



NAEP Mathematics and Reading “Effect Size” Results for Nine Idaho Demographic Groups, Grades 4 & 8, 2009-2017

Bert D. Stoneberg
 K-12 Research Idaho

Learning Point Associates (2009) has suggested that effect size statistics can provide valuable information for program evaluation and accountability systems. Effect sizes focus on the magnitude of the difference (or practical significance) between two means. This paper illustrates how effect sizes can provide us with useful descriptions of achievement in Idaho for a given demographic student group from year to year, in a given subject, and in a given grade (group/subject/grade).

Effect Size vs. Statistical Significance

“What is the relationship between *effect size* and *significance*? Effect size quantifies the size of the difference between two groups, and may therefore be said to be a true measure of the significance of the difference [i.e., the practical significance].... However, in statistics the word *significance* is often used to mean *statistical significance*, which is the likelihood that the difference between the two groups could just be an accident of sampling. If you take two samples from the same population there will always be a difference between them. The statistical significance is usually calculated as a *p-value*, the probability that a difference of at least the same size would have arisen by chance, even if there really were no difference between the two populations. For differences between the means of two groups, this p-value would normally be calculated from a *t-test*. By convention, if $p < 0.05$ (i.e. below 5%), the difference is taken to be large enough to be *significant*; if not, then it is *not significant*. There are a number of problems with using *significance tests* in this way.... The main one is that the p-value depends essentially on two things: the size of the effect and the size of the sample. One would get a [statistically] *significant* result either if the effect were very big (despite having only a small sample) or if the sample were very big (even if the actual effect size were tiny)....” (Coe, 2002)

Effect Size: Standardized Group-mean Differences

“Different measures for effect size have been developed over the decades.... Because the terminology used for describing the variety of effect-size measure has not been standardized in the literature, confusion sometimes occurs about what effect-size measure has been reported in a study.... The first category, measure of effect size, is based on standardized group-mean differences. In the most general form, *d* is expressed as follows:

$$d = \frac{\bar{X}_{\text{group 1}} - \bar{X}_{\text{group 2}}}{SD_{\text{pooled}}}$$

where SD_{pooled} represents the pooled standard deviation between the two groups.” (Fax, 2001)

Simple Effect Size

The computation formula for effect size used for this paper is the most general form, a simple effect size:

Simple Effect Size

$$\frac{\bar{X}_2 - \bar{X}_1}{\sqrt{\frac{s_1^2 + s_2^2}{2}}}$$

**Difference
Between Means**

**Pooled
Standard
Deviation**

The subscript “2” for a mean and standard deviation refers to “this year” and the subscript “1” for a mean and standard deviation refers to “last year.” When this year’s mean (“2”) is larger than last year’s mean (“1”), the effect size is positive. When last year’s mean (“1”) is larger, the effect size is negative. When the two means are equal, the effect size is zero (0), meaning achievement this year is the same as last year.

NAEP Data – Group Averages (Means) and Standard Deviations

Idaho has participated in the National Assessment of Educational Assessment (NAEP), which reported mathematics and reading average scale scores and standard deviations for grades 4 and 8 in odd-numbered years (e.g., 2009, 2011, 2013, 2015, and 2017). NAEP enables us to understand how Idaho students are performing compared to their peers in other states and in the nation’s public schools. NAEP effect sizes in this paper, however, compare Idaho students only with other Idaho students in the same group/grade/subject from year to year.

Results

Table 1 summarizes each group/grade/subject effect size in 2017 compared to its zero (0) effect size in 2009. Narratives for Table 1 include:

Positive and Negative Group Trend Effects (2009 to 2017) by Grade and Subject

- In grade 4 reading, six (6) student groups had a positive trend effect from 2009 to 2017, while three (3) student groups had a negative trend effect.
- In grade 4 mathematics, only one (1) student group had a positive trend effect from 2009 to 2017, while eight (8) student groups had a negative trend effect.

- In grade 8 reading, eight (8) student groups had a positive trend effect from 2009 to 2017, while only one (1) student group had a negative trend effect.
- In grade 8 mathematics, two (2) student groups had a positive trend effect from 2009 to 2017, while seven (7) student groups had a negative trend effect.

Student Group Trend Effects in both Reading and Mathematics

- Three student groups had a positive effect size from 2009 to 2017 in both reading and mathematics: White fourth graders, Hispanic eighth graders, and eighth graders attending suburban schools.
- Four student groups had a negative effect size from 2009 to 2017 in both reading and mathematics: Female fourth graders, fourth graders attending suburban schools, fourth graders attending town schools, and female eighth graders.

Table 1. Trend effect size in 2017 compared to zero (0) in 2009, described as higher (up arrow), same, or lower (down arrow).

Student Group	Grade 4		Grade 8	
	Reading	Mathematics	Reading	Mathematics
All	↑	↓	↑	↓
Hispanic	↑	↓	↑	↑
White	↑	↑	↑	↓
Female	↓	↓	↓	↓
Male	↑	↓	↑	↓
City	↑	↓	↑	↓
Suburb	↓	↓	↑	↑
Town	↓	↓	↑	↓
Rural	↑	↓	↑	↓

Legend: ↑ ♦ ↓

Table 1 summarizes the trend results from thirty-six (36) graphs for reading and mathematics in grades 4 and 8, over nine student groups (e.g., All students, Hispanic students, White students, Female students, Male students, City schools, Suburban schools, Town schools, and Rural schools) from NAEP 2011 to 2017. Table 1 does not indicate the magnitude of an effect size, only whether it was positive or negative.

Understanding the Effect Size Graph

The 36 effect size graphs, which also display the four bi-annual effect sizes of the group/subject/grade for each year from 2011 to 2017, are in Appendix A (page 7). Figure 1 is a sample graph that displays group/subject/grade bi-annual and trend effect sizes from 2009 to 2017. Following it are several points if which one should be aware when reading and interpreting the effect size graphs.

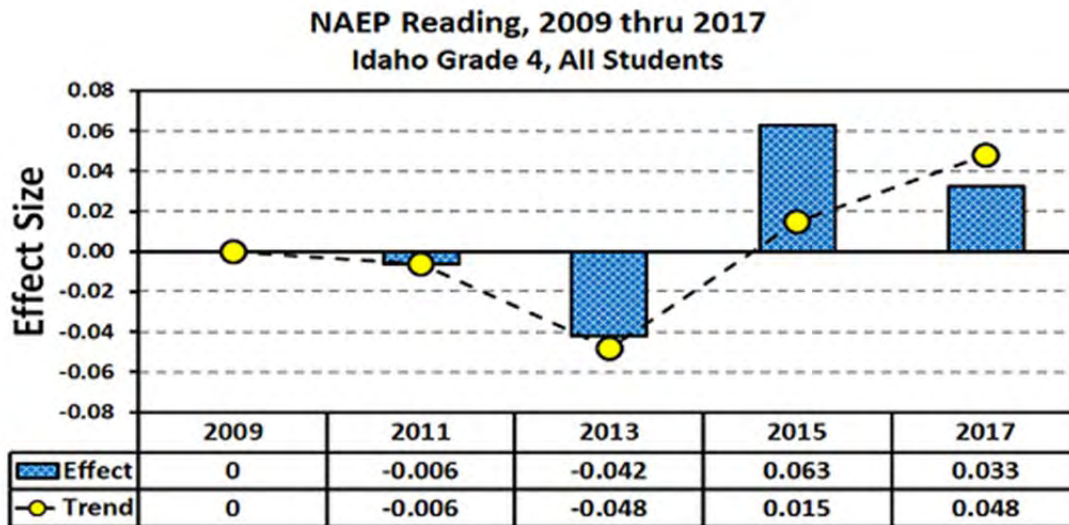


Figure 1. Effect size results for all students in Idaho fourth graders on the NAEP reading assessments from 2009 to 2017.

The State NAEP is given nationwide in odd numbered years beginning in the last week of January, ending in the week of March. Thus, the 2011 bar (“blue” for reading) reflects the impact of the instructional program from February 2009 to February 2011. Likewise, the 2015 bar represents the impact of the instructional program from February 2013 to February 2015, and so on.

The effect size can be positive, this year (posttest) higher than last year (pretest), negative (posttest lower than pretest), or zero (posttest and pretest have the same average scale score).

The size of the bar indicates the magnitude of the change. For example, on the chart above, the reading loss in 2013 was larger than the reading loss in 2011. Moreover, the reading gain in 2015 was larger than the combined losses of 2011 and 2013.

The trend effect size is defined as "zero" in 2009 (indicated by the yellow dots and the dashed line). The magnitude of effect size from each test administration is added to the magnitude of the existing trend (i.e., trend effect size + new effect size = updated trend).

CAUTION, CAUTION, CAUTION. Limit comparisons to the effect sizes exhibited on the same graph. Comparisons of the information on different graphs should be avoided beyond general observations such as both are positive, or one is positive and one is negative, etc. There are three good reasons to avoid comparing the effect sizes from two graphs:

- Reading scores are on a different scale than mathematics scores.
- The instructional objective is to take a group as it was in 2009 and to improve its performance (as measured by effect size). There is no assurance that the starting scale scores will be the same for two groups. One third-grade group may start with a high reading score, while another starts with a low reading score. It is not impossible that the group starting with the high score may have a negative effect size each test administration, and still have a higher reading score in 2017 than the group starting with a low score that earns positive effect sizes each test administration.
- The range of observed effect size scores have been scaled to ensure that the effect size range fits the graph’s plot area for 2009 to 2017. Looking at two graphs you may think that one bar on graph 1 looks to be the same size as a bar on graph 2. While they may look the same, they most likely do not represent the same effect size.

Discussion

Each year Idaho spends a considerable sum to administer and score the Idaho Standards Achievement Tests (SBAC). Unfortunately, the state typically publishes no ISAT/SBAC results other than “percent proficient or advanced” statistics. The percent above cut-score metric such as those reported by ISAT/SBAC (and NAEP) are problematic, especially when looking at trends.

The National Academies of Sciences, Engineering, and Medicine (2017) conducted a congressionally mandated evaluation of the NAEP achievement levels to determine whether they are reasonable, reliable, valid, and informative to the public, and to recommend ways that use of achievement levels can be improved. The National Academy’s report noted:

“One of the most common and unwarranted inferences using achievement level percentages involves assessing the amount of progress students have made over time, particularly by population groups. For instance, news reports often focus not only on how students are doing at a particular time, but the extent to which the percentage of students scoring Proficient or above has (or has not) improved over successive NAEP years. When these comparisons are based on the scale scores, they provide useful information. When they are based on the “percentage Proficient or above” metric and used to compare progress across groups, they can be misleading. A report by Holland on this issue focused on misinterpretations associated with using the “percent above a cut score” metric. Although this metric is widely used for NAEP, there are serious limitations to the inferences that the percent above cut score metric can support, particularly when evaluating trends over time, gaps among groups, or trends in gaps....” [*Emphasis added.*] (p. 208).

