ISSUES&ANSWERS REL 2008–No. 062

REL NORTHEAST & ISLANDS
Regional Educational Laboratory At Education Development Center, Inc.

Gender gaps in assessment outcomes in Vermont and the United States

ies NATIONAL CENTER FOR EDUCATION EVALUATION AND REGIONAL ASSISTANCE
Institute of Education Sciences
U.S. Department of Education

ISSUES & ANSWERS REL 2008–No. 062

REL NORTHEAST & ISLANDS
Regional Educational Laboratory
At Education Development
Center, Inc.

Gender gaps in assessment outcomes in Vermont and the United States

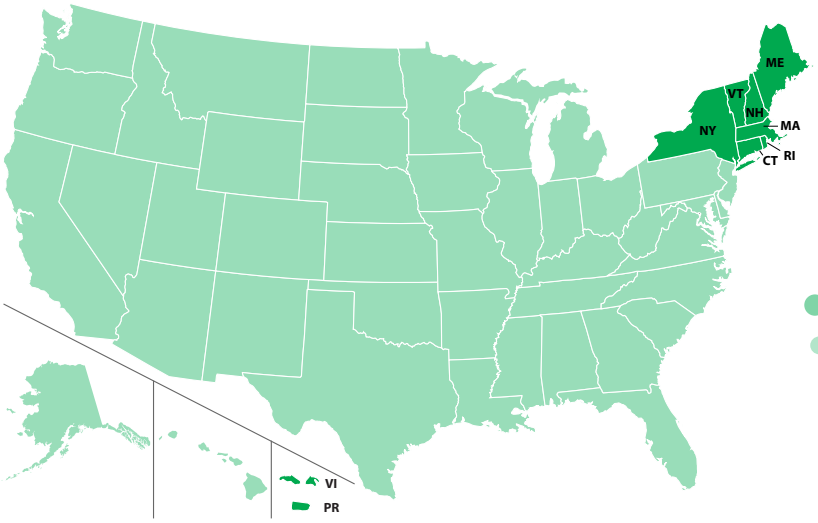
August 2008

Prepared by

Josephine Louie, Ed.D.
Education Development Center, Inc.

Stacy Ehrlich, Ph.D.
Education Development Center, Inc.

ies NATIONAL CENTER FOR EDUCATION EVALUATION AND REGIONAL ASSISTANCE
Institute of Education Sciences
U.S. Department of Education



REL
NORTHEAST
& ISLANDS
Regional Educational Laboratory
At Education Development
Center, Inc.

Issues & Answers is an ongoing series of reports from short-term Fast Response Projects conducted by the regional educational laboratories on current education issues of importance at local, state, and regional levels. Fast Response Project topics change to reflect new issues, as identified through lab outreach and requests for assistance from policymakers and educators at state and local levels and from communities, businesses, parents, families, and youth. All Issues & Answers reports meet Institute of Education Sciences standards for scientifically valid research.

August 2008

This report was prepared for the Institute of Education Sciences (IES) under Contract ED-06-CO-0025 by Regional Educational Laboratory Northeast and Islands administered by Education Development Center, Inc. The content of the publication does not necessarily reflect the views or policies of IES or the U.S. Department of Education nor does mention of trade names, commercial products, or organizations imply endorsement by the U.S. Government.

Louie, J., and Ehrlich, S. (2008). *Gender gaps in assessment outcomes in Vermont and the United States* (Issues & Answers Report, REL 2008–No. 062). Washington, DC: U.S. Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance, Regional Educational Laboratory Northeast and Islands. Retrieved from http://ies.ed.gov/ncee/edlabs.

This report is available on the regional educational laboratory web site at http://ies.ed.gov/ncee/edlabs.

Summary REL 2008–No. 062

Gender gaps in assessment outcomes in Vermont and the United States

Using data from the National Assessment of Educational Progress (NAEP) and the New England Common Assessment Program (NECAP), the report examines how gender gaps differ between Vermont NAEP scores and U.S. NAEP scores and between Vermont NAEP and NECAP scores in grades 4 and 8. Overall and disaggregated by poverty and disability status, gender achievement gaps in Vermont resembled those in the country as a whole except in a few cases.

Vermont education leaders remain concerned about the size of gender gaps in statewide assessment data. They question whether the gender gaps in Vermont differ from gaps in the country as a whole and whether they differ when measured by the state assessment (the NECAP) and by the NAEP in Vermont.

Current debates over gender gaps occur within a context of heightened national focus on achievement gaps among students. The No Child Left Behind (NCLB) Act of 2001 has increased pressure on states, districts, and public schools to close such gaps in reading and math. Under the act states must work to address achievement gaps to ensure that all students reach proficiency by 2014. States are required to conduct annual assessments in reading and math in grades 3–8 and once in high school and to report the results by student poverty, race/ethnicity, disability, and limited English proficiency status. Interest in gender gaps is thus high among many education leaders as they tackle issues of education equity and strive to improve achievement outcomes for all students.

By comparing Vermont and U.S. NAEP reading, writing, and math assessment data for 2000–07 and by examining Vermont NECAP reading and math data for 2006 and writing data for 2002, this report addresses the following questions:

- Do gender gaps differ significantly between Vermont and U.S. NAEP scores?
- How do gender gaps differ between Vermont's NECAP scores and its NAEP scores?

The study found that:

- In reading, gender gaps in Vermont and U.S. NAEP scores have typically not differed at a statistically significant level. From 2002 to 2007 grade 4 girls outscored boys by 5–8 points (or 7–10 percentiles) on average, and grade 8 girls outscored boys by 9–13 points (or 11–16 percentiles) on average, in both Vermont and nationwide.

- In writing, gender gaps in 2002 Vermont and U.S. NAEP scores did not differ at a statistically significant level. The grade 4 gender gap was 21 points (or 23 percentiles) in Vermont and 18 points (or 19 percentiles) nationwide. And the grade 8 gap was 24 points in Vermont and 21 points (or 25 and 22 percentiles) nationwide.

- In math, gender gaps in Vermont and U.S. NAEP scores did not differ at a statistically significant level. Boys outscored girls by 5 points (8 percentiles) or less on average in 2000–07.

- In all but a few cases the gender gaps in reading and math scores on the Vermont NECAP differed from those of the Vermont NAEP by 5 percentile points or less. Gender gaps in 2006 NECAP scores were larger in reading and smaller in math than gender gaps in 2007 NAEP scores by 5 percentile points or less. Accounting for disability and poverty subgroups, however, revealed several exceptions. The average NECAP and NAEP gender gaps in math scores differed by 12 percentile points among grade 8 students in poverty and by 7 percentile points among grade 8 students with disabilities.

Analysis of NAEP and NECAP data suggests areas for further research. For example, gender achievement gaps in Vermont resembled those in the country as a whole except in a few isolated cases. Future research could explore whether these instances represent broader trends or are statistical outliers. In addition, between grades 4 and 8 in both jurisdictions boys' score advantage in math shrinks and girls' score advantage in reading and writing grows. Future studies could examine whether changes in gender gaps occur between other grades and whether these changes reflect differences in development between boys and girls. Finally, gender gaps in all three content areas changed after controlling for student disability status. Among students with disabilities in both grades and in both jurisdictions the gender gap was larger in math and smaller in reading and writing. These shifts raise questions about how boys and girls are grouped in disability categories. This report outlines these issues and presents data to augment current knowledge about gender gaps in achievement in Vermont and the United States.

August 2008

TABLE OF CONTENTS

Tables

Using data from the National Assessment of Educational Progress (NAEP) and the New England Common Assessment Program (NECAP), the report examines how gender gaps differ between Vermont NAEP scores and U.S. NAEP scores and between Vermont NAEP and NECAP scores in grades 4 and 8. Overall and disaggregated by poverty and disability status, gender achievement gaps in Vermont resembled those in the country as a whole except in a few cases.

WHY THIS STUDY?

Analysis of the gender gap in statewide assessment data is an important component of the Vermont Department of Education's commitment to addressing student achievement gaps within the state.[1] Concerned by differences in boys' and girls' scores on English language arts and math assessments in Vermont and by national media reports of gender disparities in schooling, the Vermont State Board and Commissioner of Education have requested an in-depth analysis of gender gaps in reading, writing, and math assessments among Vermont public school students (see box 1 for definitions of key terms). These leaders wanted more information on the size of gender gaps among students disaggregated by poverty and disability status, on how gender gaps compare between Vermont and the country as a whole, and on how gender gaps compare between different assessments in Vermont—specifically, the statewide assessment used for meeting federal accountability requirements, the New England Common Assessment Program (NECAP), and the nationwide assessment administered biannually by the federal government, the National Assessment of Educational Progress (NAEP).

Studying gender gaps

Educators and the public have expressed concerns about differences in academic achievement between boys and girls for decades, even as the focus of debates has shifted. In the 1990s the national report *How schools shortchange girls* (American Association of University Women 1992) focused the public spotlight on the academic plight of girls (Mead 2006). The report claimed that girls had received less attention from teachers than boys had and that girls had fallen short of boys in such areas as math and self-esteem by the end of high school. Girls also averaged lower scores on standardized tests for college than boys did. And girls were much less likely than boys to pursue careers in the growing fields of science and technology, even when they excelled in those fields (American Association of University Women 1998).

BOX 1

Definitions of key terms

Effect size. A standardized measure of the difference between two group outcomes. Standardized effect sizes (often denoted in standard deviation units) help researchers compare outcomes with different units of measurement. For example, scores on the National Assessment of Educational Progress (NAEP) and the New England Common Assessment Program (NECAP) have different scales. Converting achievement gaps on the NAEP and NECAP from their NAEP or NECAP scales to effect sizes measured in standard deviation units makes it easier to compare achievement gaps from each assessment.

Gender gap. A measure of the difference between male and female outcomes. In this report a gender gap is calculated by subtracting the mean scale score of boys from the mean scale score of girls on a specific grade-level subject assessment.

Percentile. A percentile is the value below which a certain percentage of data fall. For example, if a student scores at the 23rd percentile of a test score distribution, then 23 percent of other students achieved a lower score.

Percentile difference. The percentile difference is the difference between two percentiles. For example, if boys score at the 48th percentile on an assessment, and girls score at the 52nd percentile, the percentile difference is equal to 4 percentile points. (Percentile differences are described in units of percentile points in this report.)

Scale score. A scale score is a test score that has been converted from a raw figure to a number on a common scale indicating a student's performance level. NAEP scale scores range from 0 to 500 in reading and math and from 0 to 300 in writing for grades 4, 8, and 12. NECAP scale scores range from 400 to 480 for grade 4 and from 800 to 880 for grade 8 in all content areas.

Standard deviation. Standard deviation is a measure of how widely or narrowly data are dispersed around the data mean. For example, the standard deviation of a set of student test scores is calculated by summing the squared deviations of each student's individual score from the mean, dividing this sum by the total number of students, and taking the square root of the resulting figure. A student's test score can be described in terms of standard deviation units by subtracting the mean from the student's score and dividing that figure by the standard deviation.

Standard error. Standard error is a measure of the amount of error between an estimated statistic from a sample and the true statistic for the population. For example, the mean test score for a sample of students will have a standard error that estimates the deviation between the sample mean and the mean for the entire student population. The standard error for a sample mean is calculated by dividing the standard deviation of the sample data by the square root of the number of subjects in the sample.

For more detail on these terms, see appendix A.

In recent years media attention has recognized the education plight of boys. Major newspapers, news magazines, and television news programs have featured lead stories on boys' academic performance—and whether there is a "boy crisis" in U.S. schools (Conlin 2003; Stahl 2003; Tyre 2006; Von Drehle 2007). Worrisome trends include boys' lower test scores in reading and writing, higher levels of behavioral problems, higher placement rates in special education, and smaller gains in higher education compared with girls. Some researchers argue that earlier concerns about girls may have prompted schools to alter classroom practices in ways that benefit girls over boys (Conlin 2003; Stahl 2003).

Previous findings

Gender gaps in achievement have been documented for decades. Nationwide, gender gaps in assessment outcomes have long existed in multiple content areas (Cole 1997; Coley 2001; Freeman 2004; Klecker 2006; Meadows, Land, and Lamb 2005; Nowell and Hedges 1998), with girls consistently outperforming boys in some areas, and boys outperforming girls in others. While some gaps have been persistent, others have declined over time.

In national assessments of youth literacy skills girls have outscored boys since the 1960s. Studies

of national assessment outcomes find that the gender gap in writing has changed little over time—girls' writing scores were approximately 0.4 standard deviation higher than those of boys in both 1960 and 1990 (Cole 1997). Gender gaps in NAEP reading and writing scores are negligible when students first enter school but widen as students progress through grades 4–12 (Freeman 2004).

In contrast, boys have outperformed girls in math over the past several decades. National data suggest that boys and girls display similar math achievement levels when they begin school. From grade 3 onward boys outscore girls by a steady margin (Freeman 2004). Analyses of national assessment data show that the gender gap in math shrank dramatically over 1960–90, from an effect size of 0.45 to 0.10, and has remained steady ever since (Cole 1997; McGraw, Lubienski, and Struchens 2006).

Researchers have also identified gender gaps in test score variability. In most content areas boys' scores tend to vary more than girls' scores, and the difference in the standard deviations of boys' and girls' test scores tends to grow from grade 4 to grade 12 (Willingham and Cole 1997). The group with greater variation tends to become overrepresented at the top and bottom of the score distribution. A study of NAEP data shows that in math and science boys are overrepresented in the upper tails of test score distributions and that in reading and writing girls are overrepresented at the top percentiles (Nowell and Hedges 1998).[2]

Studies show that gender gaps do not reflect achievement gaps among other student subgroups—such gaps usually persist after controlling for background variables such as race/ethnicity and socioeconomic and disability status (American Association of University Women 1998; Freeman 2004; LoGerfo, Nichols, and Chaplin 2006; McGraw, Lubienski, and Struchens 2006; Ready et al. 2005).[3] Although gender gaps have typically been smaller than racial, socioeconomic, and disability gaps on a variety of assessments (U.S. Department of Education, National Center for Education Statistics 2005, 2006), they may be of particular concern within specific student subgroups. Achievement (particularly in reading) is very low for boys in poverty (Mead 2006). Students with disabilities—a group in which boys form a disproportionate share—achieve much lower scores than their nondisabled counterparts (Mead 2006). How gender gaps interact with socioeconomic, disability, and other gaps merits deeper and ongoing investigation.

The gender gap in Vermont

In recent years Vermont education leaders have expressed a strong interest in learning about achievement gender gaps (Hayes 2007; Johnson 2002; Pandiani and Bramley 2002; Vermont Student Assistance Center 2005). Aware of existing gender gaps within the state and alarmed by national media reports on the underperformance of boys in English language arts and other areas nationwide, the Vermont State Board and Commissioner of Education asked how state and national gender gaps compare, in the aggregate and disaggregated by student poverty and disability status. Students in poverty make up more than 26 percent of all students within the state, and students with disabilities make up 11 percent (table C1 in appendix C). Because the state is accountable for the achievement of all students under the No Child Left Behind (NCLB) Act of 2001, state leaders have expressed interest in better understanding gender gaps within these two subgroups.[4]

Vermont education leaders have expressed a strong interest in learning how state and national gender gaps compare, in the aggregate and disaggregated by student poverty and disability status

Because Vermont is held accountable under the NCLB Act for student performance levels based on NECAP scores, state education leaders have expressed interest in how the NECAP has been measuring student gender gaps and how these

gaps compare with those measured by other nationally recognized assessments like the NAEP, administered to nationally representative samples of students since the 1960s.[5] A recent comparison of the NAEP and individual state assessments concludes that state assessment standards and student outcomes differ greatly across the country (U.S. Department of Education, National Center for Education Statistics 2007). Neither that study, nor any other, however, has compared NECAP scores with NAEP scores. This report addresses these issues by examining how gender gaps compare in Vermont and nationally as well as on the NECAP and the NAEP.

Research questions

This report examines two questions of interest to Vermont state education leaders about reading, writing, and math achievement for students in grades 4 and 8:

1. Do gender gaps differ significantly between Vermont's NAEP scores and U.S. NAEP scores, overall and after controlling for student poverty and disability status?

2. How do gender gaps differ between Vermont's NECAP and Vermont's NAEP scores, overall and after controlling for student poverty and disability status?

Researchers used the online database, NAEP Data Explorer (U.S. Department of Education, National Center for Education Statistics 2008), to retrieve publicly available data on NAEP scores in reading (2002, 2003, 2005, 2007), writing (2002), and math (2000, 2003, 2005, 2007) for grades 4 and 8 in Vermont and nationally, and NECAP assessment results for 2006 in grades 4 and 8, provided by the Vermont Department of Education (see box 2 and appendix A for details on study methods and limitations; see appendix B for descriptions of the NEAP and the NECAP).

In general, gender gaps in reading, writing, and math scores in Vermont and nationally do not differ, even after taking student poverty and disability status into account

DO GENDER GAPS DIFFER SIGNIFICANTLY BETWEEN VERMONT AND U.S. NATIONAL ASSESSMENT OF EDUCATIONAL PROGRESS SCORES?

In general, gender gaps in reading, writing, and math scores in Vermont and nationally do not differ, even after taking student poverty and disability status into account. The following sections provide more detailed descriptions of gender gaps from NAEP scores in each jurisdiction and differences between Vermont and national gender gaps for reading, writing, and math.

Reading

Students in grades 4 and 8 took the NAEP reading assessment in 2002, 2003, 2005, and 2007. This study examines gender gaps overall—and by poverty and disability status—for Vermont and nationally (table 1).

Gender gaps in Vermont and nationally. On average, 4th and 8th grade girls consistently outscored boys on the 2002–07 NAEP reading assessments. During that period gender gaps among grade 4 students ranged from 5 to 8 points in Vermont and 6 to 8 points nationally (see table 1). These test score gaps are statistically significant (statistically greater than zero) for each year. Expressed in effect size units, average NAEP reading scores among grade 4 girls exceeded those of boys by 0.17–0.26 standard deviation in Vermont and 0.17–0.20 standard deviation nationally. And the average girl ranked above the average boy by 7–10 percentiles in Vermont and 7–8 percentiles nationally (see appendix A for explanation of percentile difference calculations).

Aggregate gender gaps in reading were slightly larger among older students. During 2002–07 grade 8 girls outscored boys by 9–13 points in Vermont, and by 9–10 points nationally. In effect-size terms gender gaps ranged from 0.30 to 0.42 standard deviation in Vermont and from 0.28 to 0.30 nationally. Alternatively, the average grade 8 girl outranked the average grade 8 boy on the

BOX 2

Study methods and limitations

To determine gender gaps in National Assessment of Educational Progress (NAEP) reading (2002, 2003, 2005, 2007), writing (2002), and math (2000, 2003, 2005, 2007) assessment scores for students in grades 4 and 8 in Vermont and the United States, researchers used the NAEP Data Explorer online database (U.S. Department of Education, National Center for Education Statistics 2008). The database also provided Vermont NAEP reading and math results (2007), which were compared with NECAP results (2006)—provided by the Vermont Department of Education (Vermont Department of Education, Standards and Assessment, 2007)—to determine how different assessments measured gender gaps. Average NAEP and NECAP scores were examined by gender, poverty status (defined by eligibility in the National School Lunch Program), and disability status (defined by eligibility for an Individualized Education Program).

To compare gender gaps in Vermont and nationwide, researchers calculated the following outcome statistics from NAEP reading, writing, and math data: average scale scores for boys and girls (in the aggregate), gender gap estimates (measured as differences in average scale scores between girls and boys), effect sizes for gender gap estimates (in both standard deviation units and average percentile differences), and estimates of differences between Vermont and national gender gaps. Researchers disaggregated gender gaps by poverty and disability status and compared these gaps in Vermont and nationwide. Gender gaps within jurisdictions and differences in gender gaps between jurisdictions were tested for statistical significance.

To compare gender gaps measured by the Vermont NECAP and Vermont NAEP, researchers calculated the following outcome statistics from each assessment in reading and math: average scale scores for boys and girls (in the aggregate), gender gap estimates (measured as differences in average scale scores between girls and boys), and effect sizes for gender gap estimates (in both standard deviation units and average percentile differences). Researchers disaggregated gender gaps by student poverty and disability status. All cases where students took out-of-grade NECAP tests were excluded (for details on data exclusion see appendixes A and D). Gender gaps as measured by each assessment (both in the aggregate and by student poverty and disability status) were tested for statistical significance.

Three limits of this study should be noted. First, because the NECAP and NAEP measure student achievement with different scoring scales, test score gaps were converted into standardized effect sizes to compare gaps. Although reporting confidence intervals around effect sizes is becoming a recommended practice (Cumming 2001; Nix and Barnette 1998; Steiger 2004; Thompson 1998), calculating the intervals was beyond the scope of this project. This report therefore does not indicate whether gender gaps measured by the Vermont NECAP and NAEP assessments differ at statistically significant levels. Instead, gender gaps from each assessment are presented in standardized effect size units and the percentile equivalents to help readers form their own judgments.

Second, this report does not compare gender gaps in writing scores from the Vermont NECAP and NAEP assessments. Data from each assessment were available for different grade-level cohorts only and were therefore not comparable. This report compares outcomes on the Vermont NECAP and NAEP reading and mathematics assessments only. Third, more precise and detailed analyses of gender gaps across the two jurisdictions require examining student-level NAEP microdata, which are not available from the NAEP Data Explorer. Future studies may wish to explore these data to expand the analyses presented here.

At best, this report points education leaders toward areas for further analysis and investigation. This report does not assign meaning to the relative scope of gender or other types of gaps in student achievement or explain the causes behind gender gaps in NAEP or NECAP data. And it does not suggest how to close gender gaps. Because the report measures statistical significance alone, it does not draw conclusions of the substantive and practical significance of results, which may vary by context (Light, Singer, and Willett 1990). Having captured the interest of the education community, these topics now require ongoing study along with continued measurement of achievement gaps over time.

For additional details on study data, methods, and limitations, see appendix A.

