NAEP Fourth-, Eighth-, and Twelfth-Grade Reading Scores by Gender: 2005, 2007, 2009, 2011, 2013

Beverly M. Klecker

Morehead State University

b.klecker@moreheadstate.edu

Paper presented at the annual meeting of the

Mid-South Educational Research Association

Knoxville, TN

November 5, 2014

Abstract

This paper describes a secondary analysis of the National Assessment of Educational Progress (NAEP) reading scores by gender. Data were national public 4th- and 8th-grade reading scores from composite and subscales for 2005, 2007, 2009, 2011, and 2013. Twelfth-grade scores for composite and literary experience from 2005, 2009, and 2013 and gain information from 2005 were included. Differences ($p.<.001$; Cohen's *d* effect size) in reading average scale scores by gender were consistent across grade level and years with females scoring higher than males. Results are congruent with a previous study of NAEP reading by gender across fourth-, eighth, and twelfth-grade-levels for 1994, 1998, 2000, 2002, and 2003 (Klecker, 2006). Discussion includes comparisons with cross-cultural international assessments and possible explanations for the widely-observed gender difference in large-scale standardized reading assessments.

Keywords: *NAEP, reading achievement; gender*

NAEP Fourth-, Eighth-, and Twelfth-Grade Reading Scores by Gender: 2005, 2007, 2009, 2011, 2013

Background of the Study

The United States' National Assessment of Educational Achievement (NAEP) "…The Nation's Report Card… is the largest nationally representative and continuing assessment of what America's students know and can do in various subject areas…" (National Center for Educational Statistics (NCES), 2014a. para 1)

NCES (2014a) described the history of NAEP:

> After much exploration in the early 1960s, the idea of a national assessment gained impetus in 1963. NAEP planning began in 1964, with a grant from the Carnegie Corporation to set up the Exploratory Committee for the Assessment of Progress in Education (ECAPE) in June. This was followed by the appointment of the Technical Advisory Committee (TAC) in 1965.
>
> The first national assessments were held in 1969. Voluntary assessments for the states began in 1990 on a trial basis, were made a permanent feature of NAEP every two years. In 2002, selected urban districts participated in the state-level assessments on a trial basis, and continue as the Trial Urban District Assessment. (p. 1)

The data from the NAEP assessments have been available across the years for analysis by educational researchers. Workshops for national and international researchers have been conducted in the Washington, DC area and at international meetings of the American Educational Research Association. The NAEP database grew after No Child Left Behind (NCLB) (2002) legislation required NAEP participation in reading and mathematics assessment

in fourth and eighth grades by all districts that received Title I funds. Currently, NAEP data are available—with online training materials--for analysis using the NAEP Data Explorer (NCES, 2014b).

A previous study by the author (Klecker, 2006) examined national public school fourth-, eighth-, and twelfth-grade NAEP reading scores by gender for the years 1992, 1994, 1998, 2000, 2002, and 2003. Across all analyses of average scale scores by gender, girls' scores were higher than boys' ($p.<.001$). Effect sizes (Cohen's *d*) were small in fourth-grade (0.13-0.27), small to moderate in eighth-grade (0.27-0.43), and small to moderate in twelfth-grade (0.22-0.44).

Purpose of the Study

Since 2006, an additional number of research studies have examined gender differences in large-scale—national and international—reading achievement assessments. The purpose of this study was threefold: (1) to review recent related literature, (2) to repeat the NAEP reading by gender study using data from years 2005, 2007, 2009, 2011, and 2013, and (3) to examine possible explanations for gender differences in large-scale national and international assessments.

Review of Literature

The review of the literature first examines results from multiple-years of two large-scale international assessment of reading achievement and literacy. Next, summaries from meta analyses of reading by gender are presented.