Effect sizes can be and should be used to provide useful achievement trend statistics for Idaho students that are technically preferred over percent above cut-score trend statistics. Idaho’s effect size results reported in this paper were based on NAEP scale scores from 2009 to 2017. Idaho’s effect size results have also been reported for ISAT/SBAC scale scores from 2015 to 2017 (Stoneberg, 2018).

#

References

Coe, R. (2002, September). *It's the Effect Size, Stupid. What effect size is and why it is important.* Paper presented at the Annual Conference of the British Educational Research Association, University of Exeter, England. Retrieved May 1, 2018, from <https://www.leeds.ac.uk/educol/documents/00002182.htm>

Fax, X. (2001, May/June). Effect size in education research: Two sides of a coin. *Journal of Educational Research*, 94(5). Available online at http://www.jstor.org/stable/27542335?seq=1#page_scan_tab_contents

Learning Point Associates. (2009, August). *Connecting Research to Practice: Knowing Who Is Proficient Isn't Always Sufficient.* Retrieved July 14, 2016, from <http://files.eric.ed.gov/fulltext/ED509965.pdf>

National Academies of Sciences, Engineering, and Medicine. (2017). Evaluation of the achievement levels for mathematics and reading on the National Assessment of Educational Progress. Washington, DC: The National Academies Press. Available online at <https://www.nap.edu/23409>

National Center for Education Statistics. *The Status of Achievement Levels.* Retrieved April 19, 2018, from https://nces.ed.gov/nationsreportcard/tdw/analysis/2000_2001/describing_achiev_developmental.aspx

Stoneberg, B.D. (2018). *Idaho Effect Size Results for the SBAC English Language Arts and Mathematics Tests, 2015-2017, Grades 3-8 and 10.* Available on the web at <http://k12researchidaho.com/pdf/isat2017effectsizeresults.pdf>

Suggested citation for this paper:

Stoneberg, B.D. (2018, May). *NAEP mathematics and reading "effect size" results for nine Idaho demographic groups, grades 4 & 8, 2009-2017.* Available online at <http://k12researchidaho.com/pdf/naep2017effectsizeresults.pdf>

Contact Bert Stoneberg with comments or questions:
Email to bert@k12researchidaho.com or
U.S. mail to P.O. Box 5912, Boise ID 83705

Appendix A

Idaho NAEP Effect Size Graphs, 2009 to 2017

Student Group	Grade 4	Grade 8
All Students	Reading & Mathematics Grade 4, Page 8	Reading & Mathematics Grade 8, Page 9
Hispanic Students	Reading & Mathematics Grade 4, Page 10	Reading & Mathematics Grade 8, Page 11
White Students	Reading & Mathematics Grade 4, Page 12	Reading & Mathematics Grade 8, Page 13
Female Students	Reading & Mathematics Grade 4, Page 14	Reading & Mathematics Grade 8, Page 15
Male Students	Reading & Mathematics Grade 4, Page 16	Reading & Mathematics Grade 8, Page 17
City Schools	Reading & Mathematics Grade 4, Page 18	Reading & Mathematics Grade 8, Page 19
Suburb Schools	Reading & Mathematics Grade 4, Page 20	Reading & Mathematics Grade 8, Page 21
Town Schools	Reading & Mathematics Grade 4, Page 22	Reading & Mathematics Grade 8, Page 23
Rural Schools	Reading & Mathematics Grade 4, Page 24	Reading & Mathematics Grade 8, Page 25

NAEP Reading, 2009 thru 2017 Idaho Grade 4, All Students

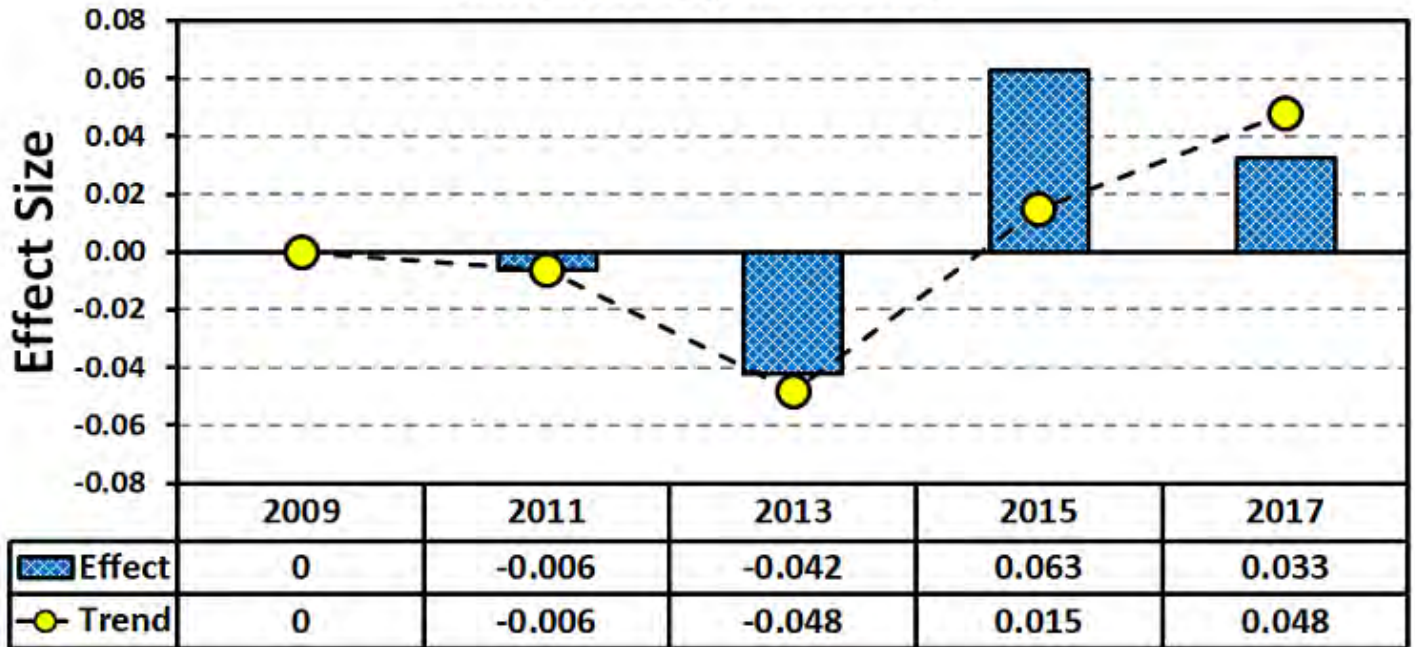


Figure A-1. Idaho effect size results on NAEP reading assessments from 2009 through 2017: Grade 4, All students.

NAEP Mathematics, 2009 thru 2017 Idaho Grade 4, All Students

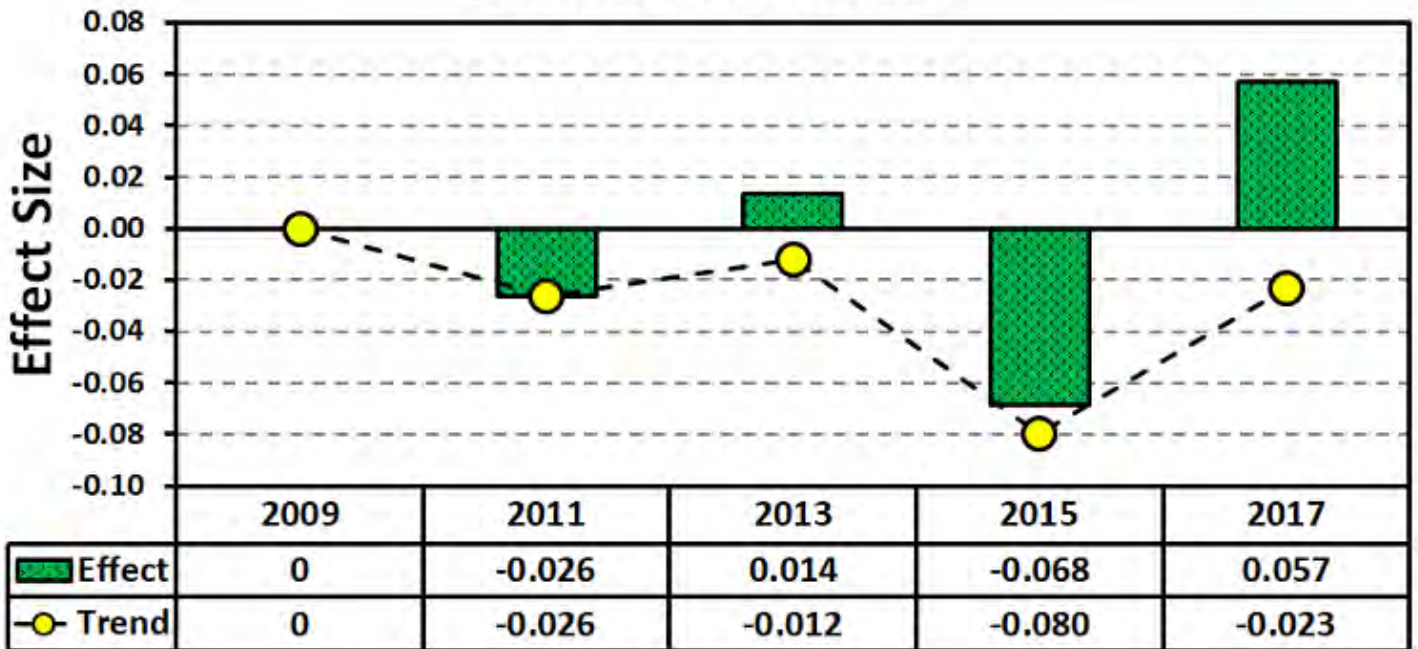


Figure A-2. Idaho effect size results on NAEP mathematics assessments from 2009 through 2017: Grade 4, All students.