TABLE 1

Grades 4 and 8 gender gaps and associated effect sizes in National Assessment of Education Progress reading scores, by student poverty and disability status, Vermont and nationally, 2002, 2003, 2005, and 2007

	Grade 4								Grade 8							
	2002		2003		2005		2007		2002		2003		2005		2007	
Students	Vermont	National	Vermont	National	Vermont	National	Vermont	National	Vermont	National	Vermont	National	Vermont	National	Vermont	National
All																
Gender gap (points)	8.4**	6.5**	5.4**	7.5**	7.1**	6.3**	7.6**	6.7**	9.3**	9.4**	11.4**	10.5**	13.4**	10.4**	10.6**	10.1**
Effect size (*d*)	0.26	0.18	0.17	0.20	0.21	0.17	0.22	0.19	0.30	0.28	0.36	0.30	0.42	0.30	0.35	0.29
Percentile difference	10	7	7	8	8	7	9	7	12	11	14	12	16	12	14	11
Students in poverty																
Gender gap (points)	5.8	7.0**	2.7	8.2**	0.5	6.6**	6.2**	7.4**	9.7**	8.5**	9.7**	10.7**	10.6**	10.1**	11.0**	10.0**
Effect size (*d*)	0.18	0.20	0.09	0.23	0.01	0.19	0.19	0.21	0.30	0.26	0.33	0.31	0.33	0.30	0.37	0.30
Percentile difference	7	8	4	9	0	7	7	8	12	10	13	12	13	12	14	12
Students not in poverty																
Gender gap (points)	8.9**	7.0**	7.4**	7.5**	9.2**	6.5**	8.2**	6.3**	8.8**	10.2**	12.3**	10.8**	14.1**	11.1**	9.9**	10.3**
Effect size (*d*)	0.29	0.22	0.25	0.23	0.30	0.20	0.26	0.20	0.30	0.34	0.40	0.34	0.47	0.35	0.35	0.33
Percentile difference	11	9	10	9	12	8	10	8	12	13	16	13	18	14	14	13
Students with disabilities																
Gender gap (points)	—	2.8	–4.2	4.5**	–5.5	1.7	–5.7	0.7	4.3	4.8**	1.6	5.2**	3.5	5.3**	–2.4	3.8**
Effect size (*d*)	—	0.07	–0.13	0.11	–0.17	0.04	–0.15	0.02	0.14	0.13	0.05	0.14	0.12	0.15	–0.07	0.10
Percentile difference	—	3	–5	4	–7	2	–6	1	6	5	2	6	5	6	–3	4
Students without disabilities																
Gender gap (points)	7.0**	5.3**	4.3**	5.8**	5.4**	4.9**	4.6**	5.3**	7.4**	7.8**	9.8**	8.4**	10.1**	8.7**	9.0**	8.4**
Effect size (*d*)	0.23	0.15	0.14	0.17	0.17	0.14	0.15	0.16	0.25	0.25	0.33	0.26	0.35	0.27	0.33	0.26
Percentile difference	9	6	6	7	7	6	6	6	10	10	13	10	14	11	13	10

**Gender gap is statistically significant (different from zero) at $p < 0.05$.

— is not available.

Note: Gender gap is defined as the mean scale score of girls minus the mean scale score of boys ($F - M$). Effect size is calculated as $d = (F - M) / SD_{pooled}$, where $SD_{pooled} = \sqrt{[(SD_F^2 + SD_M^2) / 2]}$. Percentile difference measures differences between the average girl and average boy in the test score distribution. Shaded pairs of Vermont and national gender gap estimates are statistically different from each other at $p < 0.05$.

Source: Authors' analysis based on grades 4 and 8 reading data for 2002, 2003, 2005, and 2007 from U.S. Department of Education, National Center for Education Statistics (2008).

NAEP test score distribution by 12–16 percentiles in Vermont and 11–12 percentiles nationally.

When scores were disaggregated by student poverty and disability status, there were a few shifts in the differences between girls' and boys' reading scores in both jurisdictions. Gender gaps in reading shrank, or disappeared, for some student subgroups. Gender gaps among Vermont 4th graders in poverty were not statistically significant from 2002 to 2005. Among Vermont 4th and 8th graders with disabilities gender gaps were not statistically significant through 2007. Displaying a similar pattern, national gender gaps were not statistically significant among 4th graders with disabilities in any year except 2003. Although national grade 8 gender gaps were statistically significant within all subgroups, the gaps were smaller among students with disabilities than among all 8th graders.

Differences between Vermont and national gender gaps. For all grades 4 and 8 students differences between Vermont and national gender gaps were not statistically significant. Gender gaps in reading among all Vermont 4th graders differed from those among all U.S. 4th graders by –2.1 to 0.8 points from 2002 to 2007 (see table 1). Among grade 8 students gender gaps between Vermont and the country differed by –0.1 to 3.0 points. None of these differences was statistically significant at the $p < 0.05$ level. Thus, over 2002–07 grades 4 and 8 girls outscored boys in reading by similar amounts in Vermont and nationally.

Disaggregated by poverty and disability status, differences in gender gaps in each jurisdiction have rarely differed by statistically significant amounts from 2002 to 2007. For grade 8, differences between Vermont and national gender gaps by student poverty subgroups are not statistically significant from 2002 to 2007. In both grades 4 and 8 differences between Vermont and national gender gaps for student subgroups with and without disabilities also are not statistically significant.

Among 4th graders in poverty differences between Vermont and national gender gaps were not statistically significant except in 2003, when the national gap exceeded the Vermont gap by 5.5 points. This margin is statistically significant at a 95 percent confidence level. Differences in gender gaps between the two jurisdictions were significant for 4th graders not in poverty only in 2005, when Vermont's gap exceeded the national gap by a statistically significant margin. Because of the number of differences examined (39; see table A2 in appendix A), it is possible that these two statistically significant findings may have arisen by chance. Indeed, the overall pattern of results shows that gender gaps in Vermont and nationally were very similar in grades 4 and 8 in reading, both before and after taking student poverty and disability status into account.

Writing

Grades 4 and 8 students took the national NAEP writing assessment in 2002. This study examines gender gaps overall and by poverty and disability status (table 2).

Gender gaps in Vermont and the United States. Girls outscored boys in Vermont and nationally on the 2002 NAEP writing assessment by greater margins than on the NAEP reading assessment. The grade 4 gender gap in writing was more than 21 points in Vermont and almost 18 points nationally. Among grade 8 students aggregate gender gaps were even larger—girls outperformed boys by more than 24 points in Vermont and almost 21 points nationally.

Girls outscored boys in Vermont and nationally on the 2002 NAEP writing assessment by greater margins than on the NAEP reading assessment

In effect sizes average grade 4 gender gaps were 0.61 standard deviation in Vermont and 0.50 standard deviation nationally; average grade 8 gender gaps were larger, at 0.66 standard deviation in Vermont and 0.58 standard deviation nationally.

TABLE 2

Grades 4 and 8 gender gaps and associated effect sizes in National Assessment of Educational Progress writing scores, by student poverty and disability status, Vermont and nationally, 2002

	Grade 4		Grade 8	
Students	Vermont	National	Vermont	National
All				
Gender gap (points)	21.3**	17.5**	24.2**	20.9**
Effect size (*d*)	0.61	0.50	0.66	0.58
Percentile difference	23	19	25	22
Students in poverty				
Gender gap (points)	24.1**	16.5**	26.1**	19.4**
Effect size (*d*)	0.72	0.49	0.70	0.56
Percentile difference	26	19	26	21
Students not in poverty				
Gender gap (points)	22.6**	18.9**	24.1**	22.2**
Effect size (*d*)	0.67	0.57	0.70	0.66
Percentile difference	25	22	26	25
Students with disabilities				
Gender gap (points)	6.3	9.6**	7.4	13.7**
Effect size (*d*)	0.21	0.28	0.22	0.39
Percentile difference	8	11	9	15
Students without disabilities				
Gender gap (points)	20.0**	16.5**	23.3**	18.8**
Effect size (*d*)	0.61	0.49	0.68	0.55
Percentile difference	23	19	25	21

**Gender gap is statistically significant (different from zero) at $p < 0.05$.

Note: Gender gap is defined as the mean scale score of girls minus the mean scale score of boys ($F - M$). Effect size is calculated as $d = (F - M) / SD_{pooled}$, where $SD_{pooled} = \sqrt{[(SD_F^2 + SD_M^2) / 2]}$. Percentile difference measures differences between the average girl and average boy in the test score distribution. The shaded pair of Vermont and national gender gap estimates is statistically different from each other at $p < 0.05$.

Source: Authors' analysis based on grades 4 and 8 reading data for 2002 from U.S. Department of Education, National Center for Education Statistics (2008).

In 2002 the average grade 4 girl outranked the average grade 4 boy by 23 percentiles in Vermont and 19 percentiles nationally, while the average grade 8 girl scored 25 percentiles higher than her male counterpart in Vermont and 22 percentiles higher nationally.

Disaggregated by student poverty status gender gaps in writing changed very little by grade and jurisdiction. A different pattern emerged after disaggregating by disability status. Similar to the results for reading, gender gaps in writing in both Vermont and nationally were smaller for students with disabilities than for students in the aggregate. In Vermont the gender gaps for grades 4 and 8 students with disabilities (6.3 and 7.4 points) in 2002 were not statistically significant. Nationally, girls with disabilities outscored boys with disabilities by statistically significant margins in both grades, but the gender gap was about half the size of the aggregate gender gap among 4th graders and two-thirds the aggregate gender gap among 8th graders.

Differences between Vermont and national gender gaps. Aggregate gender gaps on the 2002 NAEP writing assessment were similar in Vermont and nationally. The gender gap in Vermont exceeded

the gap nationally by 3.8 points among 4th graders and by 3.3 points among 8th graders. Neither margin was statistically significant.

In general, Vermont and national gender gaps in writing were similar after controlling for student poverty and disability status. Among students not in poverty differences between Vermont and national gender gaps (3.7 points in grade 4 and 1.9 points in grade 8) were not statistically significant. And within the subgroups of students with and without disabilities Vermont and national gender gaps never differed by statistically significant amounts. Of 10 total differences examined in writing, there was only one exception. Among grade 4 students in poverty the gender gap in Vermont exceeded the national gap by 7.6 points—a statistically significant amount.

Math

Grades 4 and 8 students took the national NAEP math assessment in 2000, 2003, 2005, and 2007. This study examines gender gaps in math scores among all students and students in poverty and disability subgroups (table 3).

Gender gaps in Vermont and nationally. Boys typically outperformed girls on the NAEP math assessment in both Vermont and nationally from 2000 to 2007. The aggregate gender gap among 4th graders ranged from 2 to 5 points (0.06 to 0.19 standard deviation) in Vermont and from 2 to 3 points (or 0.08 to 0.09 standard deviation) nationally. The average grade 4 boy outranked the average grade 4 girl in math scores by 2–8 percentiles in Vermont and 3–4 percentiles nationally.

In both jurisdictions grade 8 boys in the aggregate also outscored girls in math. But the gender gaps were smaller for 8th graders than for 4th graders, an opposite pattern to that found for reading and writing.

In both grades and jurisdictions gender gaps in math change very little after students are disaggregated by poverty status but show strong shifts after students are disaggregated by disability status. For example, Vermont boys with disabilities outscored their female counterparts in math by 8–12 points in grade 4 and 10–15 points in grade 8 over the seven-year period. Expressed differently, the average Vermont boy with disabilities outranked his female counterpart by 11–17 percentiles in grade 4 and 11–18 percentiles in grade 8. National gender gaps among students with disabilities were also larger than those for all U.S. students, particularly in grade 8 (see table 3).

Boys typically outperformed girls on the NAEP math assessment in both Vermont and nationally from 2000 to 2007

Differences between Vermont and national gender gaps. In general, aggregate gender gaps in math have not differed significantly across the two jurisdictions since 2000. An exception occurred in 2005, when grade 4 boys outscored girls by 5.1 points in Vermont and 2.4 points nationally. Vermont's gap exceeded the national gap by nearly 3 points—a statistically significant amount. In all other years the differences between Vermont and national aggregate gender gaps were not statistically significant in grades 4 or 8.

After controlling for student poverty and disability status, gender gaps in math in Vermont and nationally remain similar. From 2000 to 2007 differences between Vermont and national gender gaps ranged from 0 to 5 points in grade 4 and from 0 to 7 points in grade 8 across all poverty and disability status subgroups (see table 3 and tables E8 and E10 in appendix E). Only one of the differences examined was statistically significant. In 2005 the gender gap in math among grade 4 students without disabilities was significantly larger in Vermont than nationally. This result reflected the statistically significant finding (noted above) for all 4th graders in 2005, because in both jurisdictions populations of students without disabilities are relatively similar in size to the total student populations. Thus, of 38 differences between jurisdictions examined in math, 2 were statistically significant (see table A2).

TABLE 3

Grades 4 and 8 gender gaps and associated effect sizes in National Assessment of Educational Progress (NAEP) math scores by poverty and disability status, Vermont and nationally, 2000, 2003, 2005, and 2007

	Grade 4								Grade 8							
	2002		2003		2005		2007		2002		2003		2005		2007	
Students	Vermont	National	Vermont	National	Vermont	National	Vermont	National	Vermont	National	Vermont	National	Vermont	National	Vermont	National
All																
Gender gap (points)	–1.7	–2.5	–3.5**	–2.7**	–5.1**	–2.4**	–2.7**	–2.3**	2.7	–1.6	–0.1	–1.8	0.1	–1.4**	–1.8	–1.9**
Effect size (*d*)	–0.06	–0.08	–0.13	–0.09	–0.19	–0.08	–0.10	–0.08	0.08	–0.04	0.00	–0.05	0.00	–0.04	–0.05	–0.05
Percentile difference	–2	–3	–5	–4	–8	–3	–4	–3	3	–2	0	–2	0	–2	–2	–2
Students in poverty																
Gender gap (points)	–3.4	–0.7	–5.3**	–2.2**	–4.7**	–1.7**	–3.0	–1.3**	–0.1	0.3	–3.2	–1.6**	–0.1	–0.9**	–5.9**	–1.2**
Effect size (*d*)	–0.11	–0.02	–0.21	–0.08	–0.18	–0.06	–0.11	–0.05	0.00	0.01	–0.10	–0.05	0.00	–0.03	–0.19	–0.04
Percentile difference	–4	–1	–8	–3	–7	–2	–4	–2	0	0	–4	–2	0	–1	–8	–2
Students not in poverty																
Gender gap (points)	–1.2	–4.2**	–2.3	–3.0**	–4.3**	–2.6**	–2.5	–2.6**	3.4	–0.8	0.8	–1.2**	–0.5	–1.5**	0.5	–2.1**
Effect size (*d*)	–0.04	–0.15	–0.09	–0.12	–0.18	–0.10	–0.10	–0.10	0.10	–0.02	0.03	–0.04	–0.02	–0.04	0.02	–0.06
Percentile difference	–2	–6	–4	–5	–7	–4	–4	–4	4	–1	1	–2	–1	–2	–1	–2
Students with disabilities																
Gender gap (points)	—	–11.0**	–12.0**	–6.6**	–10.0**	–6.9**	–8.3**	–6.3**	—	–11.0**	–15.0**	–7.9**	–10.0**	–8.4**	–9.9**	–7.8**
Effect size (*d*)	—	–0.36	–0.45	–0.22	–0.36	–0.23	–0.29	–0.19	—	–0.28	–0.46	–0.22	–0.29	–0.23	–0.30	–0.21
Percentile difference	—	–14	–17	–9	–14	–9	–11	–8	—	–11	–18	–9	–11	–9	–12	–8
Students without disabilities																
Gender gap (points)	–1.8	–3.7**	–4.9**	–4.1**	–6.9**	–3.6**	–4.5**	–3.4**	–0.8	–3.3**	–1.7	–4.2**	–2.7	–3.4**	–4.4**	–3.9**
Effect size (*d*)	–0.06	–0.12	–0.20	–0.15	–0.28	–0.13	–0.19	–0.12	–0.03	–0.09	–0.06	–0.12	–0.09	–0.10	–0.14	–0.11
Percentile difference	–2	–5	–8	–6	–11	–5	–8	–5	–1	–4	–2	–5	–4	–4	–6	–4

**Gender gap is statistically significant (different from zero) at $p < 0.05$.

— is not available.

Note: Gender gap is defined as the mean scale score of girls minus the mean scale score of boys ($F - M$). Effect size is calculated as $d = (F - M) / SD_{pooled}$, where $SD_{pooled} = \sqrt{[(SD_F^2 + SD_M^2) / 2]}$. Percentile difference measures differences between the average girl and average boy in the test score distribution. Shaded pairs of Vermont and national gender gap estimates are statistically different from each other at $p < 0.05$.

Source: Authors' analysis based on NAEP math data for grades 4 and 8 for 2000, 2003, 2005, and 2007 from U.S. Department of Education, National Center for Education Statistics (2008).

HOW DO GENDER GAPS DIFFER BETWEEN VERMONT'S NEW ENGLAND COMMON ASSESSMENT PROGRAM AND VERMONT'S NATIONAL ASSESSMENT OF EDUCATIONAL PROGRESS SCORES?

In all but a few cases gender gaps in reading and math on the 2006 Vermont NECAP differed from those on the 2007 Vermont NAEP by 5 percentile points or less. There were two instances in both reading and math when gender gaps on the Vermont NECAP and NAEP differed by more than 5 percentile points. In addition, girls' advantage in reading scores was greater and boys' advantage in math scores was smaller on the NECAP than on the NAEP in all but one case.

Reading

Two observations emerged from comparing gender gaps in the 2006 NECAP and 2007 NAEP reading scores. First, aggregate gender gaps on the 2006 NECAP exceeded those on the 2007 NAEP by less than 5 percentile points. Gender gap effect sizes in grade 4 reading scores were 0.32 on the 2006 Vermont NECAP and 0.22 on the 2007 Vermont NAEP (figure 1). These statistics indicate that the average grade 4 girl outranked the average grade 4 boy in reading by 13 percentiles on the 2006 NECAP and 9 percentiles on the 2007 NAEP. The average grade 8 girl outranked her male counterpart by 16 percentiles on the 2006 NECAP (0.40 standard deviation) and 14 percentiles on the 2007 NAEP (0.35 standard deviation; figure 2).

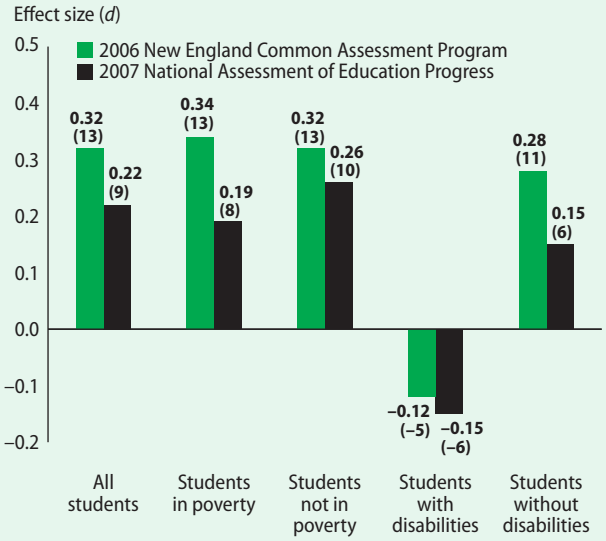
Thus, gender gaps in reading were smaller on the 2007 NAEP than on the 2006 NECAP by 4 percentile points for the grade 4 cohort and by 2 percentile points for the grade 8 cohort. The reasons for the differences in gap sizes are not clear—the differences may have arisen from dissimilar NECAP and NAEP population samples, different test instruments, or a genuine shift in the gender gap over the school year.

Second, after controlling for student poverty and disability status, Vermont gender gaps in reading on the 2006 NECAP exceeded those on the 2007 NAEP by 5 percentile points or less. Gender gaps in reading continued to be similar on the two assessments after grades 4 and 8 students were disaggregated into poverty status subgroups (see figures 1 and 2). For example, the average grade 4 girl in poverty outperformed the average grade 4 boy in poverty by 13 percentiles on the 2006 NECAP and 8 percentiles on the 2007 NAEP. And the average grade 8 girl in poverty outperformed her male counterpart by 15 percentiles on the 2006 NECAP and 13 percentiles on the 2007 NAEP. In both grades for students not in poverty the gender gaps measured by each assessment differed by 5 or fewer percentile points.

A similar story emerged after grades 4 and 8 students were disaggregated by disability status. For

FIGURE 1

Grade 4 gender gap effect sizes and percentile differences for all students and by student poverty and disability status for the Vermont 2006 New England Common Assessment Program and 2007 National Assessment of Educational Progress reading scores

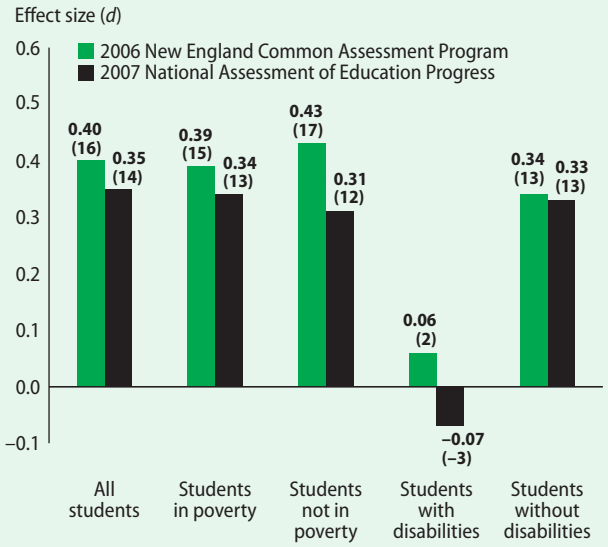


Note: Values in parentheses are percentile differences. Positive effect sizes indicate girls outperforming boys; negative effect sizes indicate boys outperforming girls.

Source: Authors' analysis based on NAEP data from the NAEP Data Explorer (U.S. Department of Education, National Center for Education Statistics 2008) and authors' tabulations of student-level NECAP data (Vermont Department of Education, Standards and Assessment 2007).

FIGURE 2

Grade 8 gender gap effect sizes and percentile differences for all students and by student poverty and disability status for the Vermont 2006 New England Common Assessment Program and 2007 National Assessment of Educational Progress student reading scores



Note: Values in parentheses are percentile differences. Positive effect sizes indicate girls outperforming boys; negative effect sizes indicate boys outperforming girls.

Source: Authors' analysis based on NAEP data from the NAEP Data Explorer (U.S. Department of Education, National Center for Education Statistics 2008) and authors' tabulations of student-level NECAP data (Vermont Department of Education, Standards and Assessment 2007).

