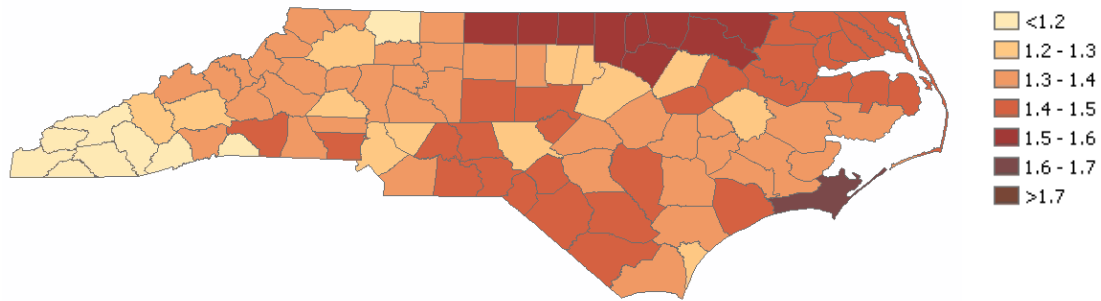


Figure 9: Gender gaps in the top 5 percent in 8th grade by area

Note: See Figure 7 for a definition of gender gaps in the top 5 percent. *Source:* North Carolina Education Research Data Center, End of Grade files.

chance is rejected at the 0.001 level. Further, gender ratios on the science assessment are always greater than one, indicating that females never outnumber males among top-scoring students on the science test in any area. Girls comprise at most 43 percent of students in the top 5 percent on the 5th grade science test and at most 42 percent of students in the top 5 percent on the 8th grade science test.

1.3.3 Proxy variables for traditional gender stereotypes across North Carolina

Section 1.2 noted that 61 percent of nearly 2.2 million voters supported Amendment One, the ballot measure to ban same-sex unions in North Carolina. However, Figures 10 and 11 illustrate that support for the Amendment ranged widely across counties. Support for the Amendment ranged from a high of 83 percent in Wilkes County to a low of 21 percent in Orange County. Support was between 70-80 percent in a plurality of the 45 areas and was greater than 80 percent in six large counties: Wilkes

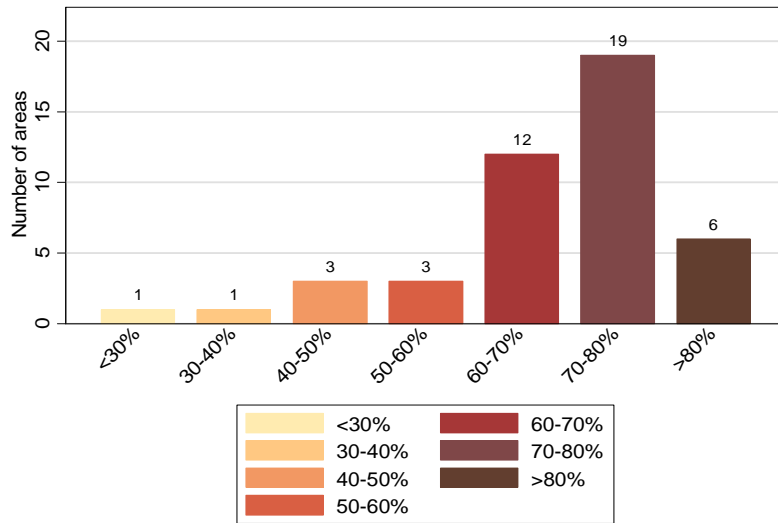


Figure 10: Distribution of support for Amendment One by area

Notes: The sample includes 39 large counties and 6 county groups. Each observation represents support for Amendment One in a large county or a county group. Support for Amendment One is defined as the proportion of voters who supported Amendment One in the primary election. *Source:* North Carolina Board of Elections.

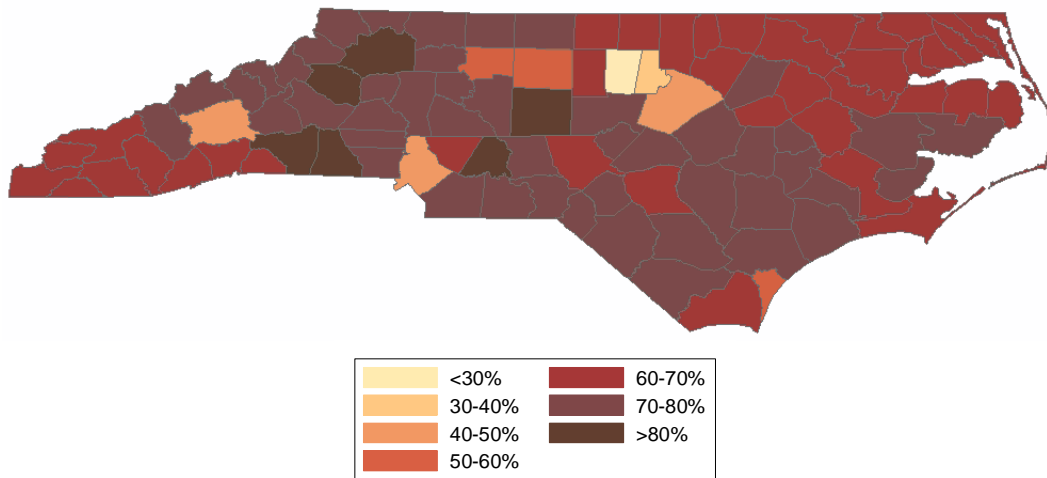


Figure 11: Support for Amendment One by area

Note: See Figure 7 for a definition of support for Amendment One. *Source:* North Carolina Board of Elections.

(83 percent), Rutherford (81 percent), Caldwell (81 percent), Stanly (80 percent), Randolph (80 percent), and Cleveland (80 percent). Two of these counties were identified previously for having high gender gaps in 5th grade: Randolph County (with a gender gap of 1.7 using the arithmetic average of gender ratios across subjects) and Stanly County (with a gender gap of 1.6).

It is important to note that support for the support for the Amendment in the county groups falls in the middle of the distribution. At the extremes, the Amendment received 68 percent support in the North Coastal county group and 74 percent support in the North Mountain county group. This small range in the middle of the distribution suggests that the sorting of small counties into county groups is not biasing the results. In fact, Section 1.5 illustrates that the county groups have a moderating effect on the results.

Section 1.2 also noted that 50 percent of North Carolina voters supported Mitt Romney in 2012, and 49 percent of North Carolina voters supported John McCain in 2008. The distribution of support for Republican presidential candidates is narrower than the distribution of support for Amendment One. Figures 12 and 13 illustrate the distribution of support by area in 2012, and Figures 14 and 15 illustrate the distribution of support by area in 2008. Over 50 percent of voters supported the Republican candidate in 32 areas during the 2012 election and in 33 areas during the 2008 election. The Republican candidates received less than 30 percent of the vote in both elections in

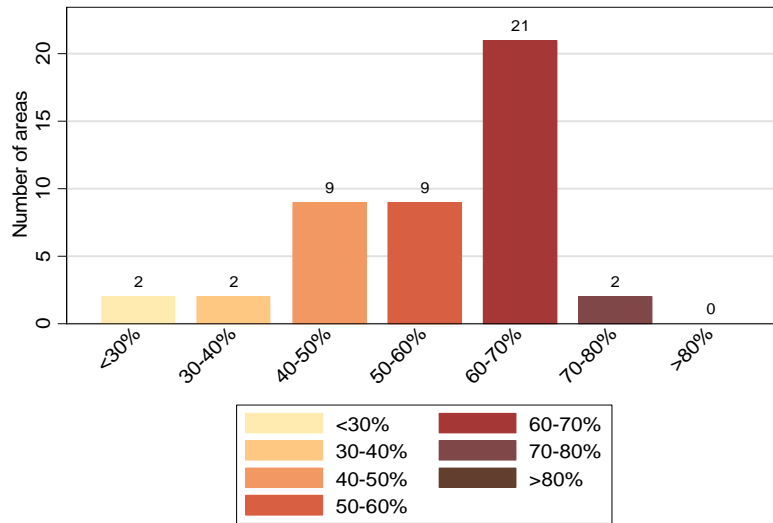


Figure 12: Distribution of support for 2012 Republican presidential candidate by area

Notes: The sample includes 39 large counties and 6 county groups. Each observation represents support for the 2012 Republican candidate in a large county or a county group. Support for the 2012 Republican candidate is defined as the proportion of voters who supported the 2012 Republican candidate in the general election. *Source:* North Carolina Board of Elections.

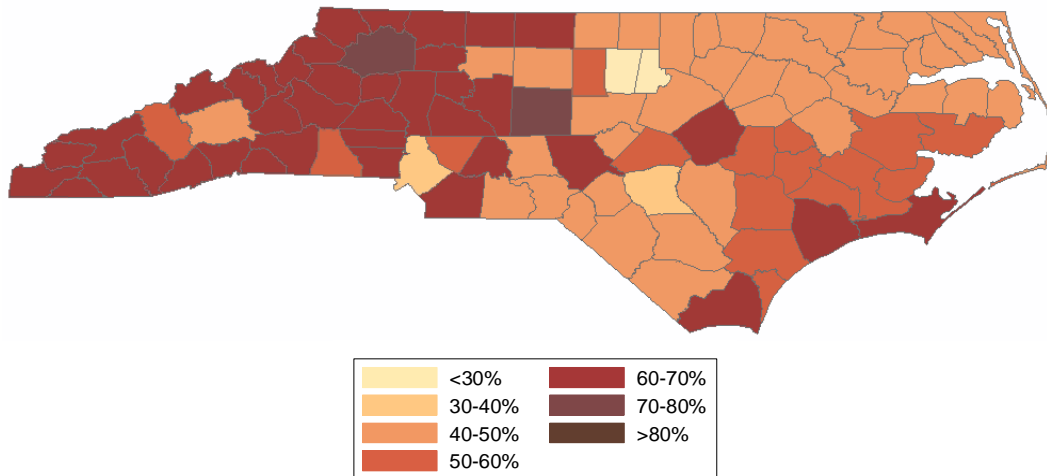


Figure 13: Support for 2012 Republican presidential candidate by area

Note: See Figure 9 for a definition of support for 2012 Republican candidate. *Source:* North Carolina Board of Elections.

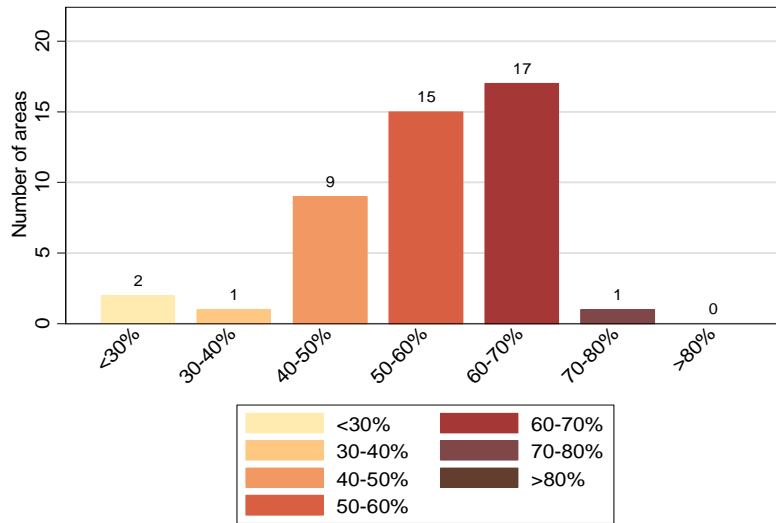


Figure 14: Distribution of support for 2008 Republican presidential candidate by area

Notes: The sample includes 39 large counties and 6 county groups. Each observation represents support for the 2008 Republican candidate in a large county or a county group. Support for the 2008 Republican candidate is defined as the proportion of voters who supported the 2008 Republican candidate in the general election. *Source:* North Carolina Board of Elections.

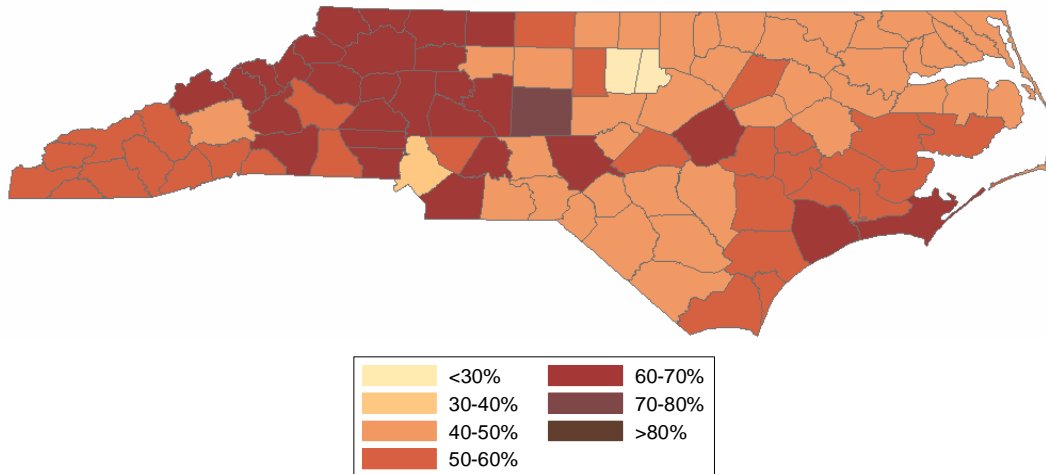


Figure 15: Support for 2008 Republican presidential candidate by area

Note: See Figure 11 for a definition of support for 2008 Republican candidate. *Source:* North Carolina Board of Elections.

two areas: Durham County (23 percent in 2012, 24 percent in 2008) and Orange County (28 percent in 2012, 27 percent in 2008). At the other extreme, the Republican candidate received more than 70 percent of the vote in at least one election in Randolph County (74 percent in 2012, 71 percent in 2008) and in Wilkes County (70 percent in 2012, 68 percent in 2008).

Again, support for the Republican candidates in the county groups falls in the middle of the distribution for both elections. The Republican candidate received between 45 percent and 65 percent of the 2012 vote in the county groups between 45 percent and 61 percent of the 2008 vote in the county groups. In both elections, the North Coastal county group exhibited the lowest support for the Republican candidate, while the North Mountain county group exhibited the highest support for the Republican candidate. Given that the range of support exhibited across the large counties is much wider than the range of support exhibited across the county groups, it seems unlikely that the county groups are biasing the subsequent results.

Over half of religious adherents in North Carolina are affiliated with an Evangelical denomination, with substantial variation across counties. Figures 16 and 17 illustrate the distribution of Evangelical adherents as a proportion of all adherents, and Figures 18 and 19 illustrate that distribution of Evangelical adherents as a proportion of non-Black Protestant adherents. Overall, the proportions of Evangelical adherents are lower than the proportions of Republican presidential candidate supporters. Using both

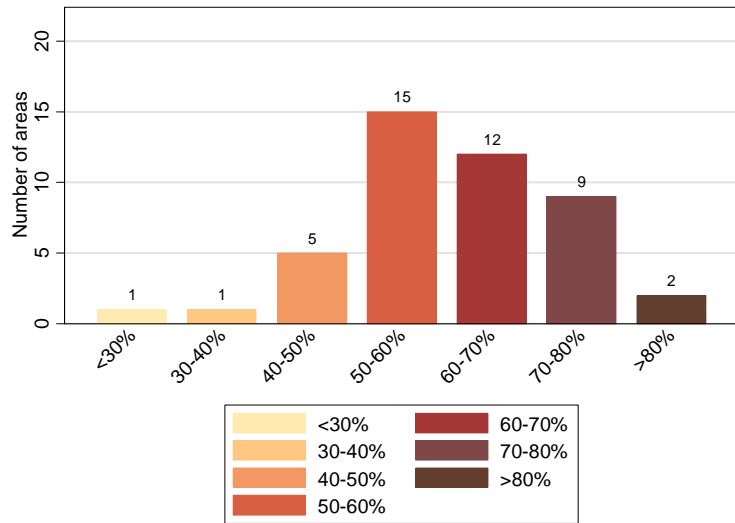


Figure 16: Distribution of Evangelical adherents as a proportion of all religious adherents by area

Notes: The sample includes 39 large counties and 6 county groups. Each observation represents Evangelical adherents as a proportion of all adherents in a large county or a county group. Evangelical adherents as a proportion of all religious adherents is defined as the proportion of all religious adherents who are classified as Evangelical Protestant in the U.S. Religious Census. *Source:* U.S. Religious Census.

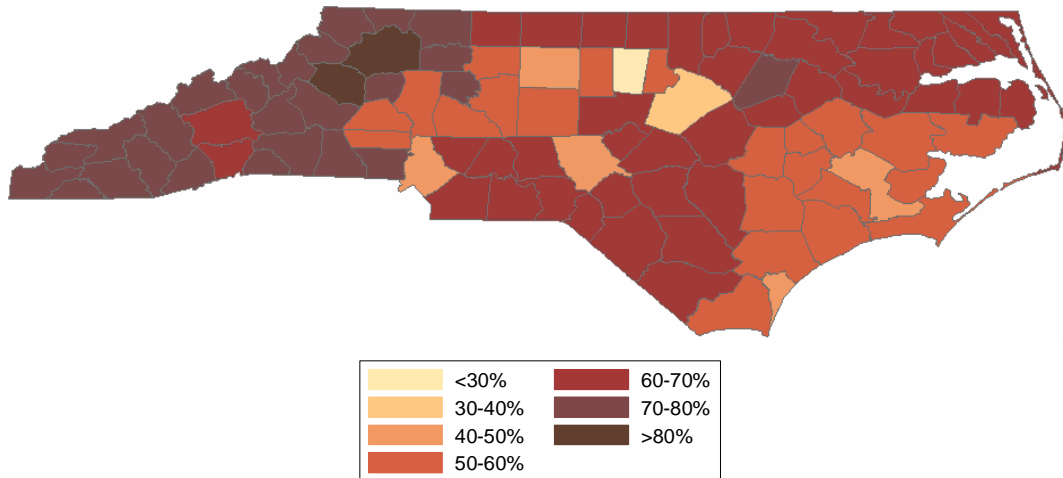


Figure 17: Evangelical adherents as a proportion of all religious adherents by area

Note: See Figure 13 for a definition of Evangelical adherents as a proportion of all religious adherents. *Source:* U.S. Religious Census.

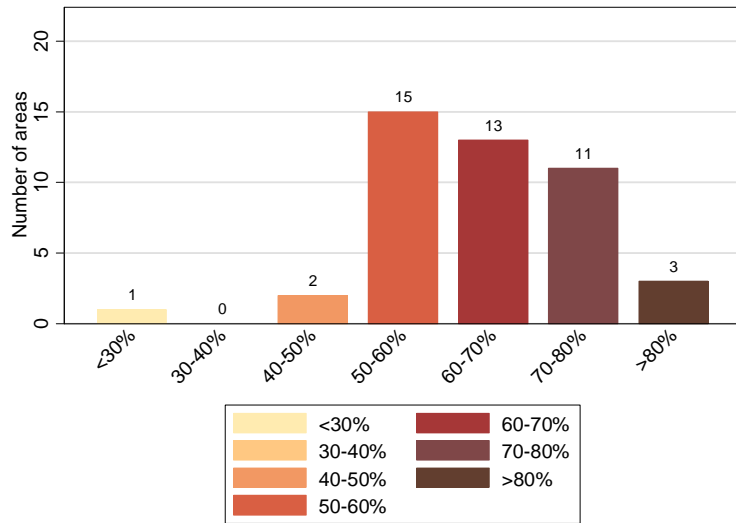


Figure 18: Distribution of Evangelical adherents as a proportion of non-Black Protestant religious adherents by area

Notes: The sample includes 39 large counties and 6 county groups. Each observation represents Evangelical adherents as a proportion of non-Black Protestant religious adherents in a large county or a county group. Evangelical adherents as a proportion of non-Black Protestant adherents is defined as the proportion of religious adherents except Black Protestants who are classified as Evangelical Protestant in the U.S. Religious Census. *Source:* U.S. Religious Census.

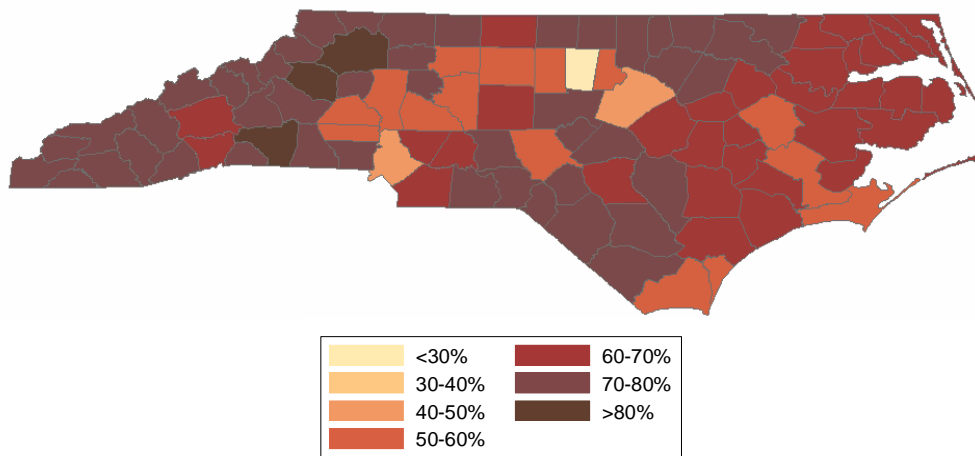


Figure 19: Evangelical adherents as a proportion of non-Black Protestant religious adherents by area

Note: See Figure 15 for a definition of Evangelical adherents as a proportion of non-Black Protestant religious adherents. *Source:* U.S. Religious Census.

calculations, Evangelical adherents represent 50-60 percent of adherents in a plurality of areas. At the high extreme, 85 percent of all adherents and 87 percent of non-Black Protestant adherents in Wilkes County are Evangelical. Caldwell and Rutherford Counties are also notable for having high proportions of Evangelical adherents: 81 percent and 78 percent of all adherents, 82 percent and 84 percent of all non-Black Protestant adherents. At the low extreme, only 28 percent of all adherents and 29 percent of non-Black Protestant adherents in Orange County are Evangelical.

The same counties tend to have extreme values for each the five proxy variables of traditional gender stereotypes. In Wilkes County, 83 percent of voters supported Amendment One, 70 percent of voters supported Mitt Romney, 68 percent of voters supported John McCain, 85 percent of all adherents are Evangelical, and 87 percent of all non-Black Protestant adherents are Evangelical. Randolph County was identified as having high support for Amendment One (80 percent) and high support for Republican presidential candidates (74 percent in 2012, 71 percent in 2008). Rutherford and Caldwell counties were identified as having high support for Amendment One and high Evangelical adherence. Meanwhile, Orange County has low values for each of the proxy variables: 21 percent support for Amendment One, 28 percent support for Mitt Romney, 27 percent support for John McCain, 28 percent Evangelical among all adherents, and 29 percent Evangelical among non-Black Protestant adherents.

Table 6 presents the correlations between the five proxy variables of traditional gender stereotypes. As expected, the correlations are positive, significant, and often large. The correlations between support for Amendment One and support for the Republican presidential candidates are 0.82 for the 2012 election and 0.85 for the 2008 election. Support for Amendment One is highly correlated with both Evangelicals as a proportion of all adherents (0.73) and Evangelicals as a proportion of non-Black Protestant adherents (0.74). Correlations between support for Republican presidential candidates and Evangelical adherence are statistically significant but much smaller, ranging from 0.44 to 0.52. The two measures of support for Republican presidential candidates and the two measures of Evangelical adherents are nearly perfectly correlated (0.996 between Republican presidential candidates and 0.984 between Evangelical adherents).

1.4 Regression results

Before turning to the regression results, Figures 20-24 present scatterplots of the correlation between the average of gender ratios in the top 5 percent and each of the five proxy variables for traditional gender stereotypes (hereafter referred to as simply gender stereotypes). The first scatterplot in each figure depicts the average of the gender ratios in the top 5 percent for 5th grade students, while the second scatterplot in each figure depicts the average of the gender ratios in the top 5 percent for 8th grade students. Each scatterplot is fit with a linear model and a 95 percent confidence interval. The figures

Table 6: Correlations between proxy variables for traditional gender stereotypes

	Support for Amendment One as a proportion of voters	Support for 2012 Republican candidate as a proportion of voters	Support for 2008 Republican candidate as a proportion of voters	Evangelical adherents as a proportion of all adherents	Evangelical adherents as a proportion of non-Black Protestant adherents
Support for Amendment One as a proportion of voters	1.000				
Support for 2012 Republican Presidential candidate as a proportion of voters	0.819 *** [N = 45]	1.000			
Support for 2008 Republican Presidential candidate as a proportion of voters	0.850 *** [N = 45]	0.996 *** [N = 45]	1.000		
72 Evangelical adherents as a proportion of all adherents	0.733 *** [N = 45]	0.486 *** [N = 45]	0.517 *** [N = 45]	1.000	
Evangelical adherents as a proportion of all adherents except Black Protestants	0.740 *** [N = 45]	0.436 *** [N = 45]	0.474 *** [N = 45]	0.984 *** [N = 45]	1.000

Notes: Entries represent pairwise correlations between each measure of political or religious beliefs. The sample for each entry includes 39 large counties and 6 county groups. Sample sizes are in brackets. * indicates significance at the 5 percent level; ** indicates significance at the 1 percent level; *** indicates significance at 0.1 percent level.

Sources: North Carolina Board of Elections; U.S. Religious Census.

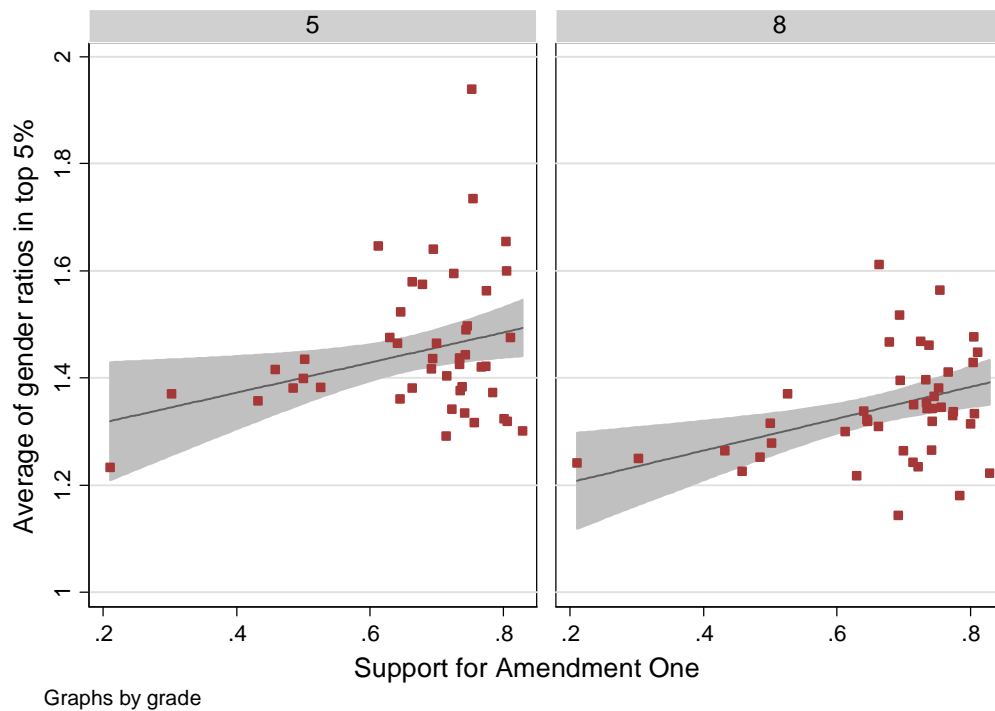


Figure 20: Scatterplots of the gender gap in the top 5 percent as a function of support for Amendment One

Notes: The sample includes 39 large counties and 6 county groups. Each observation represents support for the outcome measure in a large county or a county group. The average of gender ratios in the top 5 percent is defined as the average of the male-female ratio in the top 5 percent on the math test, the male-female ratio in the top 5 percent on the science test, and the female-male ratio in the top 5 percent on the reading test. See Figure 10 for a definition of support for Amendment One. Each graph is fit with a linear model and a 95 percent confidence interval. The relationships are weighted by the number of 5th grade students (for the first plot in each panel) or by the number of 8th grade students (for the second plot in each panel). *Sources:* North Carolina Education Research Data Center, End of Grade files; North Carolina Board of Elections.

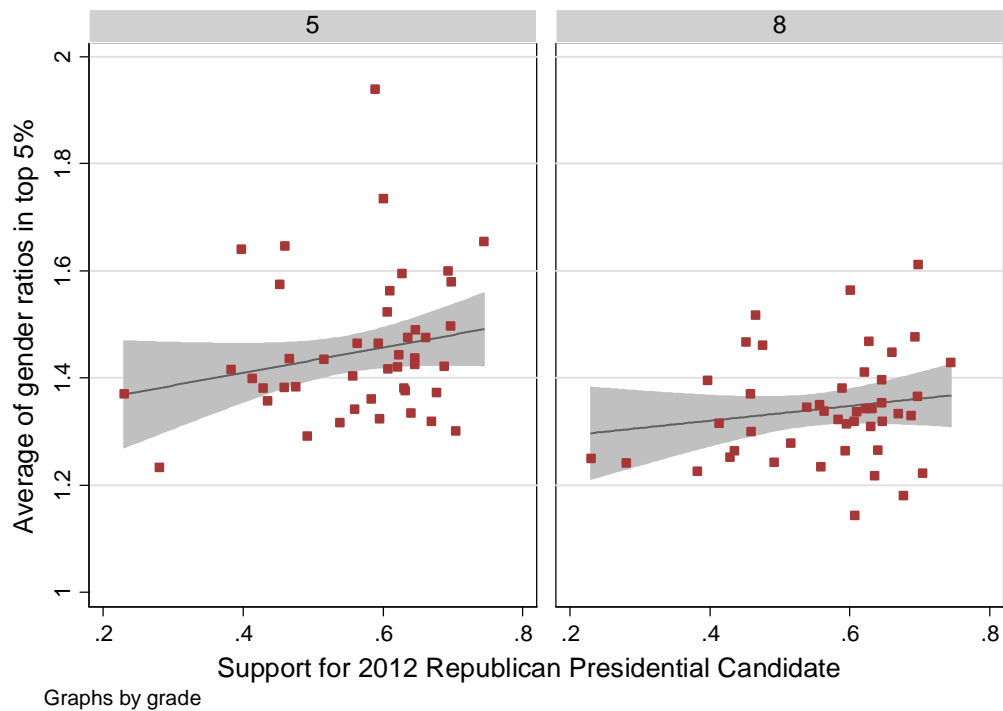


Figure 21: Scatterplots of the gender gap in the top 5 percent as a function of support for the 2012 Republican Presidential candidate

Notes: See Figure 20 for a description of the sample and for a definition of the average of gender ratios in the top 5 percent. See Figure 12 for a definition of support for the 2012 Republican candidate. Each graph is fit with a linear model and a 95 percent confidence interval. The relationships are weighted by the number of 5th grade students (for the first plot in each panel) or by the number of 8th grade students (for the second plot in each panel). *Sources:* North Carolina Education Research Data Center, End of Grade files; North Carolina Board of Elections.