Organization for Economic Co-operation and Development Assessments of Reading

Beginning in 2000, the Organization for Economic Co-operation and Development (OECD) has periodically conducted two large-scale reading assessments as measures of

international literacy: (1) the Progress in International Reading Literacy (PIRLS) and (2) the Programme for International Student Assessment (PISA). PIRLS was administered to fourth-grade students in 2001, 2006, and 2011. This assessment was conducted across international countries and units at the approximate end of the primary grades (NCES, 2014c). The Programme for International Student Assessment (PISA) is a measure of achievement—including reading—that is given to 15-year-old students at the approximate end of secondary schooling. PISA is a triennial international survey of what students know and can do. PISA was administered in 2000, 2003, 2006, 2009, and 2012 (OECD). Because of a printing error on the assessment in the United States in 2003, reading results were not available that year. Results of the PIRLS and PISA by gender are summarized below (NCES, 2014d).

Progress in International Reading Literacy (PIRLS). The following are summaries of the 2001, 2006, and 2011 PIRLS fourth-grade reading results by gender.

> **PIRLS 2001**. Ogle, et al. (April, 2003) stated:
>
> Fourth-grade girls score higher than fourth-grade boys on the combined reading literacy scale on average in every participating PIRLS 2001 country (figure 7). In the United States, on average, girls score 18 points higher than boys on the combined reading literacy scale. Internationally, the average score difference between boys and girls range from 8 points (Italy) to 27 points (Belize, Iran, and New Zealand). (p. 10)
>
> **PIRLS 2006**. Baer, Baldi, Ayotle, and Green (November, 2007) reported:
>
> In 2006, in all but two jurisdictions (Luxembourg and Spain), average scores for girls on the combined reading literacy scale were higher than average scores for boys (figure 5). In the United States, girls on average scored 10 points higher than

> boys (545 versus 535); internationally, the average score for girls was 17 points higher than the average score for boys. (p. 10)

PIRLS 2011: Mullins, Martin, Foy, and Dructer (2012) described the results:

> In nearly all of the countries and benchmarking participants, girls outperformed boys in 2011, and there has been little reduction in the reading achievement gender gap over the decade. Across the 45 countries participating at the fourth grade, girls had a 16-point advantage, on average, compared to boys. Only five countries showed no difference: Colombia, Italy, France, Spain, and Israel.
>
> The reading achievement gender gap is larger for literary than for informational reading. In literary reading, girls had higher achievement than boys in nearly every country and benchmarking participant. However, girls and boys had fewer achievement differences in informational reading. (p. 7)

Programme for International Student Assessment (PISA). The Organization for Economic Co-operation and Development (OECD) began the Programme for International Student Assessment (PISA) to measure achievement—including reading—of 15-year-old students. PISA assessments have been given every three years: 2000, 2003, 2006, 2009, and 2012. Gender differences favoring girls were found the first three years of assessments in all countries (Organization for Economic Co-operation and Development (OECD), 2014). The later 2009 PISA results were reported by OECD (2010):

> Throughout much of the 20th century, concern about gender differences in education focused on girls' underachievement. More recently, however, the scrutiny has shifted to boys' underachievement in reading. In the PISA 2009 reading assessment, girls outperform boys in every participating country by an

> average, among OECD countries, of 39 PISA score points – equivalent to more than half a proficiency level or one year of schooling. (p. 2)

PISA 2012 assessment results were reported (OECD, 2012):

> Girls outperform boys in reading almost everywhere. This gender gap is particularly large in some high-performing countries, where almost all underperformance in reading is seen only among boys. Low-performing boys face a particularly large disadvantage as they are heavily over-represented among those who fail to show basic levels of reading literacy. These low levels of performance tend to be coupled with low levels of engagement with school and – as observed in PISA 2009 – with low levels of engagement with and commitment to reading. To close the gender gap in reading performance, policy makers need to promote boys' engagement with reading and ensure that more boys begin to show the basic level of proficiency that will allow them to participate fully and productively in life. (p. 7)

Meta-Analyses of Reading by Gender Assessments

Lingard, Martino, & Mills (2009) stated, "…The underperformance of boys in the United States in comparison to girls is a relative latecomer to the debates which have been a predominant feature in educational policy in Australia, Canada and the United Kingdom (UK) for over 15 years" (cited in Skelton & Francis, 2011, p. 456). Skelton and Francis (2011) summarized some of the strategies found in the literature to address the "gender gap" including examining boys' "learning styles" and a list of "Books for Boys."