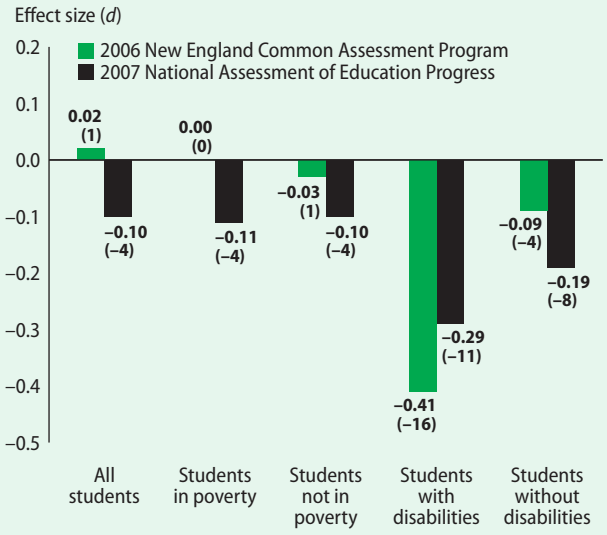
example, the average grade 4 girl with disabilities ranked below her male counterpart by 5 percentiles on the 2006 NECAP and 6 percentiles on the 2007 NAEP (see figure 1). Among grade 8 students with disabilities the average girl ranked above the average boy by 2 percentiles on the 2006 NECAP and below the average boy by 3 percentiles on the 2007 NAEP (see figure 2). In both these cases and those for students without disabilities gender gaps on the two assessments differed by no more than 5 percentile points.

Math

Two observations emerged from comparisons of gender gaps on the 2006 NECAP and 2007 NAEP math assessments in Vermont. First, aggregate gender gaps in Vermont math scores differed by 5 or fewer percentile points on the 2006 NECAP and 2007 NAEP. Grade 4 girls performed better than grade 4 boys in math by 0.02 standard deviation on the 2006 NECAP and worse than their male counterparts by 0.10 standard deviation on the 2007 NAEP (figure 3). Thus, the average grade 4 girl in Vermont ranked 1 percentile higher than her male counterpart on the 2006 NECAP and 4 percentiles lower than her male counterpart on the 2007 NAEP. Within the grade 8 cohort the average girl outscored the average boy by 3 percentiles on the 2006 NECAP and 2 percentiles on the 2007 NAEP (figure 4). For both grade-level cohorts gender gaps differed between the two assessments by 5 or fewer percentile points. As with reading scores, girls' performance relative to boys' was stronger on the NECAP than on the NAEP.

FIGURE 3

Grade 4 gender gap effect sizes and percentile differences for all students and by student poverty and disability status for Vermont 2006 New England Common Assessment Program and 2007 National Assessment of Educational Progress math scores

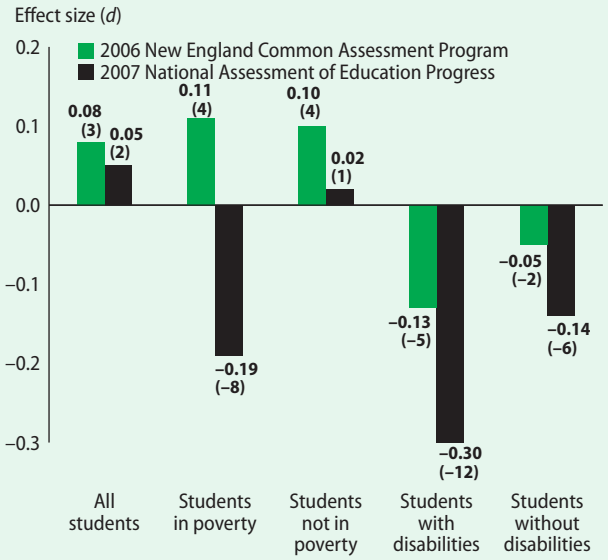


Note: Values in parentheses are percentile differences. Positive effect sizes indicate girls outperforming boys; negative effect sizes indicate boys outperforming girls.

Source: Authors' analysis based on NAEP data from the NAEP Data Explorer (U.S. Department of Education, National Center for Education Statistics 2008) and authors' tabulations of student-level NECAP data (Vermont Department of Education, Standards and Assessment 2007).

FIGURE 4

Grade 8 gender gap effect sizes and percentile differences for all students and by student poverty and disability status for Vermont 2006 New England Common Assessment Program and 2007 National Assessment of Educational Progress math scores



Note: Values in parentheses are percentile differences. Positive effect sizes indicate girls outperforming boys; negative effect sizes indicate boys outperforming girls.

Source: Authors' analysis based on NAEP data from the NAEP Data Explorer (U.S. Department of Education, National Center for Education Statistics 2008) and authors' tabulations of student-level NECAP data (Vermont Department of Education, Standards and Assessment 2007).

Second, after controlling for poverty and disability status, Vermont gender gaps differed by 5 or fewer percentile points for 4th graders, but by more than 5 percentile points among subgroups of 8th graders on the 2006 NECAP and 2007 NECAP math assessments. Among grade 4 students in poverty there was no difference in average math scores for boys and girls on the 2006 NECAP, while the average girl scored 4 percentiles below the average boy on the 2007 NAEP (see figure 3). Among grade 4 students who were not in poverty the average girl ranked below the average boy in math by 1 percentile on the 2006 NECAP and 4 percentiles on the 2007 NAEP. In both cases the size of the gender gap on each assessment differed by 5 percentile points or less.

A different pattern emerged for 8th graders after math scores were disaggregated by student poverty status. Among grade 8 students who were not in poverty, the average girl outscored the average boy by 4 percentiles on the 2006 NECAP and 1 percentile on the 2007 NAEP, for a difference in gender gaps of 3 percentile points (see figure 4). In contrast, the average grade 8 girl in poverty scored above her male counterpart by 4 percentiles on the 2006 NECAP and below her male counterpart by 8 percentiles on the 2007 NAEP, for a difference in gender gaps of 12 percentile points—a difference between assessments four times larger for 8th graders in poverty than for those not in poverty.

With math scores disaggregated by student disability status, students in grades 4 and 8 again displayed contrasting gender gaps on the two assessments. Gender gaps in math on the 2006 NECAP and 2007 NAEP differed by 5 or fewer percentiles for grade 4 students with or without disabilities (see figure 3 and table F3 in appendix F). Among grade 4 students with disabilities the underperformance of girls compared with boys was greater on the 2006 NECAP than on the 2007 NAEP—the only instance when the gender gaps on the two assessments compared in this way. In contrast, gender gaps on the two assessments diverged by larger margins among grade 8 students after controlling for disability status. Among grade 8 students with disabilities the average girl scored below the average boy by 5 percentiles on the 2006 NECAP and 12 percentiles on the 2007 NAEP—a 7 percentile point difference between the two assessments.

FINAL OBSERVATIONS AND AREAS FOR FUTURE RESEARCH

Consistent with previous research this study of publicly available NAEP data shows that over 2000–07 girls in Vermont and the country as a whole consistently outscored boys in reading and writing, while boys typically outscored girls in math. This examination of gender gaps in Vermont and U.S. NAEP scores and in recent Vermont NECAP scores suggests areas for further investigation.

The first research question asks whether gender gaps on the NAEP have varied significantly between Vermont and the country as a whole. The differences between Vermont and national gender gaps were statistically significant in only a few instances:

- In reading scores (in which girls consistently outscored boys), the gender gap was significantly smaller in Vermont than nationally in 2003 for grade 4 students in poverty and significantly larger in Vermont than nationally in 2005 for grade 4 students not in poverty (see table 1).
- In writing scores (in which girls consistently outscored boys), the gender gap in 2002 for grade 4 students in poverty was significantly larger in Vermont than nationally (see table 2).
- In math scores (in which boys typically outscored girls), the gender gaps in 2005 for grade 4 students in the aggregate and without disabilities were significantly larger in Vermont than nationally (see table 3).

Vermont and national gender gaps differed significantly in only 5 percent of all differences examined for reading and math scores and 10 percent of all differences examined for writing scores (see table A2 in appendix A). Of 87 total differences tested across all three content areas, 5 cases emerged as statistically significant. Because of the total number of comparisons made, it is possible that some of these cases arose by chance. For instance, when results are reported at a 95 percent confidence level, approximately 5 percent will be statistically significant due simply to chance. Thus, it is difficult to tell whether these cases reflect underlying trends or statistical anomalies.

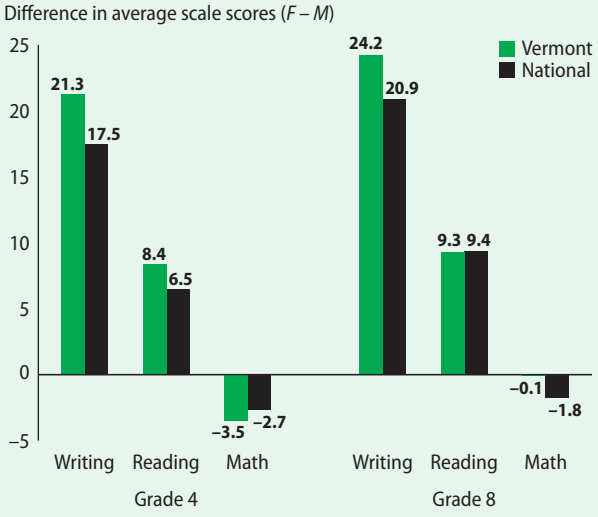
Vermont and national gender gaps differed significantly in only 5 percent of all differences examined for reading and math scores and 10 percent of all differences examined for writing scores

Because gender gaps in reading and math were not consistently larger in one jurisdiction than in the other, the isolated cases in which differences were statistically significant would not appear to signal discernable trends. By contrast, Vermont gender gaps in writing assessment scores appear to exceed national gaps consistently among both grades 4 and 8 students in the aggregate and after controlling for poverty status. Only one comparison yielded a statistically significant result, however, and the data examined were for one year. Future research could examine whether Vermont and national gender gap differences in reading, writing, and math in particular extend into subsequent years.

In both Vermont and nationally gender gaps in writing have been larger than those in reading and math. Consistent with long-term trends, Vermont and national gender gaps on the NAEP writing assessment were larger than gender gaps in reading and much larger (and in the opposite direction) than those in math (figure 5). Future research could examine these gaps in greater depth due to their scale and recorded persistence over time. Such research could explore whether gender gaps

FIGURE 5

National Assessment of Educational Progress gender gaps in grades 4 and 8, by content area and grade (points)



Note: Writing and reading data are from 2002; math data are from 2003.

Source: Authors' analysis based on NAEP data from the NAEP Data Explorer (U.S. Department of Education, National Center for Education Statistics 2008).

vary by other individual and social factors beyond poverty and disability status (such as race/ethnicity, age, and parent or teacher education levels). Studies could also probe whether gender gaps shift under different testing conditions—such as writing by hand or computer, or writing fiction or nonfiction. Further insight could also be gained by examining current gender gaps in writing at different points in the overall score distribution—such as at the mean and the upper and lower extremes. Such efforts would explore how differences between girls' and boys' scores vary by writing ability level, both in Vermont and nationally.

In both jurisdictions gender gaps in reading and writing tend to be larger in grade 8 than in grade 4. This pattern is consistent with findings from previous studies of NAEP scores (Freeman 2004). In the aggregate the reverse pattern appears to hold in math. From 2003 to 2007 gender gaps in both jurisdictions were slightly smaller in grade 8 than in grade 4. Both these patterns may reflect the same phenomenon—as students advance in grade level, scores may increase more for girls than for boys. A larger dataset—representing more grades over a longer time—could test this hypothesis, both in the aggregate and controlling for student background variables. If girls' scores increase more than boys' scores for at least some student groups as they progress through school, it is unclear why. Might some girls develop academically at a faster rate than boys do? Or might some girls learn to become better test-takers faster than boys do? If either of these propositions is true, do they hold equally for high-scoring and low-scoring students? All these questions might merit further investigation.

Gender gaps were smaller in reading and writing, and larger in math, after controlling for disability status. In both jurisdictions girls achieved higher average scale scores than did boys on the NAEP reading and writing assessments. After controlling for disability status, however, gender gaps in reading were smaller by as much as 100 percent for students with disabilities and 25 percent for students without disabilities. In writing gender gaps were as much as 100 percent smaller for students with disabilities and 10 percent smaller for students without disabilities. In math gender gaps were as much as 400 percent larger for students with disabilities and 60 percent larger for students without disabilities.

Because gender gaps in all three content areas shift markedly after students are disaggregated by disability status, the question emerges whether boys and girls are being evenly distributed into disability status categories: further research is necessary to better understand these gender gap shifts

Because gender gaps in all three content areas shift markedly after students are disaggregated by disability status, the question emerges whether boys and girls are being evenly distributed into disability status categories. Girls achieve higher average reading scores than boys in the total Vermont and national populations; this gender gap can shrink for students with disabilities if boys with Individualized Education Programs tend to have relatively high reading scores, while girls with Individualized Education Programs tend to have very low reading scores. Different score distributions among boys and girls with Individualized Education Programs can also underlie the larger gender gaps found in math after disaggregating by disability status. Further research into score distributions by gender and disability status is necessary to better understand these gender gap shifts.

With only a few exceptions gender gaps in Vermont on the 2006 NECAP have consistently differed from gender gaps on the 2007 NAEP by 5 percentile points or fewer in reading and math. This result was remarkably consistent except in two cases for grade 8 math scores. Among 8th graders in poverty boys outperformed girls in math on the 2007 NAEP by 8 percentiles, and girls outperformed boys on the 2006 NECAP by 4 percentiles, for a 12 percentile point difference between gender gaps. Girls performed better relative to boys on the NECAP than on the NAEP. Among 8th graders with disabilities boys

outperformed girls by 12 percentiles on the 2007 NAEP and 5 percentiles on the 2006 NECAP, for a 7 percentile point difference. And girls again performed better relative to boys on the NECAP than on the NAEP.

These cases underscore another consistent finding: in all but one case the degree to which girls outperformed boys in reading and neared boys' performance in math was greater on the 2006 NECAP than on the 2007 NAEP. The one exception was in math among grade 4 students with disabilities; boys outscored girls by 11 percentiles on the 2007 NAEP and 16 percentiles on the 2006 NECAP. Future research might examine why gender gaps on the NECAP appear to favor girls—especially among grade 8 students in poverty or with disabilities.

Finally, observations from both NECAP and NAEP data show that poverty and disability gaps in test scores are much larger than gender gaps. More detailed examination of poverty and disability gaps, both before and after controlling for other background student characteristics, might be interesting avenues for further research.

APPENDIX A
DATA SOURCES, RESEARCH METHODS, AND LIMITATIONS

This section describes the study's data sources, research methods, and limitations. National Assessment of Educational Progress (NAEP) data were retrieved from the NAEP Data Explorer database (U.S. Department of Education, National Center for Education Statistics 2008) and New England Common Assessment Program (NECAP) data were provided by the Vermont Department of Education (Vermont Department of Education, Standards and Assessment 2007). Researchers examined the data to measure gender gaps, differences between gender gaps, and the statistical significance of gender gaps and differences between gender gaps. Analysis of data was limited by the constraints of publicly available data, the difficulties of comparing data from different assessments, and the focus of the study.

Data sources

Researchers used the NAEP Data Explorer online database to determine gender gaps in publicly available NAEP reading (2002, 2003, 2005, 2007), writing (2002), and math (2000, 2003, 2005, 2007) assessment scores for 4th and 8th graders in Vermont and nationally. The NAEP Data Explorer also provided Vermont NAEP reading and math results (2007), which researchers compared with NECAP results (2006) (Vermont Department of Education, Standards and Assessment 2007) to determine how different assessments measured gender gaps.

Average scale NAEP and NECAP scores were examined by gender both before and after disaggregating by poverty status (defined by eligibility in the National School Lunch Program) and disability status (defined by eligibility for an Individualized Education Program). At every grade level NAEP scale scores range from 0 to 500 in reading and math and from 0 to 300 in writing. The NECAP is scored with a range of 80 points for each grade. In grade 4 scores range from 400 to 480, in grade 5 from 500 to 580, in grade 8 from 800 to 880, and so on.

Occasionally, a student in one grade will be administered the NECAP test for a lower level grade, resulting in a score that was scaled for a lower grade. All these cases—a small proportion of the total—were excluded from the analysis. Exclusion rates differed very little across student subgroups (for details on the size of the original NECAP data sets and exclusion rates, see table D1 in appendix D).

Comparison years for the New England Common Assessment Program and National Assessment of Educational Progress in reading and math. This project had access to student-level data from the Vermont NECAP reading and math assessments administered in the fall (October) of 2005 and 2006. Aggregate data for the most recent NAEP reading and math assessments, available through the NAEP Data Explorer, were for tests administered in the winter and spring (January through March) of 2005 and 2007.

This report compares gender gaps in reading and math from the NECAP administered in fall 2006 and the NAEP administered in early 2007. Students in grades 4 and 8 who took assessments in these years were from the same grade-level cohorts (table A1, cohorts A and E). Although this project had access to the 2005 Vermont NECAP and NAEP scores, comparing these two assessments was problematic and was therefore not conducted.

Specifically, a comparison of grade 4 scores on the 2005 NAEP with grade 4 scores on the 2005 NECAP would have compared outcomes from different student cohorts (see table A1). Differences in gender gaps from these comparisons could be due to variation between student cohorts as well as the two assessments. Alternatively, grade 4 scores on the 2005 NAEP and grade 5 scores on the 2005 NECAP are from the same student cohort, but differences between these two sets of scores could be due not only to variation in the two assessments but to the change in student grade levels (see cohort C in table A1).

To be clear, differences in gender gaps found from NAEP and NECAP tests administered to students

TABLE A1

Vermont reading and math assessments for the National Assessment of Educational Progress (NAEP) and the New England Common Assessment Program (NECAP) by student cohort and grade levels, 2004–07

	2004/05	2005/06	2006/07	
Student cohort	2005 NAEP[a]	2005 NECAP[b]	2006 NECAP[b]	2007 NAEP[a]
Cohort A		Grade 3	Grade 4	Grade 4
Cohort B		Grade 4	Grade 5	
Cohort C	Grade 4	Grade 5	Grade 6	
Cohort D		Grade 6	Grade 7	
Cohort E		Grade 7	Grade 8	Grade 8

Note: Shaded gender gap data were examined and compared in this report. The NAEP was administered from January through March in 2005 and 2007 and was not administered in 2006. The NAEP is administered only to students in grades 4, 8, and 12. The NECAP was administered in October of 2005 and 2006 to students in grades 3 through 8.

Source: U.S. Department of Education, National Center for Education Statistics (2008) and Vermont Department of Education, Standards and Assessment (2007).

from the same grade level and cohort cannot be attributed entirely to differences between the two tests. Other factors associated with each assessment could contribute to different student outcomes. Because student cohorts and grade levels are likely to be related to student outcomes, however, comparisons of NAEP and NECAP data without controlling for these factors were avoided.

Comparisons between New England Common Assessment Program and National Assessment of Educational Progress writing scores. The NAEP writing assessment was last administered to students in grades 4 and 8 in Vermont and across the country in 2002. Meanwhile, the NECAP writing assessment has been administered to students in grades 5 and 8 in Vermont since 2005. Comparisons of writing scores from the 2002 NAEP and 2005 or 2006 NECAP would have involved comparisons between different tests, student cohorts, and grade levels. Thus, gender gaps in writing measured by the two different assessments were not compared for this report. Instead, Vermont gender gaps in writing were calculated from the 2006 NECAP and are presented separately in table F2 in appendix F.

Research methods

Comparing gender gaps in Vermont New England Common Assessment Program and U.S. National Assessment of Educational Progress scores. Two sets of differences were calculated and tested for statistical significance: differences between girls' and boys' NAEP scores (gender gaps) within each jurisdiction and differences between Vermont and national gender gaps.

Differences between girls' and boys' National Assessment of Educational Progress scores (gender gaps) within each jurisdiction. Gender gaps were calculated as the mean scale score of girls minus the mean scale score of boys ($F - M$) in each subject. T-tests were conducted using the NAEP Data Explorer significance-testing tool to determine whether gender gaps (among all students and within poverty and disability subgroups) in each jurisdiction were significantly different from zero at the $p < 0.05$ level. In total, 177 gender gaps (79 in reading, 20 in writing, and 78 in math) were tested for statistical significance in the two jurisdictions (table A2).

To provide readers with a sense of the scale of calculated gender gaps, differences in average scale scores between girls and boys were presented two additional ways:

- *Effect sizes.* Gender gaps in each jurisdiction, subject, and student subgroup were also calculated as effect sizes to provide a standardized measure of the difference between girls' and boys' mean scale scores. Specifically, gender gap estimates (measured in NAEP scale score units) were transformed into effect sizes (measured in standard deviation units) using Cohen's *d*. Effect size was calculated as $|d| = (M_1 - M_2) / \sigma_{pooled,}$ where M_1 is the first mean, M_2 is the second mean, and σ_{pooled} is the pooled standard deviation for the distribution around each mean, or $\sigma_{pooled} = \sqrt{[(\sigma_1^2 + \sigma_2^2) / 2]}$. The "effect" of being a girl within the NAEP and NECAP was calculated by taking the difference between average girls' and boys'

TABLE A2

Total differences examined and number of statistically significant differences in grades 4 and 8 for Vermont and U.S. National Assessment of Educational Progress reading, writing, and math scores, 2000–07

Differences	Number			Percentage		
	Reading (2002–07)	Writing (2002)	Math (2000–07)	Reading (2002–07)	Writing (2002)	Math (2000–07)
Gender gaps (differences between mean girls' and boys' scores)						
Total gender gaps examined	79	20	78	100.0	100.0	100.0
Statistically significant gender gaps	66	18	51	83.5	90.0	65.4
Vermont						
All students	8	2	3	10.1	10.0	3.8
Students by poverty status	13	4	4	16.5	20.0	5.1
Students by disability status	8	2	10	10.1	10.0	12.8
National						
All students	8	2	5	10.1	10.0	6.4
Students by poverty status	16	4	13	20.3	20.0	16.7
Students by disability status	13	4	16	16.5	20.0	20.5
Differences between Vermont and national gender gaps						
Total differences between gaps examined	39	10	38	100.0	100.0	100.0
Differences between statistically significant gaps	2	1	2	5.1	10.0	5.3
All students	0	0	1	0.0	0.0	2.6
Students by poverty status	2	1	0	5.1	10.0	0.0
Students by disability status	0	0	1	0.0	0.0	2.6

Note: Statistical significance (significantly different from zero) is defined as $p < 0.05$.