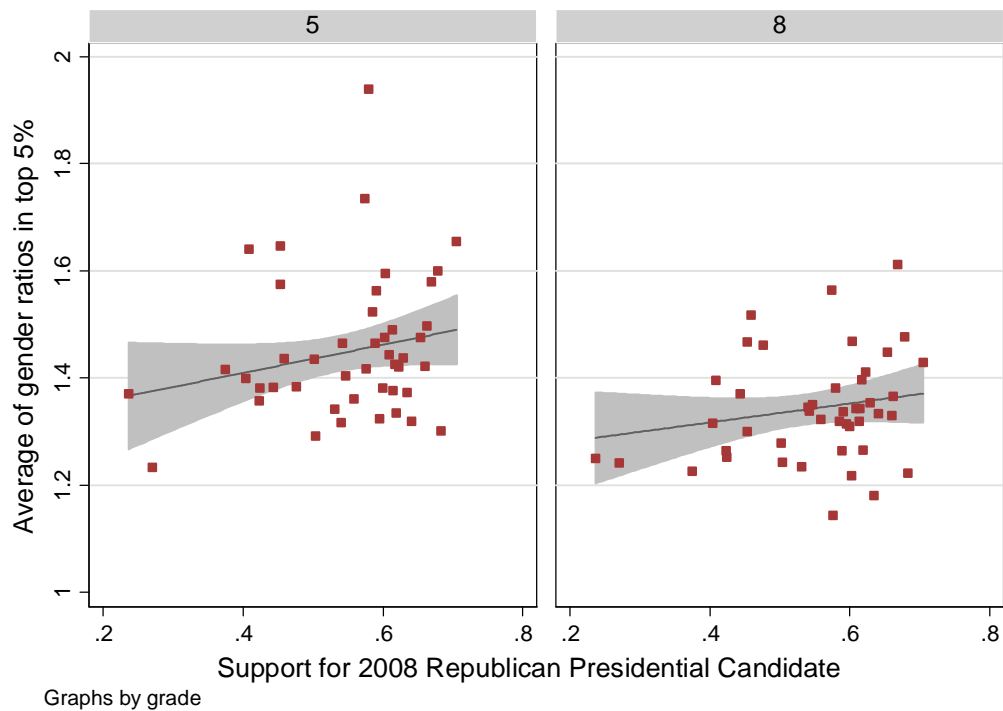


Figure 22: Scatterplots of the gender gap in the top 5 percent as a function of support for the 2008 Republican presidential candidate

Notes: See Figure 20 for a description of the sample and for a definition of the average of gender ratios in the top 5 percent. See Figure 14 for a definition of support for the 2008 Republican candidate. Each graph is fit with a linear model and a 95 percent confidence interval. The relationships are weighted by the number of 5th grade students (for the first plot in each panel) or by the number of 8th grade students (for the second plot in each panel). *Sources:* North Carolina Education Research Data Center, End of Grade files; North Carolina Board of Elections.

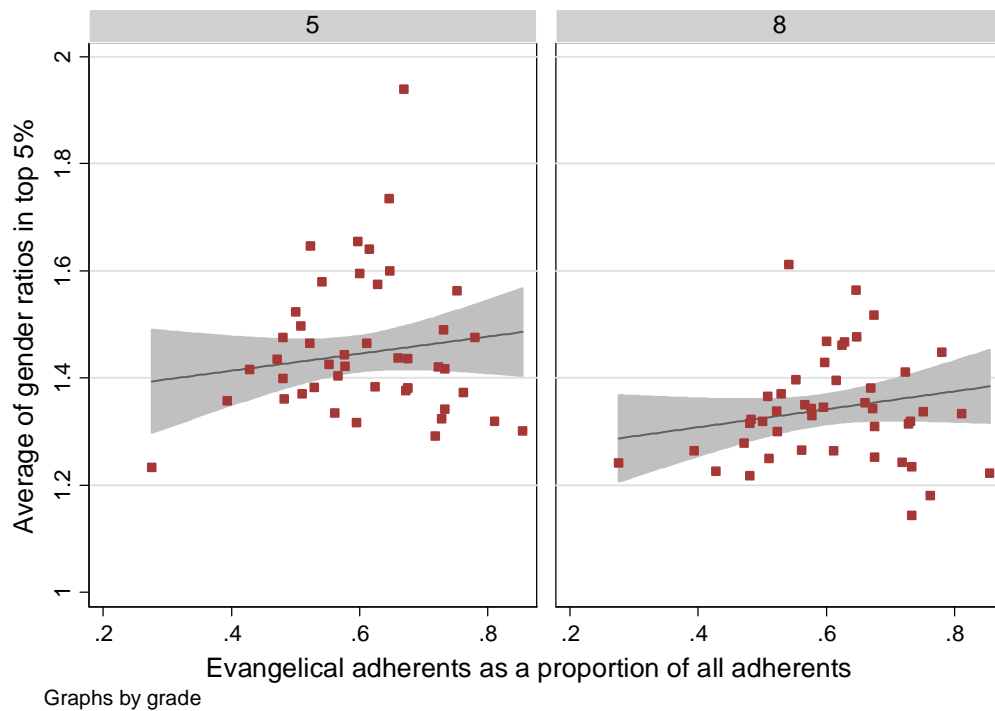


Figure 23: Scatterplots of the gender gap in the top 5 percent as a function of Evangelical adherents

Notes: See Figure 20 for a description of the sample and for a definition of the average of gender ratios in the top 5 percent. See Figure 16 for a definition of Evangelical adherents as a proportion of all adherents. Each graph is fit with a linear model and a 95 percent confidence interval. The relationships are weighted by the number of 5th grade students (for the first plot in each panel) or by the number of 8th grade students (for the second plot in each panel). *Sources:* North Carolina Education Research Data Center, End of Grade files; U.S. Religious Census.

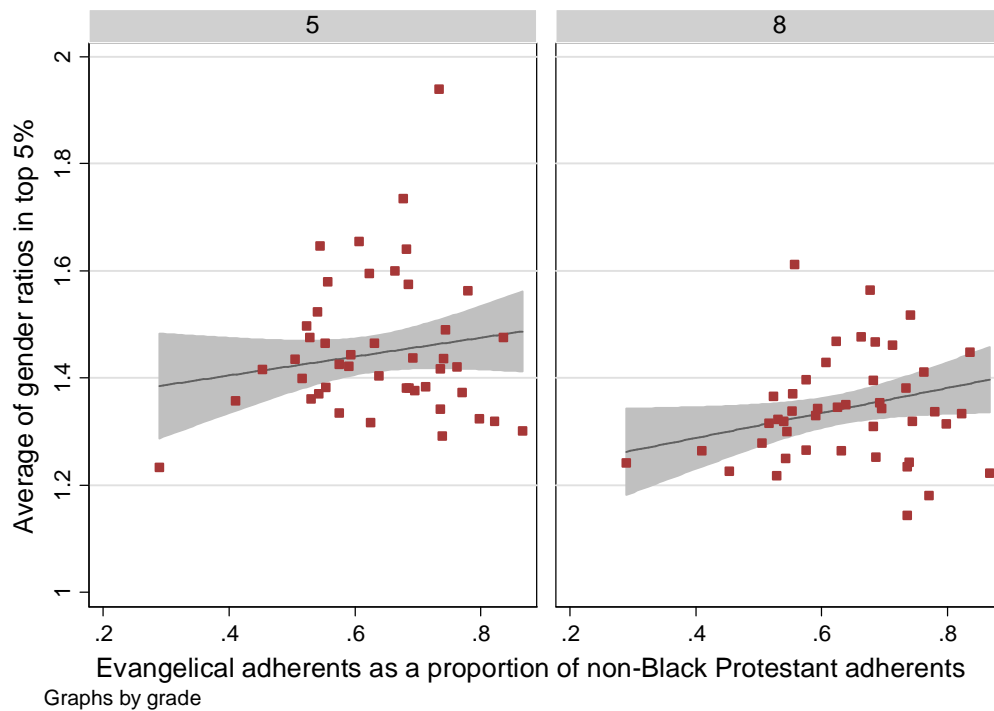


Figure 24: Scatterplots of the gender gap in the top 5 percent as a function of non-Black Protestant Evangelical adherents

Notes: See Figure 20 for a description of the sample and for a definition of the average of gender ratios in the top 5 percent. See Figure 18 for a definition of Evangelical adherents as a proportion of non-Black Protestant adherents. Each graph is fit with a linear model and a 95 percent confidence interval. The relationships are weighted by the number of 5th grade students (for the first plot in each panel) or by the number of 8th grade students (for the second plot in each panel). *Sources:* North Carolina Education Research Data Center, End of Grade files; U.S. Religious Census.

illustrate that each proxy variable is positively associated with the average of gender ratios in the top 5 percent for both grades. However, the slope appears the greatest in Figure 20, supporting the hypothesis that support for Amendment One will be the best proxy variable for gender stereotypes.

The scatterplots suggest a positive relationship between the average of gender ratios in the top 5 percent and proxy variables for gender stereotypes. Table 7 presents ordinary least squares regression estimates of these relationships, controlling for the number of students and the median household income in the area.²³ For both grades, the coefficient on support for Amendment One is positive at the 5 percent level, suggesting that students in the top 5 percent are more likely to be the “correct” gender in areas where a higher proportion of voters supported Amendment One. Controlling for the number of students and the median household income, a 10 percentage point increase in support for Amendment One is associated with an increase of 0.025 in the 5th grade gender gap and 0.023 in the 8th grade gender gap, a 19 percent standard deviation increase in 5th grade and a 23 percent standard deviation increase in 8th grade. Reframing the relationship, a one standard deviation increase in support for Amendment One corresponds to a 26 percent standard deviation increase in the 5th grade gender gap and a 31 percent standard deviation increase in the 8th grade gap.

²³ See Section 1.2.2 for a discussion of these control variables. Appendix F provides regression estimates controlling for the number of students only.

Table 7: Estimated relationship between the average of gender ratios in the top 5 percent and proxy variables for gender stereotypes

	Average of gender ratios in the top 5% across tests				
	(1)	(2)	(3)	(4)	(5)
Panel A. 5th grade					
Support for Amendment One as a proportion of voters	0.253*				
	(0.118)				
Support for 2012 Republican Presidential candidate as a proportion of voters		0.200			
		(0.158)			
Support for 2008 Republican Presidential candidate as a proportion of voters			0.223		
			(0.166)		
Evangelical adherents as a proportion of all adherents				-0.064	
				(0.200)	
Evangelical adherents as a proportion of all adherents except Black Protestants					-0.006
					(0.212)
R ²	0.072	0.051	0.053	0.027	0.024
Observations	45	45	45	45	45
Control for 5th grade enrollment?	Yes	Yes	Yes	Yes	Yes
Control for median household income?	Yes	Yes	Yes	Yes	Yes
Panel B. 8th grade					
Support for Amendment One as a proportion of voters	0.220***				
	(0.059)				
Support for 2012 Republican Presidential candidate as a proportion of voters		0.165			
		(0.110)			
Support for 2008 Republican Presidential candidate as a proportion of voters			0.203		
			(0.113)		
Evangelical adherents as a proportion of all adherents				-0.060	
				(0.138)	
Evangelical adherents as a proportion of all adherents except Black Protestants					0.009
					(0.126)
R ²	0.118	0.081	0.090	0.052	0.048
Observations	45	45	45	45	45
Control for 8th grade enrollment?	Yes	Yes	Yes	Yes	Yes
Control for median household income?	Yes	Yes	Yes	Yes	Yes

Notes: Each column represents results from a separate regression. The average of gender ratios in the top 5% across tests is defined as the average of the male-female ratio in the top 5% on the math test, the male-female ratio in the top 5% on the science test, and the female-male ratio in the top 5% on the reading test. The sample for each entry includes 39 large counties and 6 county groups. Standard errors are in parentheses. * indicates significance at the 5 percent level; ** indicates significance at the 1 percent level; *** indicates significance at 0.1 percent level.

Sources: North Carolina Education Research Data Center, End of Grade files; North Carolina Board of Elections; U.S. Religious Census.

In contrast to the significant coefficients for Amendment One, the coefficients for every other proxy variable of gender stereotypes are insignificant in both grades. The coefficients on the proxy variables measuring support for Republican presidential candidates are roughly the same magnitude as the coefficients on the proxy variable measuring support for Amendment One; a 10 percentage point increase in support for John McCain is associated with a 17 percent standard deviation increase in the 5th grade gender gap and a 20 percent standard deviation increase in the 8th grade gender gap. The relationships, however, are statistically insignificant. Interestingly, the positive relationship between gender gaps and Evangelical adherence depicted in the scatterplots becomes negative after controlling for median household income. A 10 percentage point increase in the proportion of all adherents who are Evangelical is associated with a 5 percent standard deviation decrease in the 5th grade gender gap and a 6 percent standard deviation decrease in the 8th grade gender gap. The relationship is even smaller when Evangelical adherents are measured as a proportion of non-Black Protestant adherents; a 10 percentage point increase in Evangelical adherents is associated with a 0.6 percent standard deviation increase in the 5th grade gender gap and a 0.9 percent standard deviation increase in the 8th grade gender gap. Again, none of these coefficients is statistically significant.²⁴

²⁴ The proxy variables are too highly correlated to include more than one of them in the same regression. However, to get a sense of whether support for Amendment One remains significant even after controlling for support for Republican Presidential candidates, I calculated the ratio between support for Amendment

Table 7 illustrates that greater gender gaps among top-scoring students on North Carolina state assessments are associated with greater support for Amendment One but not with greater support for Republican presidential candidates or greater adherence to Evangelical religious denominations. A one standard deviation increase in support for Amendment One corresponds to a 26 percent standard deviation increase in the 5th grade gender gap and a 31 percent standard deviation increase in the 8th grade gender gap. To put these percentages in perspective, it is helpful to consider not only how differences in gender stereotypes relate to differences in gender gaps, but also how differences in gender stereotypes relate to differences in the numbers of male and female students who score in the top 5 percent. For example, suppose that voters in Randolph County supported Amendment One by the same margin as voters in Orange County.²⁵ The coefficients in Table 5 suggest that gender gaps in Randolph County would be 1.50 in 5th grade and 1.29 in 8th grade. If Randolph County had a 5th grade gender gap of

One and support for the Republican Presidential candidate in 2012. Then, I regressed gender gaps on support for Amendment One and this ratio simultaneously. I found that the coefficient on support for Amendment One decreased slightly from 0.253 to 0.231 in the results for 5th grade and from 0.220 to 0.212 in the results for 8th grade. The coefficient on support for Amendment One retained significance at the 1 percent level in the results for 8th grade but lost significance in the results for 5th grade. Robust results for 8th grade and inconsistent results for 5th grade are also reported in Section 1.5.2 using alternate measures of gender gaps (Table 10, page 87).

²⁵ 81 percent of voters supported Amendment One in Randolph County, compared to 21 percent of voters in Orange County. I frame the relationship between gender gaps and support for Amendment One in terms of how we would expect gender gaps to change if we witnessed a change in support for Amendment One. However, I do not claim to have identified a causal relationship. I elaborate on this caveat in the following discussion section.

1.50 instead of 1.65, it could mean that 15 more girls would score in the top 5 percent on the 5th grade science test, or 66 more girls would score in the top 5 percent on the 5th grade math test, or 83 more boys would score in the top 5 percent on the 5th grade reading test. Alternatively, if the change in the gender gap was due to changes in all three gender ratios, it could mean that five more girls would score in the top 5 percent on the 5th grade science test *and* 22 more girls would score in the top 5 percent on the 5th grade math test *and* 10 more boys would score in the top 5 percent on the reading test. Further, Randolph County could boast a below-average gender gap in 8th grade rather than a gender gap in the top 20 percent. The relationship between gender gaps and gender stereotypes appears even larger when we consider that small changes in gender gaps conceal larger changes in the number of students who score in the top 5 percent.

1.5 Robustness checks

1.5.1 Alternate specifications using large counties only

As discussed in Section 1.2, small counties lack sufficient observations from top-scoring students to include in the analyses individually and are therefore combined into six county groups. Although the descriptive statistics in Table 5 do not suggest that the average county group is any different from the average large county, it is possible that the grouping of the small counties into county groups biases the results. To account for this concern, Table 8 presents the same results as Table 7 but restricts the sample to large

Table 8: Estimated relationship between the average of gender ratios in the top 5 percent and proxy variables for gender stereotypes for large counties only

	Average of gender ratios in the top 5% across tests				
	(1)	(2)	(3)	(4)	(5)
Panel A. 5th grade					
Support for Amendment One as a proportion of voters	0.256*				
	(0.115)				
Support for 2012 Republican Presidential candidate as a proportion of voters		0.220			
		(0.166)			
Support for 2008 Republican Presidential candidate as a proportion of voters			0.241		
			(0.171)		
Evangelical adherents as a proportion of all adherents				-0.081	
				(0.214)	
Evangelical adherents as a proportion of all adherents except Black Protestants					-0.010
					(0.223)
R ²	0.078	0.058	0.060	0.031	0.028
Observations	39	39	39	39	39
Control for 5th grade enrollment?	Yes	Yes	Yes	Yes	Yes
Control for median household income?	Yes	Yes	Yes	Yes	Yes
Panel B. 8th grade					
Support for Amendment One as a proportion of voters	0.232***				
	(0.051)				
Support for 2012 Republican Presidential candidate as a proportion of voters		0.280*			
		(0.104)			
Support for 2008 Republican Presidential candidate as a proportion of voters			0.313**		
			(0.107)		
Evangelical adherents as a proportion of all adherents				-0.012	
				(0.133)	
Evangelical adherents as a proportion of all adherents except Black Protestants					0.012
					(0.127)
R ²	0.131	0.150	0.158	0.039	0.039
Observations	39	39	39	39	39
Control for 8th grade enrollment?	Yes	Yes	Yes	Yes	Yes
Control for median household income?	Yes	Yes	Yes	Yes	Yes

Notes: Each column represents results from a separate regression. The average of gender ratios in the top 5% across tests is defined as the average of the male-female ratio in the top 5% on the math test, the male-female ratio in the top 5% on the science test, and the female-male ratio in the top 5% on the reading test. The sample for each entry includes 39 large counties and 6 county groups. Standard errors are in parentheses. * indicates significance at the 5 percent level; ** indicates significance at the 1 percent level; *** indicates significance at 0.1 percent level.

Sources: North Carolina Education Research Data Center, End of Grade files; North Carolina Board of Elections; U.S. Religious Census.

counties. The coefficients on support for Amendment One remain significant, positive, and similar in magnitude; a 10 percentage point increase in support is associated with exactly the same 19 percent standard deviation increase in the 5th grade gender gap and 23 percent standard deviation increase in the 8th grade gender gap. The only notable difference between Tables 7 and 8 is that the coefficients on support for the Republican presidential candidates become significant for 8th grade students. A 10 percentage point increase in support for the Republican presidential candidate in 2012 and 2008 is statistically significantly associated with a 28 percent and 32 percent standard deviation increase in the gender gap among 8th grade students when restricting the sample to large counties only.

Table 8 illustrates that the relationship between the gender gap and support for Amendment One is robust, suggesting that fewer females score in the top 5 percent on math and science test and fewer males score in the top 5 percent on reading tests in areas where individuals hold stronger gender stereotypes. Table 8 further supports the hypothesis that support for Amendment One is generally a stronger proxy variable for gender stereotypes than support for Republican presidential candidates or adherence to Evangelical religious denominations. This section also provides strong evidence that the sorting of small counties into county groups is not biasing the results.

1.5.2 Alternate specifications using different calculations for the gender gap

Pope and Sydnor (2010) is the first — and until now the only — study that combines gender ratios across subjects into a single measure. The analyses presented previously use the same calculation of gender gaps to enable comparisons between the two studies. However, it is possible that the results are sensitive to the definition of gender gaps. To test whether the relationship between gender gaps and support for Amendment One persists regardless of how gender gaps are defined, I compare three measures of gender gaps: (1) the arithmetic average of the gender ratios (as presented throughout the text), (2) the weighted average of the gender ratios where the male-female ratio in math is weighted at 25 percent, the male-female ratio in science is weighted at 25 percent, and the female-male ratio in reading is weighted at 50 percent, and (3) the geometric mean of the gender ratios. If the results remain substantively the same across these three measures, then it will suggest that the findings are robust to alternate definitions of gender gaps.

Statewide, these definitions produce similar estimates of gender gaps. Table 4 illustrates that estimates of the 5th grade gender gap are 1.41 using the arithmetic average of ratios, 1.34 using the weighted average of ratios, and 1.38 using the geometric mean of ratios. The estimates are even closer for 8th grade: 1.31 using the arithmetic average of ratios, 1.28 using the weighted average of ratios, and 1.29 using the geometric

mean of ratios. Table 9 demonstrates that the correlations across these measures is very high, ranging from 0.90 to 0.96.

Table 10 presents the results of regressions using these alternate measures of gender gaps. Each entry represents the results from a separate regression of the dependent variable reported at the top of the column on the independent variable reported to the left of the row. Each regression controls for the number of students and the median household income in the area. Three patterns are evident. First, the finding that support for Amendment One is correlated with the average of gender ratios in 5th grade is not robust to alternate measures. In fact, this definition of gender gaps, designed by Pope and Sydnor (2010), is the only definition which produces a statistically significant result. On the other hand, the finding that support for Amendment One is correlated with the average of gender ratios in 8th grade is robust across specifications. The correlation is significant at the 1 percent level for each measure of gender gaps, and the magnitude of the relationship is similar across measures. Furthermore, support for Republican presidential candidates and adherence to Evangelical religious denominations remain insignificant across specifications. In both grades, all four proxy variables of gender stereotypes are statistically unrelated to all three measures of gender gaps.

This section confirms that support for Amendment One is robustly related to the 8th grade gender gap, regardless of how the gender gap is calculated. This section also

Table 9: Correlations between measures of gender gaps

	Average of ratios, equal weighting	Average of ratios, unequal weighting	Geometric mean of ratios, equal weighting
Average of ratios, equal weighting	1.000		
Average of ratios, 25% math, 25% science, 50% reading	0.903 *** [N = 45]	1.000	
Geometric mean of ratios, equal weighting	0.960 *** [N = 45]	0.935 *** [N = 45]	1.000

Notes: Entries represent pairwise correlations between each measure of gender gaps. Each measure of the gender gap is calculated using the male-female ratio in math, the male-female ratio in science, and the female-male ratio in reading. The ratios are aggregated into a single measure as specified in each row or column. The sample for each entry includes 39 large counties and 6 county groups. Sample sizes are in brackets. * indicates significance at the 5 percent level; ** indicates significance at the 1 percent level; *** indicates significance at 0.1 percent level.

Source: North Carolina Education Research Data Center, End of Grade files.

Table 10: Estimated relationship between measures of gender gaps in the top 5 percent and proxy variables for gender stereotypes

Dependent variables:	(1) Arithmetic average of ratios	(2) Weighted average of ratios	(3) Geometric mean of ratios
Panel A. 5th grade			
Support for Amendment One as a proportion of voters	0.253* (0.118) [N = 45]	0.142 (0.102) [N = 45]	0.132 (0.105) [N = 45]
Support for 2012 Republican Presidential candidate as a proportion of voters	0.200 (0.158) [N = 45]	0.153 (0.121) [N = 45]	0.137 (0.132) [N = 45]
Support for 2008 Republican Presidential candidate as a proportion of voters	0.223 (0.166) [N = 45]	0.153 (0.129) [N = 45]	0.142 (0.139) [N = 45]
Evangelical adherents as a proportion of all adherents	-0.064 (0.200) [N = 45]	-0.074 (0.148) [N = 45]	-0.143 (0.164) [N = 45]
Evangelical adherents as a proportion of all adherents except Black Protestants	-0.006 (0.212) [N = 45]	-0.054 (0.161) [N = 45]	-0.094 (0.177) [N = 45]
Panel B. 8th grade			
Support for Amendment One as a proportion of voters	0.230*** (0.059) [N = 45]	0.210*** (0.052) [N = 45]	0.163** (0.053) [N = 45]
Support for 2012 Republican Presidential candidate as a proportion of voters	0.165 (0.110) [N = 45]	0.134 (0.091) [N = 45]	0.100 (0.102) [N = 45]
Support for 2008 Republican Presidential candidate as a proportion of voters	0.203 (0.113) [N = 45]	0.162 (0.095) [N = 45]	0.126 (0.105) [N = 45]
Evangelical adherents as a proportion of all adherents	-0.060 (0.138) [N = 45]	0.016 (0.121) [N = 45]	-0.102 (0.124) [N = 45]
Evangelical adherents as a proportion of all adherents except Black Protestants	0.009 (0.126) [N = 45]	0.082 (0.114) [N = 45]	-0.039 (0.115) [N = 45]

Notes: Each entry represents results from a separate regression. Each measure of gender gaps is calculated using the male-female ratio in math, the male-female ratio in science, and the female-male ratio in reading. The ratios are aggregated into a single measure as specified at the top of each column. Each regression controls for the median household income and the number of students in each area. The sample for each entry includes 39 large counties and 6 county groups. Standard errors are in parentheses. Sample sizes are in brackets. * indicates significance at the 5 percent level; ** indicates significance at the 1 percent level; *** indicates significance at 0.1 percent level.

Sources: North Carolina Education Research Data Center, End of Grade files; North Carolina Board of Elections; U.S. Religious Census.

confirms that other proxy variables are statistically unrelated to gender gaps in either grade. However, given the differences in results for 5th grade students, this table should caution other researchers from restricting future studies to one definition of gender gaps.

1.5.3 Alternate specifications using gender ratios on individual subject tests

The prior analyses combined gender ratios on the math, science, and reading tests for a number of reasons. A primary reason is that combining gender ratios is consistent with the hypothesis guiding this paper: that gender stereotypes affect the educational performance of both boys and girls. If I hypothesized that gender stereotypes — such as associations between women and home production — discourage girls from scoring at the top of the distribution on all subjects, then I would test this hypothesis by analyzing the male-female gender ratio in the top 5 percent on each test separately. However, I do not hypothesize simply that gender stereotypes discourage girls from scoring at the top of the distribution; I hypothesize that gender stereotypes discourage girls from scoring at the top of the distribution on math and science tests only. As Table 4 and Figure 3 illustrate, females are actually overrepresented at the top of the distribution on reading tests. I further hypothesize that gender stereotypes also discourage boys from scoring at the top of the distribution on reading tests. Combining gender ratios across subject tests ensures that the outcome measure reflects not only the effects of gender stereotypes on girls but also the effects of gender stereotypes on boys.

Some readers may, however, be interested in how proxy variables of gender stereotypes relate to gender ratios for each subject. Tables 11-13 report these results. Table 11 examines the male-female ratio in the top 5 percent on the math test; Table 12 examines the male-female ratio in the top 5 percent on the science test; and Table 13 examines the female-male ratio in the top 5 percent on the reading test. Each column reports the results of an ordinary least squares regression of the gender ratio in the top 5 percent on each proxy variable of gender stereotypes, controlling for the number of students and the median household income in an area.

Tables 11-13 illustrate that the relationship between gender gaps and support for same-sex marriage may be largely attributable to the gender ratio in the top 5 percent on the science test. Table 12 estimates that a 10 percentage point increase in support for Amendment One is associated with a 58 percent standard deviation increase in the male-female ratio in the top 5 percent on the 5th grade science test and a 66 percent standard deviation increase in the male-female ratio in the top 5 percent on the 8th grade science test. Reframing the relationship, a one standard deviation increase in support for Amendment One corresponds to a 78 percent standard deviation increase in the 5th grade science gender ratio and 88 percent standard deviation increase in the 8th grade science gender ratio.

Although Table 10 suggests that gender ratios on the science test may explain the results in Table 5, readers should be cautious emphasizing the results in this table.

Table 11: Estimated relationship between the male-female ratio in the top 5 percent on the math test and proxy variables for gender stereotypes

		Male-female ratio in the top 5% on the math test				
		(1)	(2)	(3)	(4)	(5)
Panel A. 5th grade						
Support for Amendment One as a proportion of voters	0.183 (0.153)					
Support for 2012 Republican Presidential candidate as a proportion of voters		0.133 (0.185)				
Support for 2008 Republican Presidential candidate as a proportion of voters			0.161 (0.186)			
Evangelical adherents as a proportion of all adherents				-0.226 (0.260)		
Evangelical adherents as a proportion of all adherents except Black Protestants					-0.070 (0.275)	
R ²	0.028	0.022	0.023	0.030	0.017	
Observations	45	45	45	45	45	
Control for 5th grade enrollment?	Yes	Yes	Yes	Yes	Yes	
Control for median household income?	Yes	Yes	Yes	Yes	Yes	
Panel B. 8th grade						
Support for Amendment One as a proportion of voters	-0.112 (0.115)					
Support for 2012 Republican Presidential candidate as a proportion of voters		-0.109 (0.195)				
Support for 2008 Republican Presidential candidate as a proportion of voters			-0.112 (0.201)			
Evangelical adherents as a proportion of all adherents				-0.488** (0.174)		
Evangelical adherents as a proportion of all adherents except Black Protestants					-0.457* (0.179)	
R ²	0.044	0.043	0.042	0.121	0.110	
Observations	45	45	45	45	45	
Control for 8th grade enrollment?	Yes	Yes	Yes	Yes	Yes	
Control for median household income?	Yes	Yes	Yes	Yes	Yes	

Notes: Each column represents results from a separate regression. The sample for each entry includes 39 large counties and 6 county groups. Standard errors are in parentheses. * indicates significance at the 5 percent level; ** indicates significance at the 1 percent level; *** indicates significance at 0.1 percent level.

Sources: North Carolina Education Research Data Center, End of Grade files; North Carolina Board of Elections; U.S. Religious Census

Table 12: Estimated relationship between the male-female ratio in the top 5 percent on the science test and proxy variables for gender stereotypes

		Male-female ratio in the top 5% on the science test				
		(1)	(2)	(3)	(4)	(5)
Panel A. 5th grade						
Support for Amendment One as a proportion of voters	0.768* (0.330)					
Support for 2012 Republican Presidential candidate as a proportion of voters		0.454 (0.428)				
Support for 2008 Republican Presidential candidate as a proportion of voters			0.562 (0.466)			
Evangelical adherents as a proportion of all adherents				0.135 (0.529)		
Evangelical adherents as a proportion of all adherents except Black Protestants					0.255 (0.516)	
R ²	0.122	0.074	0.081	0.054	0.058	
Observations	45	45	45	45	45	
Control for 5th grade enrollment?	Yes	Yes	Yes	Yes	Yes	
Control for median household income?	Yes	Yes	Yes	Yes	Yes	
Panel B. 8th grade						
Support for Amendment One as a proportion of voters	0.652*** (0.171)					
Support for 2012 Republican Presidential candidate as a proportion of voters		0.564* (0.247)				
Support for 2008 Republican Presidential candidate as a proportion of voters			0.681* (0.267)			
Evangelical adherents as a proportion of all adherents				0.065 (0.305)		
Evangelical adherents as a proportion of all adherents except Black Protestants					0.187 (0.284)	
R ²	0.151	0.115	0.133	0.041	0.047	
Observations	45	45	45	45	45	
Control for 8th grade enrollment?	Yes	Yes	Yes	Yes	Yes	
Control for median household income?	Yes	Yes	Yes	Yes	Yes	

Notes: Each column represents results from a separate regression. The sample for each entry includes 39 large counties and 6 county groups. Standard errors are in parentheses. * indicates significance at the 5 percent level; ** indicates significance at the 1 percent level; *** indicates significance at 0.1 percent level.

Sources: North Carolina Education Research Data Center, End of Grade files; North Carolina Board of Elections; U.S. Religious Census

Table 13: Estimated relationship between the female-male ratio in the top 5 percent on the reading test and proxy variables for gender stereotypes

		Female-male ratio in the top 5% on the reading test				
		(1)	(2)	(3)	(4)	(5)
Panel A. 5th grade						
Support for Amendment One as a proportion of voters	-0.192 (0.224)					
Support for 2012 Republican Presidential candidate as a proportion of voters		0.014 (0.214)				
Support for 2008 Republican Presidential candidate as a proportion of voters			-0.055 (0.250)			
Evangelical adherents as a proportion of all adherents				-0.102 (0.267)		
Evangelical adherents as a proportion of all adherents except Black Protestants					-0.201 (0.245)	
R ²	0.014	0.003	0.003	0.005	0.012	
Observations	45	45	45	45	45	
Control for 5th grade enrollment?	Yes	Yes	Yes	Yes	Yes	
Control for median household income?	Yes	Yes	Yes	Yes	Yes	
Panel B. 8th grade						
Support for Amendment One as a proportion of voters	0.150 (0.111)					
Support for 2012 Republican Presidential candidate as a proportion of voters		0.041 (0.154)				
Support for 2008 Republican Presidential candidate as a proportion of voters			0.040 (0.172)			
Evangelical adherents as a proportion of all adherents				0.243 (0.165)		
Evangelical adherents as a proportion of all adherents except Black Protestants					0.299 (0.178)	
R ²	0.067	0.050	0.050	0.083	0.099	
Observations	45	45	45	45	45	
Control for 8th grade enrollment?	Yes	Yes	Yes	Yes	Yes	
Control for median household income?	Yes	Yes	Yes	Yes	Yes	

Notes: Each column represents results from a separate regression. The sample for each entry includes 39 large counties and 6 county groups. Standard errors are in parentheses. * indicates significance at the 5 percent level; ** indicates significance at the 1 percent level; *** indicates significance at 0.1 percent level.