Brookhart (2006) examined gender and "in/equity" in achievement assessment in reading and language arts, mathematics, science, and multiple subjects. She described Lietz's (2006)

meta-analysis and hierarchical linear modeling statistical techniques used to examine gender differences in reading. Brookhart (2006) summarized:

> Her meta-analysis included 139 effect sizes from various studies of secondary school reading achievement between 1970 and 2003, including the International Association for the Evaluation of Educational Achievement (IEA) Reading Comprehension Study (1970-1971) and Reading Literacy Study (1990-91), PISA 2000, NAEP 1992-2003, a number of studies in Australia over the period 1975-2002 and other published studies. The overall grand mean was an effect size of 0.19, a small effect that meant girls outscored boys overall. (pp. 120-121).

Summary

From the literature reviewed for the previous study (Klecker, 2006) and the current study, it is evident that girls' average scores are higher than boys' average scores on large-scale reading assessments. This is consistent across years, grade levels, and international borders.

Method

Participants and Sampling

NCES (2014e) described the sampling and data collection protocols used for collecting NAEP fourth- and eighth-grade reading data every two years. Since NCLB (2002), participation by fourth- and eighth-grade students in reading and mathematics assessments has been mandatory in states receiving Title I funds. All states have participated in these assessments since NCLB (2002).

NAEP Sampling and Data Collection

Sampling for the 4th-grade and 8th-grade reading assessment used a multistage sampling design that sampled students from selected schools within selected

geographic areas across the country. Each assessment cycle, a sample of students in designated grades within both public and private schools throughout the United States (and sometimes specified territories and possessions) is selected for assessment.

Public School Selection in State Assessment Years

The selection of a sample of public school students for state assessment involves a complex multistage sampling design with the following stages:

Select public schools within the designated areas,

Select students in the relevant grades within the designated schools, and

Allocate selected students to assessment subjects.

The Common Core of Data (CCD) file, a comprehensive list of operating public schools in each jurisdiction that is compiled each school year by the National Center for Education Statistics (NCES), is used as the sampling frame for the selection of sample schools. The CCD also contains information about grades served, enrollment, and location of each school. In addition to the CCD list, a set of specially sampled jurisdictions is contacted to determine if there are any newly formed public schools that were not included in the lists used as sampling frames. Considerable effort is expended to increase the survey coverage by locating public schools not included in the most recent CCD file. (para 1-3)

Because state NAEP assessments do not include 12th-grade students, a grade twelve sample of schools was selected (NCES, 2014f). The sample was designed to provide national estimates of 12th-grade achievement. The sampling for the 2005, 2009, and 2013 assessments provided a nationally representative sample of 12-th grade students. The 2009 and 2013 samples

were from selected students within selected schools from eleven volunteer states: Arkansas, Connecticut, Florida, Idaho, Illinois, Iowa, Massachusetts, New Hampshire, New Jersey, South Dakota, and West Virginia (NCES, 2014f).

Data Analysis

The NAEP Data Explorer (NCES, 2014b) was used to analyze the data from the fourth-, eighth-, and twelfth- grade national public schools reading composite, gain information, and literacy average scale scores for the years 2005, 2007, 2009, 2011, and 2013 by gender. Alpha was set a priori at .001. All differences were statistically significant and effect sizes, *d* (Cohen, 1988), were hand calculated.

Results

[Table 1 about here]

Table 1 presents NAEP fourth-grade reading composite average scale scores by gender across assessment years from 2005 through 2013. The average scale scores for females increased by four scale points (220 to 2004) and the average scale scores for males increased by three points (214 to 217). In each year, females' scores were statistically significantly ($p.<.001$) than males' scores with effect sizes ranging from 0.17 to 0.20. The effect sizes are interpreted as small across the years (Cohen, 1988).