Source: Authors' analysis based on NAEP data from NAEP Data Explorer (U.S. Department of Education, National Center for Education Statistics 2008).

mean scale scores within each assessment and dividing this difference ($F - M$, the test score gender gap) by the pooled standard deviation for boys' and girls' scores. The same methodology was used to compare the effect of being a student not in poverty with a student in poverty ($NP - P$), and a student without disabilities with a student with disabilities ($ND - D$).

- *Percentile differences.* To aid in the interpretation of effect sizes, Cohen's d statistics were transformed into Cohen's U_3 statistics, and calculations were performed to translate the gender gap into a percentile difference between the average girl and the average boy in the test score distribution.

First, Cohen's U_3 statistics were determined by using a z-score table to find the area under the standard normal curve below the value of each gender gap effect size. For example, the U_3 statistic for a gender gap of $d = 0.22$ is 0.59. If, on average, girls outscored boys by 0.22 standard deviation on a specific assessment, the average girl would have ranked at the 59th percentile of the boys' score distribution.

Second, percentile differences between girls' and boys' scores were calculated as the percentile associated with the U_3 statistic, minus 50. Following the example above, the average boy would have achieved a score at the 50th percentile, assuming scores were normally

distributed. Consequently, the average girl would have achieved a score placing her 9 percentiles higher than the average boy.

This approach assumes that scores are normally distributed and that the standard deviations of scores for both groups are similar. Inspections of student-level NECAP score distributions showed reading, writing, and math scores to be normally distributed for the grade levels examined in this report. Standard deviations for boys' and girls' scores were generally quite similar on both the NECAP and NAEP (see appendixes E and F).

NAEP score distributions can be determined by examining the raw data (Mislevy, Johnson, and Muraki 1992). NAEP score distributions are not available, however, through the NAEP Data Explorer. Because the project had access to NAEP data only through the NAEP Data Explorer, this report assumes that boys' and girls' NAEP scores are normally distributed when calculating U_3 from the available NAEP data. Ultimately, the calculations performed and the distributional assumptions made do not allow this report to provide precise determinations of percentile differences between boys and girls; instead, alternative measures of gender gaps are provided to give readers an approximate gauge for the size and import of calculated gender gap effect sizes.

Differences between Vermont and national gender gaps. The difference between each pair of Vermont and national gender gaps was calculated as the Vermont gap minus the national gap (*VTgap – U.S.gap*). A series of calculations were then performed to determine whether differences between Vermont and national gender gaps were statistically significant.

First, the standard error for each gender gap estimate (SE_{F-M}) was calculated as $SE_{F-M} = \sqrt{(SE_F^2 + SE_M^2)}$, where SE_F and SE_M were the standard errors for girls' and boys' mean scale score estimates (downloaded from the NAEP Data Explorer database).

Second, the standard error for the difference between each pair of Vermont and national gender gaps ($SE_{VTgap-U.S.gap}$) was derived using the standard errors calculated for the individual Vermont and national gender gaps. The formula used was $SE_{VTgap-U.S.gap} = \sqrt{(SE_{VTgap}^2 + SE_{U.S.gap}^2)}$, where $SE_{VTgap} = SE_{F-M}$ in Vermont, and $SE_{U.S.gap} = SE_{F-M}$ in the country.

Third, t-scores were calculated for the differences between Vermont and U.S. gender gaps to determine whether these differences were statistically significant. These scores were calculated as $t_{VTgap-U.S.gap} = (VTgap - U.S.gap) / SE_{VTgap-U.S.gap.}$ Because the Vermont and U.S. NAEP samples were large and viewed as independent, the difference between Vermont and national gender gaps was considered statistically significant if the absolute value of $t_{VTgap-U.S.gap}$ was greater than 1.96.[6] This figure is the minimum standardized score that allows one to reject, at a 95 percent level of certainty, the null hypothesis that means (and differences between means) from two large and independent samples are equal.

In total, 87 differences in gender gaps (39 in reading, 10 in writing, and 38 in math) were tested for statistical significance between the two jurisdictions (see table A2). Differences between Vermont and national gender gaps were described in both scale score and percentile points.

Comparing gender gaps in Vermont New England Common Assessment Program and Vermont National Assessment of Educational Progress scores. Gender gaps in reading and math scores were calculated from the 2006 Vermont NECAP and compared with gender gaps in reading and math scores from the 2007 Vermont NAEP. Students who took the 2006 NECAP and 2007 NAEP reading and math tests were from the same grade-level cohorts (see table A1, cohorts A and E). Writing scores from the Vermont NAEP and NECAP were

not compared because data from common grade-level cohorts were not available.

The same statistics derived from the NAEP data—average scale scores, standard deviations, and standard errors—by gender (both in the aggregate and within student poverty and disability subgroups) were calculated on the 2006 Vermont NECAP data in reading and math. Researchers conducted t-tests to determine whether differences in the average scale scores of girls and boys were statistically significant at $p < 0.05$. Researchers did not measure the statistical significance of differences between gender gaps on the NAEP and NECAP. Instead, standardized effect sizes (Cohen's *d*) and their percentile equivalents were calculated for all gap estimates to compare gender gaps on the two assessments.

Limitations of the study

Constraints associated with publicly available NAEP data, the difficulties of comparing outcomes from different assessments, and the bounded nature of this empirical project placed limitations on the NAEP and NECAP analyses and the conclusions to be drawn.

This report does not indicate whether gender gaps measured by the Vermont NECAP and NAEP assessments differ from each other at statistically significant levels. The NECAP and NAEP measure student achievement with different scales; test score gaps were converted into standardized effect sizes to facilitate comparisons between the two assessments. Although the reporting of confidence intervals around effect sizes is becoming a recommended practice (Cumming 2001; Nix and Barnette 1998; Steiger 2004; Thompson 1998), calculating these intervals was beyond the scope of this project. Instead, gender gaps from each assessment are presented in standardized effect size units and their percentile equivalents to help readers form their own judgments about the extent and importance of gender gap differences.

Direct contrasts between Vermont NECAP and NAEP writing scores were not presented because available data from each assessment were drawn from different grade-level cohorts and were therefore not comparable. This report therefore responds to the project's second research question by comparing outcomes from the Vermont NECAP and NAEP assessments in reading and math only.

Finally, results are presented in terms of statistical, rather than substantive or practical, significance. In statistical tests two estimates are assumed to be equivalent and are considered different at statistically significant levels if the size of their difference could occur by chance with a probability of less than 5 percent ($p < 0.05$). Statistical significance is different from substantive or practical significance, which may vary by context (Light, Singer, and Willett 1990). This report does not make any claims about the substantive or practical significance of results.

APPENDIX B
THE VERMONT NATIONAL ASSESSMENT OF EDUCATIONAL PROGRESS AND THE NEW ENGLAND COMMON ASSESSMENT PROGRAM

This appendix describes the National Assessment of Educational Progress (NAEP) and the New England Common Assessment Program (NECAP).

The National Assessment of Educational Progress

The NAEP, also known as "The Nation's Report Card," consists of several ongoing assessment programs that are administered by the National Center for Education Statistics. The main national NAEP program has been testing students across the country in reading, writing, math, science, and four other content areas since 1969. The state NAEP program, which uses the same assessments as the national program, has been conducted on a voluntary basis by many states and jurisdictions since 1990. Under the No Child Left Behind (NCLB) Act, all states receiving Title I funds are now mandated to participate in the national assessment. The same is true for school districts receiving Title I funds. Participation in the NAEP is voluntary, however, for schools and students.

Through the NAEP program student achievement has been measured in grades 4 and 8 at the state level and in grade 12 at the national level. Reading and math assessments were administered every two to four years from 1990 to 2002, and the writing assessment was administered in grades 4 and 8 in 2002. Since 2003 reading and math assessments have occurred every other year. Vermont has participated in the state NAEP program since the mid-1990s.[7] Because the NAEP contains both state- and national-level data, NAEP data can be used to compare assessment results among different states and the country as a whole.

The NAEP does not test all students. Instead, NAEP samples a representative portion of populations within individual states and nationwide. The U.S. NAEP sample combines subsets of the students who take the individual state NAEP with a sample of students in nonparticipating states to create a nationally representative sample. The NAEP oversamples underrepresented populations (such as minorities and students from rural areas) and uses weighting procedures to generate samples that are representative of states and the country.

Each student who participates in the NAEP receives a portion of all the questions on the assessment (roughly 25 percent of the total). NAEP scores for individual students are therefore estimated by generating and then tallying plausible values for each test item, based on the overall performance of students with similar background characteristics and cognitive performance levels.

Because students' NAEP scores are estimates based on plausible values, it is not possible to know how any one student performed on an entire NAEP. The design of the test makes it possible to determine overall score estimates for groups of students but not for individual students. In addition, it is not possible to determine how many students took a specific assessment. Although the NAEP Data Explorer provides standard deviation and standard error estimates for student subgroup scores, student sample sizes calculated from these estimates are difficult to interpret because of the NAEP's sample weighting procedures. For more information on the NAEP, see http://nces.ed.gov/nationsreportcard/about/.

The New England Common Assessment Program

The NECAP is administered to all public school students in Vermont in grades 3–8. NECAP assessments were first used in Vermont in 2005. Developers of the NECAP (described as "The New England Common Test Program" in technical documentation) explain the purpose of the program as follows: "The New England Common Test Program (NECAP) is the result of collaboration among New Hampshire (NH), Rhode Island (RI), and Vermont (VT) to build a set of tests for grades 3 through 8 to meet the requirements of the No Child Left Behind Act (NCLB). The purposes

of the tests are as follows: (1) Provide data on students' achievement in reading/language arts and mathematics to meet the requirements of NCLB; (2) provide information to support program evaluation and improvement; and (3) provide to parents and the public information on the performance of students and schools. The tests are constructed to meet rigorous technical criteria, include universal design elements and accommodations so that students can access test content, and gather reliable student demographic information for accurate reporting." (Measured Progress 2007, p. 5).

APPENDIX C
STUDENT DEMOGRAPHICS FOR VERMONT AND THE COUNTRY, 2005/06

TABLE C1

Characteristics of student population in Vermont and the country, 2005/06

	Number		Percent	
Characteristic	Vermont	National[a]	Vermont	National[a]
Total	96,638	963,009		
Gender				
Male	49,861	489,697	51.6	50.9
Female	45,961	462,948	47.6	48.1
Race/ethnicity				
Asian	1,496	43,957	1.5	4.6
Black, non-Hispanic	1,424	164,252	1.5	17.1
White, non-Hispanic	91,528	544,233	94.7	56.5
Other	417	11,660	0.4	1.2
Hispanic	957	189,047	1.0	19.6
Grade				
PreK	4,061	20,323	4.2	2.1
Kindergarten	6,069	70,969	6.3	7.4
1	6,441	72,369	6.7	7.5
2	6,429	70,713	6.7	7.3
3	6,486	70,315	6.7	7.3
4	6,549	70,147	6.8	7.3
5	6,827	71,232	7.1	7.4
6	7,075	71,963	7.3	7.5
7	7,166	74,062	7.4	7.7
8	7,559	74,548	7.8	7.7
9	8,327	84,067	8.6	8.7
10	8,142	75,811	8.4	7.9
11	7,888	67,735	8.2	7.0
12	7,499	62,359	7.8	6.5
Ungraded	120	6,390	0.1	0.7
Poverty status				
Eligible for free lunch	18,820	310,723	19.5	32.3
Eligible for reduced-price lunch	6,667	70,825	6.9	7.4
English language learner status				
Limited English proficiency/ English language learner	1,775	82,806	1.8	8.6
Disability status				
With Individualized Education Program	10,915	130,940	11.3	13.6

a. National average is calculated as the total divided by the 50 states and the District of Columbia.

Source: U.S. Department of Education, National Center for Education Statistics (2006a,b,c).

APPENDIX D
NEW ENGLAND COMMON ASSESSMENT PROGRAM CASES USED IN ANALYSES

TABLE D1

Total number of available cases in Vermont's New England Common Assessment Program (NECAP) dataset and number of cases used in analyses, 2006

	Grades 4 and 5			Grade 8		
		Used			Used	
Content area	Total	Number	Percent	Total	Number	Percent
Reading						
Male students	3,351	3,338	99.6	3,761	3,739	99.4
Female students	3,124	3,119	99.8	3,474	3,462	99.7
Students in poverty	2,013	2,004	99.6	1,943	1,922	98.9
Students not in poverty	4,468	4,459	99.8	5,292	5,279	99.8
Students with disabilities	748	734	98.1	1,034	1,005	97.2
Students without disabilities	5,733	5,729	99.9	6,201	6,196	99.9
Math						
Male students	3,360	3,349	99.7	3,751	3,733	99.5
Female students	3,125	3,119	99.8	3,476	3,465	99.7
Students in poverty	2,017	2,011	99.7	1,941	1,923	99.1
Students not in poverty	4,474	4,463	99.8	5,286	5,275	99.8
Students with disabilities	751	736	98.0	1,022	998	97.7
Students without disabilities	5,740	5,738	100.0	6,205	6,200	99.9
Writing						
Male students	3,324	3,324	100.0	3,719	3,719	100.0
Female students	3,159	3,159	100.0	3,454	3,453	100.0
Students in poverty	2,031	2,031	100.0	1,909	1,908	100.0
Students not in poverty	4,452	4,452	100.0	5,264	5,264	100.0
Students with disabilities	782	782	100.0	994	993	99.9
Students without disabilities	5,701	5,701	100.0	6,179	6,179	100.0

Note: The NECAP reading and math assessments were administered to students in grades 4 and 8 in 2006. The NECAP writing assessment was administered to students in grades 5 and 8 that same year.

Source: Authors' tabulations of student-level NECAP data (Vermont Department of Education, Standards and Assessment 2007)

TABLE E1

Grade 4 National Assessment of Educational Progress average scaled reading scores by gender and student poverty and disability status in Vermont and the country, 2002, 2003, 2005, and 2007

	2002				2003				2005				2007			
	Vermont		National		Vermont		National		Vermont		National		Vermont		National	
Students	Mean	Standard error	Mean	Standard error	Mean	Standard error	Mean	Standard error	Mean	Standard error	Mean	Standard error	Mean	Standard error	Mean	Standard error
Total																
Male students	222.9	1.36	213.6	0.45	223.5	1.11	212.7	0.34	223.2	1.14	214.2	0.25	224.5	1.30	216.3	0.31
	(32.3)	1.33	(36.3)	0.31	(31.6)	0.61	(37.6)	0.17	(33.1)	1.19	(36.5)	0.15	(34.4)	0.93	(36.5)	0.18
Female students	231.3	1.53	220.1	0.53	228.9	1.24	220.2	0.30	230.3	1.19	220.5	0.27	232.1	0.94	223.0	0.28
	(32.0)	1.04	(35.6)	0.38	(30.7)	0.81	(36.4)	0.20	(33.4)	0.89	(35.6)	0.16	(33.2)	0.96	(35.0)	0.15
Gender gap (points)	8.4**	2.05**	6.5**	0.70**	5.4**	1.66**	7.5**	0.45**	7.1**	1.65**	6.3**	0.37**	7.6**	1.61**	6.7**	0.41**
Effect size (*d*)	0.26		0.18		0.17		0.20		0.21		0.17		0.22		0.19	
U_3	.60		.57		.57		.58		.58		.57		.59		.57	
Students in poverty																
Male students	210.0	2.63	198.9	0.74	212.2	1.67	197.0	0.39	210.2	2.14	199.4	0.31	209.4	2.34	201.3	0.39
	(32.1)	2.03	(35.6)	0.41	(31.8)	1.63	(36.5)	0.19	(34.5)	1.74	(35.2)	0.20	(33.9)	1.15	(35.6)	0.21
Female students	215.8	2.57	205.9	0.77	214.9	0.34	205.2	0.43	210.7	2.45	206.0	0.34	215.6	1.84	208.7	0.31
	(30.8)	1.85	(35.1)	0.54	(30.0)	0.19	(35.2)	0.34	(32.5)	1.47	(34.3)	0.19	(31.3)	1.47	(33.9)	0.22
Gender gap (points)	5.8	3.68**	7.0**	1.07**	2.7	1.70**	8.2**	0.58**	0.5	3.25**	6.6**	0.45**	6.2**	2.98**	7.4**	0.50**
Effect size (*d*)	0.18		0.20		0.09		0.23		0.01		0.19		0.19		0.21	
U_3	.57		.58		.54		.59		.50		.57		.57		.58	
Students not in poverty																
Male students	228.4	1.54	225.8	0.49	227.7	1.10	225.3	0.41	228.8	0.29	226.5	0.28	231.3	1.31	228.6	0.29
	(30.9)	1.47	(32.0)	0.38	(30.5)	0.82	(33.5)	0.23	(30.7)	0.21	(32.6)	0.18	(32.4)	1.04	(32.2)	0.21
Female students	237.3	1.87	232.8	0.48	235.1	1.27	232.8	0.36	238.0	1.28	233.0	0.28	239.5	1.19	234.9	0.29
	(30.6)	1.32	(31.1)	0.32	(29.0)	0.79	(32.3)	0.24	(30.7)	1.03	(31.6)	0.21	(31.3)	1.19	(31.2)	0.19
Gender gap (points)	8.9**	2.42**	7.0**	0.69**	7.4**	1.67**	7.5**	0.55**	9.2**	1.31**	6.5**	0.40**	8.2**	1.77**	6.3**	0.41**
Effect size (*d*)	0.29		0.22		0.25		0.23		0.30		0.20		0.26		0.20	
U_3	.61		.59		.60		.59		.62		.58		.60		.58	

(CONTINUED)

TABLE E1 (CONTINUED)

Grade 4 National Assessment of Educational Progress average scaled reading scores by gender and student poverty and disability status in Vermont and the country, 2002, 2003, 2005, and 2007

	2002				2003				2005				2007			
	Vermont		National		Vermont		National		Vermont		National		Vermont		National	
Students	Mean	Standard error	Mean	Standard error	Mean	Standard error	Mean	Standard error	Mean	Standard error	Mean	Standard error	Mean	Standard error	Mean	Standard error
Students with disabilities																
Male students	196.3	3.82	185.6	0.90	203.9	3.87	182.8	0.80	196.0	3.12	189.0	0.68	195.6	2.93	189.9	0.60
	(32.4)	3.20	(38.1)	0.50	(32.9)	2.20	(40.2)	0.38	(33.5)	2.80	(39.1)	0.45	(36.0)	1.76	(42.4)	0.45
Female students	—		188.4	1.14	199.7	3.70	187.3	1.00	190.5	3.99	190.7	0.70	189.9	4.34	190.6	0.89
	—		(38.4)	0.81	(30.7)	2.35	(39.9)	0.81	(31.9)	2.76	(39.5)	0.66	(41.8)	4.34	(42.6)	0.70
Gender gap (points)	—		2.8**	1.45**	–4.2**	5.35**	4.5**	1.28**	–5.5**	5.06**	1.7**	0.98**	–5.7**	5.23**	0.7**	1.07**
Effect size (*d*)	—		0.07		–0.13		0.11		–0.17		0.04		–0.15		0.02	
U_3			.53		.45		.54		.43		.52		.44		.51	
Students without disabilities																
Male students	226.3	1.36	216.8	0.52	226.9	1.19	216.9	0.38	228.0	1.16	217.7	0.26	231.1	1.21	220.1	0.30
	(30.7)	1.33	(34.6)	0.36	(30.0)	0.68	(35.3)	0.17	(30.6)	1.19	(34.6)	0.16	(30.4)	0.90	(33.9)	0.17
Female students	233.3	1.45	222.1	0.56	231.2	1.27	222.7	0.30	233.4	1.15	222.6	0.27	235.7	0.93	225.4	0.27
	(30.8)	0.93	(34.5)	0.43	(29.4)	0.78	(34.9)	0.20	(31.5)	0.92	(34.3)	0.15	(29.7)	0.74	(33.1)	0.13
Gender gap (points)	7.0**	1.99**	5.3**	0.76**	4.3**	1.74**	5.8**	0.48**	5.4**	1.63**	4.9**	0.38**	4.6**	1.53**	5.3**	0.40**
Effect size (*d*)	0.23		0.15		0.14		0.17		0.17		0.14		0.15		0.16	
U_3	.59		.56		.56		.57		.57		.56		.56		.56	

**Gender gap is statistically significant (different from zero) at $p < 0.05$.

— indicates that reporting standards for the NAEP were not met.

Note: Numbers in parentheses are standard deviations. Gender gaps are defined as the mean scale score of girls minus the mean scale score of boys ($F - M$). Standard errors for the gender gap estimates were calculated as $SE_{F-M} = \sqrt{(SE_F^2 + SE_M^2)}$. Effect size is calculated as $|d| = (F - M) / SD_{pooled}$, where $SD_{pooled} = \sqrt{[(SD_F^2 + SD_M^2) / 2]}$. U_3 is the proportion of scores in the group with the lower mean that fall below the noted effect size in a standard normal distribution. See appendix A for details.

Source: Authors' analysis based on 2002, 2003, 2005, and 2007 NAEP reading data for grade 4 from U.S. Department of Education, National Center for Education Statistics (2008).

