

Running head: THE GENDER GAP IN NAEP READING SCORES

The 'Gender Gap' in NAEP Fourth-, Eighth-, and Twelfth-Grade Reading Scores Across Years

Beverly M. Klecker

Morehead State University

Morehead, KY

b.klecker@morehead-st.edu

Paper presented at the annual meeting of
the Midwestern Association of Educational Research

Columbus, OH

October 15, 2005

Abstract

This paper presents a secondary analysis of the National Assessment of Educational Progress (NAEP) dataset. The study explores differences in the NAEP fourth, eighth, and twelfth grade reading scores by students' gender across the years 1992, 1994, 1998, 2000, 2002, and 2003. The study used the NAEP National Public School data. The statistically significant ($p < .01$ with effect size measured by Cohen's d) differences in reading scores by gender were consistent across grade level and years with females scoring higher than males. A discussion of the calculation and reporting of effect size with NAEP data is included as well as implications for the No Child Left Behind goals of "closing the gap." This paper presents the argument that the 'child left behind' in reading is very likely to be male--from elementary school through university.

The 'Gender Gap' in NAEP Fourth and Eighth Grade Reading Scores Across Years

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.

Research findings across time and cultures strongly support the positive correlation between student gender and reading achievement. Although the No Child Left Behind (NCLB) legislation mandated a strong focus on reading achievement in early elementary, the legislation *did not require disaggregation of school accountability test results by gender* [author's emphasis] (White House, 2001).

NCLB does, however, require state participation in the National Assessment of Educational Progress (NAEP) by any state wishing to receive Title I funding (NAEP, 2005a). The NAEP results for reading and math are reported for grades four, eight, and twelve. The NAEP results *are* disaggregated by gender in both the national data and the state data.

Background for the Study

There is an extensive body of research literature examining the relationship between gender and reading achievement. Recent studies (e.g., Cloer & Dalton, 2001; Lynch, 2002) reported that females consistently scored higher than males. Bond and Dykstra (1997) presented a meta-analysis of literature (e.g., Ballou, 1963; Carroll, 1948; Gates, 1960; Pauley, 1951; Waejen & Gramilis, 1963; cited in Bond & Dykstra, 1997) that supported the consistency of higher reading achievement in females.

Freedmon (2003) reported similar findings from her Canadian research:

The gendered results of boys in reading and writing can be seen in the achievement results of the Ontario Secondary School Literacy Test (OSSLT)

....In 2002, on the Grade 10 test, 55% of boys passed reading and writing, compared to 70% of girls...(p. 2).

Topping, Valtin, Roller, Brozo, and Dionisio (2004) studied fifteen-year-old students across 32 countries and suggested:

Schools should also consider their methods of reading instruction, to ensure that implicit cultural or gender bias are not present. Females outperformed males on the combined literacy scale in all participating countries...Females were more reflective and evaluative in their approach to reading and spent much more time reading for enjoyment than did males (p. 7)

The National Assessment of Educational Progress (NAEP)

The National Assessment of Educational Progress (NAEP) has since 1969, been the only nationally representative and continuing assessment of what America's students know in various subject areas. Demographic and questionnaire data were collected as the NAEP was administered (2005b). Students self-reported their gender.

What Does the NAEP Reading Assessment Measure?

The National Center for Educational Statistics (2005c) presented the following information on the content validity of the NAEP Reading Assessment:

NAEP measures the reading comprehension of fourth-, eighth-, and twelfth-grade students. In 2002, the reading framework was updated to provide more explicit detail about the assessment design and content. During that process, some of the terms used to describe elements of the reading assessment were changed.

The following description of the reading framework incorporates these changes. It should be noted, however, that the updating of the framework does not represent a

change in the design or content of the NAEP reading assessment that was first administered in 1992.

According to the framework, developed by the National Assessment Governing Board (NAGB), NAEP assesses three contexts for reading. In addition to reading within different contexts, NAEP reading comprehension questions are developed to engage the different approaches that readers may take in the process of trying to understand what is being read.

Method

Procedure

NAEP sampling and data collection

Sampling for the reading assessment used a multistage sampling design that sampled students from selected schools within selected geographic areas across the country. The National Center for Educational Statistics (2005d) described sampling and data collection:

The sample design had the following stages:

1. selection of geographic areas (a county, group of counties, or metropolitan statistical area),
2. selection of schools (public and nonpublic) within the selected areas, and
3. random selection of students within the selected schools.