NAEP Reading, 2009 thru 2017 Idaho Grade 8, All Students

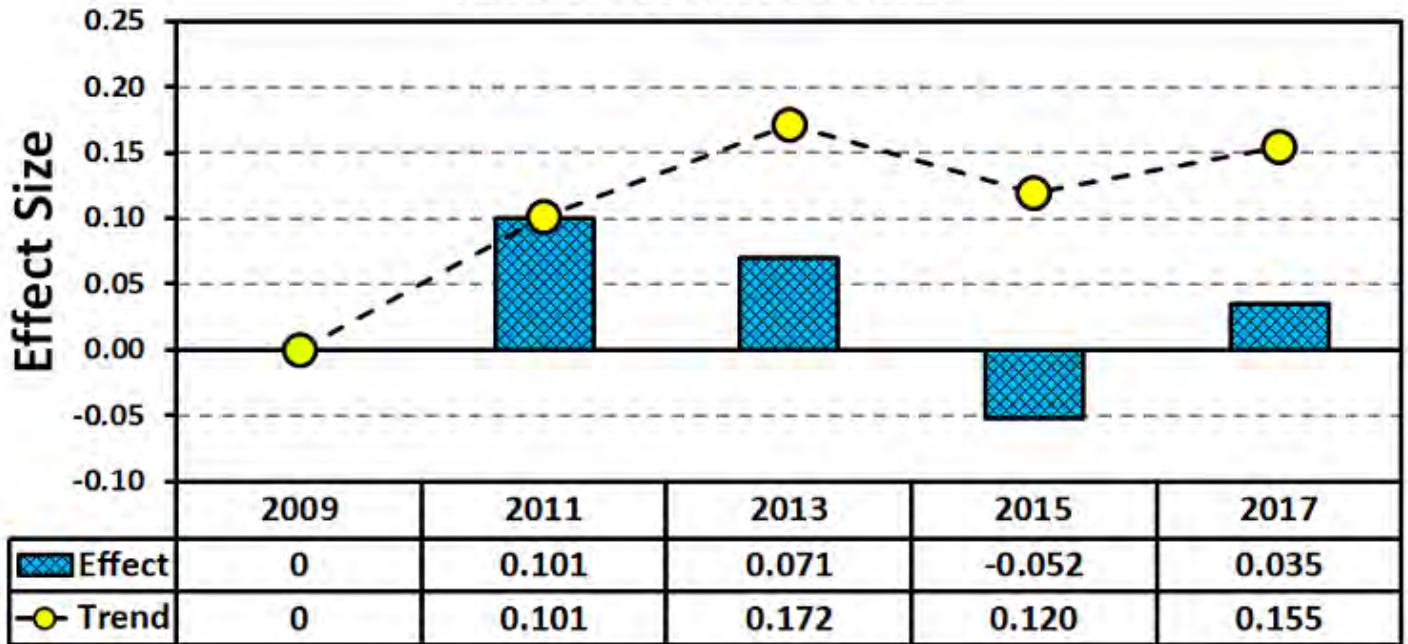


Figure A-3. Idaho effect size results on NAEP mathematics assessments from 2009 through 2017: Grade 8, All students

NAEP Mathematics, 2009 thru 2017 Idaho Grade 8, All Students

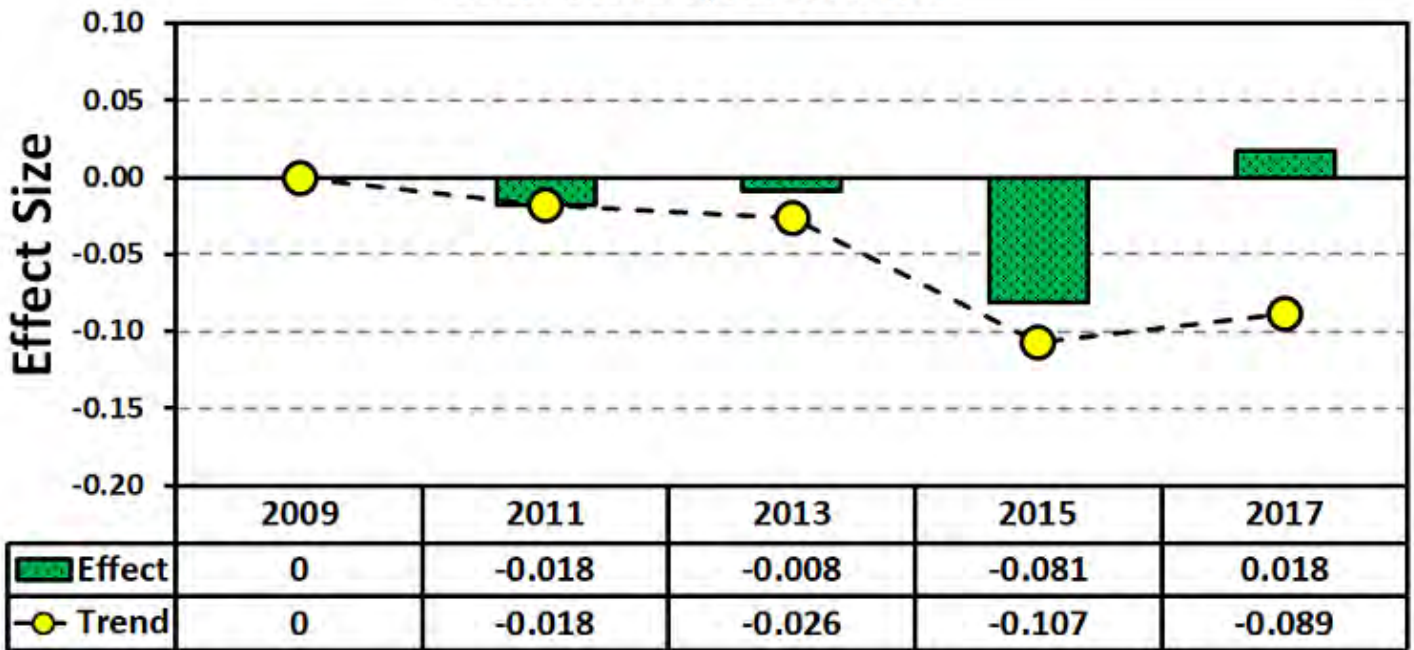


Figure A-4. Idaho effect size results on NAEP mathematics assessments from 2009 through 2017: Grade 8, All students.

NAEP Reading, 2009 thru 2017 Idaho Grade 4, Hispanic Students

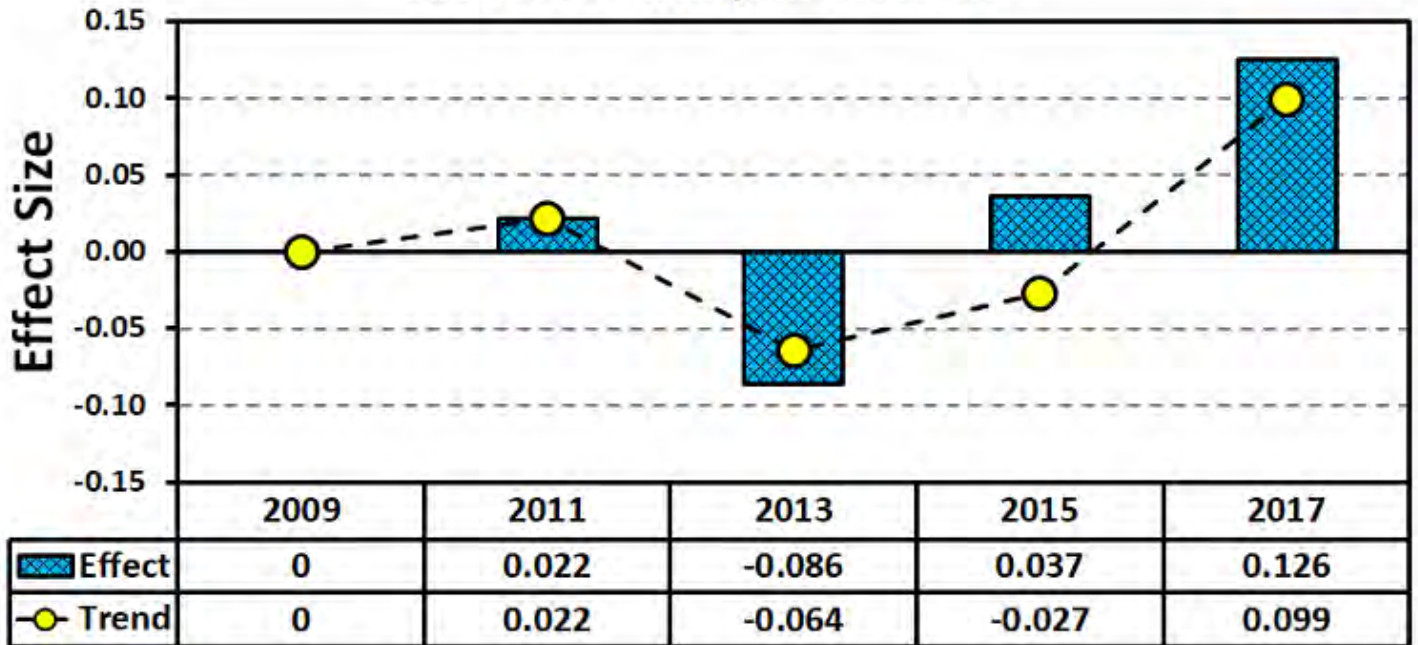


Figure A-5. Idaho effect size results on NAEP mathematics assessments from 2009 through 2017: Grade 4, Hispanic students.

NAEP Mathematics, 2009 thru 2017 Idaho Grade 4, Hispanic Students

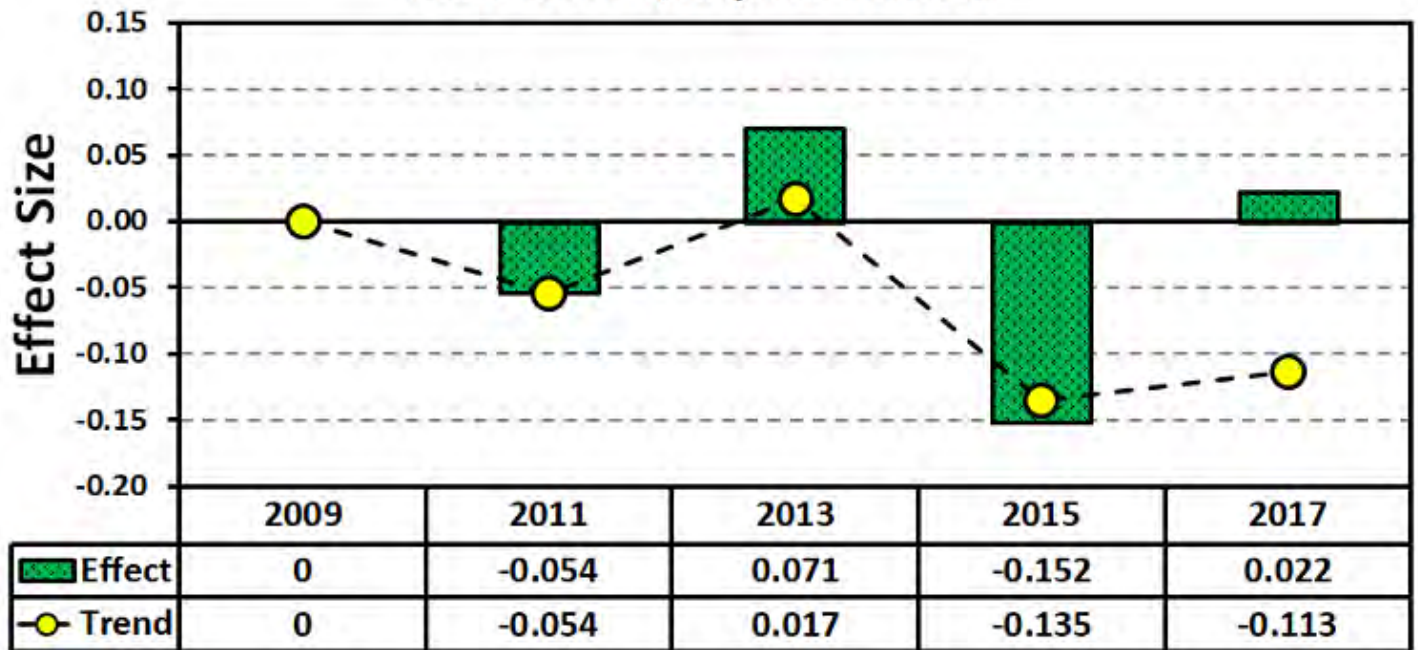


Figure A-6. Idaho effect size results on NAEP mathematics assessments from 2009 through 2017: Grade 4, Hispanic students.