Sources: North Carolina Education Research Data Center, End of Grade files; North Carolina Board of Elections; U.S. Religious Census

North Carolina has been administering End of Grade tests in math and reading since 1993, but the state just began administering an End of Grade test in science in 2008. The results in Table 10, therefore, are based on scores from the first four administrations of this test. Because the North Carolina Department of Public Instruction has had fewer opportunities to refine and revise the test, it is plausible that the science test contains more gender-biased questions than the math or reading tests. Nonetheless, examining why gender differences among top-scoring students are greater on the science test than on the math or reading tests is an important area for future research.

Table 11 also reports that the proportion of all adherents and the proportion of non-Black Protestant adherents who are Evangelical are negatively correlated with the male-female ratio in the top 5 percent on the 8th grade math test. Contrary to my hypothesis, these coefficients suggest that more females score in the top 5 percent on the 8th grade math test in areas where a higher proportion of religious adherents are Evangelical. However, the coefficients on Evangelical adherence are statistically insignificant for the 5th grade math test and for both grades in all other subjects. With 30 coefficients reported in these three tables, I cannot eliminate the possibility that these coefficients are statistically significant due to chance alone.

1.6 Discussion

It is important to emphasize three points regarding this study. First, this study examines gender gaps among top-scoring students. Recent studies find that, at the

average, girls excel relative to boys across a wide variety of educational performance and attainment measures (see Ceci, Williams, and Barnett 2009 for a recent review). These findings prompted the *Atlantic Monthly* to proclaim that “girls rule” on their May 2000 cover (Sommers 2000). However, there is overwhelming evidence that females remain underrepresented in top math and science test scores (Arden and Plomin 2006; Deary et al. 2003; Hedges and Nowell 1995; Hyde et al. 2008; Leahey and Guo 2001; Lohman and Lakin 2009; Pope and Sydnor 2010; Strand et al. 2006), in science and engineering doctoral degree recipients (National Science Foundation 2011, Table 3), and in science and engineering faculty appointments (National Science Foundation 2008, Table 5). This paper focuses exclusively on patterns among top-scoring students, who are the most likely to pursue doctoral degrees and faculty careers.²⁶ Therefore, the results should not be generalized to students at other points of the test score distribution.

Second, this study considers both girls’ underrepresentation at the top in math and science as well as boys’ underrepresentation at the top in reading. These findings, therefore, have implications not only for educating girls in math and science but also for educating boys in reading. Math and science teachers in areas with strong traditional gender stereotypes could provide greater encouragement to girls, and reading teachers

²⁶ Ideally, this paper would examine gender gaps among students who score in the top 1 percent, but with only four years of observations for the science test, the results would be based on very small sample sizes.

in areas with strong traditional gender stereotypes could provide greater encouragement to boys. Principals in areas with strong traditional gender stereotypes could hire more women as math and science teachers and more men as reading teachers to change gender norms in their communities. This paper should not be summarized as a study exploring why girls are underperforming in math and science but rather as a study identifying a relationship between traditional gender stereotypes and gender gaps.

Third, this paper does not identify a causal relationship between traditional gender stereotypes and gender gaps. This study finds robust evidence of a statistically significant relationship between gender gaps among top-scoring students and support for the constitutional amendment to prohibit same-sex marriage. However, it cannot eliminate the possibility that a variable unrelated to traditional gender stereotypes accounts for this association.

1.7 Conclusions

Although women have made admirable progress towards gender equality over the past fifty years, they remain heavily underrepresented in math and science fields. Women earned 52 percent of doctoral degrees in 2010 yet represent less than 10 percent of full professors in engineering, mathematics, and physical science departments (U.S. Department of Education 2012, Table 310; National Science Foundation 2008, Table 5). Previous research documents that females have closed gender gaps in math and science

test scores at the average but that they remain underrepresented at the top of the distribution (Arden and Plomin 2006; Ceci, Williams, and Barnett 2009; Deary et al. 2003; Hedges and Nowell 1995; Hyde et al. 2008; Leahey and Guo 2001; Lohman and Lakin 2009; Pope and Sydnor 2010; Strand et al. 2006).

A large body of research examines potential explanations for the lingering gender gaps. A hypothesis dating back to 1894 argues that differences in ability — or more precisely, differences in the variances of ability — contribute to gender gaps at the top of the distribution. This explanation received national attention in 2005 after Harvard University President Lawrence Summers presented it as a possible explanation for the underrepresentation of women in academia. However, a growing body of research documents that gender gaps vary over time (Monastersky 2005; Wai et al. 2010) and differ across countries (Feingold 1994; Guiso et al. 2008; Kane and Mertz 2010; Nosek et al. 2008). These studies present an alternative hypothesis: that differences in socialization explain at least some of the gender gaps at the top of the distribution. Recent studies find statistically significant relationships between gender gaps and implicit associations between men and science (Nosek et al. 2008), while other studies find statistically significant relationships between gender gaps and explicit agreement with statements such as “math is for boys” (Pope and Sydnor 2010).

Based on this emerging area of research, the current study hypothesizes that greater gender gaps at the top of the distribution are associated with the strength of

traditional gender stereotypes in a region. The present study further hypothesizes that the proportion of voters who oppose same-sex marriage can be used as a proxy variable to measure the strength of traditional gender stereotypes in a region. To test these hypotheses, this paper evaluates whether the proportion of voters who supported a constitutional amendment to ban same-sex marriage in North Carolina is associated with gender gaps in the top 5 percent on state assessments. The results suggest that gender gaps at the top of the distribution are strongly correlated with support for Amendment One. A one standard deviation increase in social conservatism is associated with a 26 percent standard deviation increase in gender gaps in 5th grade and a 31 percent standard deviation increase in gender gaps in 8th grade. Other possible proxy variables for traditional gender stereotypes — such as the proportion of voters who supported the Republican presidential candidate in 2012 and the proportion of religious adherents who are Evangelical — are not statistically significantly related to gender gaps.

These findings have two important implications. First, the results cast doubt on the hypothesis that differences in ability fully explain gender gaps among top-scoring students and support the hypothesis that differences in socialization contribute to the gaps. Second, the finding that students are more likely to adhere to gender norms in counties with higher traditional gender stereotypes has practical implications for educators, principals, school districts, and programs. First Lady Michelle Obama

remarked: “If we’re going to out-innovate and out-educate the rest of the world, then we have to open doors for everyone” (Office of Science and Technology Policy 2011). As a nation, we cannot allow traditional gender stereotypes to influence the educational performance or the career aspirations of high-ability students. By identifying the relationship between traditional gender stereotypes and gender gaps, individuals working towards gender equality at the top of all fields — math, science, and reading — can better target their efforts.

2. The best school district in the country? Examining the effect of the Mooresville Graded School District's Digital Conversion Initiative on student test scores

Over the course of 2008 and 2009, the Mooresville Graded School District distributed MacBook laptops to every one of its 4th through 12th grade students. This city school district in central North Carolina launched the program with the purpose of “employ[ing] technology in ways that improve teaching and learning through increased student engagement while teaching the NC Standard Course of Study focusing on reading and mathematics.” The district also emphasizes that the program is “a curriculum and instruction project” as opposed to a technology project (MGSD Digital Conversion Executive Summary).

The program, known as a one-to-one or 1:1 program because it provides one laptop per student, has garnered significant attention. President Obama visited the Mooresville Graded School District in June 2013, accompanied by U.S. Secretary of Education Arne Duncan, and commented that the results of the district's program have been “remarkable” (The White House 2013). The superintendent, Dr. Mark Edwards, was named Superintendent of the Year by the American Association of School Administrators (Porterfield et al. 2013). Scholastic's *Administrator* magazine named Mooresville as the “best district in the country” (Farrell 2013). *The New York Times* published a front-page story on the Mooresville Graded School District, heralding it as “the de facto national model of the digital school” (Schwartz 2012). An article in

Education Week reviewed the growing popularity of similar one-to-one programs and cited the district as “a model of how to do it right” (Quillen 2011). Several thousand people have visited the district to see the initiative in action. Karen Cator, then Director of Educational Technology at the U.S. Department of Education (and former Apple executive), visited the district in August 2010 to “help celebrate the district’s successes” (U.S. Department of Education 2010).

But is it time to celebrate? *The New York Times* article cited a 15 percentage point increase in the proportion of students meeting state proficiency standards on reading, math, and science assessments over the first three years of the program. However, the proportion of students meeting proficiency standards has increased in other school districts across the state. According to publicly available data from the North Carolina Department of Public Instruction, the percentage of students in 3rd through 8th grade meeting proficiency standards on math tests increased by 12.3 points for the Mooresville Graded School District and by 12.5 points for the state. Similarly, the percentage of students in the same grades meeting proficiency standards on reading tests increased by 16.0 points for the district and by 15.1 points for the state.

This study uses an administrative database containing information on all students in North Carolina to examine the effect of the Digital Conversion Initiative on 4th through 8th grade reading and math assessments. Controlling for 3rd grade test score and other demographic characteristics, the results suggest that the effect of the program

varies widely depending on the grade in which students first receive a laptop. Students who received a laptop in 4th grade scored 0.12 standard deviations higher on state reading assessments and 0.42 standard deviations higher on state math assessments after three years of exposure than students who never received a laptop. Similarly, students who received a laptop in 5th grade scored 0.22 standard deviations higher on state reading tests and 0.09 standard deviations higher on state math tests after three years of exposure than students who never received a laptop. In contrast, for students who received a laptop in 6th or 7th grade, the program had smaller effects on reading test scores and significant *negative* effects on math test scores. Compared to their unexposed peers, students who received a laptop in 6th grade scored 0.12 standard deviations lower on state math tests after three years of exposure, and students who received a laptop in 7th grade scored 0.22 standard deviations lower on state math tests after two years of exposure. These differences should serve as a warning for school districts planning to launch one-to-one programs, particularly for districts planning to implement a laptop program at the middle school level.

2.1 Conceptual framework

Superintendent Mark Edwards explains that the district “implemented the digital conversion to increase student achievement” (Farrell 2013, page 49), but there is limited research that suggests one-to-one laptop programs are effective at raising test scores. The *Journal of Technology, Learning and Assessment* devoted a 2010 volume to

examining laptop initiatives, but only two studies in the volume examine the effects of such programs on state standardized test scores. Suhr et al. (2010) examine students in a suburban southern California school district who participated in a one-to-one program in 4th and 5th grades. Although the authors find positive effects on reading test scores after two years of the program, their sample size of 54 students is very small. Shapley et al. (2010) evaluate the Texas Technology Immersion Pilot (TIP) which provided students in 6th through 8th grades at 21 treatment schools with laptops. Shapley et al. (2009) find that the program had no effect on reading test scores but that the program increased math test scores by 0.16 standard deviations at the end of two years and by 0.20 standard deviations at the end of three years (Table 6, page 86). However, the program was not consistently implemented across treatment schools. Shapley et al. (2007) report that “some schools allowed students to have unlimited home access, while others restricted out-of-school laptop use” (page 53). The remaining studies in the volume examine only the effects of one-to-one programs on subjective measures such as students’ self-reported engagement in school. As Suhr et al. (2010) explain, “there is relatively little research on measurable literacy outcomes following participation in laptop programs” (page 10).

This evidence from studies of one-to-one laptop programs is limited and mixed, suggesting that the Digital Conversion Initiative may or may not raise students’ test scores. Predicting the effect of laptop programs on test scores is further complicated

because one-to-one programs increase students' exposure to computers both at school and at home. However, many prior studies examine technology initiatives that *either* increased computer access at school or increased computer access at home.

Studies that analyze the effect of increased computer access at school yield mixed results (Angrist and Lavy 2002; Barrow et al. 2009; Dynarski et al. 2007; Goolsbee and Guryan 2006; Li, Atkins, and Stanton 2006; Morgan and Ritter 2002; Nummery et al. 2006; Rouse and Krueger 2004). Some of these studies find significant positive effects of computer use in schools (Barrow et al. 2009; Li, Atkins, and Stanton 2006; Morgan and Ritter 2002; Nunnery et al. 2006). Li, Atkins, and Stanton (2006) examine a pilot program with 122 Head Start participants in rural West Virginia. Preschoolers in the experimental group received up to 20 minutes of "developmentally appropriate educational software" access per day, while preschoolers in the control group received the standard Head Start curriculum. After six months of the program, the experimental group scored higher than the control group on a school readiness test that assessed concepts such as size, direction, time, and quantity. However, the authors do not find evidence of differences across the two groups in visual motor development (assessed by asking students to replicate a drawing of a figure), gross motor skills (assessed by asking students to run, jump, catch, etc.), or cognitive development (assessed by asking students to complete a short intelligence test).

Barrow et al. (2009) evaluate a randomized controlled experiment designed to test the efficacy of an algebra software program. Within three large urban school districts, some classrooms were randomly assigned to use the computer software, while other classrooms were randomly assigned to receive traditional instruction. The researchers estimate an “intent-to-treat” (ITT) effect of 0.21 standard deviations; that is, students who were randomly assigned to use the software program scored 0.21 standard deviations higher than students who were randomly assigned to traditional instruction (Table 2, page 64). When the researchers include student controls and teacher fixed effects, the ITT estimate increases to 0.28 standard deviations. However, the experiment was not cleanly implemented. “Treated” students spent about two-thirds of class time in computer labs and about one-third of class time in traditional classrooms, while “control” students spent about 10 percent of class time in computer labs and about 90 percent of class time in traditional classrooms (Table 3, page 66). After accounting for this “contamination,” the authors estimate a “treatment-on-treated” (TOT) effect of 0.42 standard deviations (Table 2, page 64).

Other studies fail to find significant effects of increased computer access at school (Angrist and Lavy 2002; Dynarski et al. 2007; Goolsbee and Guryan 2006; Rouse and Krueger 2004). Dynarski et al. (2007) is the product of Section 2421 of the No Child Left Behind Act, which required the U.S. Department of Education “to conduct a national study of the effectiveness of educational technology” (page xviii). The

randomized study evaluates 15 software programs: five targeting 1st grade reading skills, four targeting 4th grade reading skills, three targeting 6th grade math skills, and three targeting algebra skills, typically completed by 9th grade students. Each of these programs “were able to provide at least some evidence of effectiveness from previous research,” and twelve of the programs “have either received or been nominated to receive awards ... from trade associations, media, parents, and teachers” (page xv).

After implementing a large randomized experiment involving 9,424 students across 140 schools, the authors conclude that the software programs did not significantly increase test scores for students in any grade after one year.

Angrist and Lavy (2002), Goolsbee and Guryan (2006), and Rouse and Krueger (2004) also fail to find significant effects of increased computer access at school. Angrist and Lavy (2002) analyze the Tomorrow-98 program in Israel that provided 35,000 computers to elementary and middle schools between 1994 and 1996, while Goolsbee and Guryan (2006) examine the E-Rate program enacted by the U.S. government that subsidized spending on Internet access in schools. While the Tomorrow-98 program increased the use of computers in schools and the E-Rate program increased the number of Internet-connected classrooms, neither program raised test scores. Rouse and Krueger (2004) analyze the efficacy of *Fast ForWord*, a computer software program designed to strengthen reading skills. Using a randomized evaluation of students in 3rd

through 6th grade, they conclude that the program did not have a statistically significant effect on students' scores on state standardized reading assessments.

Studies that analyze the effect of increased computer access at home yield similarly inconclusive results (Attewell and Battle 1999; Fairlie 2005; Fairlie et al. 2010; Fairlie and London 2012; Fuchs and Woessmann 2005; Malamud and Pop-Eleches 2011; Schmitt and Wadsworth 2006; Vigdor and Ladd 2010). Fairlie and London (2012) find statistically significant positive effects of increased home computer access among community college students for some outcomes but do not find statistically significant effects for other outcomes. The researchers randomly select 141 low-income freshmen at Butte College, a large community college in Northern California, to receive a refurbished desktop computer for free. They find that, compared to students who applied for a computer and were not selected to receive one, students who applied for a computer and were randomly selected to receive one were about 3 percentage points more likely to enroll in a course for a letter grade and about 7 percentage points more likely to take a course for which the credits were transferable to a 4-year university in California (Table 4, page 737). However, students who received a computer were not significantly more likely to pass a course or to graduate with a degree or a certificate than students who did not receive a computer.

Malamud and Pop-Eleches (2011) and Vigdor and Ladd (2010) find statistically significant negative effects of home computer access on grades and test scores.

Malamud and Pop-Eleches (2011) examine a program in Romania that provided 154,474 low-income students with vouchers to purchase a computer. Students were eligible for a voucher if their family's monthly income was less than 150 RON (roughly \$65) per household member. Using a regression discontinuity design, the researchers identify that the program had a negative effect on test scores. Students who received a voucher earned grades roughly 25 to 33 percent of one standard deviation lower in math, English, and Romanian (page 990). Vigdor and Ladd (2010) examine a longitudinal sample of nearly 750,000 students who began 5th grade at a North Carolina public school at the start of the 1999-2000, 2000-2001, or 2001-2002 school years and who remained at a North Carolina public school through the end of 8th grade. They document that students scored about 1 percent of a standard deviation lower on math and reading state standardized assessments after they gained access to a home computer and scored about 3 percent of a standard deviation lower on math assessments after a broadband service provider entered their ZIP code (Table 3, page 36). Further, the negative effect of computer ownership persisted for at least three years, with no attenuation in the decline of math test scores and with little attenuation in the decline of reading test scores (Table 5, page 39).

In sum, the evidence on the effects of computers on test scores is inconclusive. Evaluations of the few existing one-to-one programs, such as the Texas TIP program, have suffered from inconsistent program implementation (Shapley et al. 2010) or from

small sample sizes (Suhr et al. 2010). Studies evaluating the effects of computer software programs in schools have reported mixed effects: statistically significant positive effects for some software programs (Barrow et al. 2009; Li, Atkins, and Stanton 2006; Morgan and Ritter 2002; Nunnery et al. 2006) and statistically insignificant effects for other software programs (Dynarski et al. 2007; Rouse and Krueger 2004). Meanwhile, studies evaluating the effects of computer access at home have yielded equally ambiguous conclusions, with some papers reporting statistically significant positive effects (Attewell and Battle 1999; Fairlie 2005; Fairlie et al. 2010; Schmitt and Wadsworth 2006), some papers reporting mixed effects (Fairlie and London 2012), and other papers reporting statistically significant negative effects (Fuchs and Woessmann 2004; Malamud and Pop-Eleches 2011; Vigdor and Ladd 2010). The present study may shed light on these mixed findings by identifying the effects of a large one-to-one laptop program on state standardized reading and math assessments.

2.2 The Digital Conversion Initiative

One of only 15 city school districts in North Carolina, the Mooresville Graded School District is located in Iredell County in the state's Piedmont region. Roughly 5,800 students were enrolled in the district during the 2012-13 school year across the district's seven schools: three elementary schools serving kindergarten through 3rd grades, two intermediate schools serving 4th through 6th grades, one middle school serving 7th and 8th grades, and one high school serving 9th through 12th grades.

Superintendent Mark Edwards launched the Digital Conversion Initiative in 2007. To test its network capacity, the district provided 400 MacBook laptops to the English department at Mooresville High School for students to use during class time only. The district also distributed laptops to all 508 teachers and encouraged them to “just try them out” (Farrell 2013, page 52). In January 2008, the district launched professional development seminars for teachers. These seminars provided teachers with training in specific programs, such as iMovie, and allowed teachers to share ideas for incorporating computers into their classrooms (Quillen 2011). The district also added five professional development half days and conducted a four-day summer institute attended by about one-quarter of teachers.

In August 2008, the district distributed laptops to all 1,650 students at the high school and all 750 students at Mooresville Intermediate School. The district expanded the program in August 2009, distributing laptops to all 900 students at the middle school and all 700 students at East Mooresville Intermediate School. Hence, every student in 4th through 12th grade had received a laptop by the start of the 2009-10 school year.

¹ Students are required to return their laptops at the end of each academic year and therefore cannot use their laptops over summer vacation.

¹ The district expanded the program further in August 2011, providing laptops to all 3rd grade students for use in school only. However, the present study evaluates test scores through June 2011, so students in this analysis are unaffected by this change.

Table 14 summarizes students' exposure to the Digital Conversion Initiative by graduating class and academic year. By the end of the 2010-11 school year, students had been exposed to the one-to-one program for periods ranging from one to three years. Two graduating classes had been exposed to the program for one year; four graduating classes had been exposed to the program for two years; and two graduating classes had been exposed to the program for three years. Additionally, exposure to the program for the graduating classes of 2015-2017 differs by school. Students who attended Mooresville Intermediate School (MIS) first received laptops in August 2008 and therefore had been exposed to the program for three years by the end of the 2010-11 school year. Meanwhile, students who attended East Mooresville Intermediate School (EMIS) first received laptops in August 2009 and therefore had been exposed for two years by the end of the 2010-11 school year.

To cover some of the costs of the program, the district charges parents \$50 per child as an annual technology usage fee, although the fee is waived for roughly 18 percent of low-income families (Schwartz 2012; personal communication with Chief Technology Officer Scott Smith). The district also provides students with a hard plastic MacBook case and a high-quality backpack with a padded laptop compartment in an effort to protect the equipment from damage. If students damage or lose their laptops, families are required to pay for the cost of repairing or replacing the equipment. If families cannot afford to pay this cost, students may borrow a laptop from the district

Table 14: Exposure to the Digital Conversion Initiative by graduating class and year

Graduating Class	Academic year						Years of exposure
	05-06	06-07	07-08	08-09	09-10	10-11	
2009	9	10	11	12			1
2010	8	9	10	11	12		2
2011	7	8	9	10	11	12	3
2012	6	7	8	9	10	11	3
2013	5	6	7	8	9	10	2
2014	4	5	6	7	8	9	2
2015	3	4	5	6	7	8	2 or 3
2016	2	3	4	5	6	7	2 or 3
2017	1	2	3	4	5	6	2 or 3
2018	K	1	2	3	4	5	2
2019		K	1	2	3	4	1

Notes: Full exposure highlighted in green. Partial exposure highlighted in red.

Sources: Presentation by Superintendent Mark Edwards in October 2012; personal communication with Chief Technology Officer Scott Smith

for use in school only. At the beginning of each school year, parents and students are required to attend a training class on operating and caring for the equipment (Farrell 2013).

Wireless internet access is ubiquitous on school grounds. To provide students with greater internet access after school, the district extended the hours that school buildings are open and collaborated with town officials to offer free wireless connections in libraries, community centers, municipal buildings, and parks. The district also partnered with MI-Connection, a broadband service provider in Mooresville, to offer reduced rates on internet service for low-income families in the district. But, to ensure

that students without internet connections at home are not disadvantaged, the district asks teachers not to assign homework that requires students to use the internet.

The district prohibits “personal use of school district technological resources for amusement or entertainment” (Responsible Use Policy, page 2). In addition to restricting inappropriate content, the district has blocked students from accessing many gaming and social media websites, including Facebook and Twitter. The district also prohibits students from using their laptops for commercial activities, such as buying or selling merchandise, and forbids students from using their laptops for political purposes. Students are allowed, however, to download music and videos, provided that they do so outside of school hours, that they do not violate copyright laws, and that the content is not deemed inappropriate. The district requires that students fully charge their laptops every night.

The district states that “no right of privacy exists in the use of technological resources” for teachers or students (Responsible Use Policy, page 5). School personnel are allowed to monitor files and communications “accessed, downloaded, created or transmitted using school district technological resources,” including e-mail messages and Internet browsing histories (Responsible Use Policy, page 5). When students are using their laptops at school, administrators can view students’ computer screens in real time. One Assistant Principal at Mooresville Middle School spends 15 minutes after lunch every day monitoring students’ computer screens.

The district reports that the Digital Conversion Initiative has prompted teachers to adopt a “flipped classroom.” In traditional classrooms, teachers lecture during school hours and students complete assignments to practice and reinforce the material after school. In flipped classrooms, teachers record lectures on videos and assign students to watch them as homework. Students then spend class time completing assignments — either individually or in small groups — while teachers circle the classroom to answer questions and to provide additional help. Because teachers record their lectures rather than deliver them to every student in their class on the same day, students can work at their own pace. One student reports: “It’s like having a personal tutor” (Schwartz 2012). As the district’s webpage explains, “we are changing the way teachers teach and students learn, while utilizing an awesome set of technological tools” (MGSD Digital Conversion Executive Summary).

These technological tools are not just the computers themselves but also the programs and resources available on the computers and on the internet. One program, icurio, provides more than 330,000 digital resources to students and teachers (Holzberg 2012). Students can access a wide array of educational resources, such as games, videos, simulations, and activities. Teachers can view lesson plans, design multimedia presentations, assign projects, and administer tests. Online discussion boards and video conferencing programs allow students and teachers to communicate after school.

Students can collaborate on group projects, and teachers can host review sessions prior to tests (Helms 2013).

These new tools also allow teachers, parents, and administrators to better track students' progress. Online assessments save teachers time that they would have otherwise spent grading and provide teachers with immediate feedback on how well their classrooms and students understand the material. Using this information, teachers can modify their lesson plans or focus their attention on particular students. Parents can use this information to monitor their child's performance. Farrell (2013) reports that "if a student finishes a test at 11 A.M., parents can log in remotely to see how their child did by 11:15" (page 53). Administrators can also use this information to identify the students who are near the threshold of passing a state assessment and to provide these students with additional help.

To make room for these new tools, the district has decreased spending on print textbooks, calculators, encyclopedias, globes, and maps. As Schwartz (2012) notes, "who needs globes in the age of Google Earth?" The district has also saved money by closing computer labs. Superintendent Edwards explained during the October 2012 tour that the Digital Conversion Initiative costs about \$1.25 per student per instructional day or about \$225 per student per year. This amount includes expenses for software programs, digital subscriptions, and maintenance.

2.3 Data and methodology

2.3.1 Data

North Carolina requires public school students in 3rd through 8th grades to take annual standardized achievement tests in math and reading, referred to as End of Grade tests. Each year, the North Carolina Department of Public Instruction compiles a database containing the test scores and demographic characteristics of these students. This database is provided to the North Carolina Education Research Data Center at Duke University, where data center administrators remove identifying student characteristics (e.g., names, birthdates, and Social Security Numbers) and replace them with unique student codes that remain with students throughout their years in the state public school system. This restricted database is then shared with researchers.

This paper combines the databases from 2006 to 2011 to yield three years of pre-Initiative test scores and three years of post-Initiative test scores. Over this period, there were 3,191,554 student-year observations on an End of Grade reading test and 3,207,458 student-year observations on an End of Grade math test for students in 4th through 8th grade. I standardized End of Grade test scores by test and year using this sample of observations. Then, I remove students who are missing 3rd grade test scores or other demographic characteristics, students who were retained or who skipped a grade, and

students who moved from the Mooresville Graded School District to another North Carolina public school district after being exposed to the Digital Conversion Initiative.²

The resulting sample contains 2,516,086 student-year observations for reading test scores and 2,530,342 student-year observations for math test scores. Tables 15 and 16 provide information on the number of student-year observations in reading and in math, respectively, that were affected by the Digital Conversion Initiative. Panel A of Tables 15 and 16 illustrate that there are about 3,800 student-year observations on End of Grade tests from students in year t who were exposed to the Initiative in year t . Panels B and C divide these 3,800 student-year observations according to how long students had been exposed to the program and by what grade students first received laptops. For example, the first column of Panel B of Table 15 illustrates that there are 2,054 observations from students in year t who were exposed to the Initiative in year t only,

² The Mooresville Graded School District began providing each 3rd grade student with a laptop for school use in August 2011, potentially making 3rd grade test scores after this date a function of the laptop program. Prior to this date, however, students' 3rd grade test scores serve as a useful proxy for ability. Because I include 3rd grade test scores as a control in all regressions, I remove 27 student-year observations for students who lack a student identifying code, as I am unable to link these students with their 3rd grade test scores. For the reading regressions, I remove 491,947 student-year observations for students who are missing a 3rd grade reading test score. For the math regressions, I remove 491,506 student-year observations for students who are missing a 3rd grade math test score. I also remove 130 student-year observations from students who are missing information on race in all years and 3 student-year observations from students who are missing information on gender in all years. Race and gender is imputed for student-year observations from year t for students who are missing these characteristics in year t but who are not missing these characteristics in another year. I remove 168,744 student-year observations from students who were ever retained and 5,077 student-year observations from students who ever skipped a grade. I remove 106 student-year observations from year t for students who were exposed to the Initiative in year $t - 1$ but who attended a public school in North Carolina in another district in year t . Last, I remove 2 student-year observations in year t from students who were exposed to the Initiative in year $t - 2$, did not attend a public school in North Carolina in year $t - 1$, and who returned to the Mooresville Graded School District in year t .

Table 15: Observations on reading tests by exposure to the Digital Conversion Initiative and by grade

	(1) All grades	(2) 4th grade	(3) 5th grade	(4) 6th grade	(5) 7th grade	(6) 8th grade
No exposure, from NC	2,512,313	567,552	528,498	492,996	469,836	453,431
No exposure, from district	5,114	1,122	1,000	911	1,046	1,035
Panel A. Any exposure						
Any exposure	3,773	987	891	830	570	495
Panel B. Length of exposure						
1 year of exposure	2,054	987	353	347	126	241
2 years of exposure	1,250	0	538	299	303	110
3 years of exposure	469	0	0	184	141	144
Panel C. Grade received						
Laptop received in 4th grade	1,709	987	538	184	0	0
Laptop received in 5th grade	793		353	299	141	0
Laptop received in 6th grade	794			347	303	144
Laptop received in 7th grade	236				126	110
Laptop received in 8th grade	241					241

Notes: The sample includes students enrolled in 4th-8th grades in a public school in North Carolina between 2006 and 2011. The sample omits students who are missing 3rd grade test scores or other demographic characteristics, students who were retained or who skipped a grade, and students who moved from the Mooresville Graded School District to another North Carolina public school district after being exposed to the Digital Conversion Initiative.

Source: North Carolina Education Research Data Center, End of Grade files

Table 16: Observations on math tests by exposure to the Digital Conversion Initiative and by grade

	(1) All grades	(2) 4th grade	(3) 5th grade	(4) 6th grade	(5) 7th grade	(6) 8th grade
No exposure, from NC	2,526,530	571,244	531,940	495,657	472,074	455,615
No exposure, from district	5,132	1,128	1,006	914	1,046	1,038
Panel A. Any exposure						
Any exposure	3,812	1,005	900	839	574	494
Panel B. Length of exposure						
1 year of exposure	2,080	1,005	356	351	126	242
2 years of exposure	1,261	0	544	302	305	110
3 years of exposure	471	0	0	186	143	142
Panel C. Grade received						
Laptop received in 4th grade	1,735	1,005	544	186	0	0
Laptop received in 5th grade	801		356	302	143	0
Laptop received in 6th grade	798			351	305	142
Laptop received in 7th grade	236				126	110
Laptop received in 8th grade	242					242

Note: See Table 15 for a description of the sample.

Source: North Carolina Education Research Data Center, End of Grade files

1,250 observations from students in year t who were exposed to the Initiative in year t and in year $t - 1$, and 469 observations from students year t who were exposed to the Initiative in year t , year $t - 1$, and year $t - 2$. The remaining columns separate these observations by grade.

Panel C provides an alternative way to describe exposure to the laptop program. In contrast to Panel B which reports how long students had been exposed to the program, Panel C reports when students first received laptops. The numbers are the same, but organizing this information in separate panels is useful for two reasons. First, the effects of the Initiative may be different for students who first received a laptop in 7th or 8th grade than for students who first received a laptop in 4th or 5th grade. If this is the case, then the results may then be easier to interpret in Panel C than in Panel B. Second, the panels are identical only because the sample removes students who were retained for a grade, students who skipped a grade, and students who switched school districts after exposure to the program. Consider, for example, a student who was retained in 4th grade after being exposed to the program. In Panel B, he would be counted in the first row and second column during his first year in 4th grade and would be counted in the second row and second column during his second year in 4th grade. In Panel C, he would be counted in the first row and second column during his first year in 4th grade and would be counted in the same cell during his second year in 4th grade. While these

students are omitted in the subsequent results, their exclusion does not materially affect the findings or conclusions of this paper.