Each selected school that participated in the assessment and each student assessed represents a portion of the population of interest. Therefore, sampling weights are needed to make valid inferences between the student samples and the respective populations from which they were drawn. Sampling weights adjust for disproportionate representation due to such oversampling. State and national

samples are drawn in the same way in odd-numbered years. In even-numbered years, national samples are drawn using the three-stage method.

Data analysis.

The NAEP Data Tool (National Center for Educational Statistics, 2005e) was used to create data tables from the fourth and eighth grade national public schools reading scores for the years 1992, 1994, 1998, 2000, 2002, and 2003 by gender (note, complete data were not available for every year). Alpha was set a priori at .01 and effect size, d (Cohen, 1992), was calculated for each statistically significant difference.

Results

Table 1 presents the differences in NAEP fourth-grade reading scores by gender across the years 2003, 2002, 2000, 1998, 1994, and 1992. In years 1994 and 1992 accommodations

[place Table 1 about here]

were not permitted for the assessment. It is not surprising to find that the observed mean differences in the scale scores were found to be statistically significantly different. NAEP samples thousands of students at each grade level each year. Effect sizes range from $d=.27$ for 1994 to $d=.13$ in 1998. The effect sizes are small (Cohen, 1997).

Table 2. presents the differences in NAEP eighth-grade reading scores by gender across

[place Table 2 about here]

the years 2003, 2002, 1998, 1994, and 1992. Again, accommodations were not permitted in either 1994, or 1992. The differences in mean scale scores by gender are statistically significant ($p<.001$). The effect sizes range from a low of $d=.27$ in 2002 to a high of $d=.43$ in 1998. The effect sizes are larger in the 8th grade data than in the 4th grade data. Cohen (1997) stated that effect sizes of $d=.50$ could be interpreted as moderate.

Table 3. presents the differences in NAEP twelfth-grade reading scores by gender across

the years 2002, 1998, 1994, and 1992. Accommodations were not permitted in either 1992 or 1994. There were statistically significant ($p < .001$) differences between mean scale scores by gender and the effect sizes ranged from a low of $d = .22$ to a high of $d = .44$.

There were consistent, statistically significant ($p < .001$) differences in the NAEP reading scores by gender across grade level (4th, 8th, and 12th) and years. Effect sizes increased from small to low moderate as data grade level increased from 4th to 8th to 12th grade. That is, as measured by effect size, differences by gender in the NAEP reading scores in the 12th grade were larger than differences in reading scores by gender in the 4th grade. The consistency of the findings in these data is remarkable.

Additionally, state level data for 4th and 8th grade NAEP reading scores are presented in the Appendix. These data further indicate the consistency of the findings across years. [Note: state data were not available for 12th grade scores.]

Conclusions and Suggestions for Further Research

This study suggests that school improvement efforts, including NCLB, should be taking a more careful look at males and reading across grades P-12. Only by requiring the disaggregation of data by gender within schools and districts (suggest amending NCLB requirements), can we begin to look at the problem in a meaningful way.

Some researchers have looked beyond correlations to examine the problem. Three varied and intriguing ideas for further research are presented below.

Freedman (2003) conducted semi-structured focus groups with boys in grades four and six. Although this study was limited to five volunteer boys in each of six schools ($N=30$), the depth of the focused interview results are very informative. This qualitative research methodology with varied samples would work well in specific schools or districts.

Johnson and Newton (2003) in their review of literature, suggested that one of the effects that colleges are now seeing is a decreasing number of male students meeting the college acceptance criteria. These authors cited Kleinfeld's (1998 cited in Johnson & Newton, 2003), statement that in some liberal arts college administrators have developed affirmative action programs for males by lowering the grade and test score requirements for them.

Li, Cohen, and Ibarra (2004) examined gender differences on a mathematics test by combining a DIF, differential item functioning, study by gender with an examination of item structural characteristics related to cognitive functions. This research would lead to a close examination of the structure of the test items. These researchers found item types that male students more frequently answered correctly, and item types that female students more frequently answered correctly. The researchers at NAEP have undoubtedly performed DIF analysis, a rather standard psychometric study, but research similar to the one described in this study, or 'think aloud' protocols would aid in understanding the measurement of 'reading.'