NAEP Reading, 2009 thru 2017 Idaho Grade 8, Hispanic Students

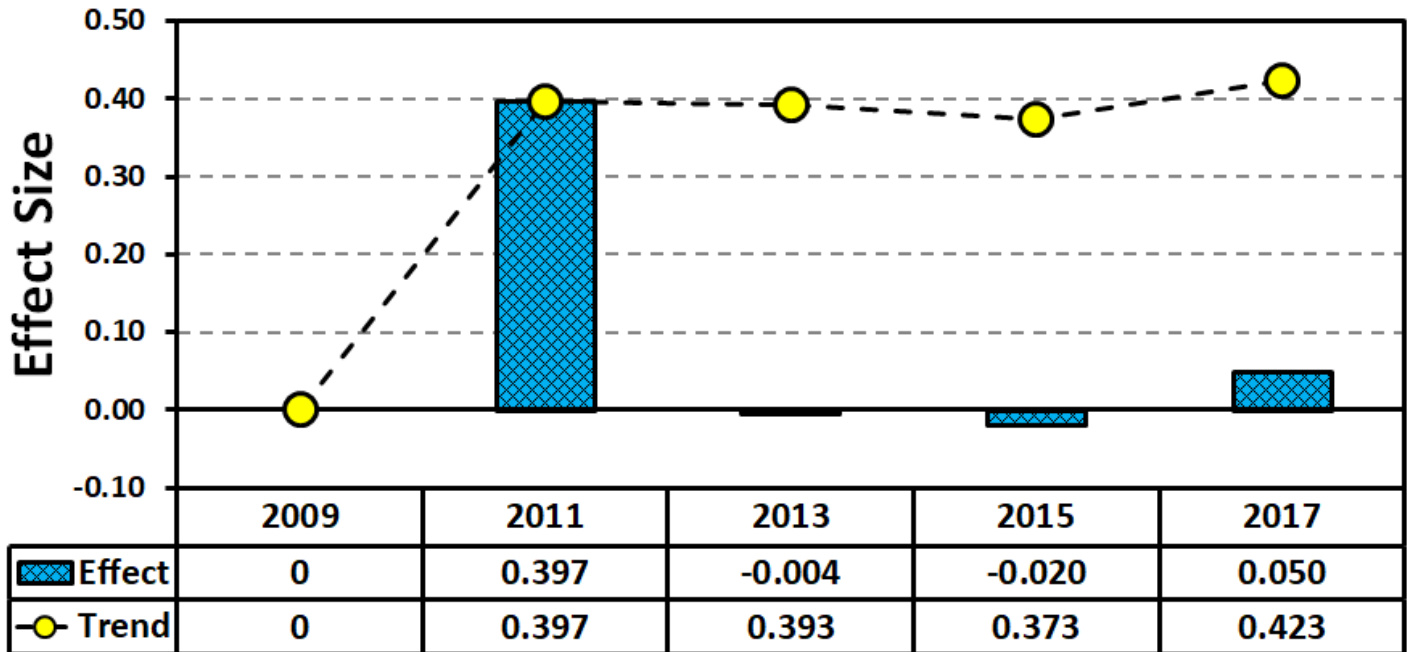


Figure A-7. Idaho effect size results on NAEP reading assessments from 2009 through 2017: Grade 8, Hispanic students.

NAEP Mathematics, 2009 thru 2017 Idaho Grade 8, Hispanic Students

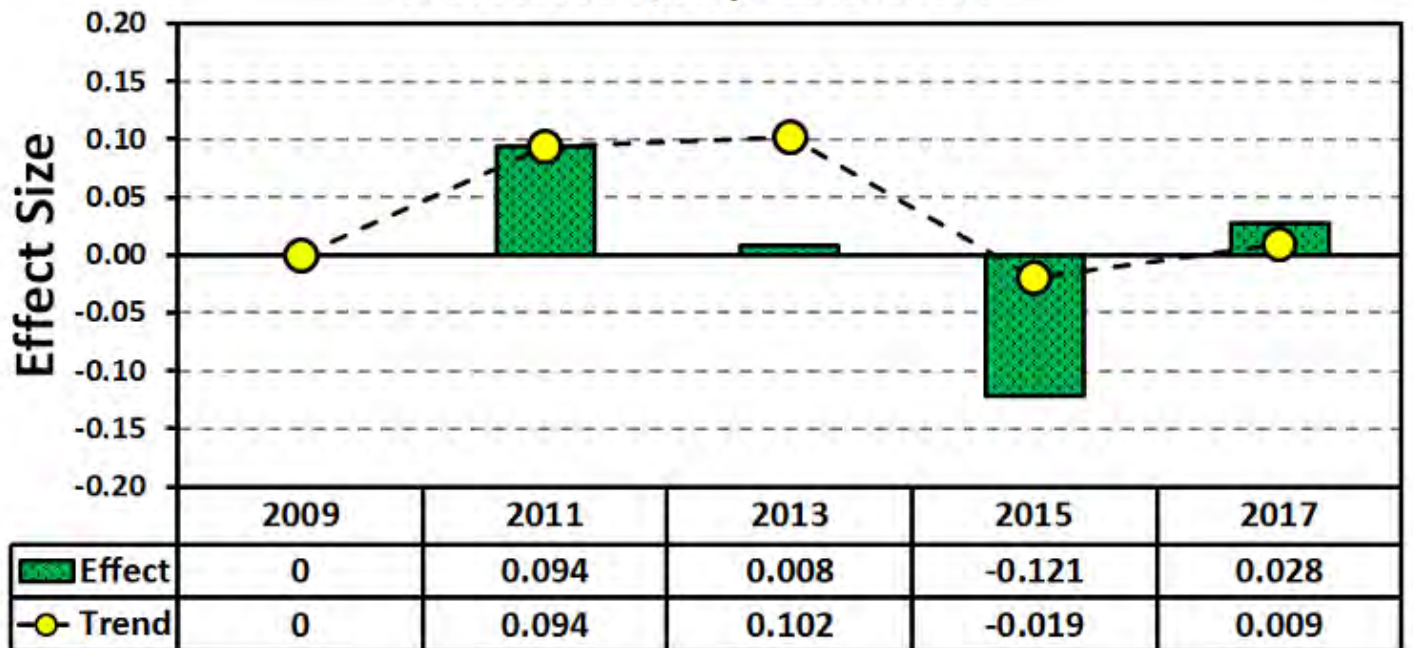


Figure A-8. Idaho effect size results on NAEP mathematics assessments from 2009 through 2017: Grade 8, Hispanic students.

NAEP Reading, 2009 thru 2017 Idaho Grade 4, White Students

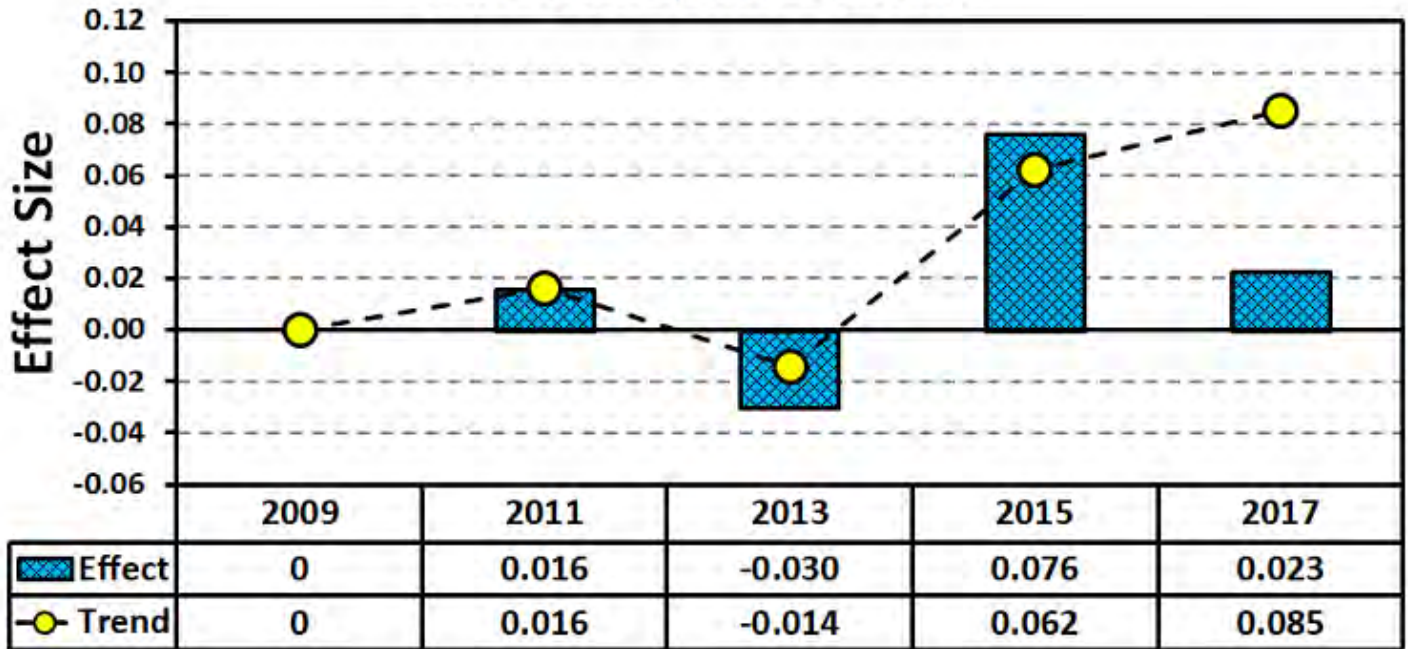


Figure A-9. Idaho effect size results on NAEP reading assessments from 2009 through 2017: Grade 4, White students.