TABLE E2

Differences in Vermont and U.S. National Assessment of Educational Progress gender gaps in grade 4 reading scores by student poverty and disability status, 2002, 2003, 2005, and 2007

	2002			2003			2005			2007		
Students	Vermont	National	Difference in gaps	Vermont	National	Difference in gaps	Vermont	National	Difference in gaps	Vermont	National	Difference in gaps
Total												
Gender gap (points)	8.4	6.5	1.9	5.4	7.5	–2.1	7.1	6.3	0.8	7.6	6.7	0.9
Standard error	2.05	0.70	2.16	1.66	0.45	1.72	1.65	0.37	1.69	1.61	0.41	1.66
Students in poverty												
Gender gap (points)	5.8	7.0	–1.2	2.7	8.2	–5.5**	0.5	6.6	–6.1	6.2	7.4	–1.2
Standard error	3.68	1.07	3.83	1.70	0.58	1.80	3.25	0.45	3.28	2.98	0.50	3.02
Students not in poverty												
Gender gap (points)	8.9	7.0	1.9	7.4	7.5	–0.1	9.2	6.5	2.7**	8.2	6.3	1.9
Standard error	2.42	0.69	2.52	1.67	0.55	1.76	1.31	0.40	1.37	1.77	0.41	1.82
Students with disabilities												
Gender gap (points)	—	2.8	—	–4.2	4.5	–8.7	–5.5	1.7	–7.2	–5.7	0.7	–6.4
Standard error	—	1.45	—	5.35	1.28	5.50	5.06	0.98	5.15	5.23	1.07	5.34
Students without disabilities												
Gender gap (points)	7.0	5.3	1.7	4.3	5.8	–1.5	5.4	4.9	0.5	4.6	5.3	–0.7
Standard error	1.99	0.76	2.13	1.74	0.48	1.81	1.63	0.38	1.68	1.53	0.40	1.58

**The difference in the Vermont and national gender gap estimates is statistically significant (different from zero) at $p < 0.05$.

— indicates that reporting standards for the NAEP were not met.

Note: Gender gaps are defined as the mean scale score of girls minus the mean scale score of boys ($F - M$). Standard errors for the difference in Vermont and national gender gap estimates were calculated as $SE_{VT\,gap\,-\,U.S.\,gap} = \sqrt{(SE_{VT\,gap}{}^2 + SE_{U.S.\,gap}{}^2)}$.

Source: Authors' analysis based on 2002, 2003, 2005, and 2007 NAEP reading data for grade 4 from U.S. Department of Education, National Center for Education Statistics (2008).

TABLE E3

Grade 8 National Assessment of Educational Progress average scaled reading scores by gender and student poverty and disability status in Vermont and the country, 2002, 2003, 2005, and 2007

	2002				2003				2005				2007			
	Vermont		National		Vermont		National		Vermont		National		Vermont		National	
Characteristic	Mean	Standard error	Mean	Standard error	Mean	Standard error	Mean	Standard error	Mean	Standard error	Mean	Standard error	Mean	Standard error	Mean	Standard error
Total																
Male students	267.2	1.38	258.1	0.53	264.8	1.25	256.1	0.26	262.2	1.08	255.2	0.23	267.6	1.27	256.0	0.25
	(31.5)	0.84	(34.1)	0.34	(31.8)	0.78	(35.6)	0.22	(32.5)	0.79	(35.4)	0.16	(30.5)	0.84	(35.0)	0.18
Female students	276.5	1.01	267.5	0.50	276.2	1.10	266.6	0.29	275.6	1.24	265.6	0.21	278.2	1.22	266.1	0.26
	(30.4)	0.63	(32.9)	0.29	(31.1)	1.10	(34.1)	0.24	(31.7)	0.86	(34.0)	0.16	(29.3)	0.90	(33.6)	0.19
Gender gap (points)	9.3**	1.71**	9.4**	0.73**	11.4**0	1.67**	10.5**	0.39**	13.4**	1.64**	10.4**	0.31**	10.6**	1.76**	10.1**	0.36**
Effect size (*d*)	0.30		0.28		0.36		0.30		0.42		0.30		0.35		0.29	
U_3	.62		.61		.64		.62		.66		.62		.64		.61	
Students in poverty																
Male students	251.9	2.89	244.8	0.68	250.5	2.21	240.5	0.43	250.0	2.12	241.4	0.36	255.0	2.23	242.2	0.33
	(33.8)	1.56	(33.9)	0.45	(30.1)	1.48	(35.5)	0.36	(32.9)	1.54	(34.7)	0.22	(29.6)	1.33	(34.5)	0.30
Female students	261.6	2.12	253.3	0.56	260.2	1.74	251.2	0.45	260.5	2.14	251.5	0.30	266.0	1.78	252.2	0.37
	(30.6)	1.69	(32.7)	0.39	(29.1)	1.69	(33.6)	0.39	(31.4)	1.59	(33.0)	0.24	(30.5)	1.92	(32.9)	0.22
Gender gap (points)	9.7**	3.58**	8.5**	0.88**	9.7**	2.81**	10.7**	0.62**	10.6**	3.01**	10.1**	0.47**	11.0**	2.85**	10.0**	0.50**
Effect size (*d*)	0.30		0.26		0.33		0.31		0.33		0.30		0.37		0.30	
U_3	.62		.60		.63		.62		.63		.62		.64		.62	
Students not in poverty																
Male students	271.7	1.38	266.3	0.59	269.6	1.34	265.4	0.31	267.4	1.18	264.2	0.23	272.4	1.33	265.6	0.30
	(29.3)	0.75	(31.1)	0.35	(31.0)	0.85	(32.1)	0.24	(30.7)	0.99	(32.9)	0.23	(29.5)	0.82	(32.0)	0.22
Female students	280.5	1.25	276.5	0.54	281.9	1.24	276.2	0.35	281.5	1.25	275.3	0.23	282.3	1.44	275.9	0.30
	(29.1)	0.73	(29.3)	0.32	(29.8)	1.38	(30.7)	0.19	(29.9)	1.03	(31.0)	0.19	(27.7)	0.93	(30.4)	0.25
Gender gap (points)	8.8**	1.86**	10.2**	0.80**	12.3**	1.83**	10.8**	0.47**	14.1**	1.72**	11.1**	0.33**	9.9**	1.96**	10.3**	0.42**
Effect size (*d*)	0.30		0.34		0.40		0.34		0.47		0.35		0.35		0.33	
U_3	.62		.63		.66		.63		.68		.64		.64		.63	

(CONTINUED)

TABLE E3 (CONTINUED)

Grade 8 National Assessment of Educational Progress average scaled reading scores by gender and student poverty and disability status in Vermont and the country, 2002, 2003, 2005, and 2007

	2002				2003				2005				2007			
	Vermont		National		Vermont		National		Vermont		National		Vermont		National	
Characteristic	Mean	Standard error	Mean	Standard error	Mean	Standard error	Mean	Standard error	Mean	Standard error	Mean	Standard error	Mean	Standard error	Mean	Standard error
Students with disabilities																
Male students	245.3	2.92	225.8	1.21	244.8	2.40	222.6	0.82	234.7	2.11	224.5	0.62	249.0	2.80	224.9	0.65
	(32.4)	1.45	(36.1)	0.91	(29.7)	1.61	(37.2)	0.52	(30.5)	1.72	(36.5)	0.37	(32.6)	1.86	(38.2)	0.54
Female students	249.6	3.30	230.6	1.51	246.4	3.81	227.8	0.96	238.2	3.57	229.8	0.80	246.6	3.97	228.7	0.77
	(28.8)	2.52	(36.4)	1.07	(30.7)	2.87	(36.0)	0.93	(29.8)	2.83	(36.1)	0.47	(35.0)	3.14	(38.3)	0.70
Gender gap (points)	4.3	4.41**	4.8**	1.93**	1.6	4.50**	5.2**	1.26**	3.5	4.15**	5.3**	1.01**	–2.4	4.86**	3.8**	1.01**
Effect size (*d*)	0.14		0.13		0.05		0.14		0.12		0.15		–0.07		0.10	
U_3	.56		.55		.52		.56		.55		.56		.47		.54	
Students without disabilities																
Male students	271.9	1.30	262.3	0.54	269.3	1.27	261.2	0.28	269.4	1.05	259.5	0.23	272.6	1.21	260.2	0.26
	(29.3)	0.83	(31.5)	0.35	(30.6)	0.93	(31.5)	0.23	(29.0)	0.77	(33.1)	0.17	(27.9)	0.84	(32.4)	0.20
Female students	279.3	1.01	270.0	0.47	279.1	1.08	269.6	0.29	279.5	1.13	268.2	0.20	281.6	1.20	268.6	0.25
	(29.1)	0.64	(31.0)	0.28	(29.6)	1.02	(32.1)	0.20	(29.2)	0.76	(32.3)	0.16	(26.5)	0.76	(31.7)	0.19
Gender gap (points)	7.4**	1.65**	7.8**	0.72**	9.8**	1.67**	8.4**	0.40**	10.1**	1.54**	8.7**	0.30**	9.0**	1.70**	8.4**	0.36**
Effect size (*d*)	0.25		0.25		0.33		0.26		0.35		0.27		0.33		0.26	
U_3	.60		.60		.63		.60		.64		.61		.63		.60	

**Gender gap is statistically significant (different from zero) at $p < 0.05$.

Note: Numbers in parentheses are standard deviations. Standard errors for the gender gap estimates were calculated as $SE_{F-M} = \sqrt{(SE_F^2 + SE_M^2)}$. Gender gaps are defined as the mean scale score of girls minus the mean scale score of boys ($F - M$). Effect size is calculated as $|d| = (F - M) / SD_{pooled}$, where $SD_{pooled} = \sqrt{[(SD_F^2 + SD_M^2) / 2]}$. U_3 is the proportion of scores in the group with the lower mean that fall below the noted effect size in a standard normal distribution. See appendix A for details.

Source: Authors' analysis based on 2002, 2003, 2005, and 2007 NAEP reading data for grade 8 from U.S. Department of Education, National Center for Education Statistics (2008).

TABLE E4

Differences in Vermont and U.S. National Assessment of Educational Progress gender gaps in grade 8 reading scores by student poverty and disability status, 2002, 2003, 2005, and 2007

	2002			2003			2005			2007		
Students	Vermont	National	Difference in gaps	Vermont	National	Difference in gaps	Vermont	National	Difference in gaps	Vermont	National	Difference in gaps
Total	9.3	9.4	–0.1	11.4	10.5	0.9	13.4	10.4	3.0	10.6	10.1	0.5
	1.71	0.73	1.86	1.67	0.39	1.71	1.64	0.31	1.67	1.76	0.36	1.80
Students in poverty												
Gender gap (points)	9.7	8.5	1.2	9.7	10.7	–1.0	10.6	10.1	0.5	11.0	10.0	1.0
Standard error	3.58	0.88	3.69	2.81	0.62	2.88	3.01	0.47	3.05	2.85	0.50	2.90
Students not in poverty												
Gender gap (points)	8.8	10.2	–1.4	12.3	10.8	1.5	14.1	11.1	3.0	9.9	10.3	–0.4
Standard error	1.86	0.80	2.03	1.83	0.47	1.88	1.72	0.33	1.75	1.96	0.42	2.01
Students with disabilities												
Gender gap (points)	4.3	4.8	–0.5	1.6	5.2	–3.6	3.5	5.3	–1.8	–2.4	3.8	–6.2
Standard error	4.41	1.93	4.81	4.50	1.26	4.68	4.15	1.01	4.27	4.86	1.01	4.96
Students without disabilities												
Gender gap (points)	7.4	7.8	–0.3	9.8	8.4	1.4	10.1	8.7	1.4	9.0	8.4	0.6
Standard error	1.65	0.72	1.80	1.67	0.40	1.72	1.54	0.30	1.57	1.70	0.36	1.74

Note: Gender gaps are defined as the mean scale score of girls minus the mean scale score of boys ($F - M$). Standard errors for the difference in Vermont and national gender gap estimates were calculated as $SE_{VT\,gap\,-\,U.S.\,gap} = \sqrt{(SE_{VT\,gap}{}^2 + SE_{U.S.\,gap}{}^2)}$. See appendix A for details.

Source: Authors' analysis based on 2002, 2003, 2005, and 2007 NAEP reading data for grade 8 from U.S. Department of Education, National Center for Education Statistics (2008).

TABLE E5

Grades 4 and 8 National Assessment of Educational Progress average scaled writing scores by gender and student poverty and disability status in Vermont and the country, 2002

	Grade 4				Grade 8			
	Vermont		National		Vermont		National	
Students	Mean	Standard error	Mean	Standard error	Mean	Standard error	Mean	Standard error
Total								
Male students	147.3	1.60	144.1	0.58	151.2	1.84	141.3	0.68
	(35.2)	1.16	(35.0)	0.51	(36.8)	1.11	(36.7)	0.38
Female students	168.6	1.77	161.6	0.43	175.4	1.34	162.2	0.63
	(35.0)	1.18	(35.3)	0.32	(36.4)	0.88	(35.8)	0.28
Gender gap (points)	21.3**	2.39**	17.5**	0.72**	24.2**	2.28**	20.9**	0.93**
Effect size (*d*)	0.61		0.50		0.66		0.58	
U_3	.73		.69		.75		.72	
Students in poverty								
Male students	129.5	2.20	132.4	1.00	130.9	3.01	126.4	0.61
	(34.1)	2.60	(33.2)	0.88	(37.8)	1.77	(35.0)	0.38
Female students	153.6	2.81	148.9	0.67	157	3.34	145.8	0.71
	(32.6)	1.71	(34.0)	0.42	(37.0)	1.86	(34.)	0.37
Gender gap (points)	24.1**	3.57**	16.5**	1.20**	26.1**	4.50**	19.4**	0.94**
Effect size (*d*)	0.72		0.49		0.70		0.56	
U_3	.76		.69		.76		.71	
Students not in poverty								
Male students	152.2	1.95	153.7	0.56	156.6	2.05	150.1	0.74
	(33.6)	1.30	(33.2)	0.43	(34.4)	1.46	(34.5)	0.32
Female students	174.8	2.19	172.6	0.51	180.7	1.38	172.3	0.79
	(33.81)	1.47	(32.7)	0.36	(34.4)	0.90	(32.3)	0.30
Gender gap (points)	22.6**	2.93**	18.9**	0.76**	24.1**	2.47**	22.2**	1.08**
Effect size (*d*)	0.67		0.57		0.70		0.66	
U_3	.75		.72		.76		.75	
Students with disabilities								
Male students	116.1	3.43	117.5	0.86	124.2	3.06	106.8	1.03
	(31.3)	2.95	(33.4)	0.61	(32.8)	1.77	(34.1)	0.68
Female students	122.4	6.49	127.1	1.29	131.6	3.36	120.5	1.68
	(29.8)	3.49	(35.2)	1.11	(35.7)	2.55	(35.6)	0.85
Gender gap (points)	6.3	7.34**	9.6**	1.55**	7.4	4.54**	13.7**	1.97**
Effect size (*d*)	0.21		0.28		0.22		0.39	
U_3	.58		.61		.59		.65	

(CONTINUED)

TABLE E5 (CONTINUED)

Grades 4 and 8 National Assessment of Educational Progress average scaled writing scores by gender and student poverty and disability status in Vermont and the country, 2002

	Grade 4				Grade 8			
	Vermont		National		Vermont		National	
Students	Mean	Standard error	Mean	Standard error	Mean	Standard error	Mean	Standard error
Students without disabilities								
Male students	152.4	1.59	147.6	0.66	156.8	1.99	146.7	0.72
	(33.1)	1.13	(33.6)	0.62	(35.0)	1.54	(34.1)	0.36
Female students	172.4	1.70	164.1	0.42	180.1	1.47	165.5	0.62
	(32.6)	0.96	(34.0)	0.32	(33.2)	1.05	(33.7)	0.31
Gender gap (points)	20.0**	2.33**	16.5**	0.78**	23.3**	2.47**	18.8**	0.95**
Effect size (*d*)	0.61		0.49		0.68		0.55	
U_3	.73		.69		.75		.71	

**Gender gap is statistically significant (different from zero) at $p < 0.05$.

Note: Numbers in parentheses are standard deviations. Gender gaps are defined as the mean scale score of girls minus the mean scale score of boys ($F - M$). Standard errors for the gender gap estimates were calculated as $SE_{F-M} = \sqrt{(SE_F^2 + SE_M^2)}$. Effect size is calculated as $|d| = (F - M) / SD_{pooled}$, where $SD_{pooled} = \sqrt{[(SD_F^2 + SD_M^2) / 2]}$. U_3 is the proportion of scores in the group with the lower mean that fall below the noted effect size in a standard normal distribution. See appendix A for details.

Source: Authors' analysis based on 2002 NAEP writing data for grades 4 and 8 from U.S. Department of Education, National Center for Education Statistics (2008).

TABLE E6

Differences in Vermont and U.S. National Assessment of Educational Progress gender gaps in grades 4 and 8 writing scores by student poverty and disability status, 2002

	Grade 4			Grade 8		
Students	Vermont	National	Difference in gaps	Vermont	National	Difference in gaps
Total						
Gender gap (points)	21.3	17.5	3.8	24.2	20.9	3.3
Standard error	2.39	0.72	2.49	2.28	0.93	2.46
Students in poverty						
Gender gap (points)	24.1	16.5	7.6**	26.1	19.4	6.7
Standard error	3.57	1.20	3.77	4.50	0.94	4.59
Students not in poverty						
Gender gap (points)	22.6	18.9	3.7	24.1	22.2	1.9
Standard error	2.93	0.76	3.03	2.47	1.08	2.70
Students with disabilities						
Gender gap	6.3	9.6	–3.3	7.4	13.7	–6.3
Standard error	7.34	1.55	7.50	4.54	1.97	4.95
Students without disabilities						
Gender gap (points)	20.0	16.5	3.5	23.3	18.8	4.5
Standard error	2.33	0.78	2.46	2.47	0.95	2.65

**The difference in the Vermont and national gender gap estimates is statistically significant (different from zero) at $p < 0.05$.

Note: Standard errors for the difference in Vermont and national gender gap estimates were calculated as $SE_{VT\,gap\,-\,U.S.\,gap} = \sqrt{(SE_{VT\,gap}{}^2 + SE_{U.S.\,gap}{}^2)}$. Gender gaps are defined as the mean scale score of girls minus the mean scale score of boys ($F - M$). See appendix A for details.

Source: Authors' analysis based on 2002 NAEP writing data for grades 4 and 8 from U.S. Department of Education, National Center for Education Statistics (2008).

TABLE E7

Grade 4 National Assessment of Educational Progress average scaled math scores by gender and student poverty and disability status in Vermont and the country, 2000, 2003, 2005, and 2007

	2002				2003				2005				2007			
	Vermont		National		Vermont		National		Vermont		National		Vermont		National	
Students	Mean	Standard error	Mean	Standard error	Mean	Standard error	Mean	Standard error	Mean	Standard error	Mean	Standard error	Mean	Standard error	Mean	Standard error
Total																
Male students	232.4	1.73	225.4	1.07	243.7	0.90	235.3	0.27	245.9	0.73	238.3	0.18	247.7	0.67	240.2	0.19
	(29.7)	1.24	(32.5)	0.53	(26.8)	0.75	(29.2)	0.15	(26.2)	0.69	(29.1)	0.11	(27.2)	0.63	(29.6)	0.13
Female students	230.7	1.87	222.9	1.06	240.2	1.02	232.6	0.24	240.8	0.76	235.9	0.18	245.0	0.72	237.9	0.21
	(29.7)	1.48	(30.2)	0.53	(26.2)	0.70	(27.7)	0.14	(26.4)	0.63	(28.0)	0.11	(26.0)	0.82	(28.0)	0.11
Gender gap (points)	–1.7	2.55**	–2.5	1.51**	–3.5**	1.36**	–2.7**	0.36**	–5.1**	1.05**	–2.4**	0.25**	–2.7**	0.98**	–2.3**	0.28**
Effect size (*d*)	–0.06		–0.08		–0.13		–0.09		–0.19		–0.08		–0.10		–0.08	
U_3	.48		.47		.45		.46		.42		.47		.46		.47	
Students in poverty																
Male students	217.7	2.32	208.6	1.13	231.5	1.42	222.6	0.33	232.6	1.57	226.1	0.22	235.9	1.43	227.7	0.25
	(29.90)	2.21	(29.42)	0.67	(26.15)	1.17	(27.35)	0.17	(25.71)	1.23	(27.58)	0.14	(27.14)	1.09	(28.37)	0.16
Female students	214.3	3.08	207.9	1.14	226.2	1.68	220.4	0.31	227.9	1.40	224.4	0.24	232.9	1.17	226.4	0.25
	(31.2)	2.68	(28.1)	0.71	(25.3)	1.13	(25.8)	0.18	(26.1)	1.03	(26.4)	0.17	(26.7)	1.40	(26.9)	0.16
Gender gap (points)	–3.4	3.86**	–0.7	1.61**	–5.3**	2.20**	–2.2**	0.45**	–4.7**	2.10**	–1.7**	0.33**	–3.0**	1.85**	–1.3**	0.35**
Effect size (*d*)	–0.11		–0.02		–0.21		–0.08		–0.18		–0.06		–0.11		–0.05	
U_3	.46		.49		.42		.47		.43		.48		.46		.48	
Students not in poverty																
Male students	238.5	2.19	236.6	1.38	248.7	0.99	245.8	0.31	251.5	0.84	248.8	0.22	253.0	0.82	250.5	0.27
	(27.28)	1.36	(29.24)	0.84	(25.54)	0.82	(26.30)	0.19	(24.47)	0.79	(26.02)	0.14	(25.52)	0.80	(26.51)	0.16
Female students	237.3	2.04	232.4	1.29	246.4	0.96	242.8	0.30	247.2	0.90	246.2	0.21	250.5	0.91	247.9	0.23
	(26.4)	1.02	(27.3)	0.67	(24.0)	0.71	(25.0)	0.19	(24.0)	0.81	(25.2)	0.16	(23.7)	0.70	(25.0)	0.14
Gender gap (points)	–1.2	2.99**	–4.2**	1.89**	–2.3	1.38**	–3.0**	0.43**	–4.3**	1.23**	–2.6**	0.30**	–2.5	1.22**	–2.6**	0.35**
Effect size (*d*)	–0.04		–0.15		–0.09		–0.12		–0.18		–0.10		–0.10			–0.10
U_3	.48		.44		.46		.45		.43		.46		.46		.46	

(CONTINUED)

TABLE E7 (CONTINUED)

Grade 4 National Assessment of Educational Progress average scaled math scores by gender and student poverty and disability status in Vermont and the country, 2000, 2003, 2005, and 2007

	2002				2003				2005				2007			
	Vermont		National		Vermont		National		Vermont		National		Vermont		National	
Students	Mean	Standard error	Mean	Standard error	Mean	Standard error	Mean	Standard error	Mean	Standard error	Mean	Standard error	Mean	Standard error	Mean	Standard error
Students with disabilities																
Male students	221.5	4.87	201.4	2.45	225.4	1.84	216.4	0.50	227.2	1.87	220.8	0.51	223.6	1.79	222.4	0.49
	(31.2)	3.13	(33.0)	1.90	(25.7)	1.40	(30.3)	0.33	(26.3)	1.66	(30.5)	0.32	(28.5)	1.38	(32.8)	0.34
Female students	—		190.3	3.44	213.9	3.00	209.8	0.57	217.2	3.06	213.9	0.55	215.3	2.52	216.1	0.50
	—		(28.9)	2.76	(26.0)	1.52	(28.4)	0.38	(28.7)	2.40	(29.8)	0.48	(28.6)	2.10	(32.2)	0.39
Gender gap (points)	—		–11.1**	4.22**	–11.5**	3.52**	–6.6**	0.76**	–10.0**	3.59**	–6.9**	0.75**	–8.3**	3.09	–6.3	.70
Effect size (*d*)	—		–0.36		–0.45		–0.22		–0.36		–0.23		–0.29		–0.19	
U_3			.36		.33		.41		.36		.41		.39		.42	
Students without disabilities																
Male students	234.7	2.04	228.9	1.02	247.9	0.93	238.6	0.27	249.9	0.83	241.4	0.19	253.1	0.77	243.2	0.20
		(28.9)	1.34	(30.9)	0.56	(25.3)	0.68	(27.7)	0.16	(24.4)	0.74	(27.7)	0.12	(23.8)	0.70	(28.0)
Female students	232.9	1.80	225.2	1.05	243.0	0.97	234.5	0.25	243.0	0.77	237.8	0.19	248.6	0.76	239.8	0.22
		(27.8)	1.06	(28.9)	0.58	(24.6)	0.66	(26.8)	0.14	(25.1)	0.68	(27.0)	0.12	(23.2)	0.65	(26.8)
Gender gap (points)	–1.8	2.72**	–3.7**	1.46**	–4.9**	1.34**	–4.1**	0.37**	–6.9**	1.13**	–3.6**	0.27**	–4.5**	1.08**	–3.4**	0.30**
Effect size (*d*)	–0.06		–0.12		–0.20		–0.15		–0.28		–0.13		–0.19		–0.12	
U_3	.48		.45		.42		.44		.39		.45		.42		.45	

**The difference in the Vermont and national gender gap estimates is statistically significant (different from zero) at $p < 0.05$.