[Table 2 about here]

Female fourth-grade students' scores on the Reading to Gain Information increased by five points from 2005 to 2009 (216-221) (Table 2). No increase was observed from 2009 to 2013. Male fourth-grade students' scores increased by three points from 2005 to 2007 (212-215) and by one point from 2009 to 2011. (216). Females' scores were higher than male scores with small effect sizes ranging from d=0.11 to 0.16.

[Table 3 about here]

Both fourth-grade female students' scores (224-227) and male students' scores (216-219) on the Literary Experience Scale increased by three points across the years 2005 to 2013 (Table 3). Female students' scores were higher with effect sizes ranging from *d*=0.21 to *d*=0.25.

[Table 4 about here]

On the Reading Composite, both eighth-grade female students' scores (266-271) and eighth-grade male students' scores (255-261) increased from 2005 to 2013 (Table 4). Female students' scores were higher than male students' scores for every year for the Composite Scale. The effect sizes for gender differences ranged from d=0.26 to d=0.32. The 2013 difference in female and male average scale score for fourth-grade students was 7 points (Table 1). The 2013 difference in female and male average score for eighth-grade students for the same scale was 10 points.

[Table 5 about here]

Both eighth-grade female students' scores (266-272) and male students' scores (257-264) on the Reading to Gain Information scale increased from 2005-2013 (Table 5). Female students' scores were higher across the years with effect sizes ranging from *d*=0.23 to *d*=0.25.

[Table 6 about here]

On the Literary Experience reading scale, both female students' scores (265-270) and male students' scores (254-258) increased from 2005 to 2013. Female students' scores were higher than male students' scores for each assessment across the five NAEP assessments during this period (Table 6). Effect sizes ranged from small to moderate (*d*=0.27 to *d*=0.33).

[Table 7 about here]

The National Assessment of Educational Progress (NAEP) assessments are taken by twelfth-grade students every four years. Data from the three assessments in 2005, 2009, and 2013 are presented in Tables 7-9

On the Reading Composite scale for grade 12 (Table 7), female students' scores increased by two points from 2005-2009 (291-291) then decreased by one point from 2009-2013 (293-292). Male students' scores increased by four points (278-282) from 2005-2013. In each assessment year, female students' scores were higher than male students' scores with effect sizes ranging from *d*=0.26 to *d*=0.35.

The comparison with effect size differences by gender for eighth-grade student scores for the Composite Reading Scale (Table 6) across the same time period are: 2005 8th-grade d=0.32; 12th grade d=0.35; 2009 8th grade d=0.29; 12th grade d=0.32; and 2013 8th grade d=0.29; 12th grade d=0.26.

[Table 8 about here]

The Gain Information Scale was revised for Grade 12 after 2005 (Table 8). The new subscale was re-named and the ability to make comparisons was not clear—thus, no data were available for this scale for 2009 and 2013. Female 12th-grade students' scores were higher than 12th grade male students' scores on this scale in 2005; the effect size is moderate *d*=0.29.

[Table 9 about here]

Twelfth-grade Literary Experience scores for female students' were higher than those of male students' (Table 9). Female students' average scale score in 2005 (285) increased three points (288) in 2009, then decreased three points (285) in 2013. Male students' average scale score in 2005 (269) increased three points in 2009 (272) and had no change in 2013. Effect sizes across the eight-year period ranged from *d*=0.28 in 2013 to *d*=0.34 in 2005.

Extended Data Tables with Data from Klecker (2006) NAEP Reading by Gender 1992-2003

Tables 10 through 12 below present data from a previous study by the author with the summarized data from the current study. In all analyses, average scale scores of female students' were higher than average scale scores of male students. Comparisons can be made using effect sizes. Table 10 presents fourth-grade Composite Scale data; Table 11 presents eighth-grade Composite Scale data; and Table 12 presents twelfth-grade Composite Scale data across the years 1992 through 2013.