NAEP Mathematics, 2009 thru 2017 Idaho Grade 4, White Students

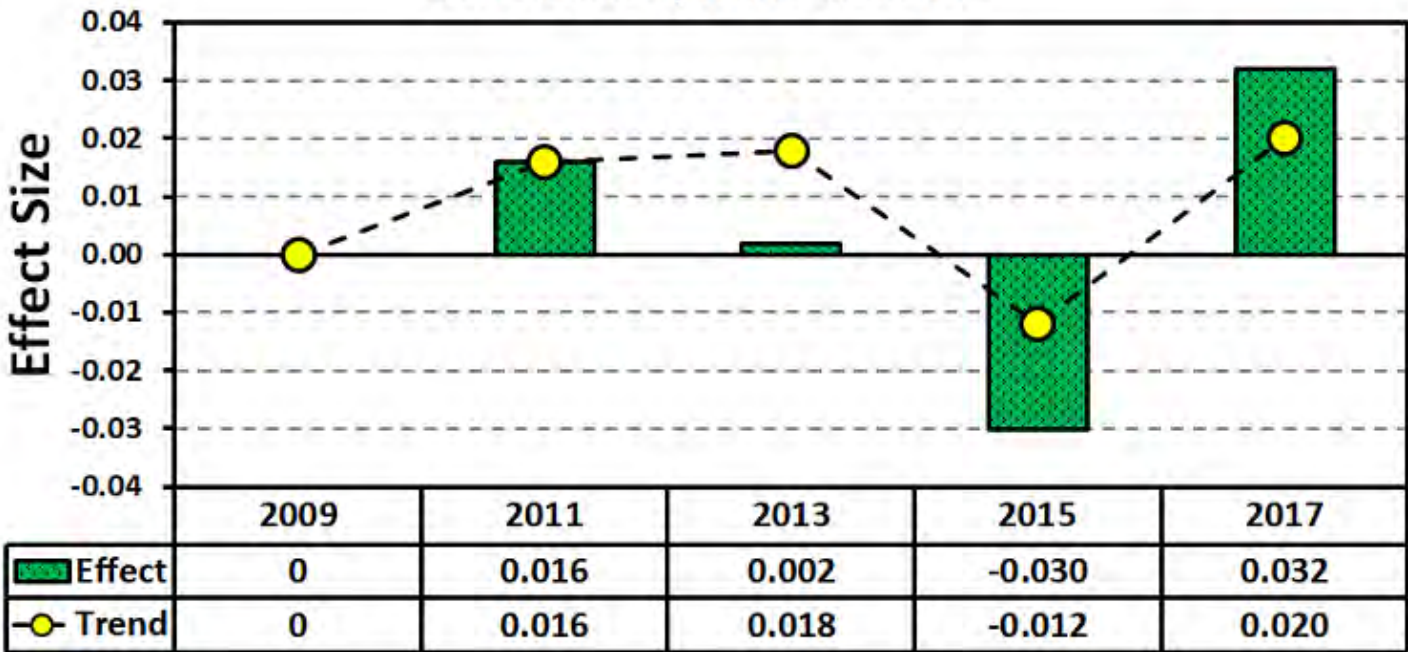


Figure A-10. Idaho effect size results on NAEP mathematics assessments from 2009 through 2017: Grade 4, White students.

NAEP Reading, 2009 thru 2017 Idaho Grade 8, White Students

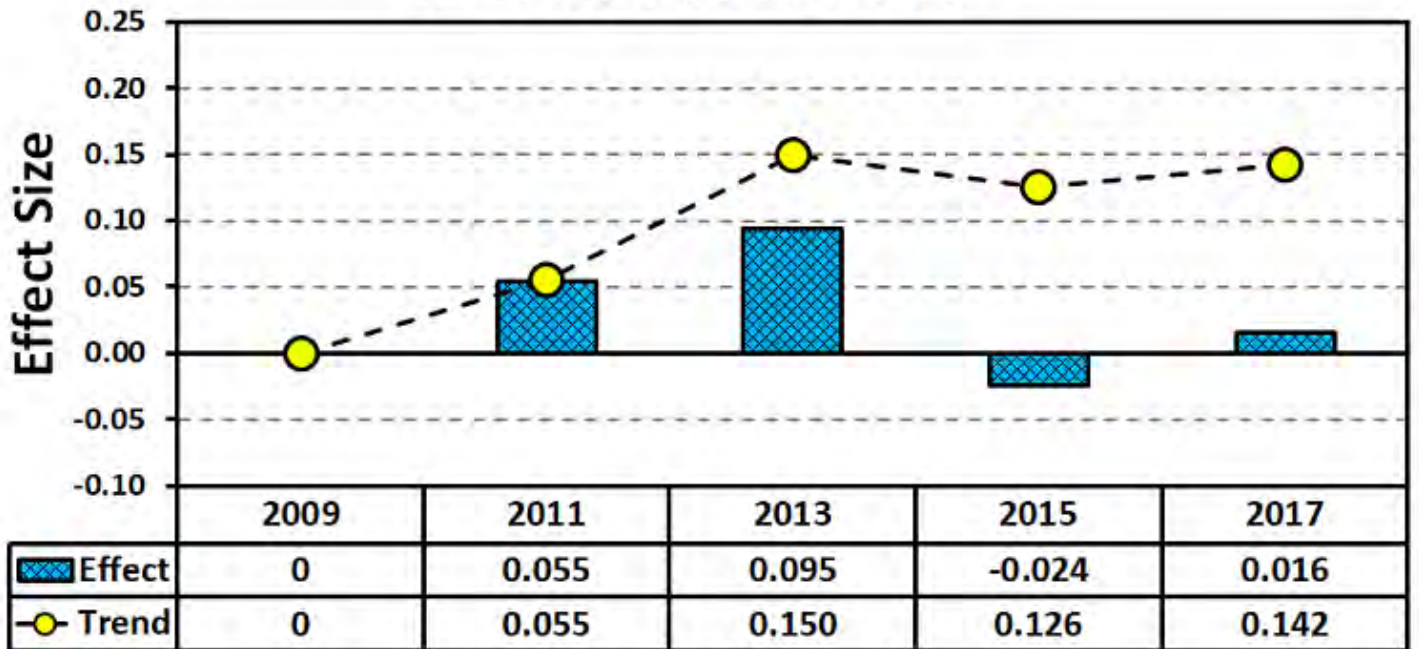


Figure A-11. Idaho effect size results on NAEP reading assessments from 2009 through 2017: Grade 8, White students.

NAEP Mathematics, 2009 thru 2017 Idaho Grade 8, White Students

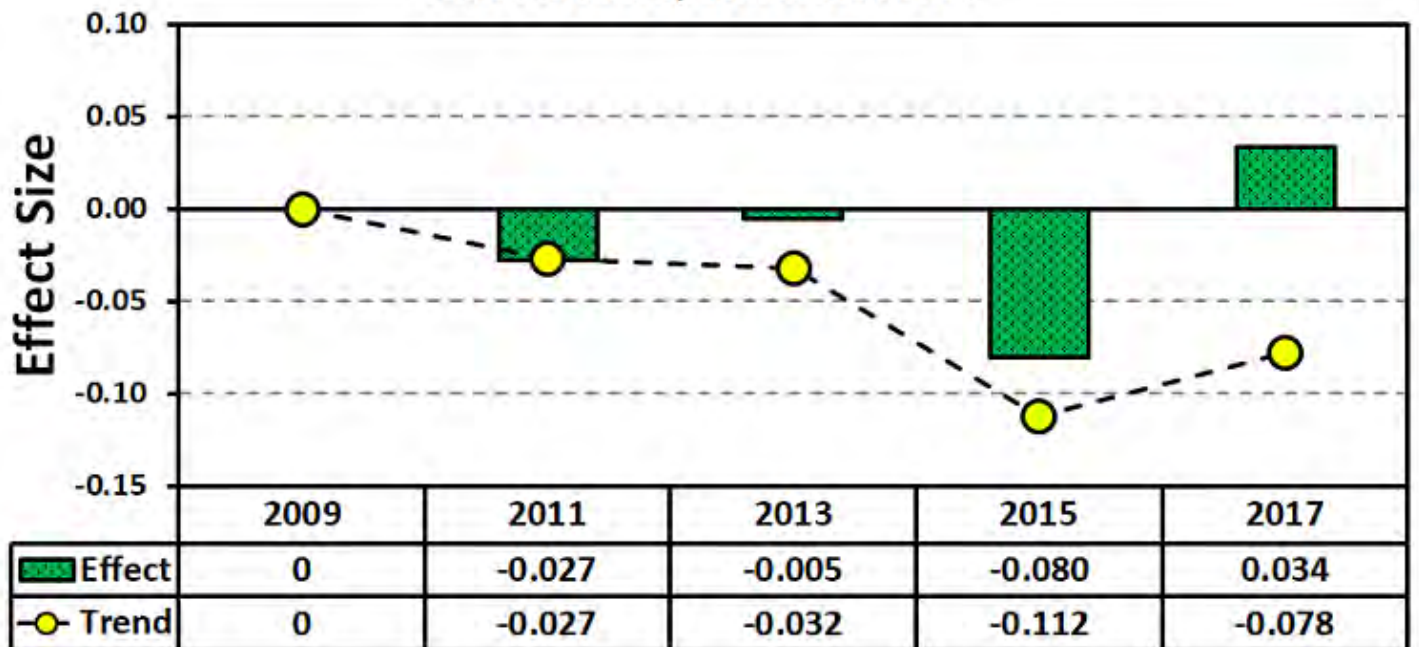


Figure A-12. Idaho effect size results on NAEP mathematics assessments from 2009 through 2017: Grade 8, White students.

NAEP Reading, 2009 thru 2017 Idaho Grade 4, Female Students

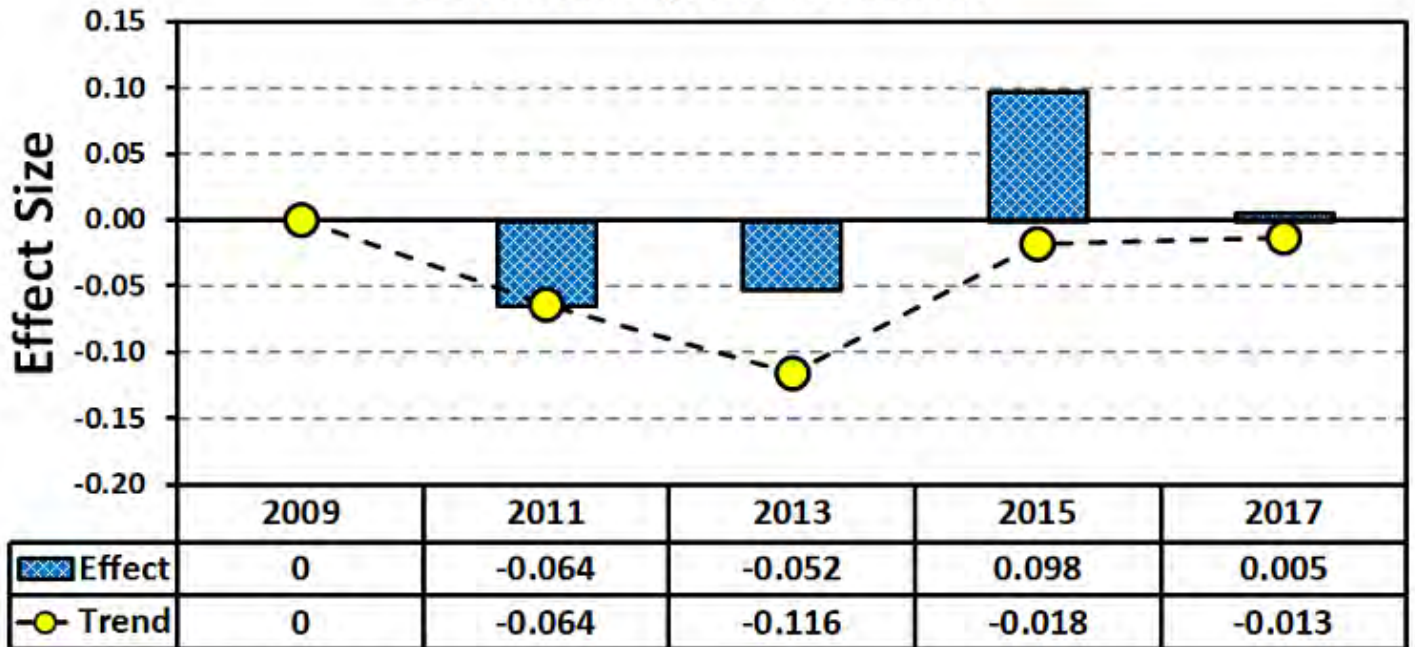


Figure A-13. Idaho effect size results on NAEP reading assessments from 2009 through 2017: Grade 4, Female students.