— indicates that reporting standards for the NAEP were not met.

Note: Numbers in parentheses are standard deviations. Gender gaps are defined as the mean scale score of girls minus the mean scale score of boys ($F - M$). Standard errors for the gender gap estimates were calculated as $SE_{F-M} = \sqrt{(SE_F^2 + SE_M^2)}$. Effect size is calculated as $|d| = (F - M) / SD_{pooled}$, where $SD_{pooled} = \sqrt{[(SD_F^2 + SD_M^2) / 2]}$. U_3 is the proportion of scores in the group with the lower mean that fall below the noted effect size in a standard normal distribution. See appendix A for details.

Source: Authors' analysis based on NAEP math data for grade 4 in 2000, 2003, 2005, and 2007 from U.S. Department of Education, National Center for Education Statistics (2008).

TABLE E8

Differences in Vermont and U.S. National Assessment of Educational Progress gender gaps in grade 4 math scores by student poverty and disability status, 2000, 2003, 2005, and 2007

	2000			2003			2005			2007		
Students	Vermont	National	Difference in gaps	Vermont	National	Difference in gaps	Vermont	National	Difference in gaps	Vermont	National	Difference in gaps
Total												
Gender gap (points)	–1.7	–2.5	0.8	–3.5	–2.7	–0.8	–5.1	–2.4	–2.7**	–2.7	–2.3	–0.4
Standard error	2.55	1.51	2.96	1.36	0.36	1.41	1.05	0.25	1.08	0.98	0.28	1.02
Students in poverty												
Gender gap (points)	–3.4	–0.7	–2.7	–5.3	–2.2	–3.1	–4.7	–1.7	–3.0	–3.0	–1.3	–1.7
Standard error	3.86	1.61	4.18	2.20	0.45	2.25	2.10	0.33	2.13	1.85	0.35	1.88
Students not in poverty												
Gender gap (points)	–1.2	–4.2	3.0	–2.3	–3.0	0.7	–4.3	–2.6	–1.7	–2.5	–2.6	0.1
Standard error	2.99	1.89	3.54	1.38	0.43	1.44	1.23	0.30	1.27	1.22	0.35	1.28
Students with disabilities												
Gender gap (points)	—	–11.1	—	–11.5	–6.6	–4.9	–10.0	–6.9	–3.1	–8.3	–6.3	–2.0
Standard error	—	4.22	—	3.52	0.76	3.60	3.59	0.75	3.66	3.09	0.70	3.17
Students without disabilities												
Gender gap (points)	–1.8	–3.7	1.9	–4.9	–4.1	–0.8	–6.9	–3.6	–3.3**	–4.5	–3.4	–1.1
Standard error	2.72	1.46	3.09	1.34	0.37	1.39	1.13	0.27	1.16	1.08	0.30	1.12

**The difference in the Vermont and national gender gap estimates is statistically significant (different from zero) at $p < 0.05$.

— indicates that reporting standards for the NAEP were not met.

Note: Gender gaps are defined as the mean scale score of girls minus the mean scale score of boys ($F - M$). Standard errors for the gender gap estimates were calculated as $SE_{F-M} = \sqrt{(SE_F^2 + SE_M^2)}$. See appendix A for details.

Source: Authors' analysis based on NAEP math data for grade 4 in 2000, 2003, 2005, and 2007 from U.S. Department of Education, National Center for Education Statistics (2008).

TABLE E9

Grade 8 National Assessment of Educational Progress average scaled math scores by gender and student poverty and disability status in Vermont and the country, 2000, 2003, 2005, and 2007

	2000				2003				2005				2007			
	Vermont		National		Vermont		National		Vermont		National		Vermont		National	
Students	Mean	Standard error	Mean	Standard error	Mean	Standard error	Mean	Standard error	Mean	Standard error	Mean	Standard error	Mean	Standard error	Mean	Standard error
Total																
Male students	279.2	2.08	272.6	1.00	285.7	1.12	277.0	0.33	287.3	1.03	278.2	0.24	291.9	1.04	281.1	0.26
	(37.4)	1.83	(39.6)	0.69	(33.3)	0.77	(37.2)	0.19	(35.6)	0.99	(37.2)	0.18	(35.2)	0.92	(37.3)	0.15
Female students	281.9	1.41	271.0	1.04	285.6	0.83	275.2	0.32	287.4	1.04	276.8	0.22	290.1	1.06	279.2	0.32
	(33.6)	1.19	(37.1)	0.64	(32.0)	0.76	(35.2)	0.15	(33.7)	0.76	(35.4)	0.16	(31.8)	0.76	(34.8)	0.19
Gender gap (points)	2.7	2.51**	–1.6	1.44**	–0.1	1.39**	–1.8	0.46**	0.1	1.46**	–1.4**	0.33**	–1.8	1.48**	–1.9**	0.41**
Effect size (*d*)	0.08		–0.04		0.00		–0.05		0.00		–0.04		–0.05		–0.05	
U_3	.53		.48		.50		.48		.50		.48		.48		.48	
Students in poverty																
Male students	261.4	3.92	252.7	1.27	270.0	2.11	259.2	0.43	272.1	2.16	261.9	0.29	280.5	2.13	265.5	0.36
	(37.7)	2.94	(37.2)	0.77	(32.1)	1.23	(35.5)	0.25	(35.7)	2.07	(35.0)	0.25	(32.6)	1.82	(35.1)	0.19
Female students	261.3	3.15	253.0	1.31	266.8	1.95	257.6	0.44	272.0	1.86	261.0	0.28	274.6	1.79	264.3	0.43
	(32.4)	2.29	(35.1)	1.05	(31.2)	1.33	(33.0)	0.27	(31.5)	1.46	(33.3)	0.24	(29.2)	1.34	(32.6)	0.29
Gender gap (points)	–0.1	5.03**	0.3	1.82**	–3.2	2.87**	–1.6**	0.62**	–0.1	2.85**	–0.9**	0.40**	–5.9**	2.78**	–1.2**	0.56**
Effect size (*d*)	0.00		0.01		–0.10		–0.05		0.00		–0.03		–0.19		–0.04	
U_3	.50		.50		.46		.48		.50		.49		.42		.48	
Students not in poverty																
Male students	284.7	1.91	283.4	1.25	291.1	1.23	287.4	0.36	293.4	1.24	288.9	0.26	295.9	1.35	291.9	0.36
	(35.4)	2.10	(36.9)	0.83	(31.9)	0.83	(34.0)	0.26	(33.7)	1.02	(34.6)	0.21	(35.2)	1.02	(34.9)	0.23
Female students	288.1	1.40	282.6	1.22	291.9	0.82	286.2	0.33	292.9	1.21	287.4	0.27	296.4	1.17	289.8	0.32
	(31.5)	1.18	(33.6)	0.66	(29.8)	0.74	(31.9)	0.22	(32.8)	0.83	(32.7)	0.18	(30.7)	1.00	(32.1)	0.20
Gender gap (points)	3.4	2.37**	–0.8	1.75**	0.8	1.48**	–1.2**	0.49**	–0.5	1.73**	–1.5**	0.37**	0.5	1.79**	–2.1**	0.48**
Effect size (*d*)	0.10		–0.02		0.03		–0.04		–0.02		–0.04		0.02		–0.06	
U_3	.54		.49		.51		.48		.49		.48		.51		.48	

(CONTINUED)

TABLE E9 (CONTINUED)

Grade 8 National Assessment of Educational Progress average scaled math scores by gender and student poverty and disability status in Vermont and the country, 2000, 2003, 2005, and 2007

	2000				2003				2005				2007			
	Vermont		National		Vermont		National		Vermont		National		Vermont		National	
Students	Mean	Standard error	Mean	Standard error	Mean	Standard error	Mean	Standard error	Mean	Standard error	Mean	Standard error	Mean	Standard error	Mean	Standard error
Students with disabilities																
Male students	245.9	6.20	232.6	2.74	262.5	2.46	244.5	0.69	260.0	2.35	246.9	0.57	265.0	2.89	248.5	0.70
		(37.7)	3.17	(39.4)	1.59	(32.4)	1.41	(36.6)	0.43	(35.2)	1.80	(37.4)	0.41	(36.2)	2.23	(38.1)
Female students	—		(38.7)	1.90	(30.4)	2.06	(35.2)	0.56	(34.0)	3.00	(35.4)	0.58	(29.1)	2.88	(36.7)	0.63
Gender gap (points)	—		–11.1**	4.04**	–14.5**	3.71**	–7.9**	1.16**	–10.0**	4.15**	–8.4**	0.89**	–9.9**	4.35**	–7.8**	1.26**
Effect size (*d*)	—		–0.28		–0.46		–0.22		–0.29		–0.23		–0.30		–0.21	
U_3			.39		.32		.41		.39		.41		.38		.42	
Students without disabilities																
Male students	286.5	1.85	277.0	0.99	291.5	1.16	282.6	0.30	294.1	1.15	283.2	0.23	298.8	1.06	285.6	0.26
	(33.1)	2.16	(37.1)	0.58	(30.8)	0.83	(34.3)	0.19	(32.3)	0.98	(34.7)	0.18	(31.5)	0.97	(34.9)	0.15
Female students	285.7	1.35	273.7	1.03	289.8	0.85	278.4	0.31	291.4	1.01	279.8	0.23	294.4	0.97	281.7	0.30
	(30.6)	1.05	(35.0)	0.58	(29.3)	0.73	(33.3)	0.16	(31.1)	0.74	(33.6)	0.15	(29.4)	0.75	(33.0)	0.17
Gender gap (points)	–0.8	2.29**	–3.3**	1.43**	–1.7	1.44**	–4.2**	0.43**	–2.7	1.53**	–3.4**	0.33**	–4.4**	1.44**	–3.9**	0.40**
Effect size (*d*)	–0.03		–0.09		–0.06		–0.12		–0.09		–0.10		–0.14		–0.11	
U_3	.49		.46		.48		.45		.46		.46		.44		.46	

**The difference in the Vermont and national gender gap estimates is statistically significant (different from zero) at $p < 0.05$.

— indicates that reporting standards for the NAEP were not met.

Note: Numbers in parentheses are standard deviations. Gender gaps are defined as the mean scale score of girls minus the mean scale score of boys ($F - M$). Standard errors for the gender gap estimates were calculated as $SE_{F-M} = \sqrt{(SE_F^2 + SE_M^2)}$. Effect size is calculated as $|d| = (F - M) / SD_{pooled}$, where $SD_{pooled} = \sqrt{[(SD_F^2 + SD_M^2 / 2]}$. U_3 is the proportion of scores in the group with the lower mean that fall below the noted effect size in a standard normal distribution. See appendix A for details.

Source: Authors' analysis based on NAEP math data for grade 8 in 2000, 2003, 2005, and 2007 from U.S. Department of Education, National Center for Education Statistics (2008).

TABLE E10

Differences in Vermont and U.S. National Assessment of Educational Progress gender gaps in grade 8 math scores by student poverty and disability status, 2000, 2003, 2005, and 2007

	2000			2003			2005			2007		
Students	Vermont	National	Difference in gaps	Vermont	National	Difference in gaps	Vermont	National	Difference in gaps	Vermont	National	Difference in gaps
Total												
Gender gap (points)	2.7	–1.6	4.3	–0.1	–1.8	1.7	0.1	–1.4	1.5	–1.8	–1.9	0.1
Standard error	2.51	1.44	2.90	1.39	0.46	1.47	1.46	0.33	1.50	1.48	0.41	1.54
Students in poverty												
Gender gap (points)	–0.1	0.3	–0.4	–3.2	–1.6	–1.6	–0.1	–0.9	0.8	–5.9	–1.2	–4.7
Standard error	5.03	1.82	5.35	2.87	0.62	2.94	2.85	0.40	2.88	2.78	0.56	2.84
Students not in poverty												
Gender gap (points)	3.4	–0.8	4.2	0.8	–1.2	2.0	–0.5	–1.5	1.0	0.5	–2.1	2.6
Standard error	2.37	1.75	2.94	1.48	0.49	1.56	1.73	0.37	1.77	1.79	0.48	1.85
Students with disabilities												
Gender gap (points)	—	–11.1	—	–14.5	–7.9	–6.6	–10.0	–8.4	–1.6	–9.9	–7.8	–2.1
Standard error	—	4.04	—	3.71	1.16	3.89	4.15	0.89	4.24	4.35	1.26	4.53
Students without disabilities												
Gender gap (points)	–0.8	–3.3	2.5	–1.7	–4.2	2.5	–2.7	–3.4	0.7	–4.4	–3.9	–0.5
Standard error	2.29	1.43	2.70	1.44	0.43	1.50	1.53	0.33	1.56	1.44	0.40	1.49

— indicates that reporting standards for the NAEP were not met.

Note: Gender gaps are defined as the mean scale score of girls minus the mean scale score of boys ($F - M$). Standard errors for the gender gap estimates were calculated as $SE_{F-M} = \sqrt{(SE_F^2 + SE_M^2)}$. See appendix A for details.

Source: Authors' analysis based on NAEP 2000, 2003, 2005, and 2007 math data for grade 8 from U.S. Department of Education, National Center for Education Statistics (2008).

TABLE E11

Grade 4 National Assessment of Educational Progress average scaled reading scores and student poverty and disability gaps in Vermont and the country, 2002, 2003, 2005 and 2007

Student status	2002				2003				2005				2007			
	Vermont		National		Vermont		National		Vermont		National		Vermont		National	
	Mean	Standard error	Mean	Standard error	Mean	Standard error	Mean	Standard error	Mean	Standard error	Mean	Standard error	Mean	Standard error	Mean	Standard error
Poverty status																
Students in poverty	212.7	2.1	202.4	0.7	213.6	1.2	201.1	0.4	210.4	1.7	202.7	0.3	212.5	1.5	205.0	0.3
	(31.6)	1.4	(35.5)	0.4	(30.9)	0.9	(36.1)	0.2	(33.5)	1.2	(34.9)	0.2	(32.8)	0.9	(35.0)	0.2
Students not in poverty	232.8	1.1	229.2	0.4	231.2	0.9	229.0	0.3	233.7	0.9	229.7	0.2	235.3	0.9	231.7	0.2
	(31.1)	1.0	(31.7)	0.3	(30.0)	0.6	(33.2)	0.2	(31.0)	0.9	(32.3)	0.1	(32.1)	0.8	(31.9)	0.2
Poverty gap	20.1**	2.4**	26.8**	0.8**	17.6**	1.5**	27.9**	0.5**	23.3**	1.9**	27.0**	0.3**	22.8**	1.7**	26.7**	0.4**
Effect size (*d*)	0.64		0.80		0.58		0.81		0.72		0.80		0.70		0.80	
U_3	.74		.79		.72		.79		.76		.79		.76		.79	
Disability status																
Students with disabilities	198.5	3.1	186.6	0.8	202.5	2.6	184.4	0.6	194.1	2.7	189.6	0.5	193.9	2.3	190.2	0.6
	(33.3)	2.4	(38.2)	0.5	(32.3)	1.6	(40.2)	0.4	(33.1)	2.0	(39.2)	0.4	(37.9)	1.7	(42.4)	0.5
Students without disabilities	229.8	1.1	219.5	0.5	229.1	1.0	219.8	0.3	230.9	0.9	220.2	0.2	233.5	0.8	222.8	0.3
	(30.9)	0.9	(34.7)	0.4	(29.8)	0.5	(35.1)	0.2	(31.2)	0.8	(34.6)	0.1	(30.1)	0.7	(33.6)	0.1
Disability gap	31.3**	3.3**	32.9**	1.0**	26.6**	2.8**	35.4**	0.7**	36.8**	2.9**	30.6**	0.6**	39.6**	2.4**	32.6**	0.6**
Effect size (*d*)	0.97		0.90		0.86		0.94		1.14		0.83		1.16		0.85	
U_3	.83		.82		.81		.83		.87		.80		.88		.80	

**The difference in average scores between students not in poverty and students in poverty, or between students without disabilities and students with disabilities, is statistically significant (different from zero) at $p < 0.05$.

Note: Numbers in parentheses are standard deviations. Standard errors for the poverty gap (students not in poverty – students in poverty, or *NP – P*) were calculated as $SE_{NP-P} = \sqrt{(SE_{NP}^2 + SE_P^2)}$. Standard errors for the disability gap (students without disabilities – students with disabilities, or *ND – D*) were calculated as $SE_{ND-D} = \sqrt{(SE_{ND}^2 + SE_D^2)}$. Effect size is calculated as $|d| = (M_1 - M_2) / SD_{pooled}$, where $SD_{pooled} = \sqrt{[(SD_1^2 + SD_2^2) / 2]}$. U_3 is the proportion of scores in the group with the lower mean that fall below the noted effect size in a standard normal distribution. See appendix A for details.

Source: Authors' analysis based on 2002, 2003, 2005, and 2007 NAEP reading data for grade 4 from U.S. Department of Education, National Center for Education Statistics (2008).

TABLE E12

Grade 8 National Assessment of Educational Progress average scaled reading scores and student poverty and disability gaps by student poverty and disability status in Vermont and the country, 2002, 2003, 2005, and 2007

Student status	2002				2003				2005				2007			
	Vermont		National		Vermont		National		Vermont		National		Vermont		National	
	Mean	Standard error	Mean	Standard error	Mean	Standard error	Mean	Standard error	Mean	Standard error	Mean	Standard error	Mean	Standard error	Mean	Standard error
Poverty status																
Students in poverty	256.7	1.8	249.0	0.5	255.4	1.6	245.9	0.4	255.0	1.4	246.5	0.3	260.3	1.4	247.2	0.3
	(32.6)	1.3	(33.6)	0.3	(30.0)	1.0	(35.0)	0.3	(32.6)	0.9	(34.2)	0.1	(30.6)	1.2	(34.1)	0.2
Students not in poverty	276.1	0.9	271.3	0.5	275.7	0.9	270.7	0.3	274.4	0.8	269.7	0.2	277.6	0.9	270.7	0.3
	(29.5)	0.5	(30.6)	0.3	(31.0)	0.8	(31.9)	0.2	(31.1)	0.7	(32.5)	0.1	(29.0)	0.6	(31.6)	0.2
Poverty gap	19.4**	2.1**	22.3**	0.7**	20.3**	1.8**	24.8**	0.5**	19.4**	1.6**	23.2**	0.3**	17.3**	1.7**	23.5**	0.4**
Effect size (*d*)	0.62		0.69		0.67		0.74		0.61		0.70		0.58		0.72	
U_3	.73		.75		.75		.77		.73		.76		.72		.76	
Disability status																
Students with disabilities	246.8	2.5	227.5	1.0	245.4	2.0	224.4	0.6	235.8	1.7	226.3	0.5	248.2	2.4	226.2	0.5
	(31.3)	1.3	(36.3)	0.7	(30.1)	1.2	(36.8)	0.5	(30.4)	1.4	(36.4)	0.3	(33.4)	1.7	(38.3)	0.5
Students without disabilities	275.7	0.8	266.2	0.5	274.5	0.8	265.5	0.2	274.7	0.7	264.0	0.2	277.5	0.8	264.5	0.2
	(29.4)	0.5	(31.5)	0.3	(30.5)	0.7	(32.5)	0.2	(29.6)	0.6	(33.0)	0.1	(27.5)	0.6	(32.3)	0.2
Disability gap	28.9**	2.6**	38.7**	1.1**	29.1**	2.2**	41.1**	0.7**	38.9**	1.9**	37.7**	0.5**	29.3**	2.6**	38.3**	0.6**
Effect size (*d*)	0.95		1.14		0.96		1.18		1.30		1.08		0.96		1.08	
U_3	.83		.87		.83		.88		.90		.86		.86		.86	

**The difference in average scores between students not in poverty and students in poverty, or between students without disabilities and students with disabilities, is statistically significant (different from zero) at $p < 0.05$.