[Table 10 about here]

[Table 11 about here]

.[Table 12 about here]

Discussion

Limitations of Correlational Research

Educational researchers have long been aware of the pitfalls of correlational studies; still the methodology continues to be popular and useful. Correlational studies cannot show cause and effect, but they can present research evidence that indicates areas for further, more controlled, in-depth studies. In gender studies, descriptive and correlational studies are all that are possible. The experimental or quasi-experimental design required to make causal statements is obviously not possible with "status" variables such as gender or socio-economic status.

What do gender differences in reading assessment scores across grade-levels, geography, and time mean? The results of these study do not mean that all girls outscore all boys nor can the results be generalized to any one girl or boy from the population. Nor do they mean that boys cannot read. It cannot be concluded that boys had different 'learning styles' or that the content of

the material on the assessment was not of interest to boys. The statistically differences are based on group mean differences with overlapping distributions of scores.

Effect sizes (measured by Cohen's *d* in this study) (Cohen, 1988) ranged from small to moderate. In the NAEP data, there is more variance in reading scores within gender groups than between gender groups. The effect sizes for the group differences examined for the required NCLB (2002) data disaggregation are all larger than those in this study. These groups are: (1) economically disadvantaged; (2) special education; (2) Limited English Proficient (LEP) students (also known as ELL-English Language Learners); and (3) students from major racial/ethnic groups.

Examining Possible Causes for Differences

Brookhart (2006) systematically examined possible causes for gender differences in assessments. Some areas included in this examination were: included (1) differences in assessment development, (2) choice of test content, (3) test-takers' behavior, and (4) scoring differences (rater effects). In summary, she stated:

> …As an educator, I believe that relative comparisons ('Who outscored whom?') are less important than change over time ('What progress is being made?'). I also believe that relative comparisons are less important than descriptions of performance capabilities: the answers to the question, 'Who is better, boys or girls?' is less important than the answer to 'What can boys and girls do now?' and 'What else could they be expected to do next?' Relative comparisons are not as useful for making instruction improvements as information about progress and performance. (p. 126)

The NAEP data in Tables 10, 11, and 12 depicting fourth-, eighth-, and twelfth-grade

NAEP reading achievement data by gender across 1992 to 2013 indicate that, with a very few exceptions, each year the means for both boys and girls were slightly higher than they were for the previous year.

Conclusions and Future Research

The NAEP data are a valuable research resource for educational researchers. However, waiting until fourth grade to measure reading achievement and NCLB (2002) defined and gender gaps may provide accountability data, however, the data come very late in a child's life for intervention planning. Chatterji (2006) examined reading achievement of 2,296 students in attending 184 schools in the Early Childhood Longitudinal (ECLS) kindergarten to first-grade sample using hierarchical linear models. Chatterji found:

> With child-level background differences controlled, significant 1st-grade reading differentials were found in African American children (_0.51 *SD* units below Whites), boys (_0.31 *SD* units below girls), and children from high-poverty households (_0.61 to _1.0 *SD* units below well-to-do children). In all 3 comparisons, the size of the reading gaps increased from kindergarten entry to 1st grade. Reading level at kindergarten entry was a significant child-level correlate, related to poverty status. At the school level, class size and elementary teacher certification rate were significant reading correlates in 1st grade. Cross-level interactions indicated reading achievement in African children was moderated by the schools students attended, with attendance rates and reading time at home explaining the variance. (p. 489)

The analyses of data from NAEP reading assessments--The Nation's Report Card, (NCES, 2014a)—in the fourth, eighth, and twelfth grades reflect the continuation of disparities in literacy that begin very early in the lives of children, The disaggregation of data throughout the

school years provides a continued focus on the need to provide a rich literacy education for all. The “gaps” in the national reading data by gender and NCLB (2002) categories across years and across grade levels clearly indicate that early literacy efforts need to be strengthened at the local, state, and national levels. Providing early reading education for all children through literacy-rich childcare and preschool is an excellent first step. Continuing education programs for adult literacy provide adults with the tools needed for life-long learning and teaching.