2.3.2 Methodology

This paper first evaluates whether exposure to the Initiative had an effect on student outcomes, regardless of the length of exposure. To determine the effect of any exposure to the program on test scores, I apply the following difference-in-difference identification strategy:

$$Score_{ist} = \beta_0 + \beta_1 AnyExposure_{ist} + \gamma X_{ist} + \tau_s + \mu_t + \varepsilon_{ist} \quad (3)$$

where $Score_{ist}$ is the standardized test score for student i in school s in year t ;

$AnyExposure_{ist}$ is an indicator variable that equals one if student i in school s was exposed to the program in year t and zero otherwise; X_{ist} is a matrix of controls for student i ; τ_s are school fixed effects; μ_t are year fixed effects; and ε_{ist} is an error term. The matrix of controls includes the following variables: 3rd grade test score, gender, race, test modifications, learning disabilities, and limited English proficiency. In regressions where the dependent variable is reading test score, I control for 3rd grade reading test score; in regressions where the dependent variable is math test score, I control for 3rd grade math test score. The coefficient β_1 represents the effect of participation in the program on test scores.

As noted above, this approach ignores the length of exposure. To capture the length of exposure on test scores, I modify equation (3) as follows:

$$Score_{ist} = \beta + \sum_{j=1} \theta_j Exposure_{ist} + \gamma X_{ist} + \tau_s + \mu_t + \varepsilon_{ist} \quad (4)$$

where $Exposure_{ist}$ is an indicator variable that equals one when student i is exposed to the program for j years and zero otherwise. As displayed in Table 12, exposure ranges from one to three years. Here, the effect of the program on test scores for a given exposure period is represented by each coefficient θ_j with $j \in [1,3]$. I hypothesize that the absolute value of the coefficients θ_j will increase as j increases, indicating that the program has larger (positive or negative) effects with longer exposure.

The results from equation (4) may be difficult to interpret if the effects of the program vary based on when students first received a laptop. To address this possibility, I include a third approach:

$$Score_{ist} = \beta + \sum_{k=4} \theta_k Received_{ist} + \gamma X_{ist} + \tau_s + \mu_t + \varepsilon_{ist} \quad (5)$$

where $Received_{ist}$ is an indicator variable that equals one if student i first receives a laptop in grade k with $k \in [4,8]$ and zero otherwise.

For all equations, I perform separate analyses for reading and math. Because the intervention occurred at the district-cohort level, I cluster standard errors by district and

cohort. To ensure that a few large school districts are not biasing the results, I exclude five school districts with greater than 80,000 students in the sample in separate results available upon request. The results are substantively identical to the results below.

2.4 Descriptive findings

Table 17 provides descriptive statistics for the Mooresville Graded School District and for the rest of the state, both for the pre-Initiative period from 2006 to 2008 and for the post-Initiative period from 2009 to 2011. Panel A reports a statistically significant increase in reading and math test scores between the pre-period and the post-period, both for the Mooresville district and for the rest of the state. The test scores in the Mooresville district increased by 0.26 standard deviations in reading and 0.24 standard deviations in math, while the test scores for the rest of the state increased by 0.21 standard deviations in reading and 0.17 standard deviations in math.³ These differences are statistically significant, both within and across groups.

Panel B provides descriptive statistics for the variables that used as controls in the regressions that follow. Third grade reading test scores increased by 0.08 standard deviations for the district and by 0.06 standard deviations for the state between the pre-

³ Test scores are standardized by test and year using all observations from the state. However, the sample described in Table 15 omits about 492,000 student-year observations from students who lack 3rd grade test scores, 168,744 student-year observations from students who were retained, 5,077 student-year observations from students who skipped a grade, and 108 student-year observations from students who transferred out of the Mooresville Graded School District after being exposed to the program. Because many of these omitted student-year observations contain below-average test scores, the average test scores for the state that are reported in Table 15 are greater than zero.

Table 17: Descriptive statistics

	Mooreville		Significant differences	State		Significant differences
	2006-2008	2009-2011		2006-2008	2009-2011	
Panel A. Outcome variables						
Reading test score	0.232	0.487	***	0.093	0.302	***
Math test score	0.310	0.545	***	0.127	0.292	***
Panel B. Independent variables						
3rd grade reading test score	0.320	0.397	***	0.120	0.176	***
3rd grade math test score	0.356	0.491	***	0.119	0.215	***
Female	51.3%	50.0%		50.5%	50.2%	***
Asian	1.6%	1.6%		2.1%	2.3%	***
Black	15.3%	13.1%	**	27.2%	25.3%	***
Hispanic	2.9%	4.3%	***	7.4%	10.3%	***
American Indian	0.1%	0.1%		1.5%	1.5%	
Multiracial	2.2%	2.8%		3.0%	3.8%	***
Any test modification	10.7%	9.8%		12.0%	11.4%	***
Ever learning disabled	6.1%	5.1%	*	8.3%	7.1%	***
Limited English proficiency	0.6%	1.3%	***	3.9%	5.2%	***

Notes: See Table 15 for a description of the sample. The across-group differences in the pre-period are significant at the 5% level for every variable except gender; that is, the difference between the district's pre-period value and the state's pre-period value is statistically significant for every variable except gender. The same is true for the across-group differences in the post-period. * indicates significance at the 5 percent level; ** indicates significance at the 1 percent level; *** indicates significance at the 0.1 percent level.

Sources: North Carolina Education Research Data Center, End of Grade and Masterbuild files

period and the post-period. Third grade math test scores also increased by 0.14 standard deviations for the district and by 0.10 standard deviations for the state. During both periods, the district had a smaller proportion of students identified as Asian, Black, Hispanic, American Indian, or multiracial than the rest of the state. The Mooresville Graded School District is disproportionately composed of white students; while 43 percent of students are identified as non-white in the state, only 22 percent of students are identified as non-white in the district.

Tables 18 and 19 provide the standardized means on reading and math test scores, respectively, by exposure to the Initiative. Students from other school districts who were not exposed to the program in year t received an average standardized score of 0.20 on the reading test, while students from the Mooresville district who were exposed to the program in year t received an average standardized score of 0.48. The effect of the program on reading test scores appears to increase as length of exposure increases, with students receiving an average standardized score of 0.45 after one year of exposure, 0.49 after two years of exposure, and 0.60 after three years of exposure. Further, this positive relationship between test scores and length of exposure appears particularly strong for students who first received a laptop in 4th or 5th grade. For example, students who first received a laptop in 5th grade received an average standardized score of 0.35 after one year of exposure, 0.51 after two years of exposure, and 0.54 after three years of exposure.

Table 18: Standardized means on reading tests by exposure and by grade

	(1) All grades	(2) 4th grade	(3) 5th grade	(4) 6th grade	(5) 7th grade	(6) 8th grade
No exposure, from NC	0.204	0.186	0.193	0.205	0.219	0.221
No exposure, from district	0.282	0.302	0.237	0.322	0.245	0.304
Panel A. Any exposure						
Any exposure	0.483	0.468	0.416	0.528	0.517	0.523
Panel B. Length of exposure						
1 year exposure	0.451	0.468	0.352	0.488	0.454	0.472
2 years exposure	0.494		0.457	0.506	0.536	0.527
3 years exposure	0.598			0.640	0.535	0.607
Panel C. Grade received						
Laptop received in 4th grade	0.483	0.468	0.457	0.640		
Laptop received in 5th grade	0.442		0.352	0.506	0.535	
Laptop received in 6th grade	0.528			0.488	0.536	0.607
Laptop received in 7th grade	0.488				0.454	0.527
Laptop received in 8th grade	0.472					0.472

Note: See Table 15 for a description of the sample.

Source: North Carolina Education Research Data Center, End of Grade files

Table 19: Standardized means on math tests by exposure and by grade

	(1) All grades	(2) 4th grade	(3) 5th grade	(4) 6th grade	(5) 7th grade	(6) 8th grade
No exposure, from NC	0.212	0.176	0.227	0.205	0.216	0.244
No exposure, from district	0.351	0.292	0.269	0.361	0.368	0.470
Panel A. Any exposure						
Any exposure	0.546	0.526	0.494	0.685	0.492	0.512
Panel B. Length of exposure						
1 year exposure	0.501	0.526	0.349	0.557	0.470	0.558
2 years exposure	0.579		0.589	0.675	0.520	0.430
3 years exposure	0.658			0.941	0.450	0.495
Panel C. Grade received						
Laptop received in 4th grade	0.591	0.526	0.589	0.941		
Laptop received in 5th grade	0.490		0.349	0.675	0.450	
Laptop received in 6th grade	0.532			0.557	0.520	0.495
Laptop received in 7th grade	0.451				0.470	0.430
Laptop received in 8th grade	0.558					0.558

Note: See Table 15 for a description of the sample.

Source: North Carolina Education Research Data Center, End of Grade files

From Table 19, the effect of any exposure to the program on math test scores appears even greater than the effect of any exposure to the program on reading test scores. Students from other school districts who were not exposed to the program in year t received an average standardized score of 0.21 on the math test, while students who were exposed to the program in year t received an average standardized score of 0.55. Similarly, the positive relationship between math test scores and length of exposure is greater than the positive relationship between reading test scores and length of exposure; students received an average standardized score of 0.50 after one year of exposure, 0.58 after two years of exposure, and 0.66 after three years of exposure. Unlike reading test scores, math test scores did not monotonically increase in Panel C as students advance. For example, students who first received a laptop in 5th grade received an average standardized score of 0.35 after one year of exposure, 0.68 after two years of exposure, and 0.45 after three years of exposure. Further, students who first received a laptop in 6th grade exhibited monotonically *decreasing* math test scores, receiving an average standardized score of 0.56 after one year of exposure, 0.52 after two years of exposure, and 0.50 after three years of exposure.

2.5 Regression results

Tables 20 and 21 report the estimated effect of the Digital Conversion Initiative on reading and math test scores, respectively. Each panel, and each column within each panel, displays the result of a separate regression. As explained in the methodology

Table 20: Estimated effect of the Digital Conversion Initiative on reading test scores

	(1)	(2)	(3)	(4)	(5)	(6)
	All grades	4th grade	5th grade	6th grade	7th grade	8th grade
Panel A. Any exposure						
Any exposure	0.034 (0.050)	-0.025 (0.013)	0.030 (0.036)	0.034 (0.031)	0.105 (0.052)	0.036 (0.027)
Panel B. Length of exposure						
1 year exposure	-0.034 (0.039)	-0.025 (0.013)	-0.030 (0.046)	-0.020 (0.022)	-0.004 (0.027)	0.009 (0.015)
2 years exposure	0.064 (0.064)		0.071* (0.025)	0.053* (0.018)	0.097 (0.050)	0.015 (0.013)
3 years exposure	0.235*** (0.043)			0.119*** (0.024)	0.220*** (0.026)	0.098*** (0.013)
Panel C. Grade received						
Laptop received in 4th grade	-0.014 (0.079)	-0.025 (0.013)	0.071* (0.025)	0.119*** (0.024)		
Laptop received in 5th grade	0.080 (0.041)		-0.030 (0.046)	0.053* (0.018)	0.220*** (0.026)	
Laptop received in 6th grade	0.071 (0.040)			-0.020 (0.022)	0.097 (0.050)	0.098*** (0.013)
Laptop received in 7th grade	-0.004 (0.042)				-0.004 (0.027)	0.015 (0.013)
Laptop received in 8th grade	0.032 (0.027)					0.009 (0.015)
R-squared	0.621	0.685	0.646	0.624	0.583	0.579
Observations	2516086	568539	529389	493826	470406	453926

Notes: Each column and each panel represents results from a separate regression. See Table 15 for a description of the sample. Standard errors are in parentheses. * indicates significance at the 5 percent level; ** indicates significance at the 1 percent level; *** indicates significance at the 0.1 percent level.

Sources: North Carolina Education Research Data Center, End of Grade and Masterbuild files

Table 21: Estimated effect of the Digital Conversion Initiative on math test scores

	(1)	(2)	(3)	(4)	(5)	(6)
	All grades	4th grade	5th grade	6th grade	7th grade	8th grade
Panel A. Any exposure						
Any exposure	0.050 (0.043)	0.035 (0.060)	0.158** (0.049)	0.157 (0.076)	-0.031 (0.044)	-0.092 (0.064)
Panel B. Length of exposure						
1 year exposure	-0.012 (0.048)	0.035 (0.060)	0.082 (0.060)	-0.013 (0.026)	-0.139*** (0.011)	-0.016 (0.035)
2 years exposure	0.084 (0.056)		0.210*** (0.041)	0.213*** (0.044)	-0.045 (0.034)	-0.222*** (0.032)
3 years exposure	0.217* (0.103)			0.422*** (0.031)	0.093*** (0.007)	-0.122** (0.032)
Panel C. Grade received						
Laptop received in 4th grade	0.023 (0.067)	0.035 (0.060)	0.210*** (0.041)	0.422*** (0.031)		
Laptop received in 5th grade	0.080 (0.066)		0.082 (0.060)	0.213*** (0.044)	0.093*** (0.007)	
Laptop received in 6th grade	0.009 (0.039)			-0.013 (0.026)	-0.045 (0.034)	-0.122** (0.032)
Laptop received in 7th grade	-0.054 (0.029)				-0.139*** (0.011)	-0.222*** (0.032)
Laptop received in 8th grade	-0.016 (0.029)					-0.016 (0.035)
R-squared	0.643	0.707	0.675	0.653	0.633	0.599
Observations	2530342	572249	532840	496496	472648	456109

Notes: Each column and each panel represents results from a separate regression. See Table 15 for a description of the sample. Standard errors are in parentheses. * indicates significance at the 5 percent level; ** indicates significance at the 1 percent level; *** indicates significance at the 0.1 percent level.

Sources: North Carolina Education Research Data Center, End of Grade and Masterbuild files

section, each regression contains 3rd grade test score as well as indicator variables for gender, race, test modifications, learning disabilities, and limited English proficiency. In addition, each regression contains indicator variables for school and year.

2.5.1 Reading test scores

Combining observations from all grades, column (1) of Panel A in Table 20 reports that exposure to the Digital Conversion Initiative in year t had a statistically insignificant overall effect on reading test scores, with students who were exposed scoring just 0.03 standard deviations higher than students who were not exposed. The remaining columns of Panel A reveal that exposure to the program in year t had statistically insignificant effects on reading test scores in all grades.

Column (1) of Panel B illustrates that the program had statistically insignificant effects on reading test scores after one and two years of exposure but that the program boosted reading test scores by a statistically significant 0.24 standard deviations after three years of exposure. When observations are separated by grade, the same pattern exists: small (always insignificant) effects of the program after one year of exposure, modest positive (sometimes significant) effects after two years of exposure, and strong positive (always significant) results after three years of exposure.

While the estimates in Panel C also exhibit this pattern, the structure of Panel C suggests a more nuanced perspective of the effects of the program on reading test scores. Students who received a laptop in 4th, 5th, or 6th grade reflect the pattern described

previously, with small negative effects of the program after one year of exposure, modest positive effects after two years of exposure, and larger positive effects after three years of exposure. After three years of exposure, students scored 0.12 standard deviations higher if they received a laptop in 4th grade, 0.22 standard deviations higher if they received a laptop in 5th grade, and 0.10 standard deviations higher if they received a laptop in 6th grade than students who never received a laptop.

For students who received a laptop in 7th grade or 8th grade, the program had no significant effect on test scores. Like students who received a laptop in an earlier grade, students who received a laptop in 7th or 8th grade did not experience a statistically significant change in test scores after one year of exposure. While students who received a laptop in 4th, 5th, or 6th grade experienced an increase of at least 0.05 standard deviations after two years of exposure, students who received a laptop in 7th grade experienced an increase of only 0.02 standard deviations after two years of exposure. Further, the coefficient is not statistically significant. Unfortunately, North Carolina does not administer End of Grade tests after 8th grade, so it is impossible to evaluate the effects of the program on End of Grade test scores after longer periods of exposure for students who received a laptop in later grades.

2.5.2 Math test scores

Table 21 presents estimated effects of the program on math test scores and reports many similar patterns with a few notable differences. First, column (1) of Panel

A reports that exposure to the Digital Conversion Initiative in year t had a statistically insignificant overall effect on math test scores, with students who were exposed scoring 0.05 standard deviations higher than students who were not exposed. The remaining columns of Panel A reveal that exposure to the program had a statistically significant effect on 5th grade test scores only, with students who were exposed to the program in 5th grade scoring 0.16 standard deviations higher than students who were not exposed to the program in 5th grade.

Panel B illustrates a similar pattern for math test scores as for reading test scores. Combining grades, the program had a negative effect on math test scores after one year of exposure, a modest positive effect on math test scores after two years of exposure, and a large positive effect on math test scores after three years of exposure. Most striking, students who had been exposed to the program for three years scored a statistically significant 0.42 standard deviations *higher* on 6th grade math tests and 0.12 standard deviations *lower* on 8th grade math tests than students who had not been exposed to the program.

The most notable differences between reading and math test scores emerge in Panel C. Students who received a laptop in 4th grade experienced large positive effects of the program after two years of exposure (0.21 standard deviations) and after three years of exposure (0.42 standard deviations). For students who received a laptop after 4th grade, the effects of the program are less encouraging. Students who received a

laptop in 5th grade experienced a large positive effect of the program after two years of exposure (0.21 standard deviations) but only a modest effect of the program after three years of exposure (0.09 standard deviations). This finding suggests that the program not only failed to improve these students' test scores during the third year of exposure but that the program inhibited these students' performance during this year. Students who received a laptop in 6th grade experienced increasingly negative effects of the program after each year of exposure, culminating in a 0.12 standard deviation test score deficit compared to their unexposed peers. Worse still, students who received a laptop in 7th grade scored a statistically significant 0.14 standard deviations lower after only one year of exposure and 0.22 standard deviations lower after two years of exposure than their peers who never received a laptop. Students who received a laptop in 8th grade also experienced a negative but insignificant effect of the program on math test scores after one year of exposure.

2.6 Discussion

The results in Tables 20 and 21 have a number of implications. First, the Digital Conversion Initiative increased both reading and math test scores for some students, illustrating that a one-to-one program can yield statistically significant positive effects on state standardized assessments. The mixed conclusions from prior research left open the possibility that the Initiative would produce no change in test scores for any grade in any subject. Further, earlier research that reported negative effects of increased

computer access at home suggested the possibility that the Initiative would decrease test scores rather than increase them (Fuchs and Woessmann 2004; Malamud and Pop-Eleches 2011; Vigdor and Ladd 2010). In contrast, the present study suggests that the one-to-one program at the Mooresville Graded School District was successful at raising some students' reading and math test scores.

Second, the results here provide robust evidence that the effects of the laptop program were generally statistically insignificant in the first year after implementation. Panel B of Tables 20 and 21 illustrate that the program had a statistically insignificant effect on students' test scores after one year of exposure in nine of ten regressions. Existing evaluations of one-to-one programs have reported similar observations (Grimes and Warschauer 2008; Silvernail and Gritter 2007; Suhr et al. 2010). For example, both Suhr et al. (2010) and the present study find that students who received a laptop in 4th grade scored lower on reading assessments at the end of 4th grade but that the difference was statistically insignificant (Table 10, page 28). To explain this (statistically insignificant) decline, Suhr et al. (2010) also report that teachers at a southern California school district spent the first year of a one-to-one program learning and teaching basic computer skills, whereas they were able to spend subsequent years focusing on content and learning (page 37). The results of the present study support the emerging consensus that one-to-one programs tend to have no effect on students' test scores at the end of the first year of implementation.

The results in Tables 20 and 21 also support earlier findings that one-to-one programs tend to increase reading test scores after multiple years of exposure. Both Suhr et al. (2010) and the present study find that students who received a laptop in 4th grade scored higher on reading assessments at the end of 5th grade than students who did not receive a laptop (Table 10, page 28).⁴ Further, both Shapley et al. (2009) and the present study find that students who received a laptop in 6th grade scored 7-10 percent of one standard deviation higher on the 8th grade state reading assessment than students who did not receive a laptop in 6th grade (Table 6.9, page 86).⁵ The positive effects of the Digital Conversion Initiative on reading test scores for students who received a laptop in 4th, 5th, or 6th grades appear consistent with these earlier findings from other one-to-one programs.

However, the results of the one-to-one program at the Mooresville school district contrast with earlier studies that identified positive effects of one-to-one programs on math test scores after multiple years of exposure. Shapley et al. (2009) report that students who received laptops in 6th grade scored 16 percent of one standard deviation higher on the 7th grade state math assessment and 20 percent of one standard deviation higher on the 8th grade state math assessment than students who did not receive a laptop in 6th grade (Table 6.9, page 86). The present study finds that students who received

⁴ Suhr et al. (2010) reports a statistically insignificant increase at the second year, whereas the present study reports a statistically significant increase at the end of the second year.

⁵ This difference is not statistically significant in Shapley et al. (2009) but is statistically significant in the present study.

laptops in 6th grade scored 5 percent of one standard deviation *lower* on the 7th grade state math assessment and 12 percent of one standard deviation *lower* on the 8th grade state math assessment.⁶ This discrepancy highlights that fact that one-to-one programs may increase test scores under some circumstances but decrease test scores under others.

While this study does not detail the circumstances that contribute to increased test scores, the mixed effects of the program by grade and by subject suggest some interesting hypotheses. Tables 20 and 21 illustrate that the one-to-one program was more successful at raising test scores when students received a laptop in 4th or 5th grades than when students received a laptop in 6th, 7th, or 8th grades. One hypothesis that fits this observation is that students who received a laptop in earlier grades were less likely to use it for entertainment purposes in later grades than students who did not receive a laptop until later grades. For example, perhaps 6th grade students who received a laptop in 4th grade were less likely to use it for entertainment purposes than 6th grade students who received a laptop that year. This hypothesis may be true, but it inappropriately compares students who had been exposed to the program for three years with students who had been exposed to the program for one year. Tables 20 and 21 provide robust evidence that the effects of the one-to-one program after one year of exposure are generally small and insignificant (see Panel B). It appears that the one-to-one program

⁶ Shapley et al. (2009) report statistically significant findings at the end of both years, whereas the present study reports statistically significant results at the end of the second year only.

may have “disrupted” the traditional educational model and that it required more than one year for students and teachers to adapt. Hence, it is not appropriate to compare 6th grade students who received a laptop in 4th grade with 6th grade students who received a laptop at the beginning of 6th grade.

Instead, it is more appropriate to compare 6th grade students who received a laptop in 4th grade with 8th grade students who received a laptop in 6th grade. Perhaps 6th grade students who received a laptop in 4th grade were less likely to use it for entertainment purposes than 8th grade students who received a laptop in 6th grade. While the district restricts access to social media and gaming websites both at school and at home, students can use their laptops to surf the internet or to purchase videos on iTunes. This hypothesis, however, does not fully explain the results in Tables 20 and 21. For one, the program had a statistically significant positive effect on reading test scores after three years of exposure *both* for students who received a laptop in 4th grade and for students who received a laptop in 6th grade (see Table 20, Panel C). Further, while students who received a laptop in 6th grade had statistically significant *higher* reading test scores after three years, they had statistically significant *lower* math test scores after three years. Since this comparison uses the same students, it is impossible to argue that these students differed in their use of laptops for entertainment purposes.

Tables 20 and 21 also illustrate that the one-to-one program was more successful at raising reading test scores than at raising math test scores. One hypothesis that fits

this observation is that math and reading teachers differed in their ability or motivation to incorporate digital technology into the classroom. Perhaps math teachers were less able to deliver lessons effectively by video or more reluctant to allow students to work independently on software programs than reading teachers. Or perhaps incorporating digital technology into math curriculum is simply more difficult than incorporating digital technology into reading curriculum, so that even if math teachers were equally able and identically motivated to incorporate digital technology as reading teachers, the one-to-one program would still be less successful at raising math test scores than at raising reading test scores. Or perhaps “flipped classrooms” are simply less effective for math instruction than for reading instruction. Anecdotal evidence suggests that students in a flipped math classroom may spend more of their class time working independently than students in a flipped reading classroom. For example, students in a flipped math classroom often work through individualized lessons on a software program, with teachers occasionally answering their questions and monitoring their progress. Schwartz (2012) observes: “One fourth grader, having to complete 10 multiplication questions in two minutes for the software to let her move on, simply consulted her times tables, making the lesson more about speed typing than mathematics” (Schwartz 2012). Students in a flipped reading classroom, meanwhile, may spend more of their class time working in small groups.

Although these hypotheses sound plausible, they also fail to explain the patterns in Tables 20 and 21. First, the program had a *larger* positive effect 5th and 6th grade math test scores after two or three years of exposure than on 5th and 6th grade reading test scores after two or three years of exposure (see columns (3) and (4) of Panel B in Tables 20 and 21). Second, the program had smaller but also positive effects on 7th grade math test scores after three years of exposure (see column (5) of Panel B in Tables 21). Only 8th grade math test scores declined after two or three years of exposure to the program (see column (6) of Panel B in Table 21). Perhaps there are differences between math and reading teachers or differences between math and reading instruction that affected the results of the one-to-one program, but such an explanation would account for the patterns observed in Tables 20 and 21 only if 8th grade math teachers or instruction were also substantively different from 5th, 6th, or 7th grade math teachers or instruction.

A hypothesis that accounts for the patterns observed in Tables 20 and 21 must explain increases in 6th, 7th, and 8th grade reading test scores after three years of exposure to the program and increases in 5th and 6th grade reading test scores after two years of exposure to the program. It must also explain increases in 6th and 7th grade math test scores after three years of exposure, increases in 5th and 6th grade math test scores after two years of exposure, and decreases in 8th grade math test scores after two and three

years of exposure.⁷ One finding worth noting is that *only* 8th grade math test scores declined after two or three years of exposure, regardless of when students received a laptop. This finding suggests that something about the one-to-one program affected 8th grade math test scores differently from how the program affected 8th grade reading test scores and from how the program affected math test scores in other grades.

There are a number of hypotheses that would account for the results reported in Tables 20 and 21 and for this observation regarding 8th grade math test scores. First, 8th grade math teachers, software, or instruction could be uniquely different from 5th, 6th, or 7th grade math teachers, software, or instruction. Students in 5th and 6th grades attend one of two intermediate schools in the district, while all students in 7th and 8th grades attend Mooresville Middle School. Although differences between the grades served by the intermediate schools and the grades served by the middle school could plausibly exist, systemic differences between 7th grade math teachers, software, and instruction and 8th grade math teachers, software, and instruction within the same school seem less likely. Further, I toured the Mooresville Middle School in October 2012 to witness the implementation of the Digital Conversion Initiative in 7th and 8th grade math classrooms. I observed no systemic differences between grades during this visit that could account for such divergent outcomes.

⁷ These observations are restricted to results from the second and third years of exposure to the program. As Panel B of Tables 20 and 21 illustrate, the program had a statistically insignificant effect on students' test scores after one year of exposure in nine of ten regressions, providing consistent evidence that the program had a roughly uniform null effect on students' test scores at the end of the first year.

A second hypothesis that could explain the difference between 8th grade math test scores and math test scores in other grades is that math software or digital instruction could be less effective in preparing students for the 8th grade state math assessment than for the state math assessments administered to students in other grades. Perhaps math software programs are more effective for rote memorization — such as mastering multiplication tables — or for less advanced material than for more advanced material. Such a finding would be consistent with the patterns observed in Tables 20 and 21, suggesting that this possibility is worthy of future research.

A third hypothesis that could explain the difference between 8th grade math test scores and math test scores in other grades is that perhaps the students who were in 6th or 7th grades when the program was implemented experienced a delayed effect from the “disruption” in learning that occurred during the first year of the program. Students who received a laptop in 6th grade scored slightly but insignificantly lower on 6th grade math tests than students who did not receive a laptop, suggesting that students who received a laptop may not have understood aspects of the 6th grade math curriculum as well as their peers. Because math courses are sequential, this slight statistically insignificant decline in 6th grade test scores may have compounded to produce the large statistically significant decline in 8th grade test scores. Along the same lines, perhaps students who received a laptop in 6th grade experienced less development of higher-order critical thinking skills during 6th grade than students who did not receive a laptop

in 6th grade. If the 6th and 7th grade assessments focus more on rote memorization but the 8th grade assessment focuses more on critical thinking, then a difference in the development of higher-order critical thinking during 6th grade would not produce a difference in test scores until 8th grade. Both of these hypotheses would also be consistent with the patterns observed in Tables 20 and 21, suggesting other potential areas of future research.

2.7 Limitations

This study evaluates the effects of the Digital Conversion Initiative on state standardized math and reading assessments and suggests areas for future research, but it does have three important limitations. First, this study cannot remove the possibility that Hawthorne effects contributed to the results. Students who were exposed to the program may have received higher test scores as a result of knowing that they were being studied, rather than as a result of the program itself. Anecdotal evidence suggests that students are acutely aware of the 60 visitors who parade through their schools five or six times per year. Schwartz (2012) reports: “After three years of computers permeating every area of their schooling, Mooresville students barely remember life before the transformation and are somewhat puzzled by the gaggle of visitors who watch them every month. (‘At times it’s kind of like being a lab rat,’ one 11th grader said.)” However, the mixed findings suggest that Hawthorne effects are unlikely to account for the results. In particular, it would be difficult to explain why Hawthorne

effects would produce an increase in 8th grade reading test scores but a decrease in 8th grade math test scores among students who received a laptop in 6th grade. While it remains possible that Hawthorne effects are contributing to the findings presented here, it seems unlikely that Hawthorne effects alone are producing the results.

A second limitation of this study is that it does not examine the effects of the Digital Conversion Initiative on other potentially important outcomes. The Mooresville Graded School District and the media outlets that have profiled the district claim that the Digital Conversion Initiative has increased student engagement, increased teacher satisfaction, and increased teacher collaboration. The district asserts on its webpage that “student engagement has increased tenfold” and that “laptop computers have significantly enhanced the level of student interest, motivation, and engagement in learning” (MGSD Digital Conversion Executive Summary). As evidence of increased teacher satisfaction, Farrell (2012) quotes Felicia Bustle, the principal at Mooresville Intermediate School, who explains: “Teaching here has been the best experience of my life” (page 55). Farrell (2012) continues: “Teachers say the data yielded by the conversion has created a far more collaborative environment with their peers” (page 53). Evaluating and quantifying the effect of the Digital Conversion Initiative on outcomes such as student engagement and teacher satisfaction is an important area for future research.

A final limitation of this study is that some of the results may not be generalizable to one-to-one programs in other school districts. When the Mooresville Digital School District implemented the Digital Conversion Initiative, it did not simply distribute laptops to teachers and students. The Initiative also included other changes that were discussed in Section 2.2 such as increased wireless internet access in the community and increased personal development for teachers. Further, the Initiative was accompanied with what Superintendent Edwards describes as “a change in culture.” One year prior to the launch of the Initiative, the district “crafted a systemic and comprehensive Strategic Plan” (MGSD Digital Conversion Executive Summary). The district also adopted the motto “every child, every day” and developed nine core values such as: “We are a community of learners” and “Every student is successful when provided high expectations and sufficient support” (MGSD Technology Plan). According to Superintendent Edwards, “Ninety percent of our visitors come here talking about hardware and leave talking about culture” (Farrell 2013). In another interview, he emphasizes: “They come to see digital but they leave talking about culture” (Helms 2013).

Quillen (2011) notes a change in culture at the high school. Todd Wirt, the principal of the high school, worked to change “what he recalls as a ‘complacent’ attitude among teachers and other staff members in a school where achievement data were average” (Quillen 2011). Principal Wirt further “established the district’s

Capturing Kids' Hearts program, in which teachers are asked to greet students with a pat or a handshake, and open the classroom to details about the good things happening in students' lives, in an effort to make the school culture less teacher-centered" (Quillen 2011). During a tour of the district in October 2012, visitors witnessed this program in action at the middle school when a 7th grade social studies teacher greeted every student at the door with a "fist bump."

It is possible that a change in culture may explain part or all of the changes in student performance on state assessments. Quillen (2011) reports that the "1-to-1 implementation made up just a part of a districtwide reform to make teaching and learning more contemporary." Farrell (2012) reports that "Edwards made it clear to parents, teachers, administrators, and students that the digital conversion wasn't about technology. It was about preparing all of the district's students for a successful and bright future" (page 49). It is important for educators to note this threat to external validity before extending the results of the present study to other one-to-one programs.