NAEP Mathematics, 2009 thru 2017 Idaho Grade 4, Female Students

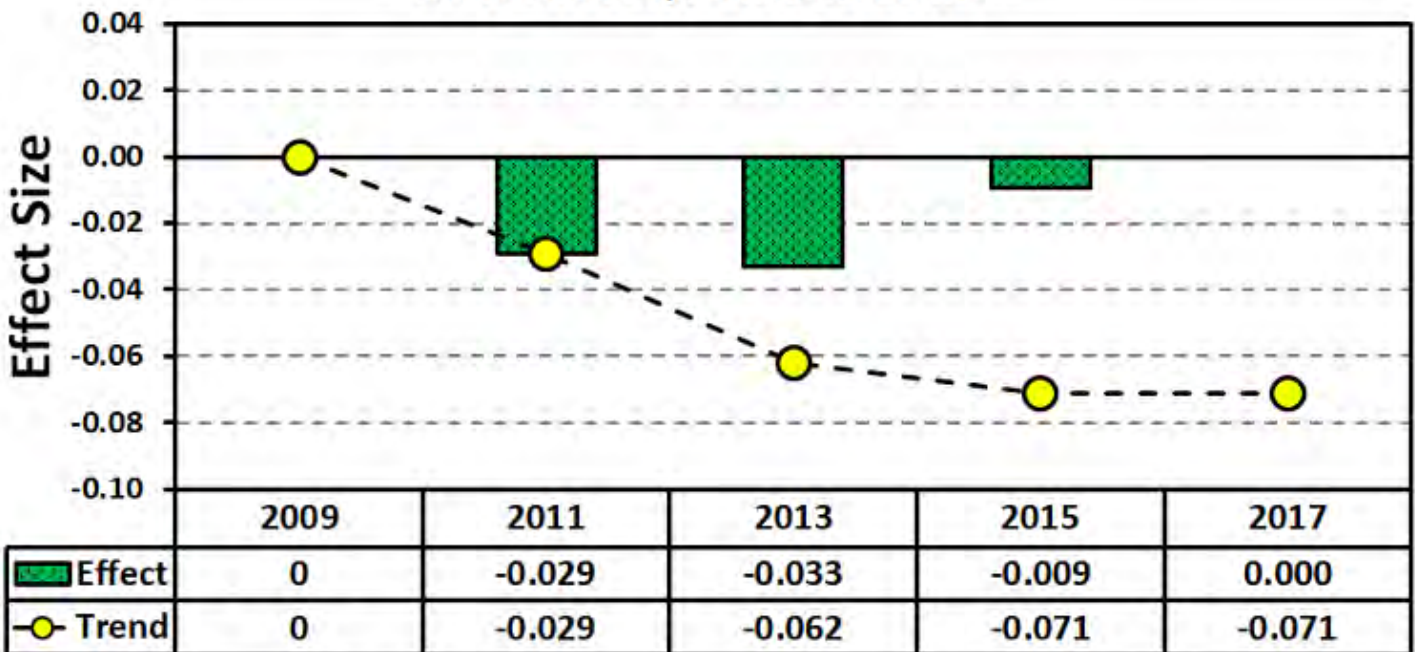


Figure A-14. Idaho effect size results on NAEP mathematics assessments from 2009 through 2017: Grade 4, Female students.

NAEP Reading, 2009 thru 2017 Idaho Grade 8, Female Students

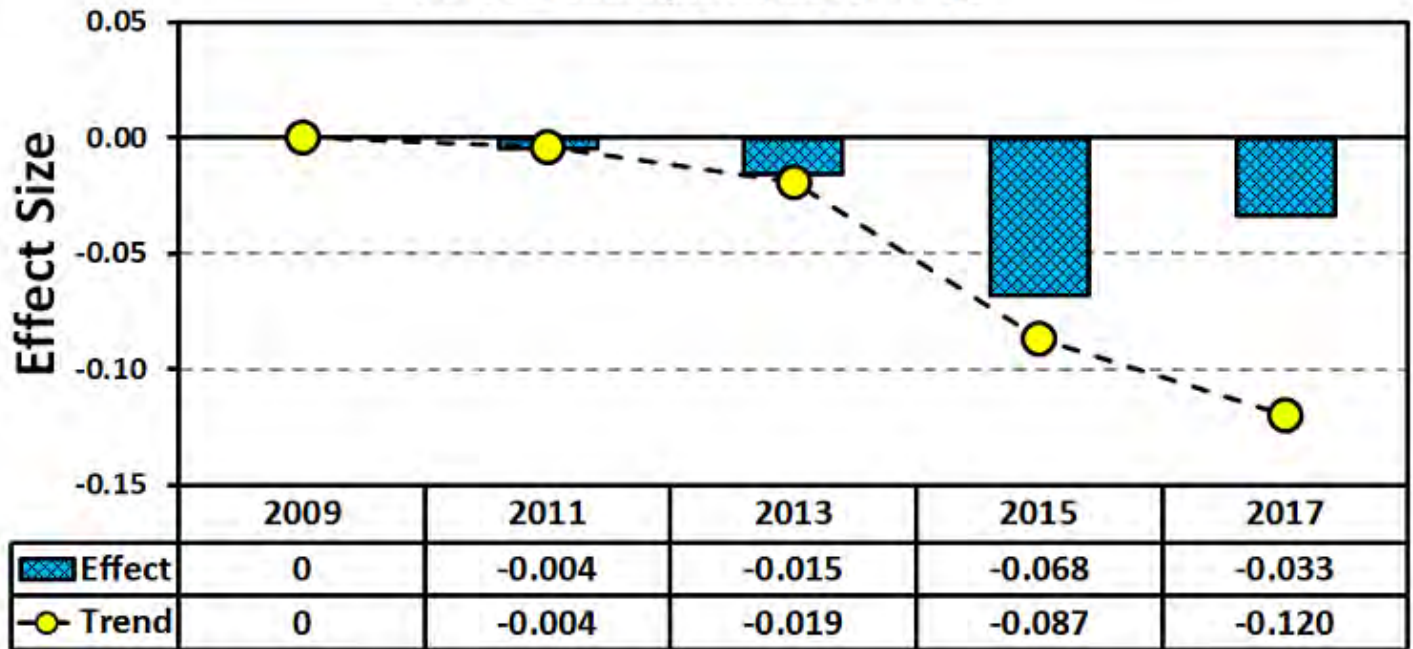


Figure A-15. Idaho effect size results on NAEP reading assessments from 2009 through 2017: Grade 8, Female students.

NAEP Mathematics, 2009 thru 2017 Idaho Grade 8, Female Students

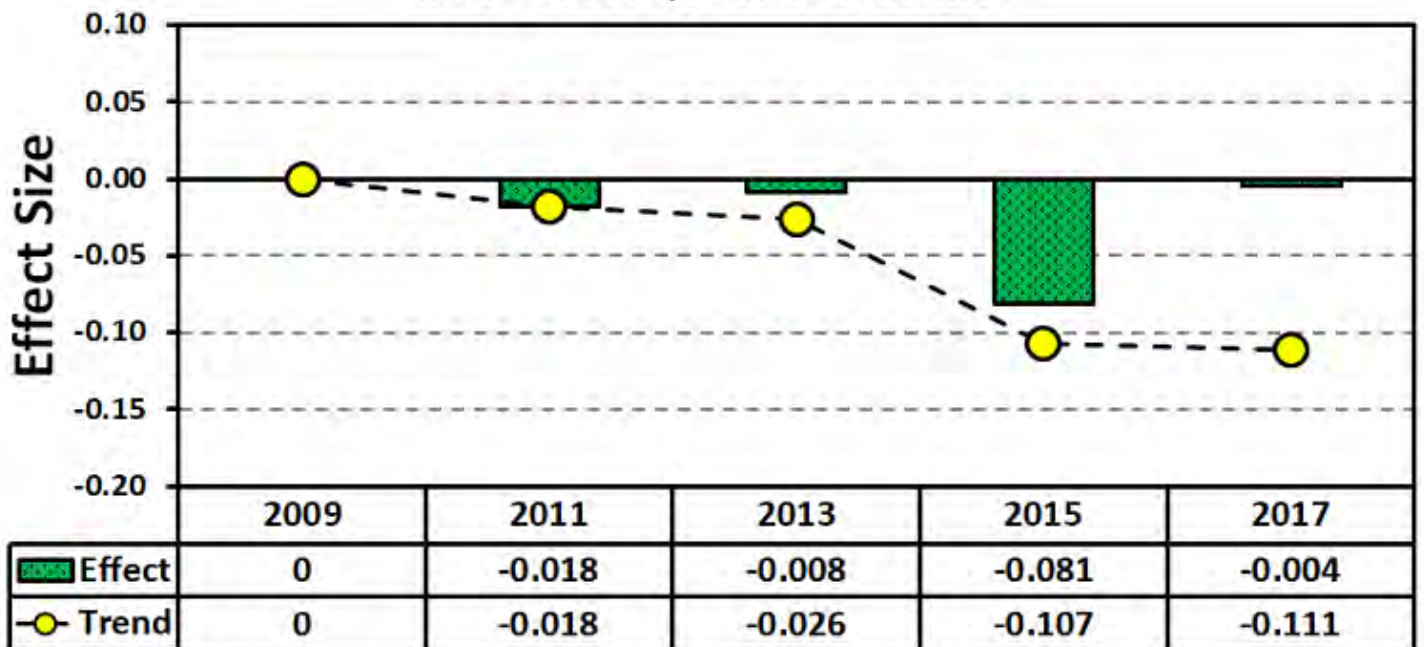


Figure A-16. Idaho effect size results on NAEP mathematics assessments from 2009 through 2017: Grade 8, Female students.