Note: Numbers in parentheses are standard deviations. Standard errors for the poverty gap (students not in poverty – students in poverty, or *NP – P*) were calculated as $SE_{NP-P} = \sqrt{(SE_{NP}^2 + SE_P^2)}$. Standard errors for the disability gap (students without disabilities – students with disabilities, or *ND – D*) were calculated as $SE_{ND-D} = \sqrt{(SE_{ND}^2 + SE_D^2)}$. Effect size is calculated as $|d| = (M_1 - M_2) / SD_{pooled}$, where $SD_{pooled} = \sqrt{[(SD_1^2 + SD_2^2) / 2]}$. U_3 is the proportion of scores in the group with the lower mean that fall below the noted effect size in a standard normal distribution. See appendix A for details.

Source: Authors' analysis based on 2002, 2003, 2005, and 2007 NAEP reading data for grade 8 from U.S. Department of Education, National Center for Education Statistics (2008).

TABLE E13

Grades 4 and 8 National Assessment of Educational Progress average scaled writing scores and poverty and disability gaps in Vermont and the country, 2002

	Grade 4				Grade 8			
	Vermont		National		Vermont		National	
Student status	Mean	Standard error	Mean	Standard error	Mean	Standard error	Mean	Standard error
Poverty status								
Students in poverty	142.6	1.9	140.5	0.8	143.7	2.5	135.9	0.5
	(34.3)	1.3	(34.6)	0.6	(39.7)	1.4	(36.2)	0.3
Students not in poverty	163.2	1.7	163.0	0.5	168.2	1.3	161.2	0.7
	(35.6)	1.1	(34.3)	0.3	(36.5)	0.9	(35.2)	0.3
Poverty gap	20.6**	2.6**	22.5**	0.9**	24.5**	2.8**	25.3**	0.9**
Effect size (*d*)	0.59		0.65		0.64		0.71	
U_3	.72		.74		.74		.76	
Disability status								
Students with disabilities	118.3	3.3	120.9	0.7	126.8	2.3	111.5	1.1
	(31.0)	2.2	(34.4)	0.6	(34.0)	1.3	(35.2)	0.6
Students without disabilities	162.8	1.4	155.9	0.5	168.5	1.3	156.3	0.6
	(34.3)	0.8	(34.8)	0.4	(36.0)	1.0	(35.1)	0.3
Disability gap	44.5**	3.6**	35.0**	0.9**	41.7**	2.6**	44.8**	1.3**
Effect size (*d*)	1.36		1.01		1.19		1.27	
U_3	.91		.84		.88		.90	

**The difference in average scores between students not in poverty and students in poverty, or between students without disabilities and students with disabilities, is statistically significant (different from zero) at $p < 0.05$.

Note: Numbers in parentheses are standard deviations. Standard errors for the poverty gap (students not in poverty – students in poverty, or *NP* – *P*) were calculated as $SE_{NP-P} = \sqrt{(SE_{NP}^2 + SE_P^2)}$. Standard errors for the disability gap (students without disabilities – students with disabilities, or *ND* – *D*) were calculated as $SE_{ND-D} = \sqrt{(SE_{ND}^2 + SE_D^2)}$. Effect size is calculated as $|d| = (M_1 - M_2) / SD_{pooled}$, where $SD_{pooled} = \sqrt{[(SD_1^2 + SD_2^2) / 2]}$. U_3 is the proportion of scores in the group with the lower mean that fall below the noted effect size in a standard normal distribution. See appendix A for details.

Source: Authors' analysis based on 2002 NAEP writing data for grades 4 and 8 from U.S. Department of Education, National Center for Education Statistics (2008).

TABLE E14

Grade 4 National Assessment of Educational Progress average scaled math scores and poverty and disability gaps in Vermont and the country, 2000, 2003, 2005, and 2007

Student status	2000				2003				2005				2007			
	Vermont		National		Vermont		National		Vermont		National		Vermont		National	
	Mean	Standard error	Mean	Standard error	Mean	Standard error	Mean	Standard error	Mean	Standard error	Mean	Standard error	Mean	Standard error	Mean	Standard error
Poverty status																
Students in poverty	216.0	2.2	208.2	0.9	228.7	1.3	221.5	0.3	230.3	1.1	225.2	0.2	234.4	1.0	227.1	0.2
	(30.6)	2.1	(28.8)	0.6	(25.9)	0.9	(26.6)	0.1	(26.0)	0.8	(27.0)	0.1	(27.0)	0.9	(27.6)	0.1
Students not in poverty	237.9	1.8	234.6	1.2	247.6	0.8	244.3	0.3	249.6	0.6	247.5	0.2	251.8	0.6	249.2	0.2
	(26.9)	1.0	(28.4)	0.6	(24.8)	0.6	(25.7)	0.2	(24.3)	0.6	(25.6)	0.1	(24.7)	0.6	(25.8)	0.1
Poverty gap	21.9**	2.8**	26.4**	1.5**	18.9**	1.5**	22.8**	0.4**	19.3**	1.2**	22.3**	0.3**	17.4**	1.2**	22.1**	0.3**
Effect size (*d*)	0.76		0.92		0.75		0.87		0.77		0.85		0.67		0.83	
U_3	.78		.82		.77		.81		.78		.80		.75		.80	
Disability status																
Students with disabilities	216.7	4.3	197.7	2.2	221.2	1.8	214.2	0.4	224.1	1.6	218.5	0.4	220.6	1.4	220.3	0.4
	(34.3)	3.9	(32.1)	1.6	(26.5)	1.1	(29.9)	0.2	(27.5)	1.3	(30.4)	0.3	(28.8)	1.2	(31.7)	0.3
Students without disabilities	233.7	1.7	227.0	1.0	245.3	0.8	236.5	0.2	246.5	0.6	239.6	0.2	250.8	0.6	241.5	0.2
	(28.3)	1.0	(30.0)	0.5	(25.0)	0.5	(27.3)	0.1	(25.0)	0.5	(27.4)	0.1	(23.6)	0.5	(27.4)	0.1
Disability gap	17.0**	4.7**	29.3**	2.4**	24.1**	2.0**	22.3**	0.5**	22.4**	1.7**	21.1**	0.4**	30.2**	1.5**	21.2**	0.4**
Effect size (*d*)	0.54		0.94		0.94		0.78		0.85		0.73		1.15		0.71	
U_3	.71		.83		.83		.78		.80		.77		.87		.76	

**The difference in average scores between students not in poverty and students in poverty, or between students without disabilities and students with disabilities, is statistically significant (different from zero) at $p < 0.05$.

Note: Numbers in parentheses are standard deviations. Standard errors for the poverty gap (students not in poverty – students in poverty, or *NP* – *P*) were calculated as $SE_{NP-P} = \sqrt{(SE_{NP}^2 + SE_P^2)}$. Standard errors for the disability gap (students without disabilities – students with disabilities, or *ND* – *D*) were calculated as $SE_{ND-D} = \sqrt{(SE_{ND}^2 + SE_D^2)}$. Effect size is calculated as $|d| = (M_1 - M_2) / SD_{pooled}$, where $SD_{pooled} = \sqrt{[(SD_1^2 + SD_2^2) / 2]}$. U_3 is the proportion of scores in the group with the lower mean that fall below the noted effect size in a standard normal distribution. See appendix A for details.

Source: Authors' analysis based on 2000, 2003, 2005, and 2007 NAEP math data for grade 4 from U.S. Department of Education, National Center for Education Statistics (2008).

TABLE E15

Grade 8 National Assessment of Educational Progress average scaled math scores and poverty and disability gaps in Vermont and the country, 2000, 2003, 2005, and 2007

Student status	2000				2003				2005				2007			
	Vermont		National		Vermont		National		Vermont		National		Vermont		National	
	Mean	Standard error	Mean	Standard error	Mean	Standard error	Mean	Standard error	Mean	Standard error	Mean	Standard error	Mean	Standard error	Mean	Standard error
Poverty status																
Students in poverty	261.4	2.9	252.9	1.2	268.4	1.4	258.4	0.3	272.1	1.4	261.5	0.2	277.4	1.5	264.9	0.3
	(35.4)	1.7	(36.2)	0.7	(31.7)	0.9	(34.3)	0.2	(33.8)	1.3	(34.2)	0.2	(31.0)	1.2	(33.9)	0.2
Students not in poverty	286.4	1.4	283.0	1.1	291.5	0.8	286.8	0.3	293.2	0.9	288.2	0.2	296.1	0.9	290.9	0.3
	(33.6)	1.4	(35.3)	0.6	(30.8)	0.6	(33.0)	0.2	(33.3)	0.7	(33.7)	0.2	(33.1)	0.7	(33.5)	0.2
Poverty gap	25.0**	3.3**	30.1**	1.6**	23.1**	1.6**	28.4**	0.5**	21.1**	1.7**	26.7**	0.3**	18.7**	1.7**	26.0**	0.4**
Effect size (*d*)	0.72		0.84		0.74		0.84		0.63		0.79		0.58		0.77	
U_3	.76		.80		.77		.80		.74		.78		.72		.78	
Disability status																
Students with disabilities	245.2	5.8	228.9	2.3	257.7	1.9	241.8	0.6	256.7	1.8	244.0	0.5	261.5	2.2	245.9	0.7
	(37.8)	3.0	(39.6)	1.3	(32.5)	1.1	(36.3)	0.3	(35.1)	1.4	(36.9)	0.3	(34.2)	1.9	(37.8)	0.4
Students without disabilities	286.1	1.3	275.4	0.9	290.7	0.8	280.4	0.3	292.7	0.8	281.5	0.2	296.5	0.7	283.6	0.2
	(31.8)	1.4	(36.1)	0.5	(30.1)	0.6	(33.9)	0.1	(31.7)	0.6	(34.2)	.13	(30.5)	0.6	(34.0)	0.1
Disability gap	40.9**	6.0**	46.5**	2.4**	33.0**	2.0**	38.6**	0.7**	36.0**	2.0**	37.5**	0.5**	35.0**	2.3**	37.7**	0.7**
Effect size (*d*)	1.17		1.23		1.05		1.10		1.08		1.05		1.08		1.05	
U_3	.88		.89		.85		.86		.86		.85		.86		.85	

**The difference in average scores between students not in poverty and students in poverty, or between students without disabilities and students with disabilities, is statistically significant (different from zero) at $p < 0.05$.

Note: Numbers in parentheses are standard deviations. Standard errors for the poverty gap (students not in poverty – students in poverty, or *NP – P*) were calculated as $SE_{NP-P} = \sqrt{(SE_{NP}^2 + SE_P^2)}$. Standard errors for the disability gap (students without disabilities – students with disabilities, or *ND – D*) were calculated as $SE_{ND-D} = \sqrt{(SE_{ND}^2 + SE_D^2)}$. Effect size is calculated as $|d| = (M_1 - M_2) / SD_{pooled}$, where $SD_{pooled} = \sqrt{[(SD_1^2 + SD_2^2) / 2]}$. U_3 is the proportion of scores in the group with the lower mean that fall below the noted effect size in a standard normal distribution. See appendix A for details.

Source: Authors' analysis based on 2000, 2003, 2005, and 2007 NAEP math data for grade 8 from U.S. Department of Education, National Center for Education Statistics (2008).

TABLE E16

Differences in Vermont and U.S. National Assessment of Educational Progress poverty and disability gaps in grades 4 and 8 reading, writing, and math scores, 2000, 2002, 2003, 2005, and 2007

Content area	2000 or 2002			2003			2005			2007		
	Vermont	National	Difference	Vermont	National	Difference	Vermont	National	Difference	Vermont	National	Difference
Reading[a]												
Grade 4												
Poverty gap	20.1	26.8	–6.7**	17.6	27.9	–10.3**	23.3	27.0	–3.7	22.8	26.7	–3.9**
Standard error	2.36	0.81	2.49	1.53	0.47	1.60	1.88	0.34	1.91	1.74	0.40	1.78
Disability gap	31.3	32.9	–1.6	26.6	35.4	–8.8**	36.8	30.6	6.2**	39.6	32.6	7.0**
Standard error	3.28	0.95	3.41	2.79	0.66	2.87	2.89	0.58	2.95	2.43	0.62	2.51
Grade 8												
Poverty gap	19.4	22.3	–2.9	20.3	24.8	–4.5**	19.4	23.2	–3.8**	17.3	23.5	–6.2**
Standard error	2.06	0.70	2.18	1.80	0.48	1.86	1.65	0.33	1.68	1.67	0.39	1.71
Disability gap	28.9	38.7	–9.8**	29.1	41.1	–12.0**	38.9	37.7	1.2	29.3	38.3	–9.0**
Standard error	2.62	1.12	2.85	2.19	0.66	2.29	1.88	0.53	1.96	2.58	0.58	2.64
Writing[b]												
Grade 4												
Poverty gap	20.6	22.5	–1.9	na	na	na	na	na	na	na	na	na
Standard error	2.55	0.89	2.71	na	na	na	na	na	na	na	na	na
Disability gap	44.5	35.0	9.5	na	na	na	na	na	na	na	na	na
Standard error	3.58	0.89	3.69	na	na	na	na	na	na	na	na	na
Grade 8												
Poverty gap	24.5	25.3	–0.8	na	na	na	na	na	na	na	na	na
Standard error	2.81	0.91	2.95	na	na	na	na	na	na	na	na	na
Disability gap	41.7	44.8	–3.1	na	na	na	na	na	na	na	na	na
Standard error	2.64	1.25	2.92	na	na	na	na	na	na	na	na	na
Math[c]												
Grade 4												
Poverty gap	21.9	26.4	–4.5	18.9	22.8	–3.9**	19.3	22.3	–3.0**	17.4	22.1	–4.7**
	2.79	1.51	3.18	1.46	0.37	1.51	1.23	0.26	1.25	1.16	0.30	1.19
Disability gap	17.0	29.3	–12.3**	24.1	22.3	1.8	22.4	21.1	1.3	30.2	21.2	9.0**
	4.66	2.44	5.26	1.99	0.46	2.05	1.70	0.44	1.76	1.52	0.43	1.58

(CONTINUED)

TABLE E16 (CONTINUED)

Differences in Vermont and U.S. National Assessment of Educational Progress poverty and disability gaps in grades 4 and 8 reading, writing, and math scores, 2000, 2002, 2003, 2005, and 2007

Content area	2000 or 2002			2003			2005			2007		
	Vermont	National	Difference	Vermont	National	Difference	Vermont	National	Difference	Vermont	National	Difference
Grade 8												
Poverty gap	25.0	30.1	–5.1	23.1	28.4	–5.3**	21.1	26.7	–5.6**	18.7	26.0	–7.3**
	3.27	1.61	3.65	1.63	0.46	1.69	1.69	0.32	1.72	1.71	0.44	1.77
Disability gap	40.9	46.5	–5.6	33.0	38.6	–5.6**	36.0	37.5	–1.5	35.0	37.7	–2.7
	5.97	2.43	6.44	2.00	0.65	2.11	2.00	0.51	2.06	2.27	0.73	2.39

**The difference in the Vermont and national poverty and disability gap estimates is statistically significant (different from zero) at $p < 0.05$.

na is not applicable because the NAEP writing assessment was not conducted in 2003 or 2005 and data for 2007 were not available at the time of this study.

a. Data are for 2002, 2003, 2005, and 2007.

b. Data are for 2002.

c. Data are for 2000, 2003, 2005, and 2007.

Note: The poverty gap is the mean scale score of students not in poverty minus the mean scale score of students in poverty ($NP - P$). The disability gap is the mean scale score of students without disabilities minus the mean scale score of students with disabilities ($ND - D$). Standard errors for the difference in Vermont and national poverty and disability gap estimates were calculated as $SE_{VT\,gap\,-\,U.S.\,gap} = \sqrt{(SE_{VT\,gap}{}^2 + SE_{U.S.\,gap}{}^2)}$. See appendix A for details.

Source: Authors' analysis based on NAEP data from U.S. Department of Education, National Center for Education Statistics (2008).

APPENDIX F
VERMONT NEW ENGLAND COMMON ASSESSMENT PROGRAM AND NATIONAL ASSESSMENT OF EDUCATIONAL PROGRESS SCALE SCORES AND EFFECT SIZES

TABLE F1

Grades 4 and 8 Vermont average scaled reading scores and effect sizes by gender and student poverty and disability status from the 2006 New England Common Assessment Program and the 2007 National Assessment of Educational Progress, 2006–07

	Grade 4				Grade 8			
	2006 NECAP		2007 NAEP		2006 NECAP		2007 NAEP	
Students and outcome statistic	Mean	Standard error	Mean	Standard error	Mean	Standard error	Mean	Standard error
Total								
Male students	442.4	0.23	224.5	1.30	842.0	0.22	267.6	1.27
	(12.3)		(34.4)		(13.3)		(30.5)	
Female students	446.5	0.23	232.1	0.94	847.2	0.22	278.2	1.22
	(12.9)		(33.2)		(12.9)		(29.3)	
Difference (percentile points)	4.1**	0.33**	7.6**	1.61**	5.2**	0.31**	10.6**	1.76**
Effect size (*d*)	0.32		0.22		0.40		0.35	
U_3	.63		.59		.66		.64	
Students in poverty								
Male students	436.6	0.44	209.4	2.34	835.8	0.42	255.0	2.23
	(14.2)		(33.9)		(13.3)		(33.9)	
Female students	441.2	0.42	215.6	1.84	840.9	0.43	266.0	1.78
	(13.2)		(31.3)		(13.0)		(31.3)	
Difference (percentile points)	4.6**	0.61**	6.2**	2.98**	5.1**	0.60**	11.0**	2.85**
Effect size (*d*)	0.34		0.19		0.39		0.34	
U_3	.63		.58		.65		.63	
Students not in poverty								
Male students	445.0	0.25	231.3	1.31	844.2	0.24	272.4	1.33
	(12.0)		(32.4)		(12.6)		(32.4)	
Female students	448.9	0.26	239.5	1.19	849.5	0.24	282.3	1.44
	(12.1)		(31.3)		(12.0)		(31.3)	
Difference (percentile points)	3.9**	0.36**	8.2**	1.77**	5.3**	0.34**	9.9**	1.96**
Effect size (*d*)	0.32		0.26		0.43		0.31	
U_3	.63		.60		.67		.62	

(CONTINUED)

TABLE F1 (CONTINUED)

Grades 4 and 8 Vermont average scaled reading scores and effect sizes by gender and student poverty and disability status from the 2006 New England Common Assessment Program and the 2007 National Assessment of Educational Progress, 2006–07

	Grade 4				Grade 8			
	2006 NECAP		2007 NAEP		2006 NECAP		2007 NAEP	
Students and outcome statistic	Mean	Standard error	Mean	Standard error	Mean	Standard error	Mean	Standard error
Students with disabilities								
Male students	427.5	0.67	195.6	2.93	827.2	0.45	249.0	2.80
	(15.1)		(36.0)		(11.7)		(32.6)	
Female students	425.7	0.95	189.9	4.34	827.9	0.64	246.6	3.97
	(14.5)		(41.8)		(11.4)		(35.0)	
Difference (percentile points)	–1.8**	1.16**	–5.7	5.23**	0.7	0.78**	–2.4	4.86**
Effect size (*d*)	–0.12		–0.15		0.06		–0.07	
U_3	.45		.44		.52		.47	
Students without disabilities								
Male students	445.0	0.21	231.1	1.21	845.3	0.20	272.6	1.21
	(11.0)		(30.4)		(11.3)		(27.9)	
Female students	448.2	0.21	235.7	0.93	849.1	0.20	281.6	1.20
	(11.2)		(29.7)		(11.3)		(26.5)	
Difference (percentile points)	3.2**	0.29**	4.6**	1.53**	3.9**	0.29**	9.0**	1.70**
Effect size (*d*)	0.28		0.15		0.34		0.33	
U_3	.61		.56		.63		.63	

**Gender gap is statistically significant (different from zero) at $p < 0.05$.

Note: Numbers in parentheses are standard deviations. Effect size is calculated as $|d| = (M_1 - M_2) / SD_{pooled}$, where $SD_{pooled} = \sqrt{[(SD_1^2 + SD_2^2) / 2]}$. U_3 is the proportion of scores in the group with the lower mean that fall below the noted effect size in a standard normal distribution. See appendix A for details.

Source: Authors' analysis based on NECAP data from Vermont Department of Education, Standards and Assessment (2007) and NAEP data from U.S. Department of Education, National Center for Education Statistics (2008).

TABLE F2

Grades 5 and 8 Vermont average scaled writing scores and effect sizes by gender and student poverty and disability status from the 2006 New England Common Assessment Program, 2006

Students and outcome statistics	Grade 5		Grade 8	
	Mean	Standard error	Mean	Standard error
Total				
Male students	536.7	0.27	835.8	0.23
	(15.3)		(13.9)	
Female students	543.3	0.29	844.5	0.23
	(16.2)		(13.8)	
Difference (percentile points)	6.6**	0.39**	8.7**	0.33**
Effect size (*d*)	0.42		0.63	
U_3	.66		.74	
Students in poverty				
Male students	530.4	0.44	829.6	0.41
	(14.4)		(12.9)	
Female students	536.7	0.50	838.2	0.43
	(15.5)		(13.2)	
Difference (percentile points)	6.2**	0.67**	8.6**	0.60**
Effect size (*d*)	0.42		0.66	
U_3	.66		.75	
Students not in poverty				
Male students	539.5	0.31	838.0	0.26
	(14.9)		(13.5)	
Female students	546.3	0.34	846.8	0.26
	(15.6)		(13.3)	
Difference (percentile points)	6.7**	0.46**	8.8**	0.37**
Effect size (*d*)	0.44		0.66	
U_3	.67		.75	
Students with disabilities				
Male students	520.9	0.56	820.9	0.44
	(12.4)		(11.4)	
Female students	521.3	0.76	826.6	0.66
	(12.8)		(11.6)	
Difference (percentile points)	0.4	0.94**	5.7**	0.79**
Effect size (*d*)	0.03		0.49	
U_3	.51		.69	

(CONTINUED)

TABLE F2 (CONTINUED)

Grades 5 and 8 Vermont average scaled writing scores and effect sizes by gender and student poverty and disability status from the 2006 New England Common Assessment Program, 2006

Students and outcome statistics	Grade 5		Grade 8	
	Mean	Standard error	Mean	Standard error
Students without disabilities				
Male	539.5	0.27	839.1	0.22
	(14.1)		(12.1)	
Female	545.4	0.28	845.3	0.23
	(14.9)		(12.6)	
Difference	6.0**	0.38**	6.2**	0.31**
Effect size (*d*)	0.41		0.50	
U_3	.66		.69	

**The difference in average female scores and male scores (*F* – *M*) is statistically significant (different from zero) at $p < 0.05$.