In any case, the Digital Conversion Initiative is receiving national attention for being a "model" for digital technology (Porterfield et al. 2013; Quillen 2011; Schwartz 2012). Superintendent Edwards estimates that several thousand individuals have visited the district since the launch of the program. Five or six times each year, the district offers a tour to 60 individuals who want to learn more about the Digital

Conversion Initiative.⁸ These tours are booked for over a year in advance, with a waiting list of over 200 individuals for the upcoming fall 2013 tours (Helms 2013; Message from the Superintendent 2013). The district has also hosted about 1,250 educators since the launch of its Summer Connection program in July 2010 (personal communication with Chief Technology Officer Scott Smith).⁹ This annual three-day conference, priced at \$450 per participant, offers content training for teachers, strategic planning assistance for administrators, and logistical training for technology staff. Farrell (2013) summarizes: “For educators with grand plans to transform their schools through technology, Mooresville has become a mecca of sorts” (page 50). Further, the district has been heralded not just as an exemplary school district for digital technology but also as an exemplary school district overall (Farrell 2013). Educators from other school districts may be using the Mooresville Graded School District as a model for incorporating laptops into the classroom and as a model for “changing the culture of instruction” (Schwartz 2012). Hence, it is even more important to identify both the successes and the limitations of changes that have occurred at the Mooresville Graded School District over the past few years.

⁸ Visitors spend four hours touring two schools of their choice and are encouraged to speak with administrators, teachers, and students. Afterwards, visitors are treated to lunch (courtesy of Apple) and listen to presentations by members of the district’s central administration, including Superintendent Edwards and Chief Technology Officer Scott Smith.

⁹ According to the district’s website, the conference will “motivate you to implement a successful 1:1” and will “get your team and district equipped for immediate success.”

2.8 Conclusions

The Digital Conversion Initiative had mixed success at raising students' test scores in the Mooresville Graded School District. On the one hand, the program was highly effective at raising reading and math test scores when students received a laptop in 4th or 5th grades. Students who received a laptop in 4th grade scored 0.12 standard deviations higher on reading tests and 0.42 standard deviations higher on math tests after three years of exposure than their unexposed peers. Students who received a laptop in 5th grade also experienced increases compared to their unexposed peers, scoring 0.22 standard deviations higher on reading tests and 0.09 standard deviations higher on math tests after three years of exposure.

On the other hand, the program also lowered math test scores when students received a laptop in 6th or 7th grades. While students who received a laptop in 6th grade experienced a 0.10 standard deviation increase in reading test scores after three years of exposure, these students experienced a 0.12 standard deviation decrease in math test scores after three years of exposure. Furthermore, students who received a laptop in 7th grade scored 0.14 standard deviations lower on state math assessments after one year of exposure and 0.22 standard deviations lower on state reading assessments after two years of exposure to the program than students who were unexposed. Students who received a laptop in 8th grade were generally unaffected by the program, although this result likely reflects these students' limited exposure to the program. Overall, the results

suggest that the success of the laptop program depended highly on *when* students first received a laptop.

During his visit to Mooresville Middle School in June 2013, President Obama commented: “There is no reason why we can’t replicate the success you’ve found here” (The White House 2013). However, the Digital Conversion Initiative had only mixed success at raising reading and math test scores when students first received a laptop in 6th, 7th, or 8th grades. School districts hoping to “replicate the success” of the program may also struggle to duplicate the “change in culture” that accompanied the one-to-one program. Although the district claims that “providing technology for every child has been a ‘win win’ for the students, school system and community” (MGSD Digital Conversion Initiative Executive Summary), the mixed effects of the program on state assessments and the threats to external validity should serve as potent warnings to school districts modeling one-to-one programs after the Digital Conversion Initiative.

3. Does having a daughter affect state legislators' beliefs? Evidence from a survey of state legislative candidates

"All great change in America begins at the dinner table."

—Former U.S. President Ronald Regan, January 1989

For years, Senator Rob Portman (R-OH) was a strong opponent of same-sex marriage. In 1996, he cosponsored the Defense of Marriage Act, and he made headlines for his views on gay marriage in 2011 when one-fifth of the University of Michigan Law School graduating class walked out during his commencement address (Snider 2011). Less than two years later, Senator Portman made headlines for his views on gay marriage again — when he announced that he had changed his mind. In a column in the *Columbus Dispatch*, he explains that he began reconsidering his stance on same-sex marriage when he learned that his son is gay. He recounts: "Knowing that my son is gay prompted me to consider the issue from another perspective: *that of a dad* who wants all three of his kids to lead happy, meaningful lives with the people they love" (Portman 2013, emphasis added). He is the only current Republican Senator to openly support gay marriage.

Senator Portman's experience suggests that children affect their parents' views. More specifically, studies have explored the relationship between children's gender and

parents' behavior. Warner and Steel (1999) find that parents with only daughters are significantly more likely to support policies to promote gender equity (such as Title IX) than parents with only sons (Table 2, page 512). Washington (2008) applies these results to explore whether legislators with daughters are more supportive of women's rights legislation than legislators with sons. She documents that members of the U.S. House of Representatives with at least one daughter are significantly more likely to support legislation pertaining to abortion, birth control, or contraceptives than U.S. Representatives with only sons (Table 5, page 325). Her paper sparked a number of other studies, such as Oswald and Powdthavee (2010) who find that parents with daughters have more liberal views than parents with only sons. Dahl et al. (2012) also extend the results from Washington (2008) to explore the effects of daughters on corporate executives. They find that CEOs pay their employees less if the CEO's newborn child is a boy but not if the CEO's newborn child is a girl.

Given how Washington's paper shaped the direction of future research, the present study aims to replicate Washington's findings for a sample of state legislators and to explore the potential mechanisms to explain her results. The present study examines legislators' responses to questions on abortion and abstinence-only education programs from the 2012 National Candidate Study, a survey distributed to all state legislative candidates during the 2012 election. This survey asked candidates whether they agreed or disagreed with two statements: (1) abortion should always be illegal, and

(2) school sex education programs should teach abstinence only. The results from Washington (2008) suggest that legislative candidates with daughters would be less likely to agree with these statements than candidates with only sons, but the present study finds that legislative candidates with daughters feel the same about abortion and abstinence-only education as their colleagues with only sons. This finding suggests that changes in legislators' views do not explain the changes in legislators' voting that Washington identifies. This paper also serves as a cautionary tale to researchers who may be quick to generalize the conclusions from prior work. Although legislators' own characteristics produce robust differences in voting, it appears that legislators' children's characteristics may only matter in some circumstances.

3.1 Conceptual framework

Prior research supports the idea that legislators might behave differently based on the gender of their children. First, studies suggest that legislators vote differently based on their own gender. Swers (1998) analyzes voting records for members of the U.S. House of Representatives during the 103rd Congress. She examines 14 bills that were identified by the American Association for University Women (AAUW) or by the Congressional Caucus for Women's Issues (CCWI) as pertaining to women's issues. She estimates that women were 8 percent more likely to vote according to the position advocated by the AAUW or the CCWI than men. Among a subset of five bills that "deal most directly with women's reproductive and other women's health issues" (page 438),

she finds that women were 11 percent more likely to vote according to the position advocated by the AAUW or the CCWI than men. She finds that, for both outcomes, the effect is driven by Republican women. Republican women were 24 percent more likely to support women's issues and 29 percent more likely to support women's health issues than men (Table 1, page 441).

Not only are female legislators more likely to support bills relating to women's issues than male legislators, but female legislators are also more likely to introduce bills relating to women's issues. Bratton and Haynie (1999) examine all bills introduced in six state houses in 1969, 1979, and 1989.¹ They find that female legislators are statistically significantly more likely to introduce bills relating to women's issues (such as an equal pay act) than male legislators. They also find that women are significantly more likely to introduce bills pertaining to children's policy, health care, welfare policy, education policy, and government spending than men, but the likelihood that a female introduced these bills was smaller than the likelihood that a female introduced a bill relating to women's issues (Table 1, page 668).

Other studies examine how the gender composition of legislative bodies affects public spending. Bolzendahl and Brooks (2007) examine welfare expenditure in 12 OECD countries throughout the 1980s and 1990s. They estimate that countries with a greater proportion of women in national legislatures spend a higher proportion of GDP

¹ The six states are Arkansas, California, Illinois, Maryland, New Jersey, and North Carolina.

on public spending for social programs, such as unemployment benefits and state sponsored childcare, than countries with lower proportion of women in national legislatures (Table 2, page 1521). Chattopadhyay and Duflo (2004) find that, among the leaders of village councils in India, female leaders invest in more public works projects that benefit women than male leaders. For example, because women are responsible for collecting drinking water in most households, female leaders invested more in building or repairing water facilities than male leaders; on average, female leaders built or repaired 24 water facilities, compared to male leaders who built or repaired only 15 facilities (Table V, page 1432).

Studies have also identified other personal characteristics that influence legislators, such as race (Barrett 1995; Canon 1999; Dye and Renick 1981; Meier and England 1984; Pande 2003; Tate 2003; Welch and Karnig 1979; Whitby 1997), religion (Hibbing and Marsh 1987), age (Hibbing and Marsh 1987), class (Carnes 2012), and educational background (Thiele et al. 2012). Bratton and Haynie (1999) find that black legislators are more likely to introduce bills relating to black interests (such as a bill designed to decrease racial discrimination) than legislators of other races (Table 1, page 668). Pande (2003) reports that public transfers to disadvantaged minorities in India increased with their representation in state and national legislatures increased. He estimates that a one percentage point increase in the representation of minority groups increased total state spending by a statistically significant 0.02 percentage points and

increased the share of state spending allocated to public transfers to these groups by a statistically significant 0.86 percentage points (Tables 6 and 7, pages 1143-1144).²

Hibbing and Marsh (1987) examine voting among members of Britain's House of Commons and find statistically significant effects of religion and age on "free votes" (votes where political parties do not dictate how members of Parliament should vote).

Carnes (2012) examines legislators from the U.S. House of Representatives. He finds that legislators whose last nonpolitical job was as a businessperson (e.g. a banker or a real estate agent) or another private-sector professional (e.g. an economist, a doctor, or an engineer) were more likely to support conservative economic policies than legislators whose last nonpolitical job was as a worker (e.g. a laborer, a soldier, or a union officer) (Table 1, page 19).

These studies find that the personal characteristics of legislators influence their voting records. There is reason to believe that characteristics of legislators' children may also affect their voting records. First, studies suggest that children shape their parents' values, just as parents' shape their children's values. Warner (1991) first proposed and explored this idea by randomly surveying residents in Detroit and Toronto on their

² The results described here are for the members of scheduled tribes. Pande (2003) explains that the constitutional orders of 1950 identified groups as "scheduled tribes" by using the following criteria: "1. Tribal origin, 2. Primitive ways of life and habitation in remote and less accessible areas, and 3. General backwardness in all respects" (Table 2, page 1138). Scheduled tribes represent 8 percent of the Indian population, with 52 percent of scheduled tribe members falling below the poverty line (Table 3, page 1138).

agreement with six statements designed to measure feminist values.³ Respondents also provided the number and sex of their children, as well as other demographic characteristics. Among residents in Toronto, she finds that respondents who had daughters only or respondents who had both sons and daughters expressed statistically significantly higher support for feminist values than individuals who had sons only. These findings were true for both men and women in Toronto. However, respondents in Detroit did not display the same pattern. Among residents in Detroit, women who had daughters only expressed statistically significantly more support for feminist values than women who had sons only, but men who had daughters only did not differ in their support for feminist values from men who had sons only. Additionally, Detroit residents of both genders who had both sons and daughters did not differ in their support for feminist values from respondents who had sons only (Table 2, page 1055). Warner (1991) concludes that “sex of children does seem to make a difference (specifically the female only versus the male only) for all except American men” (page 1055). She attributes the difference in results to different “cultural contexts,” but she notes that her sample is restricted to only two cities.

³ She measured feminist values using responses to six statements: (1) “It is more important for a wife to help her husband than to have a career herself,” (2) “Women who volunteer for active duty in the armed forces should be exempted from combat duty,” (3) “A married woman should be able to obtain credit without the consent of her husband,” (4) “A working mother can establish just as warm and secure a relationship with her children as a mother who does not work,” (5) “If a wife earns more than her husband, the marriage is headed for trouble,” and (6) “It is much better for everyone involved if the man is the achiever outside the home and the woman takes care of the home and family” (pages 1053-1054).

In a subsequent study, Warner and Steel (1999) surveyed residents in Oregon and Washington to explore differences in support for policies that promote gender equity. The survey asked respondents to indicate their support for the following policies: “pay equity, comparable worth, affirmative action (with regard to gender and employment), Title IX, maternity leave, and subsidized day care for working parents” (page 509). Warner and Steel (1999) combine these responses into a single measure of support for gender equity policies. They also separate respondents into four groups: respondents with only daughters, respondents with only sons, respondents with both daughters and sons, and respondents without children. Controlling for other demographic characteristics, they find that women with only daughters express statistically significantly higher support for these policies relative to women in all other groups. They find no statistically significant differences between women with only sons, women with both daughters and sons, and women without children. Among male respondents, they find that men with only daughters also express statistically significantly higher support for gender equity policies but that men with only sons express statistically significant lower support for gender equity policies, relative to men in all other groups (Table 2, page 512). The authors hypothesize that “men may see gender equity as a more ‘personal’ issue when it has the potential to affect their children” and that men with only sons “are more likely to oppose policies that are perceived to disadvantage” their male offspring (page 513).

Citing these findings, Washington (2008) hypothesizes that legislators with daughters may vote differently from legislators with only sons. To test this hypothesis, she compiles information on members of the U.S. House of Representatives during the 105th through 108th Congresses. Although a Representative's decision to have a child may be endogenous to his voting behavior, Washington notes that the gender of a Representative's child is a plausibly exogenous source of variation. By conditioning on a Representative's number of children, she estimates the effect of having an additional daughter, relative to having an additional son, on various outcomes. She finds that U.S. Representatives with an additional daughter receive higher scores from women's rights advocacy groups than Representatives with an additional son (Table 2, page 320). When she examines votes on individual pieces of legislation, she finds that the effect of having an additional daughter is the largest and the most robust for votes on reproductive rights; Representatives' number of daughters has a statistically significant effect in 46 of the 78 votes that pertain to abortion, birth control, or contraceptives in her sample (Table 5, page 325).⁴ Representatives with an additional daughter were 3.5 percentage points less likely to support an abortion ban, 3.7 percentage points more likely to support teen access to abortions, 4.9 percentage points more likely to support teen access to

⁴ The coefficient on number of daughters is statistically significant in 36 percent of votes that pertain to campaigns and elections, 24 percent of votes on labor, 21 percent of votes on health, and 21 percent of votes on government operations. It is statistically significant in less than 20 percent of votes that pertain to other issues.

contraceptives, and 4.7 percentage points more likely to support contraceptive use than Representatives with an additional son (Table 4, page 323).⁵

These findings prompted other researchers to analyze the effects of having daughters on other outcomes. Referencing Washington's study as motivation, Oswald and Powdthavee (2010) hypothesize that individuals with daughters are more likely to support liberal political parties than individuals with only sons (page 213). They test this hypothesis using the British Household Panel Survey (BHPS), a nationally representative study of over 10,000 British adults. Because this survey provides panel data, the authors use a fixed effects model to "difference out the unobservable personal characteristics" (page 221) and to examine how respondents' support for liberal political parties changes after the birth of a daughter. They find that the birth of an additional daughter increases the probability of voting for a liberal political party by 2.7 percentage points relative to the birth of an additional son (page 223). They report roughly equal effects for men and women (Table 1, page 222). They also examine panel data from the German Socioeconomic Panel (GSOEP). They find strikingly similar point estimates for the full sample, but they find much larger effects for men than for women (Table 8, page 225). Oswald and Powdthavee (2010) summarize: "Our work builds on, and attempts to generalize, innovative research by Rebecca Warner on children's influence on parents' views on feminist issues and affirmative action and by Ebonya Washington on

⁵ These effects were all significant at the 0.05 or 0.10 level.

children's influence on congresspersons' views on issues such as reproductive rights" (page 225).

Studies also report other effects based on children's gender. Ananat and Michaels (2008) and Dahl and Moretti (2008) document small but statistically significant higher divorce rates among couples whose first child is a daughter than among couples whose first child is a son. Dahl and Moretti (2008) estimate that couples whose first child is a daughter are 1-3 percent more likely to divorce than couples whose first child is a son (Table 1, page 1089). Dahl and Moretti (2008) further estimate that, among women who have an ultrasound during pregnancy, unmarried women who learn that they are pregnant with a girl are 4-5 percent less likely to be married at the time of the child's birth than unmarried women who learn that they are pregnant with a boy (Table 2, page 1092). Dahl et al. (2012) analyze panel data from Denmark's Integrated Database for Labor Market Research to explore the effects of children on male chief executive officers. The authors find that male chief executive officers pay themselves 4.9 percent more after the birth of a child and pay their employees 0.2 percent less after the birth of a child (Tables 3 and 5, pages and 685). However, these effects differ depending on whether the child is a son or a daughter. If their newborn is a boy, CEOs pay themselves 6.4 percent more and pay their employees 0.4 percent less. In contrast, CEOs pay themselves only 3.5 percent more and do not pay their employees any less if their

newborn is a girl. Each of these studies illustrates differential effects of parenting daughters versus parenting sons.

Washington (2008) is notable because it was the first study to suggest that legislators are influenced not only by their own personal characteristics but also by the characteristics of their children. Additionally, Washington's study motivated other researchers to explore the effects of having daughters on other behaviors, such as supporting liberal political parties and determining employees' wages. Given the significance of Washington's paper in shaping the direction of future research, the present study aims to replicate Washington's findings for a sample of state legislators. There are reasons to believe that Washington's sample of members from the U.S. House of Representatives and this sample of state legislators should be comparable. Many members of the U.S. House of Representatives begin their political careers as state legislators. Further, Oswald and Powdthavee (2010) find that children's gender affects individuals' political views. If children's gender influences U.S. Representatives' voting and individuals' views, it seems likely that children's gender affects state legislators' behavior as well. It is possible, however, that the two samples will yield different results simply because state legislators are different from U.S. Representatives along some unmeasured or unobserved dimension. State legislators are somewhat more representative of the U.S. population than Representatives, although women and minorities remain substantially underrepresented among both groups. While 50.8

percent of the U.S. population is female, only 24.2 percent of state legislators and 17.9 percent of U.S. Representatives are women (U.S. Census Bureau 2013; National Conference of State Legislatures; Center for American Women and Politics 2012).

A second aim of this paper is to explore one potential mechanism to explain Washington's findings. Washington examines the effect of children's gender on legislators' voting records, which is an important outcome measure. However, Washington cannot identify a mechanism to account for the difference in voting between legislators with daughters and legislators with sons. She explains: "While the study is motivated by research that suggests an attitudinal shift arises from parenting daughters, alternative explanations are possible. For example, parenting daughters may increase the cost of voting conservatively on reproductive rights legislation. ... Separating a 'true' preference shift from a cost-based change in voting patterns is beyond the scope of this study" (Washington 2008; page 313).

The current study measures legislators' preferences by using survey responses. The survey asked legislators whether they agree or disagree with the statements "abortion should always be illegal" and "school sexual education programs should teach abstinence only." Ideally, this approach accomplished what Washington's approach does not; it separates a "true" difference in preferences from other mechanisms that could explain her results. There is ample evidence that voting behavior is a noisy signal of legislators' preferences. Studies find that a host of external

factors, such as party platforms (Cox and McCubbins 1993), constituent preferences (Stadelmann et al. 2013; Hayes et al. 2010; Arnold 1990; Mayhew 1974), and interest group pressures (Hall and Wayman 1990), affect voting behavior. Because demonstrating party allegiance may affect legislators' committee assignments, legislators may vote according to their party's position even if their individual preference differs. Similarly, legislators may vote according to their constituents' preferences to increase their probability of reelection. The present study disentangles the effect of personal characteristics on preferences from the effect of personal characteristics on voting. If the results here are comparable to Washington's results, then this study will support the hypothesis that parenting daughters causes a shift in preferences, as well as a shift in voting. If the findings here are not comparable to Washington's study, then this paper will cast doubt on the hypothesis that differences in views explain differences in voting.

A third aim of this study is to extend the existing literature by incorporating children's ages. Children's ages alone are potentially endogenous to legislators' beliefs because legislators who choose to have their first child in their 20s may hold systematically different views than legislators who choose to have their first child in their 40s. To account for this possibility, the current paper applies Washington's approach to determine the effect of having an additional daughter of a given age compared to having an additional son of the same age. Washington conditions on

number of children to determine the effect of having daughters compared to having sons; the results here condition on number of children and age of children to determine the effect of having daughters of a certain age compared to having sons of the same age. In other words, conditional on the number and age of children, this paper identifies the marginal effect of gender. By employing an analogous identification strategy as Washington, the estimates here can be interpreted as causal.

This extension may shed light on *when* the gender of legislators' children begins to affect their behavior. Washington finds that having additional daughters accounts for differences in U.S. Representatives' voting on reproductive rights issues, but because she lacks information on children's ages, she cannot determine when this effect emerges. Does a Representative with a one year old daughter differ in his views of reproductive rights from a Representative with a one year old son? Or does a legislator with a daughter feel the same about reproductive rights as a legislator with a son, until their children become teenagers?

Prior work suggests that a difference in behavior between legislators with daughters and legislators with sons may emerge when children are in utero. As mentioned previously, Dahl and Moretti (2008) identify differences in marriage rates between mothers who learn that their unborn child is a girl and mothers who learn that their unborn child is a boy (Table 2, page 1092). Their study suggests that children's gender may affect legislators' behavior even before their children are born.

There is also more specific evidence to suggest that legislators with teenage daughters may feel differently about questions of reproductive rights and sex education than legislators with teenage sons. The National Parent Survey (2009) found that parents of 12 to 18 year old boys were over 60 percent more likely to think that it was acceptable for their child to have sex in high school than parents of 12 to 18 year old girls (Exhibit 3-10, page 42). Similarly, a study conducted in 2011 reported that teenage boys between 15 and 18 years old were slightly less likely to report that their parents would disapprove if they had sex than teenage girls (Planned Parenthood 2012, page 3). Other studies suggest that the differences between parenting teenage boys and parenting teenage girls extend beyond permissiveness of sexual behavior. A study of children between 12 and 15 years old conducted in 2000 documented that 44 percent of girls reported speaking with their parents “frequently” about values and beliefs, compared to only 38 percent of boys (White House Conference on Teenagers). When their teenager children are female, parents communicate more frequently about values and are more restrictive in their views of acceptable sexual behavior.

Studies of views on reproduction tend to sample either from the U.S. population at large or from parents of teenagers. There is little evidence from studies that have specifically surveyed the views of parents with younger children or parents with older children. However, there is some evidence that parents’ views differ depending on children’s age even within the teenage years. The National Parent Survey (2009) found

that parents of 13 year old children were less likely to agree that it was acceptable for their child to have sex in certain circumstances than parents of 18 year old children but more likely to agree that it was acceptable for their child to have sex in other circumstances (Table 3-4, page 39). Parents with children in 7th grade were far less likely to agree that it was acceptable for their child to have sex if their child used birth control or if they were dating someone for at least one year than parents with children in 12th grade.⁶ However, parents with a child in 12th grade were *less* likely to agree that it was acceptable for their child to have sex in high school than parents with a child in 7th grade.⁷ If parents' views on acceptable sexual behavior for their children differ depending on their children's age, then perhaps there is also a contrast in parents' views on reproductive rights and sex education.

These studies suggest the following hypotheses. First, the evidence suggests that legislators with daughters will be less likely to support abortion bans and abstinence-only programs than legislators with sons. Because Washington (2008) finds that legislators with an additional daughter were 3.5 percentage points less likely to support an abortion ban and 3.7 percentage points more likely to support teen access to abortions, it seems likely that legislators with an additional daughter will be more likely

⁶ 3.2% of parents with children in 7th grade agreed that it would be acceptable for their child to have sex if their child used birth control, compared to 17.4% of parents with children in 12th grade. While no parent with a child in 7th grade agreed that it would be acceptable for their child to have sex if their child had been dating someone for at least one year, 26.7% of parents with a child in 12th grade agreed with this statement.

⁷ While 12.2% of parents with children in 7th grade agreed with this statement, only 7.6% of parents with children in 12th grade agreed with this statement.

to disagree with the statement that “abortion should always be illegal” and more likely to agree with the statement that “school sexual education programs should teach abstinence only.” Second, the evidence suggests that legislators with daughters may express different views from legislators with sons as early as when children are in utero. Prior research by Dahl and Moretti (2008) suggests that a child’s gender affects his or her parents’ decision to marry prior to his or her birth. Given this result, it is not outside the realm of possibility that a child’s gender may affect his or her parents’ views on abortion and abstinence-only programs prior to his or her birth (presuming, of course, that mothers have an ultrasound during pregnancy to determine the sex of their child). Third, the evidence suggests that differences in views on abortion and abstinence-only programs between legislators with sons and legislators with daughters will exist when children are adolescents. The National Parent Survey (2008) reported that parents of teenage boys were 60 percent more likely to think that it was acceptable for their child to have sex in high school than parents of teenage girls. Although this survey does not examine differences in views on abortion or abstinence-only education specifically, it does suggest that parents’ views on issues relating to sexual behavior differ depending on the gender of their children. The following sections explore each of these three hypotheses.

3.2 Data and methodology

3.2.1. Data

This paper evaluates candidate responses from the 2012 National Candidate Study, a survey fielded by researchers at four universities and distributed to nearly all of the 10,131 state legislative candidates during the 2012 election.⁸ In August, all 7,444 candidates with email addresses received three requests to complete the electronic survey, yielding 1,318 responses. In September, 5,000 candidates who did not complete the electronic survey were randomly selected to receive a physical copy of the survey, yielding an additional 589 responses. A total of 1,907 candidates completed the survey, representing a 19.5 percent response rate.⁹ Candidates were informed that their responses would remain confidential.

In the survey, candidates were asked whether they agreed or disagreed with the following statements: “abortion should always be illegal” and “school sexual education programs should teach abstinence only.” Of the 1,907 respondents, 1,650 (87 percent) answered the question on abortion and 1,639 (86 percent) answered the question on abstinence-only programs. Candidates were also asked a series of questions about their children, and 1,797 candidates (94 percent) responded to the questions. I remove 257

⁸ The sampling frame for the study included 9,825 of the 10,131 state legislative candidates who appeared on the ballot in this election. The remaining 306 candidates (6%) lacked both physical and electronic mailing addresses.

⁹ Prior research has failed to identify statistically significant non-response bias, with the exception that “candidates who were running unopposed were slightly less responsive” (Broockman and Skovron 2013, page 9). These authors attribute this difference to the decreased frequency with which unopposed candidates likely checked their campaign mailboxes.

candidates who did not express their views on either statement or who did not disclose information on their children, leaving a final sample of 1,650 candidates.

Table 22 reports summary statistics on the final sample. About 26 percent of candidates agree that abortion should always be illegal, but this proportion varies dramatically by party affiliation. Republicans are more than three times as likely to agree that abortion should always be illegal than Democrats, with 44 percent of Republicans and only 13 percent of Democrats supporting this position. It is interesting to compare these findings to results from the most recent Gallup Values and Beliefs poll, which surveyed U.S. adults on whether they agreed that abortion should be illegal in all circumstances (Saad 2013). In contrast to the 44 percent of Republican candidates who believe abortion should always be illegal, only 28 percent of Republican adults believe that abortion should be illegal in all circumstances. Meanwhile, support for this position among Democratic adults (12 percent) is comparable to support for this position among Democratic candidates (13 percent). Consistent with Broockman and Skovron (2013), the present study finds that Republican state legislative candidates are far more supportive of the most extreme pro-life view than Republican U.S. adults.

On the statement that school education programs should teach abstinence only, 17 percent of candidates agree with this position, compared to 15 percent of the U.S.

Table 22: Summary statistics

		Full sample		Republicans		Democrats	
	Sample	Number of observations	Mean	Number of observations	Mean	Number of observations	Mean
Panel A. Outcome variables							
Agree: "Abortion should always be illegal"	1,617	425	0.26	282	0.44	104	0.13
Agree: "Schools should teach abstinence only"	1,607	272	0.17	222	0.35	27	0.03
Panel B. Independent variables							
Candidate's children							
Any children	1,650	1,306	0.79	547	0.82	664	0.79
Any daughters	1,650	1,019	0.62	426	0.64	517	0.62
Number of children	1,650	3,553	2.15	1,590	2.39	1,675	1.99
Number of daughters	1,650	1,754	1.06	750	1.13	857	1.02
Total number of children							
None	1,650	344	0.21	117	0.18	176	0.21
One	1,650	173	0.10	44	0.07	110	0.13
Two	1,650	543	0.33	228	0.34	288	0.34
Three	1,650	316	0.19	136	0.20	163	0.19
Four	1,650	150	0.09	78	0.12	56	0.07
Five	1,650	62	0.04	27	0.04	27	0.03
Six	1,650	32	0.02	18	0.03	9	0.01
Seven or more	1,650	30	0.02	16	0.02	11	0.01
Age of children							
0-4 years	1,650	130	0.08	54	0.08	61	0.07
5-9 years	1,650	170	0.10	54	0.08	61	0.07
10-13 years	1,650	163	0.10	78	0.12	77	0.09
14-17 years	1,650	202	0.12	81	0.12	68	0.08
18-24 years	1,650	403	0.24	102	0.15	85	0.10

25-34 years	1,650	566	0.34	187	0.28	182	0.22
35+ years	1,650	462	0.28	231	0.35	296	0.35
Candidate characteristics							
Female	1,111	312	0.28	71	0.15	218	0.40
White	1,636	1,480	0.90	618	0.94	734	0.88
Age							
18-25	1,650	37	0.02	20	0.03	10	0.01
26-34	1,650	138	0.08	49	0.07	65	0.08
35-44	1,650	219	0.13	92	0.14	109	0.13
45-54	1,650	359	0.22	156	0.23	176	0.21
55-64	1,650	521	0.32	204	0.31	273	0.33
65+	1,650	373	0.23	142	0.21	205	0.24
Religion							
Protestant	1,128	191	0.17	96	0.21	80	0.14
Catholic	1,128	115	0.10	47	0.10	62	0.11
Other Christian	1,128	134	0.12	67	0.14	53	0.10
Other religion	1,128	23	0.02	5	0.01	16	0.03
None or missing	1,128	665	0.59	253	0.54	345	0.62

Notes: The sample includes 1,650 candidates who completed the survey, who answered at least one of the questions on abortion or abstinence-only education, and who disclosed information on their children. Dependent variables may have fewer observations due to nonresponse. Independent variables may have fewer observations due to missing data.

Source: National Candidate Study, 2012

adult population (National Public Radio et al. 2004a, page 14).¹⁰ The differences by political party are even greater than the differences on the statement concerning abortion. Republicans were nearly 12 times more likely to agree with the statement on abstinence-only programs than Democrats; only 3 percent of the Democratic candidates believed that schools should teach abstinence-only education, compared with 35 percent of the Republican candidates. These figures are comparable to prior estimates in the literature. Bleakley et al. (2006, page 1154) find that 47 percent of Republican U.S. adults “somewhat” or “strongly” support abstinence-only programs.¹¹

Candidates were also surveyed about the number, gender, and age of their children. The survey asked candidates to report the number of sons and the number of daughters in each of the following age categories: 0-4 years, 5-9 years, 10-13 years, 14-17 years, 18-24 years, 25-34 years, and 35 years or older. As Table 22 illustrates, 79 percent of candidates had at least one child and 62 percent of candidates had at least one daughter. The average number of children was 2.15, and the average number of daughters was 1.06.¹² The number of children ranged from zero to 11, where 10 percent

¹⁰ Additionally, only 6% of parents with children between 15-18 years old think that birth control is an inappropriate topic in high school sex education programs (Planned Parenthood 2012, page 3). However, roughly 27% of U.S. public middle and high schools provide abstinence-only education and 10% of public schools provide no education program at all (National Public Radio et al. 2004, pages 4 and 7).

¹¹ The results reported by Bleakley et al. (2006) are somewhat higher than the results recorded in the 2012 National Candidate Study, but the difference is likely attributable to the fact that Bleakley et al. (2006) recorded responses on a 5-point scale while the 2012 National Candidate Study recorded responses on a binary scale.

¹² Among candidates with at least one child, the average number of children was 2.70 and the average number of daughters was 1.33.

of candidates had one child, 33 percent of candidates had two children, 19 percent of candidates had three children, and 17 percent of candidates had four or more children. The ages of children also differed widely: 27 percent of candidates had at least one child younger than 18 years, 24 percent of candidates had at least one child between 18-24 years, 34 percent of candidates had at least one child between 25-34 years, and 28 percent of candidates had at least one child 35 years or older.