NAEP Reading, 2009 thru 2017 Idaho Grade 4, Male Students

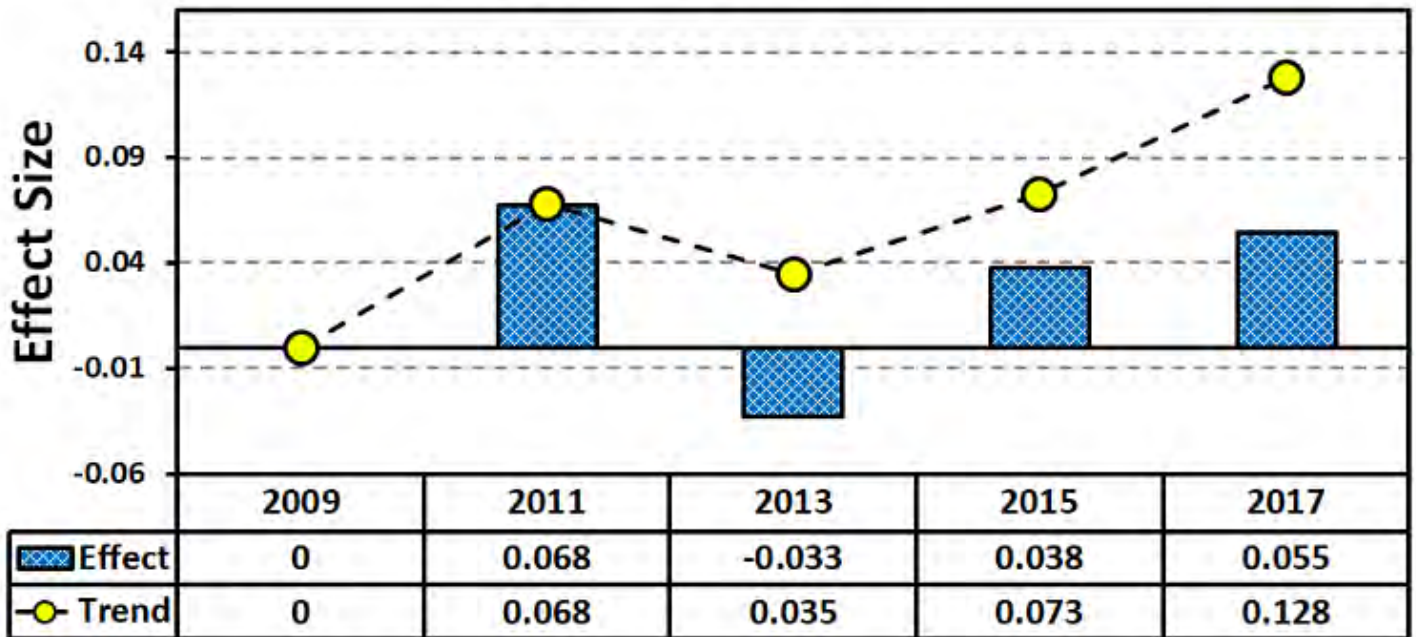


Figure A-17. Idaho effect size results on NAEP reading assessments from 2009 through 2017: Grade 4, Male students.

NAEP Mathematics, 2009 thru 2017 Idaho Grade 4, Male Students

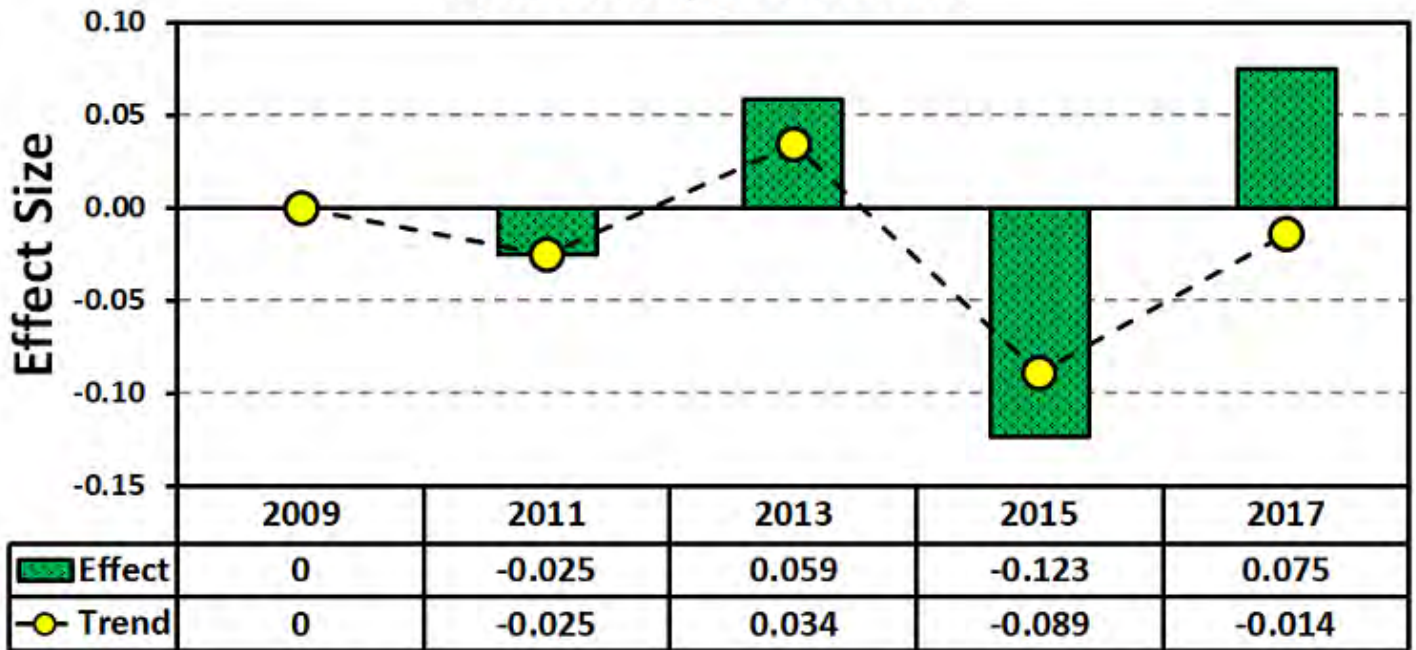


Figure A-18. Idaho effect size results on NAEP mathematics assessments from 2009 through 2017: Grade 4, Male students.

NAEP Reading, 2009 thru 2017 Idaho Grade 8, Male Students

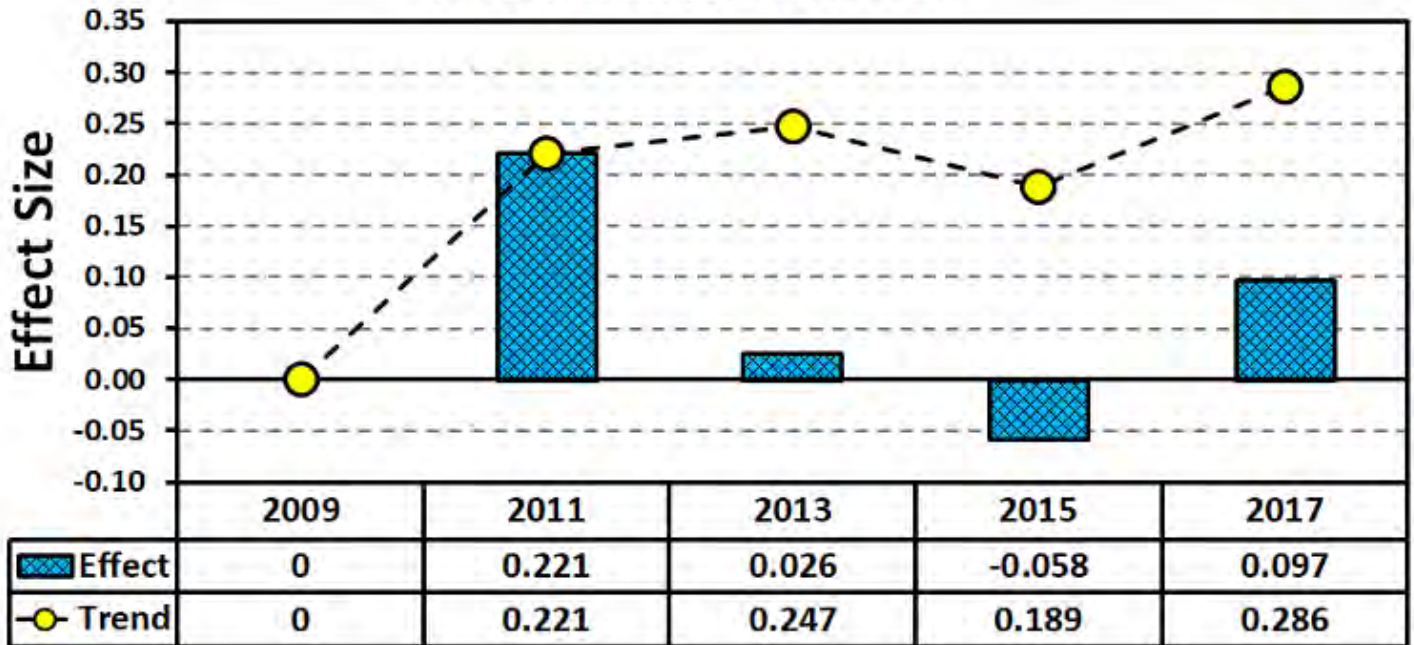


Figure A-19. Idaho effect size results on NAEP reading assessments from 2009 through 2017: Grade 8, Male students.

NAEP Mathematics, 2009 thru 2017 Idaho Grade 8, Male Students

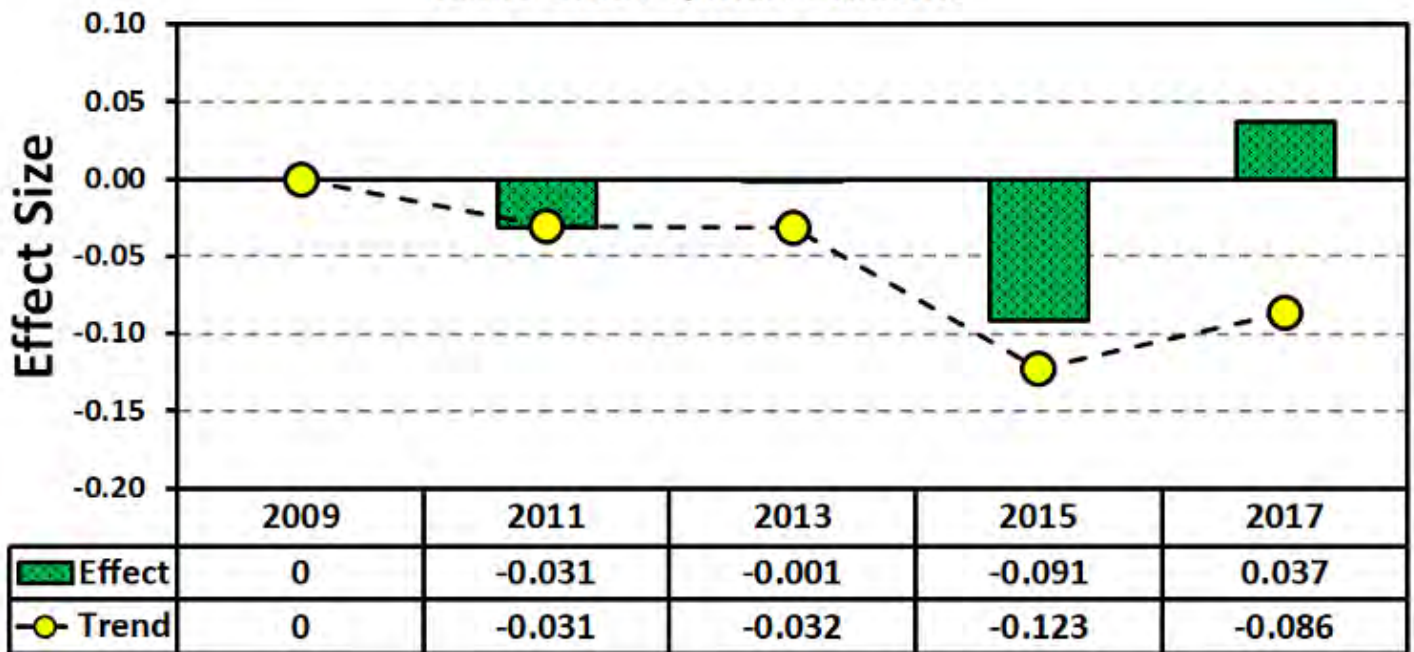


Figure A-20. Idaho effect size results on NAEP mathematics assessments from 2009 through 2017: Grade 8, Male students.