References

Baer, J., Baldi, S., Ayotte, K., and Green, P. (2007). *The reading literacy of U.S. fourth-grade students in an international context: Results from the 2001 and 2006 Progress in International Reading Literacy Study (PIRLS)* (NCES 2008–017). National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education. Washington, DC. Retrieved from http://nces.ed.gov/pubsearch/pubsinfo.asp?pubid=2008017

Brookhart, S. M. (2009). Assessment, gender, and in/equity. In C.Wyatt-Smith & J.J.Cumming (Eds.), (pp. 119-136). *Educational assessment in the 21st century*. Springer Science and Business Media. DOI 10 1007/978-1-4020-9964-7

Chatterji, M. (2006). Reading achievement gaps, correlates, and moderators of early reaching achievement: Evidence form the Early Childhood Longitudinal Study (ECLS) Kindergarten to first grade sample. *Journal of Educational Psychology, 98* (3) 489-507. DOI: 10.1037/0022-0663.98.3.489

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum.

Klecker, B.M. (2006). The gender gap in NAEP fourth-, eighth-, and twelfth-grade reading scores across years. *Reading Improvement, 43,* 50-56.

Leitz, P. (2006). A meta-analysis of gender differences in reading achievement at the secondary school level. *Studies in Educational Evaluation, 32,* 317-344.

Mullis, I.V.S., Martin, M.O., Foy, P., & Drucker, K.T. (2012). *The PIRLS 2011 International Results in Reading: Executive Summary.* TIMMS & PIRLS International Study Center. Boston College, Chestnut Hill, MA. Retrieved from

http://timssandpirls.bc.edu/pirls2011/downloads/P11_IR_Executive%20Summary.pdf

National Center for Educational Statistics (NCES). (2014a). *Measuring student progress since 1964*. Retrieved from http://nces.ed.gov/nationsreportcard/about/naephistory.aspx

NCES (2014b). *NAEP Data Explorer* Retrieved from http://nces.ed.gov/nationsreportcard/naepdata/dataset.aspx

NCES (2014c). *Progress in International Reading Literacy* (*PIRLS).* Retrieved from http://nces.ed.gov/surveys/pirls/

NCES (2014d). *Program for International Student Assessment (PISA)*. Retrieved from http://nces.ed.gov/surveys/pisa/

NCES (2014e). *NAEP assessment sample design 4th and 8th grades*. Retrieved from http://nces.ed.gov/nationsreportcard/tdw/sample_design/

NCES (2014f). *NAEP assessment sample design 12th grade*. Retrieved from http://www.nationsreportcard.gov/reading_math_g12_2013/#/about#naep_samples

No Child Left Behind (NCLB) (2002). Act of 2001, Pub. L. No. 107-110, § 115, Stat. 1425 (2002) Retrieved from http://www2.ed.gov/policy/elsec/leg/esea02/107-110.pdf

Organization for Economic Co-operation and Development (OECD). (2014). *About PISA*. Retrieved from http://www.oecd.org/pisa/aboutpisa/

OECD (2010), *PISA 2009 results: Executive summary*. Retrieved from http://www.oecd.org/pisa/pisaproducts/46619703.pdf

Ogle, L. T.; Sen, A.; Pahlke, E.; Jocelyn, L.; Kastberg, D.; Roey, S.; Williams, T. (April, 2003). *International comparisons in fourth-grade reading literacy: Findings from the Progress in International Reading Literacy Study (PIRLS) of 2001*, National Center for Education Statistics 2003073 Retrieved from http://nces.ed.gov/pubsearch/pubsinfo.asp?pubid=2003073

Skelton, C., & Francis, B. (2011). Successful boys and literacy: Are "literate boys" challenging or repackaging hegemonic masculinity? *Curriculum Inquiry, 41*, (4) 456-479. DOI: 10.1111/j.1467-873X.2011.00559.x