These figures are higher than estimates typically reported by the Census Bureau, but the discrepancy emerges partly because the Census Bureau only maintains a record of children living with the householder (most commonly children under 18 years). After adjusting for this difference, two observations emerge. First, candidates are less likely to have children under 18 than an average U.S. household. The average number of children under 18 per household is 0.61 (U.S. Census Bureau 2012, Table AVG1), while the average number of children under 18 per candidate is 0.52.¹³ Second, candidates with children under 18 have more children than an average household with children under 18. The average number of children among families with children under 18 was 1.88 (U.S. Census Bureau 2012a, Table FM-3), while the average number of children among candidates with children under 18 was 1.95.

¹³ This difference is likely attributable to the fact that the sample is older than the U.S. population, as discussed subsequently.

Democrats were less likely to have children and, conditional on having children, more likely to have fewer children than Republicans. Democrats were 3 percentage points less likely to have any children and had an average of 0.40 fewer children than Republicans. As a result, they were also less likely to have at least one daughter, and if they had at least one daughter, they tended to have a fewer number of daughters. They were 6 percentage points more likely to have one child and 9 percentage points less likely to have four or more children. These findings are consistent with Washington (2008), who reports that Democrats were 7 percentage points more likely to have one child and 10 percentage points more likely to have four or more children (Table 1, page 316). Further, I observe that Democrats were less likely to have a child in almost every age bracket.¹⁴

Compared to the U.S. adult population, candidates in the sample were more likely to be male, white, and older (U.S. Census Bureau 2013). Females comprise 51 percent of the adult population but only 28 percent of the sample, while racial minorities represent 22 percent of the population but only 10 percent of the sample. On the other hand, individuals older than 65 comprise 13 percent of the population but are nearly doubly represented in this sample of state legislative candidates. Within the sample, candidates who were male or white were also more likely to be Republican than Democrat. These results are consistent with the literature on the underrepresentation of

¹⁴ Democrats were equally likely to have a child aged 35 years or older as Republicans.

women as legislators (Center for American Women and Politics 2012) and with the literature on the propensity for women to affiliate with the Democratic Party (Edlund and Pande 2002).

Most variables were generated from survey responses, but gender, religious affiliation, and party identification were collected from Project Vote Smart, a non-profit organization that gathers and publishes information on political candidates and elected officials. While the organization has reliable data on gender and party identification, it has less reliable observations on religious affiliation. Of the 1650 candidates, Project Vote Smart lacks information on religious affiliation for 68 percent of the sample. One possible explanation for this data limitation is that few candidates seem willing to report that they are unaffiliated, as only 9 candidates (2 percent) report that they are unaffiliated or nonreligious. To account for this data limitation, I treat “none or missing” as its own category.

3.2.2 Methodology

To determine whether candidates with at least one daughter view abortion and abstinence-only education differently from candidates with only sons, I estimate the following probit model:

$$\Pr(S_j) = \Phi [\alpha + \beta \textit{Daughter}_j + \theta X_j + \gamma_j] \tag{6}$$

where $\Pr(S_j)$ is the probability that state legislative candidate j expressed support for the statement on abortion or abstinence-only education; $\Phi [.]$ is the cumulative normal distribution function; $Daughter_j$ is an indicator variable that equals one if candidate j has a daughter and zero otherwise; X_j is a vector of candidate j 's personal characteristics; and γ_j is a series of fixed effects for candidate j 's total number of children. The coefficient β represents the effect of having at least one daughter compared to having only sons, conditional on having the same number of children.

Second, to determine whether candidates with one additional daughter view abortion and abstinence-only education differently than candidates with one additional son, I substitute the indicator variable for daughters in equation (6) with a count variable for the number of daughters to yield the following model:

$$\Pr(S_j) = \Phi [\alpha + \beta DaughterCount_j + \theta X_j + \gamma_j] \quad (7)$$

where $DaughterCount_j$ is a count variable that represents candidate j 's number of daughters. Here, the coefficient β represents the effect of having one additional daughter compared to having one additional son, conditional on having the same number of children.

Last, to determine whether candidates with at least one daughter of a given age view abortion and abstinence-only education differently from candidates with only sons of the same age, I estimate the following model:

$$\begin{aligned} \Pr(S_j) = \Phi [& \alpha + \beta_1 \text{Daughter0-4}_j + \beta_2 \text{Daughter5-9}_j + \beta_3 \text{Daughter10-13}_j + \beta_4 \text{Daughter14-17}_j \quad (8) \\ & + \beta_5 \text{Daughter18-24}_j + \beta_6 \text{Daughter25-34}_j + \beta_7 \text{Daughter35} + \beta_8 \text{Child0-4}_j + \beta_9 \text{Child5-9}_j \\ & + \beta_{10} \text{Child10-13}_j + \beta_{11} \text{Child14-17}_j + \beta_{12} \text{Child18-24}_j + \beta_{13} \text{Child25-34}_j + \beta_{14} \text{Child35} \\ & + \theta X_j + \gamma_j] \end{aligned}$$

where *Daughter0-4_j* is an indicator variable which equals one if candidate *j* has a daughter between 0 and 4 years old; *Daughter5-9_j* is an indicator variable which equals one if candidate *j* has a daughter between 5 and 9 years old; and so on. Additionally, *Child0-4_j* is an indicator variable which equals one if candidate *j* has a child between 0 and 4 years old; *Child5-9_j* is an indicator variable which equals one if candidate *j* has a child between 5 and 9 years old; and so on. In short, the *Child* coefficients are a series of fixed effects for the ages of candidate *j*'s children. The coefficient β_1 represents the effect of having at least one daughter between 0 and 4 years old compared to having only sons between 0 and 4 years old, conditional on having a child between 0 and 4 years old and conditional on having the same number of children. The coefficients β_2 through β_7 have analogous interpretations.

For all equations, it is crucial to condition on the number of children because a candidate's number of children is potentially endogenous to his views on abortion and abstinence-only education. For example, Catholic candidates are likely to have more children and to express greater support for the statements on abortion and abstinence. However, the gender of children is plausibly exogenous, an outcome randomly assigned by nature.¹⁵ Thus, I can compare candidates with the same number of children — some with no daughters, some with few daughters, and some with many daughters — to determine whether the act of parenting daughters affects candidates' beliefs differently than the act of parenting sons.

For equation (8), it is important to condition on age of children as well as number of children. Without age of children fixed effects, the coefficients on daughters' ages would represent the effect of having a daughter in the given age range compared to not having a daughter in the stated range, conditional on having the same number of children. This approach is flawed because children's ages are not randomly assigned; instead, children's ages are strongly predicted by legislators' ages. While 36 percent of candidates between 26-34 years old have at least one child between 0-4 years old, less than 1 percent of candidates older than 45 years have a child between 0-4 years old. Even after controlling for legislators' ages, it is likely that the choice to have a child

¹⁵ I could not use such an approach in countries with high rates of sex-selective abortion, such as in China. In the United States, the sex ratio among infants is "squarely within biologically normal parameters," suggesting that sex-selective abortions are infrequent or nonexistent (Barot 2012).

earlier or later in life is correlated with views on abortion and abstinence. By including age of children fixed effects, the coefficients on daughters' ages estimate the effect of having at least one daughter of a given age compared to having only sons of the same age, conditional on having at least one child of that age and having the same number of children. Because the gender of children is plausibly exogenous, I can compare candidates with the same number and ages of children — some without a daughter in a given range and some with a daughter in a given range — to determine whether act of parenting a daughter in that age group affects candidates' beliefs differently than the act of parenting a son in that age group.

Because Dahl and Moretti (2008) document that children's gender affects parents' behavior when children are still in utero, I test the hypothesis that candidates' views differ soon after the birth of a child. I estimate whether candidates with at least one daughter between 0 and 4 years old differ in their support of the statements on abortion and abstinence-only education from candidates with only sons between 0 and 4 years old. If differences exist between these two groups, then this finding will suggest that the effects reported by Washington (2008) occur in utero, at birth, or soon after the birth of a daughter.

A number of studies identify differences between parents with 12 to 18 year old daughters and parents with 12 to 18 year old sons (National Parent Survey 2009; Planned Parenthood 2012; White House Conference on Teenagers). To test the

hypothesis that candidates' views differ when children are teenagers, I compare candidates with at least one daughter between 14 and 17 years old to candidates with only sons in that age range. This approach will identify whether differences exist among the parents of teenagers. I also compare candidates with at least one daughter between 10 and 13 years old to candidates with only sons in that age range. This comparison will identify whether differences among the parents of preteens are comparable to differences among the parents of teenagers.

Last, to test the hypothesis that candidates' views differ when children are other ages, I examine differences between candidates with at least one daughter and candidates with only sons in the following age ranges: 5 to 9 years old, 18 to 24 years old, and 25 years or older. While there is limited research on how views differ for parents with children in elementary school or parents with children older than 17, the evidence presented in Section II suggests that it is worthwhile to consider the possibility of differential effects for parents with children in these age ranges.

In each table, I present results both with and without controls. In the specifications with controls, I include indicator variables for gender, race, and party. I also include a series of indicator variables for candidates' ages and a series of indicator variables for candidates' religious affiliations. Existing research indicates that each of

these variables is correlated with views on abortion (Pew Research Center 2012, page 8).¹⁶ In alternate specifications, I estimate the results separately by party and by gender.

3.3 Descriptive findings

3.3.1 Support for the statement that abortion should always be illegal

Figures 25-29 illustrate the proportion of candidates who agree that abortion should always be illegal, separated by party. Figure 25 reveals that support for the statement is greater among candidates with at least one daughter than among candidates without a daughter. 40 percent of Republican candidates and 10 percent of Democratic candidates without a daughter agree with the statement, while 46 percent of Republican candidates and 14 percent of Democratic candidates with at least one daughter agree. However, because this figure does not condition on the number of children, the effect of having at least one daughter is confounded with the effect of having at least one child.

Figure 26 removes this confounding effect by restricting the sample to candidates with exactly one child. Among Republican candidates, there is a large effect of having one daughter compared to having one son: 46 percent of Republicans with one daughter

¹⁶ Men are three percentage points more likely to agree that abortion should be illegal in all or most cases than women. Hispanics are 14 percentage points more likely to agree than Caucasians, while African Americans are two percentage points less likely to agree than Caucasians. Conservative Republicans are 47 percentage points more likely to agree than liberal Democrats. Adults between 50-64 years old are four to six percentage points less likely to agree than adults in all other age ranges. Protestants and Catholics are 24 and 21 percentage points more likely to agree than religiously unaffiliated people, while Jewish people are 13 percentage points less likely to agree than religiously unaffiliated people.

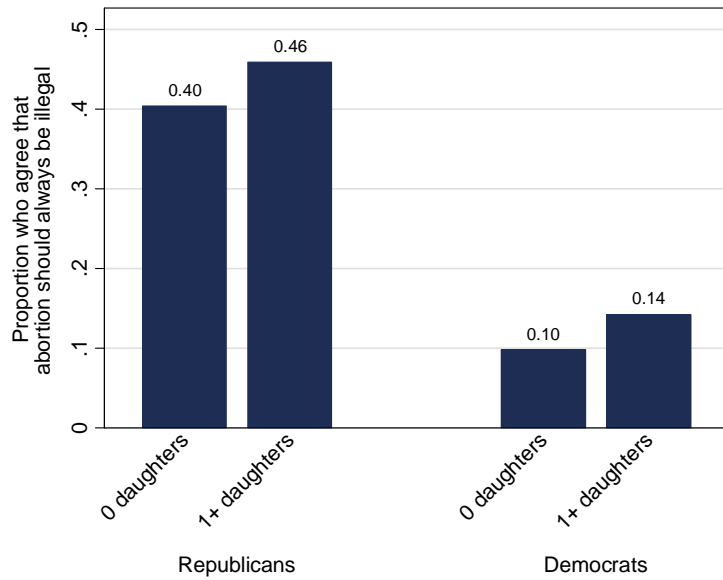


Figure 25: Proportion of all candidates who agree that abortion should always be illegal

Notes: See Table 22 for a description of the sample. The sample sizes for each column are as follows: 230, 412, 319, and 512.

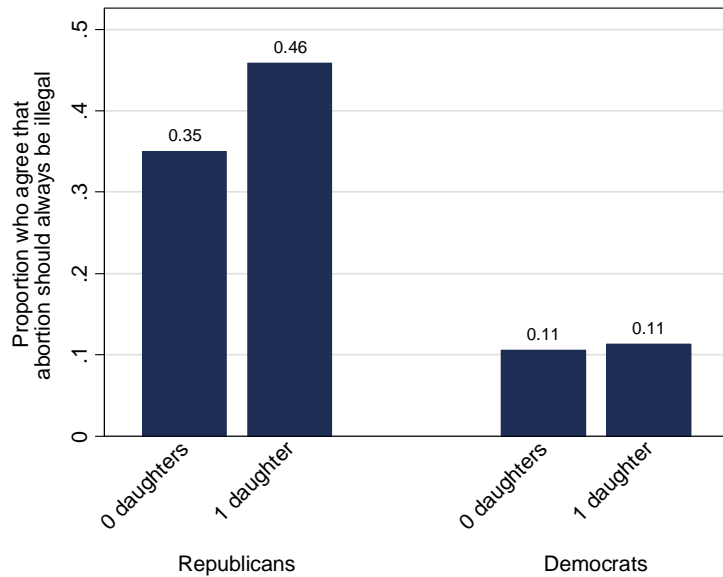


Figure 26: Proportion of candidates with one child who agree that abortion should always be illegal

Notes: See Table 22 for a description of the sample. The sample sizes for each column are as follows: 20, 24, 57, and 53.

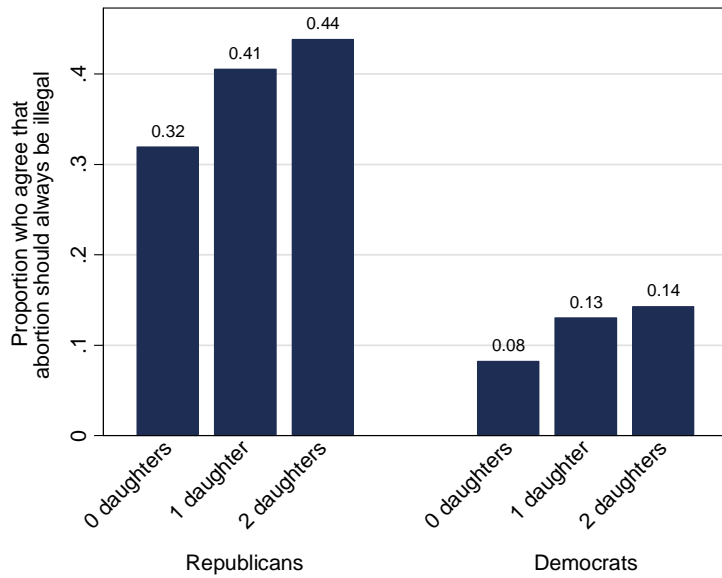


Figure 27: Proportion of candidates with two children who agree that abortion should always be illegal

Notes: See Table 22 for a description of the sample. The sample sizes for each column are as follows: 72, 106, 41, 61, 146, and 77.

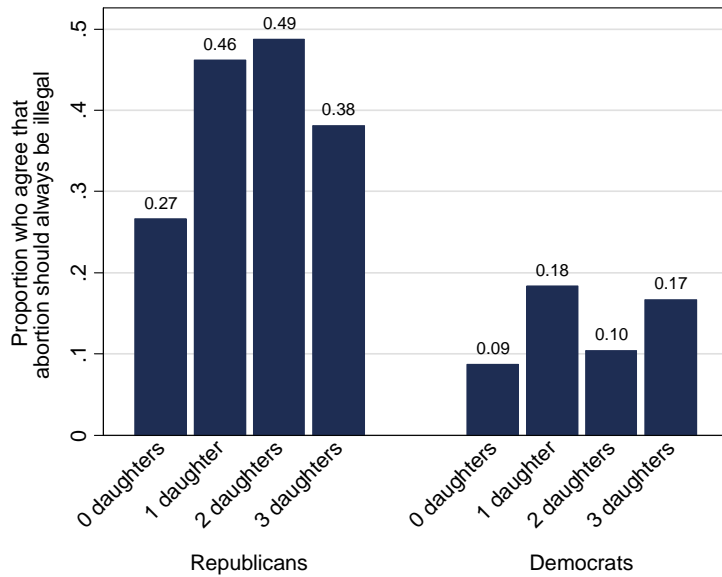


Figure 28: Proportion of candidates with three children who agree that abortion should always be illegal

Notes: See Table 22 for a description of the sample. The sample sizes for each column are as follows: 15, 52, 41, 21, 23, 60, 48, and 30.

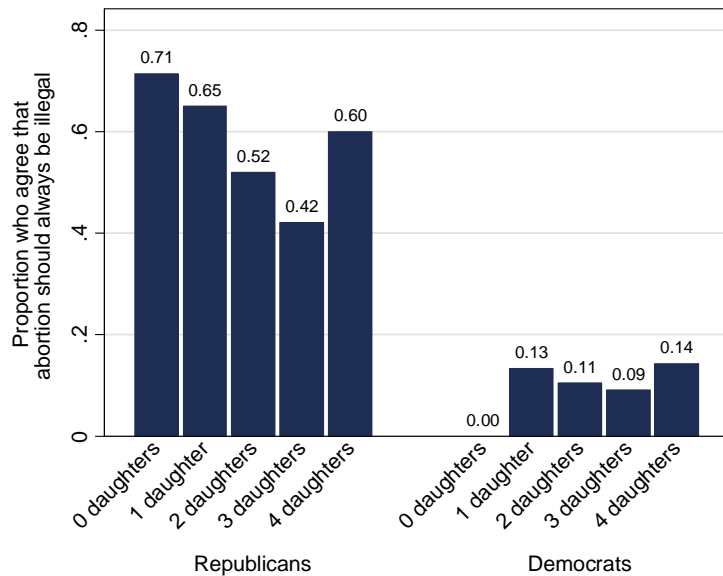


Figure 29: Proportion of candidates with four children who agree that abortion should always be illegal

Notes: See Table 22 for a description of the sample. The sample sizes for each column are as follows: 7, 20, 25, 19, 5, 4, 15, 19, 11, and 7.

agreed with the statement compared to 35 percent of Republicans with one son. Among Democratic candidates, there is no difference in support for the statement between candidates with one daughter and candidates with one son.

Figures 27-29 restrict the sample to candidates with exactly two, three, and four children, respectively. Figure 27 depicts a consistent pattern across parties that candidates with more daughters express higher support for the statement on abortion than candidates with fewer daughters. 32 percent of Republican candidates with two sons support the statement on abortion, 41 percent of Republican candidates with one son and one daughter support the statement, and 44 percent of Republican candidates

with two daughters support the statement. A smaller but consistent pattern exists among Democrats.

Figures 25-27 provide consistent evidence that candidates with more daughters are more likely to agree that abortion should always be illegal. However, the pattern becomes much less clear when I restrict the sample to candidates with three or four children in Figures 28 and 29. Among Republicans with three children, there is a positive relationship between support for the statement on abortion and number of daughters when candidates have zero, one, or two daughters. Then, there is a sudden decrease in support when candidates have exactly three daughters. Among Republicans with four children, the pattern is reversed. There is a negative relationship between support for the statement and number of daughters when candidates have zero, one, two, or three daughters, but there is a sudden increase in support when candidates have exactly four daughters. Among Democrats, Figures 28 and 29 offer mixed results with no clear pattern.

Two points are important. First, some of these findings are based on small sample sizes, as noted in the notes of each figure. However, my sample is nearly four times as large as the sample in Washington (2008). For example, Washington (2008) observes only five Republicans and seven Democrats with three daughters among U.S. Representatives with three children, whereas I observe 21 Republicans and 30 Democrats with three daughters among candidates with three children. Second, the

pattern that does emerge in Figures 25-27 contrasts with the pattern described by Washington (2008). Her results are consistent with the hypothesis that legislators with an additional daughter would express lower support for the statements on abortion and abstinence-only education than legislators with an additional son (conditional on number of children). In contrast, Figures 25-29 provide little evidence of any relationship between a legislator's views on abortion or abstinence-only programs and the gender of his children.

3.3.2 Support for the statement that school sex education programs should teach abstinence-only

Figures 30-34 provides descriptive results on the proportion of candidates who agree that school sex education programs should teach abstinence-only. Figure 30 reports greater support for abstinence-only education among Republican candidates with at least one daughter (37 percent) than among Republican candidates without a daughter (32 percent). However, the trend reverses when I condition on the candidate having exactly one child. Figure 31 reveals that 44 percent of Republican candidates and 9 percent of Democratic candidates with one son support abstinence-only education, while only 40 percent of Republican candidates and no Democratic candidates with one daughter support abstinence-only education.

The remaining figures do not suggest a consistent pattern between support for abstinence-only education and number of daughters. For example, among Republican

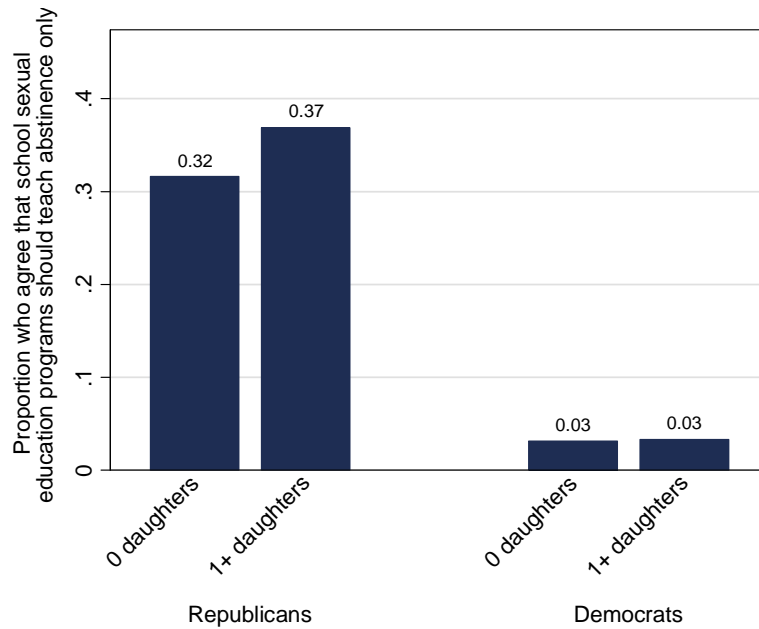


Figure 30: Proportion of all candidates who agree that school sex education programs should teach abstinence only

Notes: See Table 22 for a description of the sample. The sample sizes for each column are as follows: 228, 407, 321, and 511.

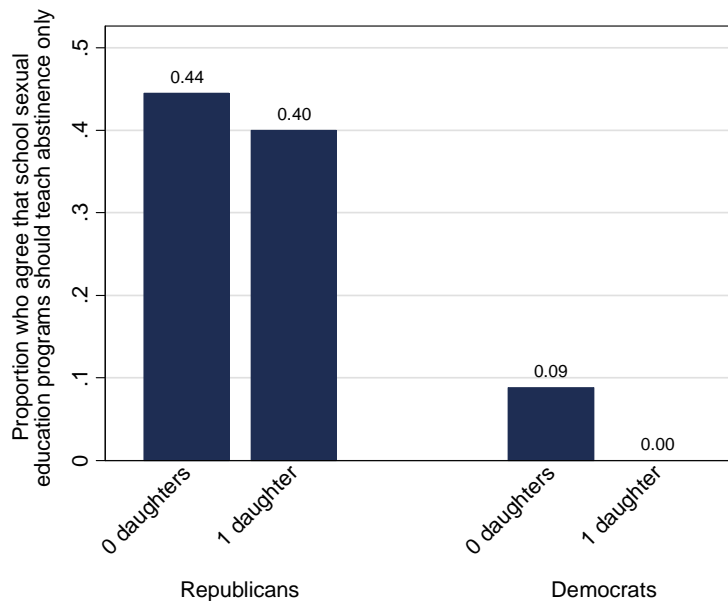


Figure 31: Proportion of candidates with one child who agree that school sex education programs should teach abstinence only

Notes: See Table 22 for a description of the sample. The sample sizes for each column are as follows: 18, 20, 57, and 52.

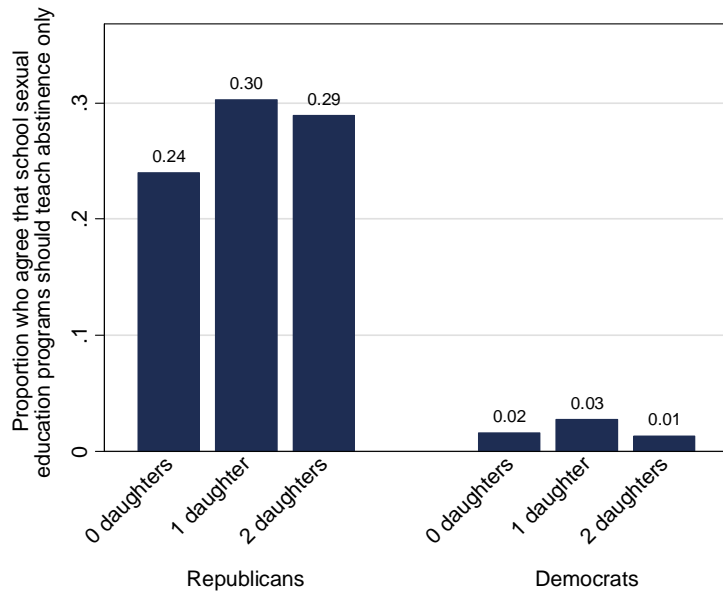


Figure 32: Proportion of candidates with two children who agree that school sex education programs should teach abstinence only

Notes: See Table 22 for a description of the sample. The sample sizes for each column are as follows: 75, 109, 38, 62, 145, and 77.

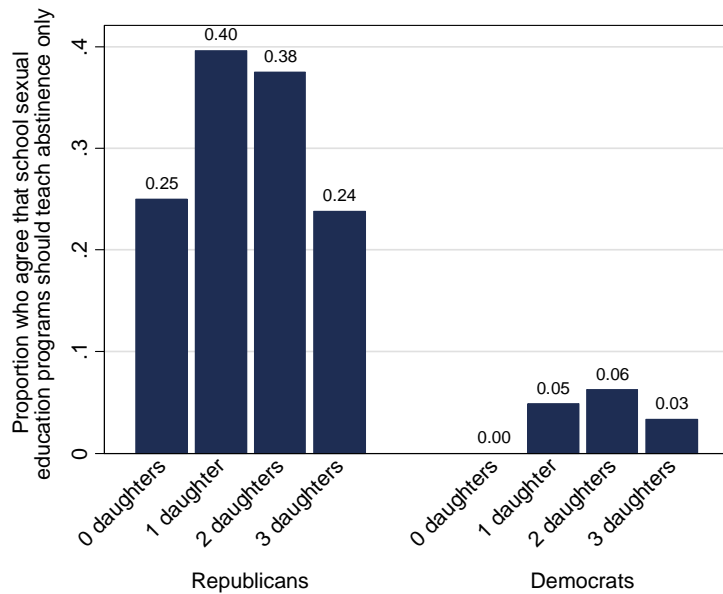


Figure 33: Proportion of candidates with three children who agree that school sex education programs should teach abstinence only

Notes: See Table 22 for a description of the sample. The sample sizes for each column are as follows: 16, 53, 40, 21, 23, 62, 48, and 30.

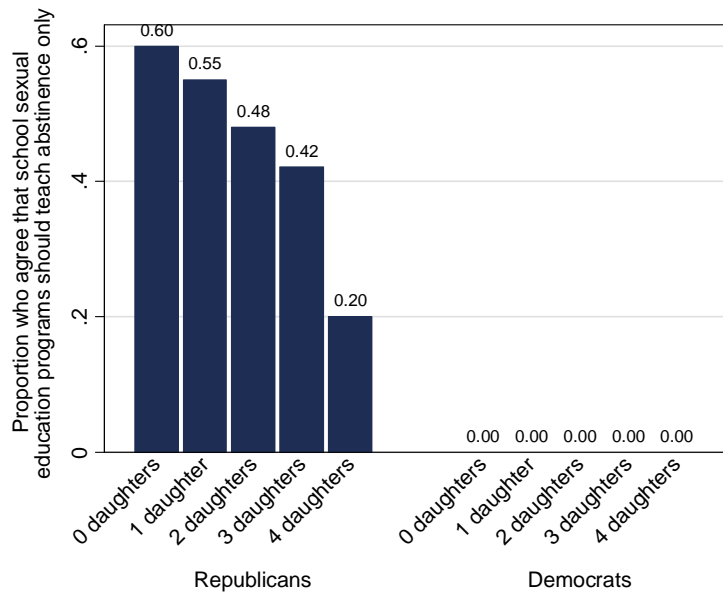


Figure 34: Proportion of candidates with four children who agree that school sex education programs should teach abstinence only

Notes: See Table 22 for a description of the sample. The sample sizes for each column are as follows: 5, 20, 25, 19, 5, 4, 15, 19, 11, and 7.

candidates with three children, candidates exhibit similar support for abstinence-only education regardless of whether they have no daughters or three daughters: 25 percent of candidates with no daughters and 24 percent of candidates with three daughters support abstinence-only education. Only Figure 34 provides suggestive evidence of a pattern: among Republican candidates with four children, support for abstinence-only education declines from 60 percent for candidates with no daughters, to 56 percent for candidates with one daughter, to 48 percent for candidates with two daughters, to 42 percent for candidates with three daughters, to 20 percent for candidates with four

daughters. However, not one Democratic candidate with four children expressed support for abstinence-only education, regardless of their number of daughters.

Together, Figures 25-34 do not suggest a consistent pattern between a candidates' number of daughters and his support for the statements on abortion or abstinence-only education, after conditioning on his number of children. These descriptive results are surprising given that they contradict Washington (2008), which suggests that candidates' support for these statements should decrease as their number of daughters increases.

3.4 Regression results

The regressions reported in Tables 23 and 24 also suggest that candidates with daughters tend to have the same views on abortion and abstinence-only education as candidates with sons. Each panel and each column present results from a separate regression. Panel A presents results from equation (1) which include an indicator variable that equals one if a candidate has at least one daughter; Panel B presents results from equation (2) which include a continuous variable that equals a candidate's number of daughters; and Panel C presents results from equation (3) which include a series of indicator variables that equal one if a candidate has at least one daughter in the stated age range. All regressions include number of children fixed effects, and regressions in Panel C also include age of children fixed effects. Column (1) presents results without controlling for candidate characteristics; columns (2) through (6) present results after

Table 23: Estimated effects of daughters on agreement that abortion should always be illegal

	(1)	(2)	(3)	(4)	(5)	(6)
	All	All	<i>Party</i>		<i>Gender</i>	
	candidates	candidates	Republicans	Democrats	Men	Women
Panel A. Effect of at least one daughter						
Daughter	0.138 (0.100)	0.251 (0.136)	0.302 (0.172)	0.212 (0.214)	0.328* (0.159)	0.014 (0.254)
Pseudo R-squared	0.011	0.136	0.050	0.048	0.127	0.178
Observations	1614	976	433	525	694	272
Panel B. Effect of one additional daughter						
Number of daughters	0.004 (0.043)	0.055 (0.058)	0.042 (0.077)	0.096 (0.097)	0.054 (0.066)	0.026 (0.123)
Pseudo R-squared	0.010	0.134	0.046	0.048	0.122	0.178
Observations	1614	976	433	525	694	272
Panel C. Effect of daughter by age range						
Daughter ages 0-4	0.197 (0.249)	0.223 (0.397)	0.695 (0.545)	†	0.146 (0.462)	†
Daughter ages 5-9	-0.007 (0.210)	-0.077 (0.271)	-0.012 (0.375)	†	-0.122 (0.307)	†
Daughter ages 10-13	-0.146 (0.215)	-0.195 (0.291)	-0.272 (0.361)	†	-0.164 (0.313)	†
Daughter ages 14-17	-0.041 (0.194)	0.157 (0.257)	0.203 (0.342)	0.322 (0.540)	0.118 (0.287)	0.427 (0.571)
Daughter ages 18-24	0.182 (0.142)	0.535** (0.193)	0.714** (0.245)	0.388 (0.317)	0.540* (0.221)	0.814 (0.445)
Daughter ages 25 or older	0.015 (0.112)	0.056 (0.150)	0.094 (0.199)	0.010 (0.217)	0.132 (0.172)	-0.121 (0.298)
Pseudo R-squared	0.012	0.142	0.071	0.054	0.132	0.220
Observations	1614	976	433	502	694	272
Include controls?	No	Yes	Yes	Yes	Yes	Yes
Include number of children FE?	Yes	Yes	Yes	Yes	Yes	Yes
Include age of children FE?	Panel C	Panel C	Panel C	Panel C	Panel C	Panel C

Notes: Each panel and each column represent results from a separate regression. The dependent variable in each regression is an indicator variable that equals one if the respondent agrees with the statement "abortion should always be illegal" and zero otherwise. All regressions include number of children fixed effects. Regressions reported in Panel C also include age of children fixed effects. Regressions reported in columns (2)-(6) include the following controls: candidate gender, race, party, age, and religion. Standard errors are in parentheses. * indicates significance at the 5 percent level; ** indicates significance at the 1 percent level; *** indicates significance at the 0.1 percent level. † indicates three or fewer affirmative responses to the dependent variable statement among candidates in this group.