References

- Bond, G. L., & Dykstra, R. (1997). The cooperative research program in first-grade reading instruction. *Reading Research Quarterly, 32*, (4), 348-427.
- Cloer, T., & Dalton, S. (2001). Gender and grade differences in reading achievement and self-concept as readers. *Journal of Reading Achievement, 26*, 31-36.
- Cohen, J. (1992) A power primer. *Psychological Bulletin, 112*, 155-159.
- Freedman, B. (2003). *Boys and literacy: Why boys? Which boys? Why now?* Paper presented at the annual meeting of the American Educational Research Association, Chicago, IL (ERIC Reproduction Service Document No. ED477857)
- Johnson, R., & Newton, R.M. (2003). *What about the other gender? Male bias in English studies*. EDRS Research Report (ERIC Reproduction Service Document No. ED480870)
- Li, Y., Cohen, A.S., Ibarra, R.A. (2004). Characteristics of mathematics items associated with gender DIF. *International Journal of Testing, 4*, (2), 115-136.
- Lynch, J. (2002). Parents' self-efficacy beliefs, parents' gender, children's reader self-perceptions, reading achievement, and gender. *Journal in Reading, 25*, 54-67.
- National Center for Educational Statistics (2005a). *NAEP and No Child Left Behind*. Retrieved September 9, 2005 from <http://nces.ed.gov/nationsreportcard/nclb.asp>
- National Center for Educational Statistics (2005b) *The nation's report card: Reading* Retrieved September 9, 2005 from <http://nces.ed.gov/nationsreportcard/reading/>
- National Center for Educational Statistics (2005c) *The NAEP reading assessment*. Retrieved September 9, 2005 from <http://nces.ed.gov/nationsreportcard/reading/whatmeasure.asp>

National Center for Educational Statistics (2005d) The NAEP reading assessment sample.

Retrieved September 9, 2005

<http://nces.ed.gov/nationsreportcard/reading/sampledesign.asp>

National Center for Educational Statistics (2005e) The NAEP data tool. Retrieved July 10, 2005

from <http://nces.ed.gov/nationsreportcard/naepdata/>

Topping, K., Vaitin, R., Roller, C., Brozo, W., & Dionisio, M.L. (April, 2003). Policy and practice implications of the program for international student assessment (PISA) 2000.

Report of the International Reading Association PISA Task Force. Newark, DE:

International Reading Association (ERIC Reproduction Service Document No.

ED478184)

Table 1.

Differences in NAEP Fourth Grade Reading Scores by Gender Across Years

Year	Female		Male		<i>p. value</i>	<i>Effect Size</i> <i>Cohen's d</i>
	Average Scale Score	SD	Average Scale Score	SD		
2003	220	36	213	38	<i>p</i> < .001	<i>d</i> = .19
2002	220	36	214	36	<i>p</i> < .001	<i>d</i> = .16
2000	217	40	206	43	<i>p</i> < .001	<i>d</i> = .26
1998	215	39	210	39	<i>p</i> < .001	<i>d</i> = .13
1994 ⁿ	218	39	207	42	<i>p</i> < .001	<i>d</i> = .27
1992 ⁿ	219	35	211	36	<i>p</i> < .001	<i>d</i> = .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), 2003, 2002, 2000, 1998, 1994, and 1992.

Table 2.

Differences in NAEP Eighth Grade Reading Scores by Gender Across Years

Year	Female		Male		<i>p. value</i>	<i>Effect Size</i>
	Average Scale Score	SD	Average Scale Score	SD		
2003	267	34	256	36	<i>p</i> < .001	<i>d</i> = .31
2002	267	33	258	34	<i>p</i> < .001	<i>d</i> = .27
1998	268	33	253	36	<i>p</i> < .001	<i>d</i> = .43
1994 ⁿ	265	35	250	37	<i>p</i> < .001	<i>d</i> = .42
1992 ⁿ	264	35	251	36	<i>p</i> < .001	<i>d</i> = .37

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), 2003, 2002, 1998, 1994, and 1992.

Table 3.

Differences in NAEP Twelfth Grade Reading Scores by Gender Across Years

Year	Female		Male		<i>p. value</i>	<i>Effect Size</i> <i>Cohen's d</i>
	Average Scale Score	SD	Average Scale Score	SD		
2002	293	37	277	36	<i>p</i> < .001	<i>d</i> = .44
1998	297	36	280	39	<i>p</i> < .001	<i>d</i> = .44
1994 ⁿ	293	36	279	36	<i>p</i> < .001	<i>d</i> = .39
1992 ⁿ	219	35	211	36	<i>p</i> < .001	<i>d</i> = .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) 2002, 1998, 1994, and 1992.

APPENDIX

NAEP 4th and 8th Grade Reading

Gender Gap by State

<http://nces.ed.gov/nationsreportcard/reading/results2003/stategendergaps-4g.asp>

<http://nces.ed.gov/nationsreportcard/reading/results2003/stategendergaps-8g.asp>

[note hard copy of above links will be in presented paper]