Note: Numbers in parentheses are standard deviations. Effect size is calculated as $|d| = (M_1 - M_2) / SD_{pooled}$, where $SD_{pooled} = \sqrt{[(SD_1^2 + SD_2^2) / 2]}$. U_3 is the proportion of scores in the group with the lower mean that fall below the noted effect size in a standard normal distribution. See appendix A for details.

Source: Authors' tabulations based on NECAP data from Vermont Department of Education, Standards and Assessment (2007) and NAEP data from U.S. Department of Education, National Center for Education Statistics (2008).

TABLE F3

Grades 4 and 8 Vermont average scaled math scores and effect sizes by gender and student poverty and disability status from the 2006 New England Common Assessment Program and the 2007 National Assessment of Educational Progress, 2006–07

	Grade 4				Grade 8			
	2006 NECAP		2007 NAEP		2006 NECAP		2007 NAEP	
Students and outcome statistics	Mean	Standard error	Mean	Standard error	Mean	Standard error	Mean	Standard error
Total								
Male students	443.3	0.22	247.7	0.67	840.9	0.20	291.9	1.04
	(12.5)		(27.2)		(12.1)		(35.2)	
Female students	443.1	0.22	245.0	0.72	841.9	0.19	290.1	1.06
	(12.3)		(26.0)		(11.0)		(31.8)	
Difference (percentile points)	0.2	0.31**	–2.7**	0.98**	0.9**	0.27**	1.8	1.48**
Effect size (*d*)	0.02		–0.10		0.08		0.05	
U_3	.51		.46		.53		.52	
Students in poverty								
Male students	437.7	0.40	235.9	1.43	835.9	0.39	280.5	2.13
	(12.8)		(27.1)		(12.2)		(32.7)	
Female students	437.6	0.40	232.9	1.17	836.7	0.36	274.6	1.79
	(12.3)		(26.7)		(11.0)		(29.2)	
Difference (percentile points)	0.0	0.56**	–3.0	1.85**	1.8	0.53**	–5.9**	2.78**
Effect size (*d*)	0.00		–0.11		0.11		–0.19	
U_3	.50		.46		.54		.42	
Students not in poverty								
Male students	445.9	0.24	253.0	0.82	842.7	0.22	295.9	1.35
	(11.5)		(25.5)		(11.6)		(35.2)	
Female students	445.5	0.25	250.5	0.91	843.8	0.21	296.4	1.17
	(11.6)		(23.7)		(10.4)		(30.7)	
Difference (percentile points)	–0.3	0.34**	–2.5	1.22**	1.1**	0.30**	0.5	1.79**
Effect size (*d*)	–0.03		–0.10		0.10		0.02	
U_3	.49		.46		.54		.51	
Students with disabilities								
Male students	431.4	0.62	223.6	1.79	827.7	0.47	265.0	2.89
	(13.9)		(28.5)		(12.3)		(36.2)	
Female students	426.0	0.84	215.3	2.52	826.2	0.65	255.1	3.25
	(12.9)		(28.6)		(11.6)		(29.1)	
Difference (percentile points)	–5.4**	1.05**	–8.3**	3.09**	–1.5	0.81**	–9.9**	4.35**
Effect size (*d*)	–0.41		–0.29		–0.13		–0.30	
U_3	.34		.39		.45		.38	

(CONTINUED)

TABLE F3 (CONTINUED)

Grades 4 and 8 Vermont average scaled math scores and effect sizes by gender and student poverty and disability status from the 2006 New England Common Assessment Program and the 2007 National Assessment of Educational Progress, 2006–07

	Grade 4				Grade 8			
	2006 NECAP		2007 NAEP		2006 NECAP		2007 NAEP	
Students and outcome statistics	Mean	Standard error	Mean	Standard error	Mean	Standard error	Mean	Standard error
Students without disabilities								
Male students	445.4	0.21	253.1	0.77	843.9	0.18	298.8	1.06
	(10.9)		(23.8)		(9.9)		(31.5)	
Female students	444.5	0.21	248.6	0.76	843.4	0.17	294.4	0.97
	(11.2)		(23.2)		(9.7)		(29.4)	
Difference (percentile points)	–1.0**	0.29**	–4.5**	1.08**	–0.4	0.25**	–4.4**	1.44**
Effect size (*d*)	–0.09		–0.19		–0.05		–0.14	
U_3	.46		.42		.48		.44	

**The difference in average female scores and male scores ($F - M$) is statistically significant (different from zero) at $p < 0.05$.

Note: Numbers in parentheses are standard deviations. Effect size is calculated as $|d| = (M_1 - M_2) / SD_{pooled}$, where $SD_{pooled} = \sqrt{[(SD_1^2 + SD_2^2) / 2]}$. U_3 is the proportion of scores in the group with the lower mean that fall below the noted effect size in a standard normal distribution. See appendix A for details.

Source: Authors' analysis of NECAP data from Vermont Department of Education, Standards and Assessment (2007) and NAEP data from U.S. Department of Education, National Center for Education Statistics (2008).

TABLE F4

Grades 4 and 8 Vermont average scaled reading scores and effect sizes by poverty status and disability status from the 2006 New England Common Assessment Program and the 2007 National Assessment of Educational Progress, 2006–07

	Grade 4				Grade 8			
	2006 NECAP		2007 NAEP		2006 NECAP		2007 NAEP	
Student status and outcome statistics	Mean	Standard error	Mean	Standard error	Mean	Standard error	Mean	Standard error
Poverty status								
Students in poverty	438.8	0.31	212.5	1.48	838.3	0.31	260.3	1.39
	(13.9)		(32.8)		(13.4)		(30.6)	
Students not in poverty	446.9	0.18	235.3	0.92	846.7	0.17	277.6	0.92
	(12.2)		(32.1)		(12.6)		(29.0)	
Difference (percentile points)	8.0**	0.36**	22.8**	1.74**	8.5**	0.35**	17.3**	1.67**
Effect size (*d*)	0.61		0.70		0.65		0.58	
U_3	.73		.76		.74		.72	
Disability status								
Students with disabilities	426.9	0.55	193.9	2.30	824.5	0.37	248.2	2.44
	(14.9)		(37.9)		(11.6)		(33.4)	
Students without disabilities	446.6	0.15	233.5	0.80	847.2	0.15	277.5	0.83
	(11.3)		(30.1)		(11.4)		(27.5)	
Difference	19.7**	0.57**	39.6**	2.43**	22.7**	0.39**	29.3**	2.58**
Effect size (*d*)	1.49		1.16		1.97		0.96	
U_3	.93		.88		.98		.83	

**The difference in average scores between students not in poverty and in poverty (*NP* – *P*), or students without disabilities and with disabilities (*ND* – *D*), is statistically significant (different from zero) at $p < 0.05$.

Note: Numbers in parentheses are standard deviations. Effect size is calculated as $|d| = (M_1 - M_2) / SD_{pooled}$, where $SD_{pooled} = \sqrt{[(SD_1^2 + SD_2^2) / 2]}$. U_3 is the proportion of scores in the group with the lower mean that fall below the noted effect size in a standard normal distribution. See appendix A for details.

Source: Authors' analysis of NECAP data from Vermont Department of Education, Standards and Assessment (2007) and NAEP data from U.S. Department of Education, National Center for Education Statistics (2008).

TABLE F5

Grades 5 and 8 Vermont average scaled writing scores and effect sizes by poverty status and disability status from the 2006 New England Common Assessment Program, 2006

	Grade 5		Grade 8	
Student status	Mean	Standard error	Mean	Standard error
Poverty status				
Students in poverty	533.4	0.34	833.8	0.32
	(15.3)		(13.8)	
Students not in poverty	542.8	0.23	842.2	0.19
	(15.6)		(14.1)	
Difference (percentile points)	9.4**	0.41**	8.5**	0.37**
Effect size (*d*)	0.61		0.61	
U_3	.73		.73	
Disability status				
Students with disabilities	521.0	0.45	822.7	0.37
	(12.6)		(11.8)	
Students without disabilities	542.5	0.20	842.8	0.16
	(14.8)		(12.9)	
Difference (percentile points)	21.5**	0.49**	20.1**	0.41**
Effect size (*d*)	1.57		1.62	
U_3	.94		.95	

**The difference in average scores between students not in poverty and in poverty (*NP* – *P*), or students without disabilities and with disabilities (*ND* – *D*), is statistically significant (different from zero) at $p < 0.05$.

Note: Numbers in parentheses are standard deviations. Effect size is calculated as $|d| = (M_1 - M_2) / SD_{pooled}$, where $SD_{pooled} = \sqrt{[(SD_1^2 + SD_2^2) / 2]}$. U_3 is the proportion of scores in the group with the lower mean that fall below the noted effect size in a standard normal distribution. See appendix A for details.

Source: Authors' analysis of NECAP data from Vermont Department of Education, Standards and Assessment (2007) and NAEP data from U.S. Department of Education, National Center for Education Statistics (2008).

TABLE F6

Grades 4 and 8 Vermont average scaled math scores and effect sizes by poverty and disability status from the 2006 New England Common Assessment Program and the 2007 National Assessment of Educational Progress, 2006–07

	Grade 4				Grade 8			
	2006 NECAP		2007 NAEP		2006 NECAP		2007 NAEP	
Student status	Mean	Standard error	Mean	Standard error	Mean	Standard error	Mean	Standard error
Poverty status								
Students in poverty	437.6	0.28	234.4	0.97	836.3	0.27	277.4	1.45
	(12.5)		(27.0)		(11.7)		(31.0)	
Students not in poverty	445.7	0.17	251.8	0.63	843.2	0.15	296.1	0.91
	(11.5)		(24.7)		(11.0)		(33.1)	
Difference (percentile difference)	8.1**	0.33**	17.4**	1.16**	6.9**	0.31**	18.7**	1.71**
Effect size (*d*)	0.67		0.67		0.61		0.58	
U_3	.75		.75		.73		.72	
Disability status								
Students with disabilities	429.7	0.51	220.6	1.40	827.7	0.12	261.5	2.15
	(13.8)		(28.8)		(12.1)		(34.2)	
Students without disabilities	444.9	0.15	250.8	0.59	843.6	0.38	296.5	0.74
	(11.1)		(23.6)		(9.8)		(30.5)	
Difference (percentile points)	15.2**	0.53**	30.2**	1.52**	15.9**	0.40**	35.0**	2.27**
Effect size (*d*)	1.22		1.15		1.45		1.08	
U_3	.89		.87		.93		.86	

**The difference in average scores between students not in poverty and in poverty (*NP* – *P*), or students without disabilities and with disabilities (*ND* – *D*), is statistically significant (different from zero) at $p < 0.05$.

Note: Numbers in parentheses are standard deviations. Effect size is calculated as $|d| = (M_1 - M_2) / SD_{pooled}$, where $SD_{pooled} = \sqrt{[(SD_1^2 + SD_2^2) / 2]}$. U_3 is the proportion of scores in the group with the lower mean that fall below the noted effect size in a standard normal distribution. See appendix A for details.

Source: Authors' tabulations of NECAP data from Vermont Department of Education, Standards and Assessment (2007) and NAEP data from U.S. Department of Education, National Center for Education Statistics (2008).

NOTES

1. The authors thank Jane Nesbitt and the Vermont Department of Education for assistance with this project. The authors also acknowledge the contributions of Craig Hoyle, Katie Buckley, Karen Clay, Denise Lamb, Jane Donnelly, Rebecca Carey, Maria-Paz Avery, Natalie Lacireno-Paquet, Sarah Chace, Katie Culp, Jessica Brett, Bailey Triggs, and Michelle LaPointe.

2. Gender differences in aggregate math scores have been relatively small in recent years. However, due to the greater variability of male scores and the underrepresentation of female students in the upper tail of the math score distribution, the gender gap in math has been larger among high-achieving students bound for higher education (Lewis and Willingham 1995). These larger gender gaps in math among populations that aim for postsecondary education may help to explain the existing underrepresentation of women in science, technology, engineering, and math (National Science Foundation 2006).

3. For decades scholars have debated the origin of gender gaps in academic and other outcomes. The most widely accepted current view is that gender differences arise from complex interactions between biology and the environment (Halpern 2000, 2004; Kimura 1999, 2004; Neisser et al. 1996).

4. Although national data show that gender gaps may vary in important ways by race/ethnicity and limited English proficiency status, Vermont education leaders chose not to examine gender gaps within these subgroups because of their low proportions within the state (see table C1 in appendix C).

5. The NAEP is administered to representative samples of grade 4 and 8 students across Vermont. Because the NAEP does not assess all students, it cannot be used to measure adequate yearly progress in reading and math as required by the act (appendix B).

6. Samples for the Vermont and U.S. NAEP assessments each include more than 1,000 students and are therefore large. See NAEP Technical Documentation at http://nces.ed.gov/nationsreportcard/tdw/sample_design/. Although the Vermont sample forms a subset of the larger national sample, the Vermont and U.S. NAEP samples were treated as independent because the Vermont fraction of the national sample is likely to be extremely small. The NAEP does not report the proportion of Vermont students within the total national sample, but according to 2005/06 data from the U.S. Department of Education, National Center for Education Statistics (2006a), Vermont had a total enrollment of 96,638 students—less than 2 percent of the national total (49.1 million). The proportion of Vermont students within the U.S. NAEP sample should therefore also be less than 2 percent. Any correlation between NAEP scores in Vermont and the United States is therefore likely to be trivial.

7. The Vermont NAEP reading and writing assessments were first administered in 2002. The state's NAEP math assessment was first administered in Vermont in 1996, but it differed from the one administered after 2000. This project therefore focused on NAEP math data from 2000 to 2007.

REFERENCES

American Association of University Women. (1992). *The AAUW report: how schools shortchange girls.* Washington, D.C.: American Association of University Women Educational Foundation.

———. (1998). *Where schools still fail our children* [Executive summary]. Washington, D.C.: American Association of University Women Educational Foundation.

Cole, N. (1997). *The ETS gender study: how females and males perform in educational settings*. Princeton, NJ: Educational Testing Service.

Coley, R. J. (2001). *Differences in the gender gap: comparisons across racial/ethnic groups in education and work*. Princeton, NJ: Educational Testing Service.

Conlin, M. (2003). The gender gap [Electronic Version]. *Business Week*. May 26, 2003. Retrieved January 29, 2008, from http://www.businessweek.com/magazine/content/03_21/b3834001_mz001.htm.

Cumming, G. F., S. (2001). A primer on the understanding, use and calculation of confidence intervals based on central and noncentral distributions. *Educational and Psychological Measurement, 61*(4), 530–72.

Freeman, C. E. (2004). *Trends in educational equity of girls and women: 2004*. Washington, D.C.: National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education.

Halpern, D. F. (2000). *Sex differences in cognitive abilities*. Mahwah, NJ: Lawrence Erlbaum Associates.

Halpern, D. F. (2004). A cognitive-process taxonomy for sex differences in cognitive abilities. *Current Directions in Psychological Science, 13*(4), 135–39.

Halpern, D. F., Benbow, C. P., Geary, D. C., Gur, R. C., Hyde, J. S., and Gernsbacher, M. A. (2008). Sex, math, and scientific achievement. *Scientific American Mind, 18*(6), 44–51.

Hayes, S. (2007). *Vermont NAEP 2007 results: achievement gap report*. Montpelier, VT: Vermont Department of Education.

Johnson, K. A. (2002). *Equity gaps in Vermont, 2002. Closing the gap: fulfilling the promise of student success through equity, diversity, and character education*. Retrieved June 5, 2008, from http://www.vermontinstitutes.org/equity/closing_the_gap_503.ppt.

Kimura, D. (1999). *Sex and cognition*. Cambridge, MA: MIT Press.

Kimura, D. (2004). Human sex differences in cognition, fact, not predicament. *Sexualities, Evolution & Gender, 6*(1), 45–53.

Klecker, B. M. (2006). The gender gap in NAEP fourth-, eighth-, and twelfth-grade reading scores across years. *Reading Improvement, 43*(1), 50–6.

Lewis, C., and Willingham, W. W. (1995). *The effects of sample restriction on gender differences* (ETS-RR-95-13). Princeton, NH: Educational Testing Service.

Light, D., Singer, J., and Willett, J. (1990). *By design: planning research on higher education*. Cambridge, MA: Harvard University Press.

LoGerfo, L., Nichols, A., and Chaplin, D. (2006). *Gender gaps in math and reading gains during elementary and high school by race and ethnicity*. Washington, DC: The Urban Institute. Available from http://www.urban.org/url.cfm?ID=411428.

McGraw, F., Lubienski, S. T., and Struchens, M. E. (2006). A closer look at gender in NAEP mathematics achievement and affect data: intersections with achievement, race/ethnicity, and socioeconomic status. *Journal for Research in Mathematics Education, 37*(2), 129–50.

Mead, S. (2006). *The evidence suggests otherwise: the truth about boys and girls*. Washington. D.C.: The Education Sector.

Meadows, S. O., Land, K. D., and Lamb, V. L. (2005). Assessing Gilligan vs. Sommers: gender-specific trends in child and youth well-being in the United States, 1985–2001. *Social Indicators Research, 70*(1), 1–52.

Measured Progress. (2007). *New England Common Test Program: 2006–2007 Technical Report*. Dover, NH: Measured Progress.

Mislevy, R. J., Johnson, E. G., and Muraki, E. (1992). Scaling procedures in NAEP. *Journal of Educational Statistics, 17*(2), 131–54.

National Science Foundation. (2006). *Science and Engineering Degrees: 1966–2004* (NSF 07-307). Retrieved May 20, 2008, from http://www.nsf.gov/statistics/nsf07307/.

Neisser, U., Boodoo, G., Bouchard, T. J., Boykin, W. A., Brody, N., Ceci, S. J., Halpern, D. F., Loehlin, J. C., Perloff, R., Sternberg, R. J., Urbina, S. (1996). Intelligence: knowns and unknowns. *American Psychologist, 51*(2), 77–101.

Nix, T. W., and Barnette, J. J. (1998). The data analysis dilemma: ban or abandon. A review of null hypothesis significance testing. *Research in the Schools, 5*(2), 3–14.

Nowell, A., and Hedges, L. V. (1998). Trends in gender differences in academic achievement from 1960 to 1994: an analysis of differences in the mean, variance, and extreme scores. *Sex Roles, 39*(112), 21–42.

Pandiani, J., and Bramley, J. (2002). Memorandum to the Vermont Mental Health Performance Indicator Project Advisory Group and Interested Parties. Retrieved June 5, 2008, from http://www.healthvermont.gov/mh/docs/pips/2002/pip071202.pdf.

Ready, D. D., LoGerfo, L. F., Burkham, D. T., and Lee, V. E. (2005). Explaining girls' advantage in kindergarten literacy learning: do classroom behaviors make a difference? *The Elementary School Journal, 106*(1), 21–38.

Spelke, E. (2005, December). Sex differences in intrinsic aptitude for mathematics and science? A critical review. *American Psychologist, 60*(9), 950–58.

Stahl, L. (2003). The gender gap: boys lagging [Electronic Version]. *CBS News: 60 Minutes*. May 25, 2003. Retrieved January 29, 2008, from http://www.cbsnews.com/stories/2002/10/31/60minutes/main527678.shtml?source=search_story.

Steiger, J. (2004). Beyond the F Test: effect size confidence intervals and tests of class fit in the analysis of variance and contrast analysis. *Psychological Methods, 9*(2), 164–82.

Thompson, B. (1998). Statistical significance and effect size reporting: portrait of a possible future. *Research in the Schools, 5*(2), 33–8.

Tyre, P. (2006). The trouble with boys [Electronic Version]. *Newsweek*. January 30, 2006. Retrieved January 29, 2008, from http://www.newsweek.com/id/47522.

U.S. Department of Education, National Center for Education Statistics. (2005). *The condition of education 2005* (No. NCES 2005-071). Washington, D.C.: U.S. Government Printing Office.

U.S. Department of Education, National Center for Education Statistics. (2006a). *Common Core of Data. Public Elementary/Secondary School Universe Survey 2005–06* [Data file]. Available from http://nces.ed.gov/ccd/.

U.S. Department of Education, National Center for Education Statistics. (2006b). *Common Core of Data. Local Education Agency Universe Survey, 2005-06* [Data file]. Available from http://nces.ed.gov/ccd/.

U.S. Department of Education, National Center for Education Statistics. (2006c). *Common Core of Data. State Nonfiscal Survey of Public Elementary/Secondary Education, 2005-06* [Data file]. Available from http://nces.ed.gov/ccd/.

U.S. Department of Education, National Center for Education Statistics. (2006d).*Common Core of Data. Numbers and Types of Public Elementary and Secondary Schools, School year 2005–06* [Data file]. Available from http://nces.ed.gov/ccd/.

U.S. Department of Education, National Center for Education Statistics. (2006e). *The condition of education, 2006* (No. NCES 2006-071). Washington, DC: U.S. Government Printing Office.

U.S. Department of Education, National Center for Education Statistics. (2007). *Mapping 2005 state proficiency standards onto the NAEP scales* (No. NCES 2007-482).

Washington, D. C.: National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education.

U.S. Department of Education, National Center for Education Statistics. (2008). *The NAEP Explorer Database*. Retrieved March 15, 2008, from http://nces.ed.gov/nationsreportcard/tdw/database/.

Valentine, J. C., Cooper, H. (2003). *Effect size substantive interpretation guidelines: issues in the interpretation of effect sizes.* Washington, D.C.: What Works Clearinghouse.

Vermont Department of Education, Standards and Assessment. (2007). *New England Common Assessment Program, 2005 and 2006* [Data file]. Received from the Vermont Department of Education, February 25, 2007.

Vermont Student Assistance Center. (2005). *Gender gap in achievement a worry*. Retrieved June 5, 2008, from http://services.vsac.org/ilwwcm/resources/file/ebd81b4cb79204e/Newsline%20spring%202005.pdf.

Von Drehle, D. (2007). The boys are all right. *Time, 170*(August 6), 38–47.

Willingham, W., and Cole, N. S. (1997). *Gender and fair assessment*. Mahwah, NJ: Lawrence Erlbaum Associates.