Source: National Candidate Study, 2012

controlling for candidate gender, race, age, religion, and party.

Table 23 reports no statistically significant effects of having at least one daughter (Panel A) or of having an additional daughter relative to having an additional son (Panel B). The effect of having at least one daughter is greater for Republicans than for Democrats; Republicans with at least one daughter are 11.3 percentage points more likely to support the statement on abortion than Republicans with only sons, whereas Democrats with at least one daughter are only 4.2 percentage points more likely to support the statement on abortion than Democrats with only sons. These differences, however, are statistically insignificant. The only statistically significant effect in Panels A and B is the coefficient on the indicator variable for having a daughter when the sample is restricted to male candidates. Interestingly, a male candidate with at least one daughter is 9.9 percentage points *more* likely to support the statement that abortion should always be illegal than a male candidate with only sons, a finding which contradicts the existing literature.

Table 24: Estimated effects of daughters on agreement that school sex education programs should teach abstinence only

	(1)	(2)	(3)	(4)	(5)	(6)
	All	All	<i>Party</i>		<i>Gender</i>	
	candidates	candidates	Republicans	Democrats	Men	Women
Panel A. Effect of at least one daughter						
Daughter	0.015 (0.109)	0.187 (0.162)	0.207 (0.188)	0.037 (0.225)	0.300 (0.191)	-0.067 (0.305)
Pseudo R-squared	0.018	0.256	0.082	0.121	0.235	0.352
Observations	1604	949	427	379	674	268
Panel B. Effect of one additional daughter						
Number of daughters	-0.044 (0.045)	-0.015 (0.068)	-0.081 (0.081)	-0.015 (0.129)	0.004 (0.078)	-0.01 (0.162)
Pseudo R-squared	0.019	0.255	0.081	0.121	0.232	0.351
Observations	1604	949	427	379	674	268
Panel C. Effect of daughter by age range						
Daughter ages 0-4	0.312 (0.280)	0.580 (0.582)	0.748 (0.627)	†	0.380 (0.614)	†
Daughter ages 5-9	0.015 (0.221)	0.344 (0.328)	0.441 (0.393)	†	0.351 (0.360)	†
Daughter ages 10-13	-0.195 (0.233)	-0.062 (0.342)	-0.127 (0.403)	†	0.109 (0.392)	†
Daughter ages 14-17	0.117 (0.209)	-0.165 (0.292)	-0.178 (0.370)	†	-0.136 (0.341)	†
Daughter ages 18-24	0.046 (0.156)	-0.006 (0.229)	0.039 (0.261)	†	0.052 (0.264)	0.615 (0.451)
Daughter ages 25 or older	-0.050 (0.122)	0.166 (0.179)	0.126 (0.211)	0.242 (0.253)	0.285 (0.203)	-0.281 (0.487)
Pseudo R-squared	0.022	0.267	0.106	0.165	0.252	0.400
Observations	1604	949	427	332	674	268
Include controls?	No	Yes	Yes	Yes	Yes	Yes
Include number of children FE?	Yes	Yes	Yes	Yes	Yes	Yes
Include age of children FE?	Panel C	Panel C	Panel C	Panel C	Panel C	Panel C

Notes: Each panel and each column represent results from a separate regression. The dependent variable in each regression is an indicator variable that equals one if the respondent agrees with the statement "school sexual education programs should teach abstinence only" and zero otherwise. All regressions include number of children fixed effects. Regressions reported in Panel C also include age of children fixed effects. Regressions reported in columns (2)-(6) include the following controls: candidate gender, race, party, age, and religion. Standard errors are in parentheses. * indicates significance at the 5 percent level; ** indicates significance at the 1 percent level; *** indicates significance at the 0.1 percent level. † indicates three or fewer affirmative responses to the dependent variable statement among candidates in this group.

Source: National Candidate Study, 2012

Panel C of Table 23 reports few statistically significant effects of having a daughter in any given age range. There only statistically significant effect arises when candidates have a daughter between 18-24 years old. Candidates with at least one daughter in this age range are 16.5 percentage points more likely to support the statement on abortion than candidates with only sons in this age range, conditional on having the same number of children. Looking across specifications, this effect increases in magnitude and remains statistically significant after restricting the sample to Republican candidates or to male candidates. Republican candidates are 25.7 percentage points more likely to support the statement on abortion when they have a daughter in this age range than when they have only sons in this age range, and male candidates are 17.4 percentage points more likely to support the statement when they have a daughter in this age range than when they have only sons in this age range. Although Washington (2008) also identified large and statistically significant effects for males, she did not find differential effects by party.

Table 24 offers no evidence that the gender or age of candidates' children influences their views on abstinence-only education. There are no statistically significant effects of having at least one daughter, of having an additional daughter, or of having at least one daughter in any given age range. Republicans with at least one daughter are 6.9 percentage points more likely to support the statement on abortion than Republicans with only sons, whereas Democrats with at least one daughter are only 0.3 percentage points more likely to support the statement on abortion than Democrats with only sons. Both of these differences are, however, statistically insignificant. There are also no statistically significant effects when separating the results by party in specifications (3) and (4), or when separating the results by gender in specifications (5) and (6). The overall conclusions from Tables 23 and 24 are evident: the gender and ages of candidates' children has little or no effect on their views on abortion and abstinence-only education.

To determine if the survey responses to the questions on abortion and abstinence only sex education are correlated with other legislator characteristics, Appendix G provides the full results of the regressions reported in the second column of Tables 23 and 24. Democrats are 30 percentage points less likely to agree with the statement on abortion and 31 percentage points less likely to agree with the statement on abstinence only sex education than Republicans. Non-Catholic Christians are also 10 percentage points more likely to agree with the statement on abstinence only sex education than

Protestants (the omitted religious category). Other legislator characteristics are statistically insignificant.¹⁷

3.5 Discussion

The finding that children's gender has little effect on state legislative candidates' views runs counter to the existing evidence presented by Washington (2008). These results may differ for a few different reasons. First, the results here are based on a sample of 1,650 state legislative candidates while Washington's results were based on a sample of about 430 members of the U.S. House of Representatives. Our samples are different in three respects: (1) this sample is nearly four times as large as Washington's sample, (2) this sample includes candidates as well as elected officials, and (3) this sample consists of state public officials rather than federal public officials. To eliminate the possibility that unelected candidates are affecting these results, Tables 25 and 26 restrict the sample to elected state legislators. The results remain substantively the same, suggesting that children's characteristics do not affect candidates for public office

¹⁷ Washington (2008) finds that Republicans and Catholics are statistically significantly less likely to vote in support of women's rights legislation. Meanwhile, she finds that female legislators, legislators with "other" religious beliefs (i.e., legislators who are not Protestant, Catholic, other Christian, or not religious), and legislators from Congressional districts with a higher proportion of Democrats are statistically significantly more likely to vote in support of women's rights legislation (Table 2, page 320).

Table 25: Estimated effects of daughters on agreement that abortion should always be illegal for elected legislators only

	(1)	(2)	(3)	(4)	(5)	(6)
	All	All	<i>Party</i>		<i>Gender</i>	
	candidates	candidates	Republicans	Democrats	Men	Women
Panel A. Effect of at least one daughter						
Daughter	0.145 (0.137)	0.290 (0.160)	0.269 (0.202)	0.330 (0.268)	0.420* (0.192)	0.048 (0.290)
Pseudo R-squared	0.023	0.135	0.084	0.057	0.131	0.156
Observations	814	644	293	338	443	193
Panel B. Effect of one additional daughter						
Number of daughters	0.051 (0.060)	0.102 (0.069)	0.034 (0.092)	0.163 (0.117)	0.124 (0.080)	0.020 (0.147)
Pseudo R-squared	0.022	0.134	0.080	0.058	0.127	0.156
Observations	814	644	293	338	443	193
Panel C. Effect of daughter by age range						
Daughter ages 0-4	0.222 (0.429)	5.838*** (0.375)	6.008*** (0.639)	†	5.665*** (0.434)	†
Daughter ages 5-9	-0.002 (0.337)	-0.258 (0.395)	-0.013 (0.499)	†	-0.143 (0.461)	†
Daughter ages 10-13	-0.301 (0.339)	0.051 (0.412)	-0.240 (0.509)	†	0.234 (0.472)	†
Daughter ages 14-17	0.415 (0.278)	0.273 (0.329)	0.339 (0.414)	†	0.423 (0.361)	†
Daughter ages 18-24	0.255 (0.198)	0.507* (0.237)	0.703* (0.291)	0.050 (0.386)	0.604* (0.267)	0.810 (0.572)
Daughter ages 25 or older	0.090 (0.149)	0.113 (0.171)	0.174 (0.227)	0.020 (0.256)	0.224 (0.202)	-0.060 (0.338)
Pseudo R-squared	0.027	0.147	0.115	0.075	0.146	0.197
Observations	814	644	293	318	443	185
Include controls?	No	Yes	Yes	Yes	Yes	Yes
Include number of children FE?	Yes	Yes	Yes	Yes	Yes	Yes
Include age of children FE?	Panel C	Panel C	No	No	No	No

Notes: Each panel and each column represent results from a separate regression. The dependent variable in each regression is an indicator variable that equals one if the respondent agrees with the statement "abortion should always be illegal" and zero otherwise. All regressions include number of children fixed effects. Regressions reported in Panel C also include age of children fixed effects. Regressions reported in columns (2)-(6) include the following controls: candidate gender, race, party, age, and religion. Standard errors are in parentheses. * indicates significance at the 5 percent level; ** indicates significance at the 1 percent level; *** indicates significance at the 0.1 percent level. † indicates 3 or fewer affirmative responses to the dependent variable statement among candidates in this group.

Source: National Candidate Study, 2012

differentially from elected public officials.¹⁸ It remains possible that children's characteristics do affect federal officials differentially from state officials. However, given that many U.S. Representatives begin their political careers as state legislators, this explanation seems unsatisfying.

Another potential explanation for the difference between Washington's results and these results is that this paper measured candidates' views rather than their voting. As discussed previously, candidates' voting is affected by many factors other than their views, such as party positions, constituent preferences, and interest groups. Washington identified a statistically significant effect of parenting daughters on legislators' voting, but she was unable to identify the mechanism. She discusses two different hypotheses:

¹⁸ The only difference is that the coefficient on having a daughter between 0-4 years old is statistically significant in Tables 25 and 26, whereas it is statistically insignificant in Tables 23 and 24. However, these coefficients in Tables 25 and 26 are based on a very small sample of candidates who agree with the dependent variable statement. There are a total of six candidates who agree with the statement on abortion in specification (2) of Table 25 and only four candidates who agree with the statement on abstinence-only education in specification (2) of Table 26. Dividing this sample by gender and party yields even fewer observations. This result could also be attributable to chance; across the 36 coefficients in Panel C of each table, two coefficients could be statistically significant at the 0.95 level even when the independent variables are completely unrelated to the dependent variables.

Table 26: Estimated effects of daughters on agreement that school sex education programs should teach abstinence only for elected legislators only

	(1)	(2)	(3)	(4)	(5)	(6)
	All	All	<i>Party</i>		<i>Gender</i>	
	candidates	candidates	Republican	Democrats	Men	Women
Panel A. Effect of at least one daughter						
Daughter	0.085 (0.149)	0.193 (0.191)	0.225 (0.224)	-0.235 (0.356)	0.401 (0.230)	-0.534 (0.416)
Pseudo R-squared	0.028	0.294	0.113	0.291	0.268	0.442
Observations	806	618	283	195	423	188
Panel B. Effect of one additional daughter						
Number of daughters	0.012 (0.063)	0.058 (0.085)	-0.032 (0.100)	0.091 (0.187)	0.127 (0.093)	-0.178 (0.230)
Pseudo R-squared	0.028	0.294	0.111	0.290	0.265	0.438
Observations	806	618	283	195	423	188
Panel C. Effect of daughter by age range						
Daughter ages 0-4	0.853 (0.503)	5.288*** (0.518)	5.545*** (0.654)	†	5.912*** (0.610)	†
Daughter ages 5-9	-0.022 (0.341)	0.597 (0.457)	0.737 (0.527)	†	0.833 (0.549)	†
Daughter ages 10-13	-0.519 (0.375)	†	†	†	0.641 (0.635)	†
Daughter ages 14-17	0.244 (0.286)	0.216 (0.372)	0.302 (0.441)	†	0.299 (0.461)	†
Daughter ages 18-24	-0.022 (0.213)	0.036 (0.279)	0.036 (0.321)	†	0.016 (0.331)	†
Daughter ages 25 or older	0.074 (0.163)	0.211 (0.209)	0.236 (0.252)	0.148 (0.416)	0.384 (0.242)	-0.176 (0.579)
Pseudo R-squared	0.039	0.311	0.142	0.286	0.306	0.457
Observations	806	618	283	149	423	176
Include controls?	No	Yes	Yes	Yes	Yes	Yes
Include number of children FE?	Yes	Yes	Yes	Yes	Yes	Yes
Include age of children FE?	Panel C	Panel C	Panel C	Panel C	Panel C	Panel C

Notes: Each panel and each column represent results from a separate regression. The dependent variable in each regression is an indicator variable that equals one if the respondent agrees with the statement "school sexual education programs should teach abstinence only" and zero otherwise. All regressions include number of children fixed effects. Regressions reported in Panel C also include age of children fixed effects. Regressions reported in columns (2)-(6) include the following controls: candidate gender, race, party, age, and religion. Standard errors are in parentheses. * indicates significance at the 5 percent level; ** indicates significance at the 1 percent level; *** indicates significance at the 0.1 percent level. † indicates 3 or fewer affirmative responses to the dependent variable statement among candidates in this group.

Source: National Candidate Study, 2012

(1) the "preferences" hypothesis, where parenting daughters prompts a change in views, or (2) the "cost-based" hypothesis, where parenting daughters increases the cost of voting against reproductive rights legislation. Because this study directly surveys legislators on their views, the findings cast doubt on the "preferences" hypothesis. However, the "cost-based" hypothesis is consistent with both her results and the results presented here. Perhaps parenting daughters increases the cost of voting against reproductive rights legislation if daughters disagree with their fathers about the legislation or if stakeholders urge legislators to "think about your daughters" when casting their votes. It is further possible that interest groups are targeting their outreach to U.S. Representatives with daughters before votes on women's rights legislation or that constituents are more likely to contact their Representatives before votes on women's rights legislation if he has a daughter. Studies find that women are more aware of legislators' gender than men (Rosenthal 1995) and that constituents are more likely to contact legislators of the same race than legislators of a different race (Barrett

1995; Gay 2002; Tate 2003), so it is not implausible to consider that U.S. Representatives' family composition may affect the behavior of external actors. Any mechanism whereby voting differs but views remain the same would reconcile the difference between Washington's findings and the findings presented earlier.

A final possible explanation is that Washington's findings may be ungeneralizable to other samples. This caveat does not diminish the importance of her results. Parenting daughters influenced the voting records of U.S. Representatives during the 105th, 107th, and 108th Congresses. However, the present study should serve as a cautionary tale to other researchers who may try to generalize the results — from this study or from her study — to other samples and circumstances. The effect of children's gender on state legislators' views is simply not the same as the effect of children's gender on U.S. Representatives' voting. Although legislators' own characteristics produce robust differences in voting, it appears that legislators' children's characteristics may only matter in some circumstances.

3.6 Conclusions

Contrary to existing literature, this paper finds that the gender or age of children has little statistically significant effect on state legislative candidates' views on abortion or abstinence-only education. The divergence between these results and Washington's findings is most likely attributable to either (1) differences between views and voting, or (2) differences between state legislators and U.S. Representatives. If the divergence is

attributable to differences between views and voting, then this study casts doubt on the “preferences” mechanism (that parenting daughters prompts a change in views) and supports the “cost-based” mechanism (that parenting daughters increases the cost of voting against certain pieces of legislation). If the divergence is attributable to differences between state legislators and U.S. Representatives, then these results should caution other researchers from generalizing the findings of either study to other samples or circumstances.

When Anil Dash learned that Senator Rob Portman changed his stance on gay marriage, he tweeted: “Eventually one of these Republican congressmen is going to find out his daughter is a woman, and then we’re all set” (Drum 2013). If only it were that simple.

Appendix A

Table 27: Gender ratios in the top 5 percent and variance ratios across countries

Test	M-F ratio in top 5% in math	F-M ratio in top 5% in reading	Variance ratios on math tests			Variance ratios on reading tests
	PISA 2003 †	PISA 2003 †	PISA 2003†	TIMSS 2003‡	TIMSS 2007‡	PISA 2003†
Australia	1.67	1.72	1.18	1.16	1.20	1.25
Austria	1.80	2.22	1.20			1.23
Belgium	2.06	1.82	1.19	1.06		1.23
Brazil	1.74	1.59	1.20			1.23
Canada	1.77	1.64	1.24	1.06	1.19	1.26
Czech Republic	1.64	1.79	1.07		1.03	1.05
Denmark	1.39	1.72	0.99			1.11
Finland	1.72	2.78	1.22			1.26
France	1.59	2.50	1.19			1.22
Germany	1.76	1.75	1.12			1.19
Greece	2.14	1.56	1.23			1.33
Hong Kong	2.03	1.82	1.36	1.12	1.29	1.45
Hungary	1.92	1.67	1.10	1.08	1.07	1.11
Iceland	1.04	2.86	1.24			1.31
Indonesia	0.91	2.04	0.95	1.02	0.98	1.00
Ireland	1.79	2.04	1.07			1.09
Italy	2.16	1.89	1.27	1.14	1.08	1.31
Japan	2.25	1.19	1.29	1.19	1.02	1.25
Korea	2.55	1.15	1.10	1.10	1.14	1.09
Latvia	1.57	2.13	1.19	1.18		1.25
Lichtenstein	6.45	1.45	1.39			1.18
Luxembourg	1.98	1.72	1.16			1.23
Macao	2.26	1.09	1.26			1.18

2003

Mexico	1.43	1.47	1.08			1.07
Netherlands	1.34	1.52	1.00	1.03		1.08
New Zealand	1.84	1.41	1.16	1.20		1.14
Norway	1.63	2.13	1.20	1.09	1.17	1.26
Poland	1.81	1.89	1.27			1.29
Portugal	1.85	1.75	1.30			1.31
Russian Federation	1.88	1.41	1.20	1.14	1.16	1.28
Serbia	1.92	2.08	1.30	1.14	1.12	1.28
Slovakia	2.24	1.96	1.10	1.16		1.08
Spain	1.76	1.85	1.20	1.28	1.23	1.25
Sweden	1.44	2.08	1.10	1.06	1.12	1.11
Switzerland	2.09	1.82	1.11			1.15
Thailand	0.92	2.38	1.10			1.17
Tunisia	1.63	1.64	1.03	0.93	0.91	1.01
Turkey	1.90	1.00	1.22			1.28
United Kingdom	1.08	2.08	1.06	1.05	1.07	1.13
United States	1.72	1.67	1.19	1.11	1.08	1.17
Uruguay	1.60	1.61	1.10			1.20

Notes: The M-F ratio represents the male-female ratio in the top 5 percent on the math assessment. The F-M ratio represents the female-male ratio in the top 5 percent on the reading assessment. Variance ratios are calculated by dividing the variance of test scores among males by the variance of test scores among females. Variance ratios greater than one represent greater male variance. Countries with fewer than three entries across the six columns were excluded.

Sources: † indicates data are from Machin and Pekkarinen (2008), Tables S1 and S2 of the supporting online materials. ‡ indicates data are from Kane and Mertz (2012), Table 2.

Appendix B

In 1970, Richard Baker and James McConnell applied for a marriage license at the Hennepin County District Court in Minnesota. A clerk denied their request, and the men took their case to the courts. After a trial court ruled against them, they appealed the case to the Minnesota Supreme Court. In October 1971, the Minnesota Supreme Court upheld the trial court's ruling. The Court ruled, first, that the Minnesota statute governing marriage did not authorize same-sex marriage. While acknowledging that the statute did not expressly forbid same-sex marriage, the Court argued that the statute employed the term "marriage" as "one of common usage" and argued that the common usage of the term meant "the state of union between persons of the opposite sex" (291 Minn. 310). The Court cited definitions in Webster's Third New International Dictionary and Black's Law Dictionary as evidence, both of which described marriage as a union a man and a woman.

The Court also ruled that the statute was not unconstitutional, either on the grounds of due process or equal protection. The petitioners invoked the U.S. Supreme Court's *Loving v. Virginia* decision to support their case. In Chief Justice Warren's opinion for the Court in the *Loving* decision, he described marriage as "one of the 'basic civil rights of man,' fundamental to our very existence and survival" (503 U.S. 946). However, the Minnesota Supreme Court argued that "in commonsense and in a constitutional sense, there is a clear distinction between a marital restriction based

merely upon race and one based upon the fundamental difference in sex” (291 Minn. 310).

In sum, the Minnesota Supreme Court ruled that same-sex marriages were prohibited in the state. The Court explained: “The institution of marriage as a union of man and woman, uniquely involving the procreation and rearing of children within a family, is as old as the book of Genesis. ... This historic institution manifestly is more deeply founded than the asserted contemporary concept of marriage and societal interests for which petitioners contend” (291 Minn. 310). The plaintiffs appealed the decision to the U.S. Supreme Court, but the U.S. Supreme Court dismissed the appeal “for want of substantial federal question” (409 U.S. 810).

After the Baker decision, same-sex couples brought the issue of same-sex marriage to other state courts. During the 1970s and 1980s, state courts in Kentucky, Washington, and Pennsylvania heard cases on same-sex marriage and unanimously ruled that same-sex couples were prohibited from marrying (501 S.W.2d 588; 11 Wn. App. 247; 328 Pa. Super. 181). However, the tide changed in 1993 when the Supreme Court of Hawaii overturned a lower court’s decision to dismiss a case on same-sex marriage. The plaintiffs in the Baehr v. Lewin case alleged that the state’s refusal to grant marriage licenses to same-sex couples violated the Hawaii Constitution. The Supreme Court of Hawaii agreed and explained that the state’s marriage statute “denies same-sex couples access to the marital status and its concomitant rights and benefits,

thus implicating the equal protection clause” of the state’s constitution (74 Haw. 530). The Supreme Court of Hawaii vacated the lower court’s ruling and remanded the case back to the trial court “for further proceedings” (74 Haw. 530).

In response to the Baehr decision, the federal government passed the Defense of Marriage Act (DOMA) in 1996. The House committee report on the legislation described that “state courts in Hawaii appear to be on the verge of requiring that State to issue marriage licenses to same-sex couples” (House Report 104-664, page 2). Because the United States Constitution stipulates that “Full faith and credit shall be given in each state to the public acts, records, and judicial proceedings of every other state,” other states would have been required to recognize same-sex couples who obtained marriage licenses in Hawaii (U.S. Constitution, Article IV, Section 1). To prevent this situation, the Defense of Marriage Act stipulated that states were not required to abide by the Full Faith and Credit Act for “a relationship between persons of the same sex that is treated as a marriage under the laws of such other State” (110 Stat. 2419, page 1). The Defense of Marriage Act further defined marriage to include “only a legal union between one man and one woman as husband and wife” (110 Stat. 2419, page 1), thereby prohibiting same-sex marriages from being legally recognized by the federal government. The Act was overwhelmingly supported by both chambers of Congress — passing the House with a vote of 342-67 and the Senate with a vote of 85-14 — and was signed by President Clinton in September 1996.

Many states responded to the Defense of Marriage Act by amending their state constitutions to restrict same-sex marriages. McVeigh and Diaz (2009) report that ballot initiatives to prohibit same-sex marriage passed in 29 of the 30 states where the issue appeared on the ballot between 1998 and 2008.¹ However, same-sex marriage advocates were more successful in the courts than at the voting booths. In 2003, the Massachusetts Supreme Court ruled that the state could not “deny the protections, benefits, and obligations conferred by civil marriage to two individuals of the same sex who wish to marry” (440 Mass. 2309). In 2006, the New Jersey Supreme Court ruled that “under the equal protection guarantee of Article I, Paragraph 1 of the New Jersey Constitution, committed same-sex couples much be afforded on equal terms the same rights and benefits enjoyed by opposite-sex couples under the civil marriage statutes” (188 N.J. 415). Other state supreme courts followed suit, including Connecticut in 2008 (289 Conn. 135) and Iowa in 2009 (763 N.W.2d 862).

Perhaps the most notable court ruling on same-sex marriage, however, occurred in 2013 when the U.S. Supreme Court declared that the Defense of Marriage Act is unconstitutional under the due process clause of the U.S. Constitution. As Justice Kennedy explains in the *United States v. Windsor* decision, “DOMA’s principal effect is to identify and make unequal a subset of state-sanctioned marriages” (570 U.S.). Although states remain free to withhold marriage licenses to same-sex couples if they

¹ Voters in Arizona were the only exception (McVeigh and Diaz 2009: Table 1, page 892).

choose, the federal government is required to recognize marriages between gay and lesbian couples. Currently, 13 states and the District of Columbia allow same-sex marriage, while 35 states prohibit same-sex marriage through constitutional or statutory provisions (National Conference of State Legislatures).

Appendix C

Since Pope and Sydnor (2010) use NAEP test scores, I attempt to replicate their results using North Carolina standardized test scores to ensure the comparability of the two assessments. Pope and Sydnor document in their paper that North Carolina has a gender gap of 1.6, which distinguishes the state for having the third-lowest gender gap after only Massachusetts and New Mexico. Further, the authors generously provided detailed ratios for the state upon request. For North Carolina, they found a male-female ratio of 1.23 for math, a male-female ratio of 1.68 for science, and a female-male ratio of 1.91 for reading in the top 5 percent of the test score distribution.

To mirror Pope and Sydnor's approach as closely as possible, I pool observations on the math test from 2000, 2003, and 2005 and observations on the reading test from 2003 and 2005. However, North Carolina only began administering science tests in 2008, so I use science test scores from 2008 instead of pooling science test scores from 2000 and 2003.

Table 20 compares my findings to Pope and Sydnor's results. While I find exactly the same male-female ratio in math, I find a higher male-female ratio in science and a lower female-male ratio in reading. Averaging these three ratios, I find a gender gap of 1.53, slightly lower than Pope and Sydnor's gender gap of 1.61. However, given that my observations for science are from 2008 while Pope and Sydnor's observations for science are from 2000 and 2003, I also create an alternate measure of the gender gap that

Table 28: Comparison of gender ratios on NAEP tests and NC End of Grade tests

	Pope and Sydnor (2010)	Duch
<i>Gender ratios by test</i>		
Math (M-F)	1.23	1.23
Science (M-F)	1.68	1.94
Reading (F-M)	1.91	1.43
<i>Measures of gender gaps</i>		
Average of ratios, equal weighting	1.61	1.53
Average of math and reading ratios, equal weighting	1.57	1.33
Average of ratios, 25% math, 25% science, 50% reading	1.68	1.51
Geometric mean	1.58	1.51

Notes: The sample for Pope and Sydnor (2010) includes white students with scores in the top 5 percent on the 8th grade NAEP tests. Observations are pooled from 2000, 2003, and 2005 for the math test; from 2000 and 2003 for the science test; and from 2003 and 2005 for the reading test. The sample for Duch includes white students with scores in the top 5 percent on the 8th grade NC End of Grade tests. Observations are pooled from 2000, 2003, and 2005 for the math test; from 2008 for the science test; and from 2003 and 2005 for the reading test. Both samples omit test scores from students whose genders are missing.

Sources: Private communication with Devin Pope; North Carolina Education Research Data Center, End of Grade files.

averages the ratios from the math and reading tests only. I find a gender gap of 1.33 using math and reading tests only, while Pope and Sydnor would find a gender gap of 1.57 using math and reading tests only. This discrepancy is driven by the difference in reading test scores; it appears that 8th grade females have less of an advantage over males on the North Carolina EOG reading test than on the NAEP reading test.

Appendix D

Table 29: Descriptive statistics by area

	Number of counties	Number of students	Math, top 5%			Science, top 5%			Reading, top 5%			Average of ratios, top 5%
			(a) Male	(b) Female	(c) M-F	(a) Male	(b) Female	(c) M-F	(a) Male	(b) Female	(c) F-M	
Panel A. 5th grade												
Counties (average)	39	4,741	231	185	1.27	303	166	1.95	188	214	1.14	1.45
North Mountain	11	11,434	388	309	1.26	652	318	2.05	307	357	1.16	1.49
South Mountain	8	5,982	164	141	1.16	401	224	1.79	168	218	1.30	1.42
North Piedmont	9	7,992	224	188	1.19	375	188	1.99	182	204	1.12	1.44
South Piedmont	11	11,235	351	306	1.15	525	283	1.86	272	312	1.15	1.38
North Coastal	14	7,922	219	178	1.23	415	183	2.27	200	245	1.23	1.57
South Coastal	8	6,031	204	162	1.26	333	192	1.73	170	207	1.22	1.40
Panel B. 8th grade												
Counties (average)	39	4,665	211	197	1.06	276	168	1.73	201	240	1.22	1.33
North Mountain	11	11,489	382	386	0.99	464	274	1.69	392	499	1.27	1.32
South Mountain	8	5,939	144	176	0.82	216	158	1.37	179	223	1.25	1.14
North Piedmont	9	8,152	188	147	1.28	304	163	1.87	200	282	1.41	1.52
South Piedmont	11	11,025	266	231	1.15	345	169	2.04	301	359	1.19	1.46
North Coastal	14	7,847	194	151	1.28	289	151	1.91	240	289	1.20	1.47
South Coastal	8	6,089	188	168	1.12	225	134	1.68	187	234	1.25	1.35

Notes: See Table 3 for a description of the sample. Figures in the rows for counties reflect the average across the 39 counties. Figures in the rows for regions reflect the actual numbers from each region. For example, the columns (a) and (b) for each subject reflect the average number of males or females who score in the top five percent in the rows for the counties and the actual number of males and females who score in the top five percent in the rows for the regions.

Source: North Carolina Education Research Data Center, End of Grade files.

Appendix E

Table 30: Descriptive statistics

	Mean	Standard deviation	Minimum	Maximum
Panel A. Dependent variables for 5th grade				
Arithmetic average of ratios	1.45	0.13	1.23	1.94
Weighted average of ratios	1.38	0.11	1.19	1.70
Geometric mean of ratios	1.40	0.11	1.23	1.78
Math ratio (M:F)	1.26	0.19	0.90	2.03
Science ratio (M:F)	1.95	0.33	1.35	2.80
Reading ratio (F:M)	1.15	0.20	0.60	1.66
Panel B. Dependent variables for 8th grade				
Arithmetic average of ratios	1.34	0.10	1.14	1.61
Weighted average of ratios	1.31	0.08	1.17	1.52
Geometric mean of ratios	1.30	0.09	1.12	1.52
Math ratio (M:F)	1.06	0.17	0.76	1.56
Science ratio (M:F)	1.73	0.22	1.37	2.36
Reading ratio (F:M)	1.23	0.13	0.99	1.73
Panel C. Independent variables				
Support for Amendment One as a proportion of voters	0.68	0.13	0.21	0.83
Support for 2012 Republican Presidential candidate as a proportion of voters	0.57	0.12	0.23	0.74
Support for 2008 Republican Presidential candidate as a proportion of voters	0.55	0.11	0.24	0.71
Evangelical adherents as a proportion of all adherents	0.61	0.12	0.28	0.85
Evangelical adherents as a proportion of all adherents except Black Protestants	0.64	0.12	0.29	0.87
Number of students	5,200	3,174	2,017	17,783
Median household income (\$1,000s)	42.709	6.026	34.886	64.486

Notes: The sample includes 39 large counties and 6 county groups. The math ratio is defined as the male-female ratio in the top 5 percent; the science ratio is defined as the male-female ratio in the top 5 percent; and the reading ratio is defined as the female-male ratio in the top 5 percent. The arithmetic average of ratios is the arithmetic average of the math, science, and reading ratios. The weighted average of ratios is the average of the math, science, and reading ratios, with the math and science ratios weighted at 25 percent and the reading ratio weighted at 50 percent. The geometric mean of ratios is the geometric mean of the math, science, and reading ratios.