Table 1. NAEP Fourth-Grade Reading Composite Average Scores by Gender by Year

	Female		Male			
Year	Average Scale Score	SD	Average Scale Score	SD	*p* value	Effect Size
2013	224	36	217	38	*p<.001*	*d*=0.19
2011	223	35	217	37	*p<.001*	*d*=0.19
2009	223	35	216	36	*p*.<.001	*d*=0.20
2007	223	35	216	36	*p<.001*	*d*=0.20
2005	220	36	214	36	*p*.<.001	*d*=0.17

Table 2. NAEP 4th- Grade Reading to Gain Information Scores by Gender by Year

	Female		Male			
Year	Average Scale Score	SD	Average Scale Score	SD	*p* value	Effect Size
2013	221	37	216	39	*p*.<.001	*d*=0.13
2011	221	36	216	38	*p*.<.001	*d*=0.14
2009	221	37	215	38	*p*.<.001	*d*=0.16
2007	220	36	215	38	*p*.<.001	*d*=0.14
2005	216	37	212	38	*p*.<.001	*d*=0.11

Table 3. NAEP 4th- Grade Literary Experience Average Scale Scores by Gender by Year

	Female		Male			
Year	Average Scale Score	SD	Average Scale Score	SD	*p* value	Effect Size
2013	227	37	219	39	*p*.<.001	*d*=0.21
2011	226	36	218	38	*p*.<.001	*d*=0.22
2009	225	36	217	37	*p*.<.001	*d*=0.22
2007	226	36	217	37	*p*.<.001	*d*=0.25
2005	224	37	216	37	*p*.<.001	*d*=0.25

Table 4. NAEP 8th-Grade Reading Composite Average Scores by Gender by Year

	Female		Male			
Year	Average Scale Score	SD	Average Scale Score	SD	*p* value	Effect Size
2013	271	34	261	34	*p*.<.001	*d*=0.29
2011	268	33	259	34	*p*.<.001	*d*=0.27
2009	267	33	258	35	*p*.<.001	*d*=0.26
2007	266	34	256	35	*p*.<.001	*d*=0.29
2005	266	34	255	35	*p*.<.001	*d*=0.32

Table 5. NAEP 8th- Grade Reading to Gain Information Scores by Gender by Year

	Female		Male			
Year	Average Scale Score	SD	Average Scale Score	SD	*p* value	Effect Size
2013	272	34	264	35	*p*.<.001	*d*=0.23
2011	269	34	261	35	*p*.<.001	*d*=0.23
2009	268	35	260	36	*p*.<.001	*d*=0.23
2007	267	35	258	37	*p*.<.001	*d*=0.25
2005	266	35	257	37	*p*.<.001	*d*=0.25

Table 6. NAEP 8th- Grade Literary Experience Average Scores by Gender by Year

	Female		Male			
Year	Average Scale Score	SD	Average Scale Score	SD	*p* value	Effect Size
2013	270	36	258	37	*p*.<.001	*d*=0.33
2011	267	35	256	37	*p*.<.001	*d*=0.31
2009	266	36	256	37	*p*.<.001	*d*=0.27
2007	265	36	255	37	*p*.<.001	*d*=0.27
2005	265	36	254	37	*p*.<.001	*d*=0.30

Table 7. NAEP 12th-Grade Reading Composite Average Scale Scores by Gender by Year

	Female		Male			
Year	Average Scale Score	SD	Average Scale Score	SD	*p* value	Effect Size
2013	292	37	282	39	*p*.<.001	*d*=0.26
2009	293	36	281	39	*p*.<.001	*d*=0.32
2005	291	37	278	38	*p*.<.001	*d*=0.35

Table 8. NAEP Twelfth-Grade Gain Information Average Scores by Gender by Year

	Female		Male			
Year	Average Scale Score	SD	Average Scale Score	SD	*p* value	Effect Size
2013	———	———	———	———	———	———
2009	———	———	———	———	———	———
2005	267	33	258	35	*p*.<.001	*d*=0.29