Sources: North Carolina Education Research Data Center, End of Grade files; North Carolina Board of Elections; U.S. Religious Census.

Appendix F

Table 31: Estimated relationship between the average of gender ratios in the top 5 percent and proxy variables without controls for median household income

	Average of gender ratios in the top 5% across tests				
	(1)	(2)	(3)	(4)	(5)
Panel A. 5th grade					
Support for Amendment One as a proportion of voters	0.246*				
	(0.112)				
Support for 2012 Republican Presidential candidate as a proportion of voters		0.218			
		(0.164)			
Support for 2008 Republican Presidential candidate as a proportion of voters			0.242		
			(0.172)		
Evangelical adherents as a proportion of all adherents				0.016	
				(0.186)	
Evangelical adherents as a proportion of all adherents except Black Protestants					0.062
					(0.186)
R ²	0.072	0.046	0.048	0.013	0.016
Observations	45	45	45	45	45
Control for 5th grade enrollment?	Yes	Yes	Yes	Yes	Yes
Control for median household income?	No	No	No	No	No
Panel B. 8th grade					
Support for Amendment One as a proportion of voters	0.257***				
	(0.063)				
Support for 2012 Republican Presidential candidate as a proportion of voters		0.200			
		(0.101)			
Support for 2008 Republican Presidential candidate as a proportion of voters			0.240*		
			(0.102)		
Evangelical adherents as a proportion of all adherents				0.055	
				(0.129)	
Evangelical adherents as a proportion of all adherents except Black Protestants					0.109
					(0.122)
R ²	0.114	0.050	0.061	0.004	0.017
Observations	45	45	45	45	45
Control for 8th grade enrollment?	Yes	Yes	Yes	Yes	Yes
Control for median household income?	No	No	No	No	No

Notes: Each column represents results from a separate regression. See Table 5 for a definition of the average of gender ratios in the top 5 percent. The sample for each entry includes 39 large counties and 6 county groups. Standard errors are in parentheses. * indicates significance at the 5 percent level; ** indicates significance at the 1 percent level; *** indicates significance at 0.1 percent level.

Sources: North Carolina Education Research Data Center, End of Grade files; North Carolina Board of Elections; U.S. Religious Census

Appendix G

Table 32: Estimated effects of daughters on agreement with statements on abortion and abstinence only sex education

Dependent variable	(1) Statement on abortion	(2) Statement on abstinence only
Sample	All candidates	All candidates
Number of daughters	0.055 (0.058)	-0.015 (0.068)
<i>Other legislator characteristics</i>		
Female	-0.115 (0.110)	-0.157 (0.144)
White	-0.197 (0.156)	-0.248 (0.196)
Democrat	-0.978*** (0.099)	-1.535*** (0.137)
18-25 years old	0.039 (0.389)	-0.685 (0.519)
26-34 years old	0.228 (0.216)	-0.245 (0.285)
35-44 years old	-0.008 (0.157)	-0.116 (0.197)
45-54 years old	0.001 (0.136)	-0.046 (0.164)
55-65 years old	0.124 (0.122)	0.147 (0.146)
No religion	-0.060 (0.130)	-0.253 (0.149)
Catholic	0.332 (0.177)	0.119 (0.213)
Other Christian	0.186 (0.176)	0.447* (0.193)
Other religion	-0.183 (0.372)	†

Pseudo R-squared	0.134	0.255
Observations	976	949
Include number of children FE?	Yes	Yes
Include age of children FE?	No	No

Notes: Each column represents results from a separate regression. The dependent variable for the regression reported in column (1) is an indicator variable that equals one if the respondent agrees with the statement "abortion should always be illegal" and zero otherwise. The dependent variable for the regression reported in column (2) is an indicator variable that equals one if the respondent agrees with the statement "school sexual education programs should teach abstinence only" and zero otherwise. Both regressions include number of children fixed effects. Standard errors are in parentheses. * indicates significance at the 5 percent level; ** indicates significance at the 1 percent level; *** indicates significance at the 0.1 percent level. † indicates 3 or fewer affirmative responses to the dependent variable statement among candidates in this group.

Source: National Candidate Study, 2012

References

- American Association of University Women. 2008. *Where the Girls Are: The Facts about Gender Equity in Education*. Washington DC: American Association of University Women Educational Foundation.
- Ananat, Elizabeth O., and Guy Michaels. 2008. "The effect of marital breakup on the income and poverty of women with children." *Journal of Human Resources*, 43(3), 611-629.
- Angrist, Joshua and Victor Lavy. 2002. "New evidence on classroom computers and pupil learning." *The Economic Journal*, 112, 735-765.
- Arden, Rosalind and Robert Plomin. 2006. "Sex differences in variance of intelligence across childhood." *Personality and Individual Differences*, 41, 39-48.
- Arnold, R. Douglas. 1990. *The Logic of Congressional Action*. Yale University Press: New Haven, CT.
- Attewell, Paul and Juan Battle. 1999. "Home computers and school performance." *The Information Society*, 15(1), 1-10.
- Association of Statisticians of American Religious Bodies. 2012. "U.S. Religious Census: Religious Congregations and Membership Study, 2010." <<http://www.thearda.com/Archive/Files/Descriptions/RMSCY10.asp>>
- Baker, David P., and Deborah Perkins Jones. 1993. "Creating gender equality: Cross-national gender stratification and mathematical performance." *American Sociological Association*, 66(2), 91-103.
- Barot, Sneha. 2012. "A problem-and-solution mismatch: Son preference and sex-selective abortion bans." *Guttmacher Policy Review*, 15(2), 18-22.
- Barrett, Edith J. 1995. "The policy priorities of African American women in state legislatures." *Legislative Studies Quarterly*, 20(2), 223-247.
- Barrow, Lisa, Lisa Markman, and Cecilia Rouse. 2009. "Technology's edge: The educational benefits of computer-aided instruction." *American Economic Journal: Economic Policy*, 1(1), 52-74.

- Benbow, Camilla P. 1992. "Academic achievement in math and science of students between 13 and 23: Are there differences among students in the top 1% of math ability?" *Journal of Educational Psychology*, 84, 51-61.
- Benbow, Camilla P., David Lubinski, Daniel L. Shea, and Hossain Eftekhari-Sanjani. 2000. "Sex differences in mathematical reasoning ability: Their status 20 years later." *Psychological Science*, 11(6), 474-480.
- Benbow, Camilla P. and Julian C. Stanley. 1983. "Sex differences in mathematical reasoning ability: More facts." *Science*, 222(4627), 1029-1031.
- . 1980. "Sex differences in mathematical ability: Fact or artifact?" *Science*, 210(4475), 1262-1264.
- Berkman, Michael B., and Robert E. O'Connor. 1993. "Do women legislators matter? Female legislators and state abortion policy." *American Politics Quarterly*, 21, 102-124.
- Bettinger, Eric and Bridget Terry Long. 2005. "Do faculty serve as role models? The impact of instructor gender on female students." *American Economic Review Papers and Proceedings*, 152-157.
- Bleakley, Amy, Michael Hennessy, and Martin Fishbein. 2006. "Public opinion on sex education in US schools." *Archives of Pediatrics & Adolescent Medicine*, 160(11), 1151-1156.
- Bolzendahl, Catherine, and Clem Brooks. 2007. "Women's political representation and welfare state spending in twelve capitalist democracies." *Social Forces*, 85, 1509-1534.
- Brandon, Paul R., Barbara J. Newton, and Ormond W. Hammond. 1987. "Children's mathematics achievement in Hawaii: Sex differences favoring girls." *American Educational Research Journal*, 24(3), 437-461.
- Bratton, Kathleen A., and Kerry L. Haynie. 1999. "Agenda setting and legislative success in state legislatures: The effects of gender and race." *Journal of Politics*, 61, 658-679.
- Brewer, Paul R. 2003. "The shifting foundations of public opinion about gay rights." *Journal of Politics*, 65(4), 1208-20.

- Brint, Steven, and Seth Abrutyn. 2010. "Who's right about the Right? Comparing competing explanations of the link between White Evangelicals and conservative politics in the United States." *Journal for the Scientific Study of Religion*, 49(2), 329-350.
- Broockman, David E., Nicholas Carnes, Melody Crowder-Meyer, and Christopher Skovron. 2013. "2012 National Candidate Study." Computer file.
- Broockman, David E., and Christopher Skovron. 2013. "What politicians believe about their constituents: Asymmetric misperceptions and prospects for constituency control." Working Paper.
<http://www.ocf.berkeley.edu/~broockma/broockman_skovron_asymmetric_misperceptions.pdf>
- Burden, Barry. 2007. *Personal Roots of Representation*. Princeton University Press: Princeton, NJ.
- Campbell, David, and Carin Larson. 2007. "Religious coalitions for and against gay marriage: The culture war wages on." In Craig A. Rimmerman and Clyde Wilcox (eds.), *The Politics of Same-Sex Marriage*. University of Chicago Press: Chicago, IL. 131-54.
- Canon, David T. 1990. *Race, Redistricting, and Representation: The Unintended Consequences of Black Majority Districts*. The University of Chicago Press: Chicago, IL.
- Carnes, Nicholas. 2012. "Does the numerical underrepresentation of the working class in Congress matter?" *Legislative Studies Quarterly*, 37(1), 5-34.
- Carrell, Scott, Marianne Page, and James West. 2010. "Sex and science: how professor gender perpetuates the gender gap." *Quarterly Journal of Economics*, 125(3), 1101-1144.
- Ceci, Stephen J. and Wendy M. Williams. 2011. "Understanding current causes of women's underrepresentation in science." *Proceedings of the National Academy of Sciences*, 108(8), 3157-1362.
- Ceci, Stephen J., Wendy W. Williams, and Susan M. Barnett. 2009. "Women's underrepresentation in science: Sociocultural and biological considerations." *Psychological Bulletin*, 135(2), 218-261.

- Center for American Women and Politics. 2012. "Women who will be serving in 2013." <http://cawp.rutgers.edu/fast_facts/elections/2013_womenserving.php>
- Charles, Maria and Karen Bradley. 2006. "A matter of degrees: Female underrepresentation in computer science programs cross-nationally." In Joanne McGrath Cohoon and Bill Aspray, eds. *Women and Information Technology: Research on the Reasons for Underrepresentation*. Cambridge, MA: MIT Press. 183-203.
- Chattopadhyay, Raghavendra, and Esther Duflo. 2004. "Women as policy makers: evidence from a randomized policy experiment in India." *Econometrica*, 72(5), 1409-1443.
- Cox, Gary W., and Mathew D. McCubbins. 1993. *Legislative Leviathan: Party Government in the House*. University of California Press: Berkeley, CA.
- Dahl, Michael S., Cristian L. Dezso, and David G. Ross. 2012. "Fatherhood and managerial style: How a male CEO's children affect the wages of his employees." *Administrative Science Quarterly*, 57(4), 669-693.
- Dahl, Gordon B., and Enrico Moretti. 2004. "The demand for sons." *Review of Economic Studies*, 75(4), 1085-1120.
- Dar-Nimrod, Ilan, and Steven J. Heine. 2006. "Exposure to scientific theories affects women's math performance." *Science*, 314(5798), 435.
- Davies, Paul G., Steven J. Spencer, Dianne M. Quinn, and Rebecca Gerhardstein. 2002. "Consuming images: How television commercials that elicit stereotype threat can restrain women academically and professionally." *Personality and Social Psychology Bulletin*, 28(12), 1615-28.
- Deary, Ian, Graham Thorpe, Valerie Wilson, John Starr, and Lawrence Whalley. 2003. "Population sex differences in IQ at age 11: the Scottish mental survey 1932." *Intelligence*, 31, 533-542.
- Dee, Thomas S. 2005. "A teacher like me: Does race, ethnicity, or gender matter?" *American Economic Review: Papers and Proceedings*, 95(2), 158-165.
- Downey, Douglas B., Pamela B. Jackson, and Brian Powell. 1994. "Sons versus daughters: Sex composition of children and maternal views on socialization." *Sociological Quarterly*, 35(1), 33-50.

- Drum, Kevin. 2013. "Members of Congress with daughters vote better on women's issues." *Mother Jones*, 15 March. <<http://www.motherjones.com/kevin-drum/2013/03/members-congress-daughters-vote-better-womens-issues>>
- Duckitt, John, Boris Bizumic, Stephen W. Krauss, and Edna Heled. 2010. "A tripartite approach to right-wing authoritarianism: The authoritarianism-conservatism-traditionalism model." *Political Psychology*, 31(5), 685-715.
- Dye, Tomas R., and James Renick. 1981. "Political power and city jobs: Determinants of minority employment." *Social Science Quarterly*, 62, 475-486.
- Dynarski, Mark et al. 2007. "Effectiveness of reading and mathematics software products: findings from the first student cohort." U.S. Department of Education: National Center for Education Evaluation and Regional Assistance.
- Edlund, Lena C., and Rohini Pande. 2002. "Why have women become left-wing? The political gender gap and the decline in marriage." *Quarterly Journal of Economics*, 117(3), 917-961.
- Fairlie, Robert. 2005. "The effects of home computers on school enrollment." *Economics of Education Review*, 24(5), 533-547.
- Fairlie, Robert, Daniel Beltran, and Kuntal Das. 2010. "Home computers and educational outcomes: Evidence from the NLSY97 and CPS." *Economic Inquiry*, 48(3), 771-792.
- Fairlie, Robert and Rebecca London. 2012. "The effects of home computers on educational outcomes: Evidence from a field experiment with community college students." *Economic Journal*, 122(561), 727-753.
- Farrell, Elizabeth. 2013. "10 lessons from the best district in the county." *Administrator*, spring edition.
<http://www.nxtbook.com/nxtbooks/scholastic/administrator_2013spring/#/50>
- Federal Election Commission. 2012. "Official 2012 Presidential General Election Results."
<<http://www.fec.gov/pubrec/fe2012/2012presgeresults.pdf>>
- . 2008. "2008 Official Presidential General Election Results."
<<http://www.fec.gov/pubrec/fe2008/2008presgeresults.pdf>>

- Feingold, Alan. 1994. "Gender differences in variability in intellectual abilities: A cross-cultural perspective." *Sex Roles*, 30(1-2), 81-92.
- . 1992. "Sex differences in variability in intellectual abilities: A new look at an old controversy." *Review of Educational Research*, 62, 61-84.
- Figlio, David. 2005. "What Barbie says 'math is hard.'" Working paper, University of Florida and National Bureau of Economic Research.
- Friedman, Lynn. 1989. "Mathematics and the gender gap: A meta-analysis of recent studies on sex differences in quantitative tasks." *Review of Educational Research*, 59(2), 185-213.
- Fuchs, Thomas, and Ludger Woessmann. 2005. "Computers and student learning: Bivariate and multivariate evidence on the availability and use of computers at home and at school." Ifo Working Paper No. 8.
- Gaines, N. Susan, and James C. Garand. 2010. "Morality, equality, or locality: Analyzing the determinants of support for same-sex marriage." *Political Research Quarterly*, 63(3), 553-567.
- Gay, Claudine. 2002. "Spirals of trust? The effect of descriptive representation on the relationship between citizens and their government." *American Journal of Political Science*, 46(4), 717-732.
- 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.
- Goolsbee, Austan and Jonathan Guryan. 2006. "The impact of Internet subsidies in public schools." *The Review of Economics and Statistics*, 88(2), 336-347.
- Guiso, Luigi, Ferdinando Monte, Paola Sapienza, and Luigi Zingales. 2008. "Culture, Gender, and Math." *Science*, 320(5880), 1164-65.
- Hall, Richard L., and Frank W. Wayman. 1990. "Buying time: Moneyed interests and the mobilization of bias in Congressional committees." *American Political Science Review*, 84, 797-820.

- Hayes, Matthew, Matthew V. Hibbing, and Tracy Sulkin. 2010. "Redistricting, responsiveness, and issue attention." *Legislative Studies Quarterly*, 35(1), 91-115.
- Hedges, Larry V. and Amy Nowell. 1995. "Sex differences in mental test scores, variability, and numbers of high-scoring individuals." *Science*, 269(5220), 41-45.
- Helms, Ann Doss. 2013. "Mooresville's Mark Edwards named school superintendent of year." *Charlotte Observer*, 15 May.
- Hibbing, John and David Marsh. 1987. "Accounting for the voting patterns of British MP's on free votes." *Legislative Studies Quarterly*, 12(2), 275-297.
- Hoffmann, Florian and Philip Oreopoulos. 2009. "A professor like me: the influence of instructor gender on university achievement." *Journal of Human Resources*, 44(2), 479-494.
- Hyde, Janet S. 2005. "The gender similarities hypothesis." *American Psychologist*, 60(6), 581-592.
- Hyde, Janet S., Sara M. Lindberg, Marcia C. Linn, Amy B. Ellis, and Caroline C. Williams. 2008. "Gender similarities characterize math performance." *Science*, 321, 494-495.
- Hyde, Janet S. and Marcia C. Linn. 2006. "Mathematics and Science." *Science*, 314, 599-600.
- Hyde, Janet S. and Janet E. Mertz. 2009. "Gender, culture, and mathematics performance." *Proceedings of the National Academy of Sciences of the United States of America*, 106(22), 8801-8807.
- Inzlicht, Michael, and Talia Ben-Zeev. 2000. "A threatening environment: Why females are susceptible to experiencing problem-solving deficits in the presence of males." *Psychological Science*, 11(5), 365-371.
- Judd, Charles M. and Bernadette Park. 1993. "Definition and assessment of accuracy in social stereotypes." *Psychological Review*, 100(1), 109-128.
- Kane, Jonathan M. and Janet E. Mertz. 2012. "Debunking myths about gender and mathematics performance." *Notices of the AMS*, 59(1), 10-21.

- Knuckey, Jonathan. 2005. "A new front in the culture war? Moral traditionalism and voting behavior in U.S. House Elections." *American Politics Research*, 33(5), 645-671.
- Leahey, Erin and Guang Guo. 2001. "Gender differences in math trajectories." *Social Forces*, 80(2), 713-732.
- Levitt, Steven. 1996. "How do Senators vote? Disentangling the role of voter preferences, party affiliation, and Senate ideology." *American Economic Review*, 86(3), 425-441.
- Li, Xiaoming, Melissa Atkins, and Bonita Stanton. 2006. "Effects of home and school computer use on school readiness and cognitive development among Head Start children: a randomized controlled pilot trial." *Merrill-Palmer Quarterly*, 52(2), 239-263.
- Lohman, David F. and Joni M. Lakin. 2009. "Consistencies in sex differences on the cognitive abilities test across countries, grades, test forms, and cohorts." *British Journal of Educational Psychology*, 79(2), 389-407.
- Malamud, Ofer and Christian Pop-Eleches. 2011. "Home computer use and the development of human capital." *The Quarterly Journal of Economics*, 126, 987-1027.
- Mayhew, David R. 1974. *Congress: The Electoral Connection*. Yale University Press: New Haven, CT.
- Meer, Jonathan and Harvey Rosen. 2009. "Altruism and the child-cycle of alumni donations." *American Economic Journal: Economic Policy*, 1(1), 258-286.
- Meier, Kenneth J., and Robert E. England. 1984. "Black representation and educational policy: Are they related?" *The American Political Science Review*, 78, 392-403.
- Monastersky, Rich. 2005. "Primed for numbers." *Chronicle of Higher Education*, 4 March.
- Morgan, Pat and Steven Ritter. 2002. "An experimental study of the effects of Cognitive Tutor Algebra I on student knowledge and attitude."
<http://www.carnegielearning.com/web_docs/morgan_ritter_2002.pdf>
- National Conference of State Legislatures. "Legislator demographics."
<<http://www.ncsl.org/legislatures-elections/legisdata/legislator-demographics.aspx>>

- National Public Radio, Kaiser Family Foundation and John F. Kennedy School of Government. 2004. "Sex education in America: Principals survey." <<http://www.npr.org/programs/morning/features/2004/jan/kaiserpoll/principalsfinal.pdf>>
- . 2004a. "Sex education in America: General public/parents survey." <<http://www.npr.org/programs/morning/features/2004/jan/kaiserpoll/publicfinal.pdf>>
- National Science Foundation, Division of Science Resources Statistics. 2011. "S&E Degrees: 1966-2008." <<http://www.nsf.gov/statistics/nsf11316/>>
- . 2008. "Thirty-Three Years of Women in S&E Faculty Positions." <<http://www.nsf.gov/statistics/infbrief/nsf08308/>>
- Nosek, Brian A. et al. 2008. "National differences in gender-science stereotypes predict national sex differences in science and math achievement." *Proceedings of the National Academy of Sciences*, 106(26), 10593-10597.
- Nunnery, Ross, and McDonald. 2006. "A randomized experimental evaluation of the impact of Accelerated Reader/Reading Renaissance implementation on reading achievement in grades 3 to 6." *Journal of Education for Students Placed at Risk*, 11(1), 1-18.
- Office of Science Technology Policy. 2011. "Women in STEM." <<http://www.whitehouse.gov/administration/eop/ostp/women>>
- Oswald, Andrew J. and Nattavudh Powdthavee. 2010. "Daughters and left-wing voting." *The Review of Economics and Statistics*, 92(2), 213-227.
- Pande, Rohini. 2003. "Can mandated political representation increase policy influence for disadvantaged minorities? Theory and evidence from India." *American Economic Review*, 94(4), 1132-1151.
- Penner, Andrew M. 2008. "Gender differences in extreme mathematics achievement: An international perspective on biological and social factors." *American Journal of Sociology*, 114(S1), S138-S170.
- . 2008a. "Appendix A." Available upon request from author. 146-166.

- Penner, Andrew M. and Marcel Paret. 2008. "Gender differences in mathematics achievement: Exploring the early grades and the extremes." *Social Science Research*, 37, 239-253.
- Pew Research Center. 2012. "More support for gun rights, gay marriage than in 2008, 2004." <<http://www.people-press.org/2012/04/25/more-support-for-gun-rights-gay-marriage-than-in-2008-or-2004/>>
- Pew Forum on Religion and Public Life. 2008. "U.S. Religious Landscape Survey: Summary of key findings." <<http://religions.pewforum.org/pdf/report2religious-landscape-study-key-findings.pdf>>
- Pope, Devin and Justin Sydnor. 2010. "Geographic variation in the gender differences in test scores." *Journal of Economic Perspectives*, 24(2), 95-108.
- Porterfield, Kitty, Karen Cutler, Kelly Banaszak, and Mary Ducuma. 2013. "Mark A. Edwards named AASA 2013 Superintendent of the Year." *American Association of School Administrators*, 21 February. <<http://aasa.org/content.aspx?id=27088>>
- Portman, Rob. 2013. "Gay couples also deserve chance to get married." *The Columbus Dispatch*, 15 March. <<http://www.dispatch.com/content/stories/editorials/2013/03/15/gay-couples-also-deserve-chance-to-get-married.html>>
- Quillen, Ian. 2011. "Building the digital district." *Education Week*, 19 October.
- Reingold, Beth. 1992. "Concepts of representation among female and male state legislators." *Legislative Studies Quarterly*, 17, 509-537.
- Reis, Sally M. and Sunghee Park. 2001. "Gender differences in high-achieving students in math and science." *Journal for the Education of the Gifted*, 25(1), 52-73.
- Rosenthal, Cindy. 1995. "The role of gender in descriptive representation." *Political Research Quarterly*, 48(3), 599-611.
- Rouse, Cecilia, and Alan Krueger. 2004. "Putting computerized instruction to the test: A randomized evaluation of a 'scientifically based' reading program." *Economics of Education Review*, 23(4), 323-338.

- Saad, Lydia. 2013. "Americans' abortion views steady amid Gosnell trial." *Gallup: Politics*. <<http://www.gallup.com/poll/162374/americans-abortion-views-steady-amid-gosnell-trial.aspx>>
- Schmitt, John, and Jonathan Wadsworth. 2006. "Is there an impact of household computer ownership on children's educational attainment in Britain?" *Economics of Education Review*, 25(6), 659-673.
- Schratz, Marjorie M. 1978. "A developmental investigation of sex differences in spatial (visual-analytic) and mathematical skills in three ethnic groups." *Developmental Psychology*, 14(3), 263-267.
- Schwartz, Alan. 2012. "Mooresville's shining example (it's not just about the laptops)." *The New York Times*, 12 February.
- Shapley, Kelly, Daniel Sheehan, Catherine Maloney, and Fanny Caranikas-Walker. 2010. "Evaluating the implementation fidelity of technology immersion and its relationship with student achievement." *Journal of Technology, Learning, and Assessment*, 9(4), 1-69.
- . 2009. "Evaluation of the Texas Technology Immersion Project: Final outcomes for a four-year study (2004-05 to 2007-08)." Texas Center for Educational Research. <<http://www.tcer.org/research/etxtip/>>
- . 2007. "Evaluation of the Texas Technology Immersion Project: Findings from the second year." Texas Center for Educational Research. <<http://www.tcer.org/research/etxtip/>>
- Sommers, Christina Hoff. 2000. "The war against boys." *The Atlantic Magazine*. 1 May. <<http://www.theatlantic.com/magazine/archive/2000/05/the-war-against-boys/304659>>
- Spelke, Elizabeth S. 2005. "Sex differences in intrinsic aptitude for mathematics and science? A critical review." *American Psychologist*, 60(9), 950-958.
- Spencer, Steven J., Claude M. Steele, and Dianne M. Quinn. 1999. "Stereotype threat and women's math performance." *Journal of Experimental Social Psychology*, 35(1), 4-28.

- Stadelmann, David, Marco Portmann, and Reiner Eichenberger. 2013. "Quantifying parliamentary representation of constituents' preferences with quasi-experimental data." *Journal of Comparative Economics*, 41(1), 170-180.
- Steele, Claude M. 1997. "A threat in the air: How stereotypes shape intellectual identity and performance." *American Psychologist*, 52(6), 613-629.
- Steele, Claude M. and Joshua Aronson. 1995. "Stereotype threat and the intellectual test performance of African Americans." *Journal of Personality and Social Psychology*, 69(5), 797-811.
- Strand, Steve, Ian J. Deary, and Pauline Smith. 2006. "Sex differences in cognitive abilities test scores: A UK National Picture." *British Journal of Educational Psychology*, 76(3), 463-480.
- Stratmann, Thomas. 2000. "Congressional voting over legislative careers: Shifting position and changing constraints." *American Political Science Review*, 94(3), 665-676.
- Snider, Patricia. 2011. "Law students protest commencement speaker." *The Michigan Daily*, 14 May. <<http://www.michigandaily.com/news/law-school-commencement-sees-protests>>
- Suhr, Kurt, David Hernandez, Douglas Grimes, and Mark Warschauer. 2010. "Laptops and fourth-grade literacy: assisting the jump over the fourth-grade slump." *Journal of Technology, Learning, and Assessment*, 9(5).
- Swers, Michele L. 2002. *The Difference Women Make: The Policy Impact of Women in Congress*. The University of Chicago Press: Chicago, IL.
- . 1998. "Are women more likely to vote for women's issue bills than their male colleagues?" *Legislative Studies Quarterly*, 23, 435-448.
- Swiatek, Mary Ann, Ann Lupkowski-Shoplik, and Cathleen C. O'Donoghue. 2000. "Gender differences in above-level EXPLORE scores of gifted third through sixth graders." *Journal of Educational Psychology*, 92(4), 718-723.
- Tate, Katherine. 2003. *Black Faces in the Mirror: African Americans and Their Representatives in the U.S. Congress*. Princeton University Press: Princeton, NJ.

- Thiele, Megan, Kristen Shorette, and Catherine Bolzendahl. 2012. "Exploring the link between legislators' public school degrees and state spending on higher education." *Sociological Inquiry*, 82(2), 305-328.
- Thomas, Sue. 1991. "The impact of women on state legislative policies." *The Journal of Politics*, 53, 958-976.
- Thomas, Sue, and Susan Welch. 1991. "The impact of gender on activities and priorities of state legislators." *Western Political Quarterly*, 44, 445-456.
- U.S. Census Bureau. 2013. "State & County QuickFacts." <<http://quickfacts.census.gov/qfd/states/00000.html>>
- . 2012. "Table AVG1. Average number of people per household, by race and Hispanic origin, marital status, age, and education of householder: 2012." *Families and Living Arrangements*. <<http://www.census.gov/hhes/families/data/cps2012.html>>
- . 2012a. "Table FM-3. Average number of own children under 18 per family, by type of family: 1955 To Present." *Families and Living Arrangements*. <<http://www.census.gov/hhes/families/data/families.html>>
- U.S. Department of Education. 2012. Digest of Education Statistics. Washington, D.C.: National Center for Education Statistics, Institute of Education Sciences.
- . 2010. "U.S. Department of Education senior official to discuss National Educational Technology Plan at Mooresville Graded School District convocation." *U.S. Department of Education Press Office*, 19 August.
- Vigdor, Jacob and Helen Ladd. 2010. "Scaling the digital divide: home computer technology and student achievement." NBER Working Paper No. 16078.
- Walton, Gregory M., and Geoffrey L. Cohen. 2003. "Stereotype lift." *Journal of Experimental Social Psychology*, 39(5), 456-467.
- Washington, Ebonya. 2008. "Female socialization: How daughters affect their legislator fathers' voting on women's issues." *American Economic Review*, 98(1), 311-332.
- Warner, Rebecca L. 1991. "Does the sex of your child matter? Support for feminism among women and men in the United States and Canada." *Journal of Marriage and the Family*, 53(4), 1051-1056.

- Warner, Rebecca L., and Brent S. Steel. 1999. "Child rearing as a mechanism for social change: The relationship of child gender to parents' commitment to gender equity." *Gender and Society*, 13(4), 503-517.
- Weinberger, Catherine. 2005. "Is the science and engineering workforce drawn from the far upper tail of the math ability distribution?" Working paper, University of California-Santa Barbara.
- Welch, Susan, and Albert K. Karnig. 1979. "The impact of black elected officials on urban social expenditures." *Policy Studies Journal*, 7, 707-714.
- Whitby, Kenny J. 1997. *The Color of Representation: Congressional Behavior and Black Interests*. The University of Michigan Press: Ann Arbor, MI.
- White House Conference on Teenagers. 2000. "Talking with teens: The YMCA Parent and Teen Survey final report." *The White House*.
<http://clinton4.nara.gov/WH/EOP/First_Lady/html/teens/survey.html>
- Wilcox, Clyde, Paul R. Brewer, Shauna Shames, and Celinda Lake. 2007. "If I bend this far I will break? Public opinion about same-sex marriage." In Craig A. Rimmerman and Clyde Wilcox (eds.), *The Politics of Same-Sex Marriage*. University of Chicago Press: Chicago, IL. 215-242.
- Xie, Yu and Kimberlee Shauman. 2003. *Women in Science: Career Processes and Outcomes*. Cambridge, MA: Harvard University Press.

Biography

Katherine Duch was born in Buffalo in January 1987. She graduated from Cornell University with a Bachelor of Science degree in Industrial and Labor Relations in 2008 and with a Master of Science degree in Labor Economics in 2009. She compiled and analyzed financial data from colleges and universities to examine endowment spending rates in her thesis, "Determinants of endowment spending rates at institutions of higher education." She was elected by the student body to serve on the Cornell University Board of Trustees from 2007 to 2009, where she was a voting member of the Finance, Buildings and Properties, Academic Affairs, Student Life, and Alumni Affairs committees.

At Duke, Katherine was a University Scholar and a James B. Duke fellow. She served on the Sanford School of Public Policy Board of Visitors as the doctoral student representative from 2012 to 2013. She was also elected to serve on the Duke University Board of Trustees from 2013 to 2015, where she is a member of the Academic Affairs committee.