Table 9. NAEP Twelfth-Grade Literary Experience Average Scores by Gender by Year

	Female		Male			
Year	Average Scale Score	SD	Average Scale Score	SD	*p* value	Effect Size
2013	285	46	272	48	*p*.<.001	*d*=0.28
2009	288	47	272	50	*p*.<.001	*d*=0.33
2005	285	47	269	48	*p*.<.001	*d*=0.34

Table 10. NAEP 4th-Grade Reading Composite Scores by Gender Years 1992-2013

	Female		Male			
Year	Average Scale Score	SD	Average Scale Score	SD	*p* value	Effect Size
2013	**224**	**36**	**217**	**38**	***p*<.001**	***d*=0.19**
2011	**223**	**35**	**217**	**37**	***p*<.001**	***d*=0.19**
2009	**223**	**35**	**216**	**36**	***p*.<.001**	***d*=0.20**
2007	**223**	**35**	**216**	**36**	***p*<.001**	***d*=0.20**
2005	**220**	**36**	**214**	**36**	***p*.<.001**	***d*=0.17**
2003	220	36	213	38	*p*.<.001	*d*=0.19
2002	220	36	214	36	*p*.<.001	*d*=0.16
2000	217	40	206	43	*p*.<.001	d=0.26
1998	215	39	210	39	*p*.<.001	d=0.13
1994[n]	218	39	207	42	*p*.<.001	*d*=0.27
1992[n]	219	35	211	36	*p*.<.001	*d*=0.22

Note: [n] Accommodations were not permitted for this assessment
Source: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 2013, 2011, 2009, 2007, 2005, 2003, 2002, 2000, 1998, 1994, and 1992.
Data from 2013 analysis is in boldface type added to data table from Klecker (2006).

Table 11. NAEP 8th- Grade Reading Composite Scores by Gender Years 1992-2013

	Female		Male			
Year	Average Scale Score	SD	Average Scale Score	SD	*p* value	Effect Size
2013	**271**	**34**	**261**	**34**	***p*.<.001**	***d*=0.29**
2011	**268**	**33**	**259**	**34**	***p*.<.001**	***d*=0.27**
2009	**271**	**34**	**261**	**34**	***p*.<.001**	***d*=0.29**
2007	**266**	**34**	**256**	**35**	***p*.<.001**	***d*=0.29**
2005	**266**	**34**	**255**	**35**	***p*.<.001**	***d*=0.31**
2003	267	34	256	36	*p*.<.001	*d*=0.31
2002	267	33	258	36	*p*.<.001	*d*=0.27
1998	268	33	253	36	*p*.<.001	d=0.43
1994[n]	265	35	250	37	*p*.<.001	*d*=0.42
1992[n]	264	35	251	36	*p*.<.001	*d*=0.37

Note: 2000 Data not available for grades 8 and 12.

Note: [n] Accommodations were not permitted for this assessment

Source: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 2013, 2011, 2009, 2007, 2005, 2003, 2002, 1998, 1994, and 1992.

Data from 2013 analysis is in boldface type added to data table from Klecker (2006)

Table 12. NAEP 12th- Grade Reading Composite Scores by Gender Years 1992-2013

	Female		Male			
Year	Average Scale Score	SD	Average Scale Score	SD	*p* value	Effect Size
2013	**292**	**37**	**282**	**39**	***p*.<.001**	***d*=0.26**
2009	**293**	**36**	**281**	**39**	***p*.<.001**	***d*=0.32**
2005	**291**	**37**	**278**	**38**	***p*.<.001**	***d*=0.35**
2002	293	36	277	37	*p*.<.001	*d*=0.44
1998	292	36	280	39	*p*.<.001	d=0.32
1994[n]	297	35	281	38	*p*.<.001	*d*=0.44
1992[n]	295	32	285	32	*p*.<.001	*d*=0.31

Note: 2000 Data not available for grades 8 and 12.

Note: [n] Accommodations were not permitted for this assessment

Source: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 2013, 2011, 2009, 2007, 2005, 2003, 2002, 1998, 1994, and 1992.

Data from 2013 analysis is in boldface type added to data table from Klecker (2006).