NAEP Reading, 2009 thru 2017 Idaho Grade 4, City Schools

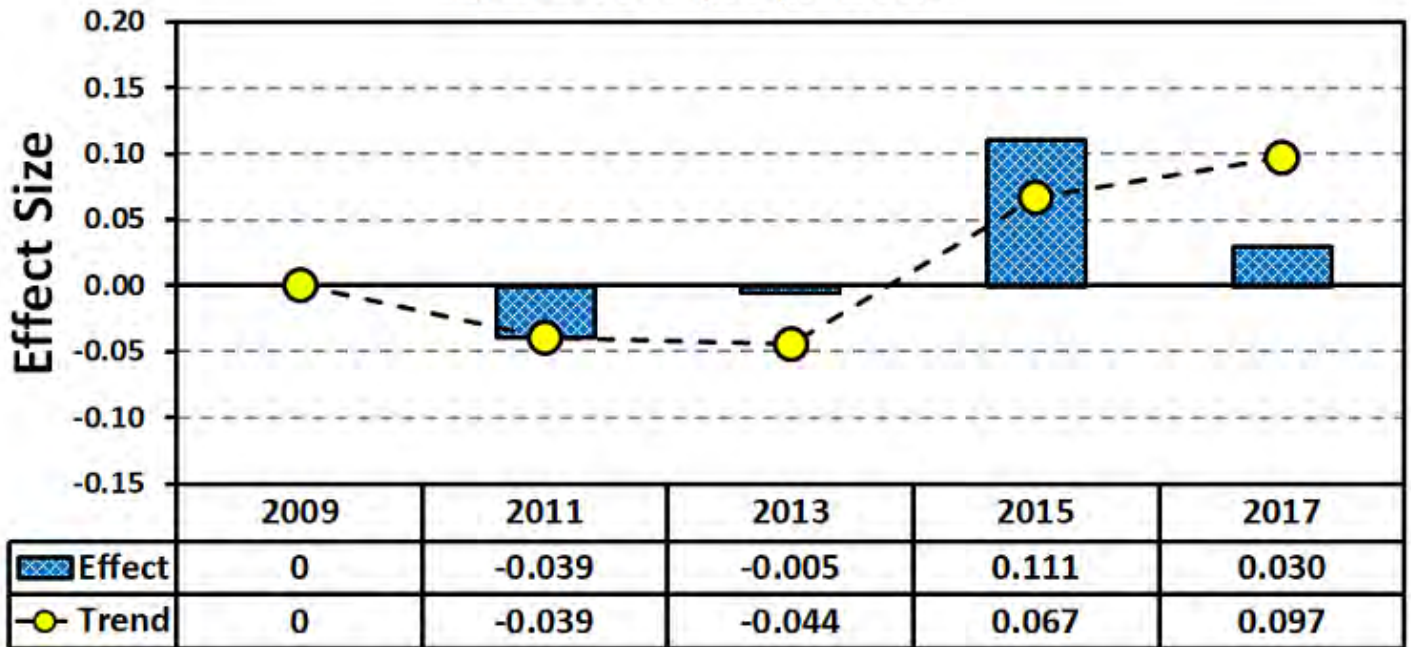


Figure A-21. Idaho effect size results on NAEP reading assessments from 2009 through 2017: Grade 4, City schools.

NAEP Mathematics, 2009 thru 2017 Idaho Grade 4, City Schools

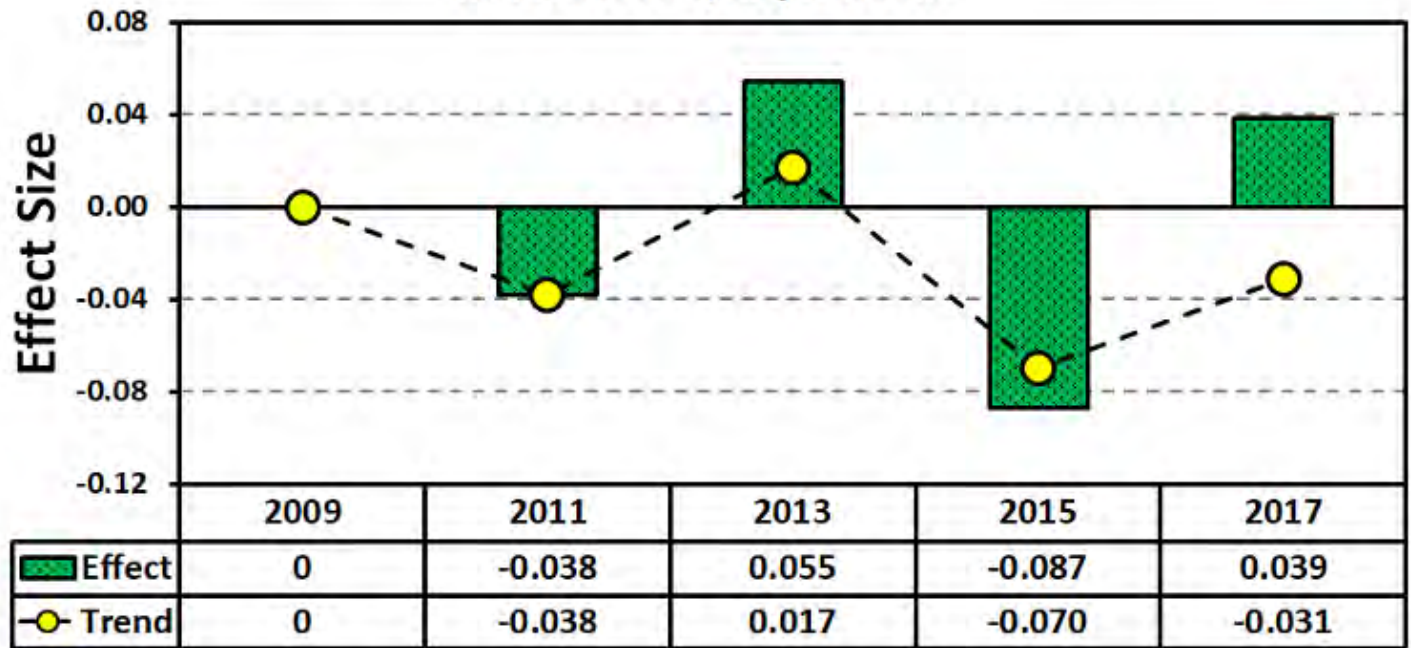


Figure A-22. Idaho effect size results on NAEP mathematics assessments from 2009 through 2017: Grade 4, City schools.

NAEP Reading, 2009 thru 2017 Idaho Grade 8, City Schools

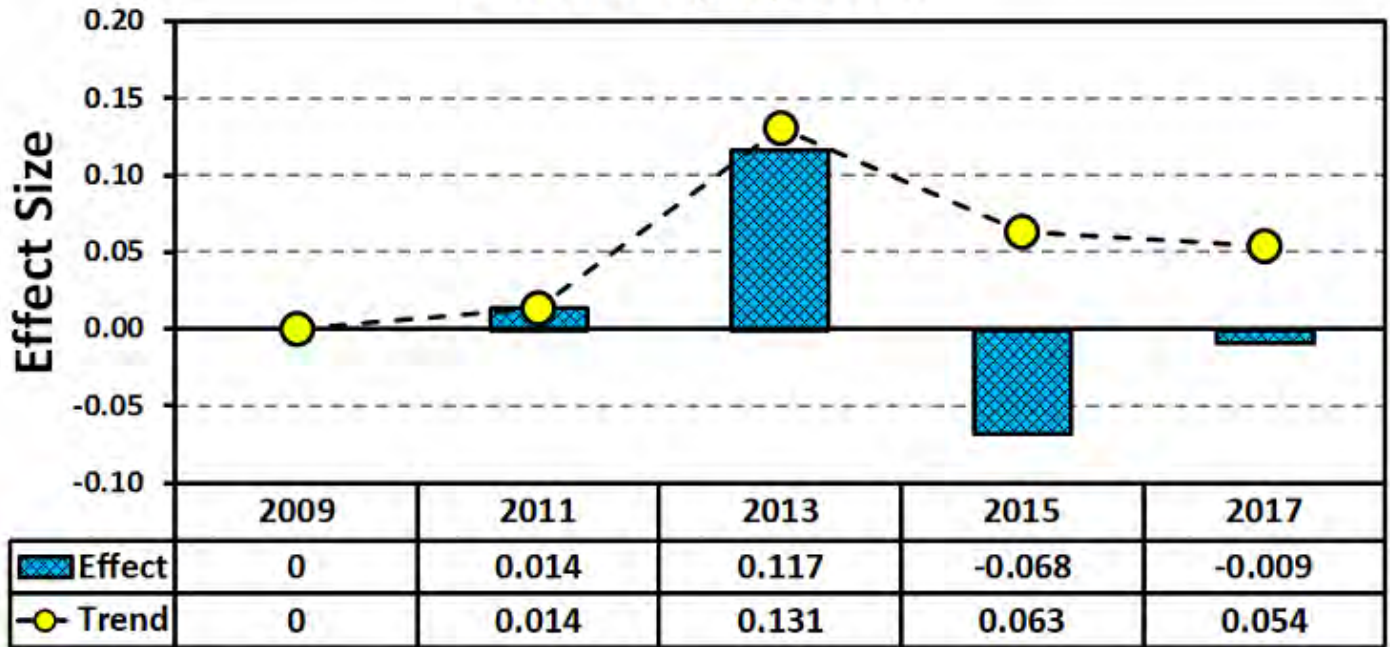


Figure A-23. Idaho effect size results on NAEP reading assessments from 2009 through 2017: Grade 8, City schools.

NAEP Mathematics, 2009 thru 2017 Idaho Grade 8, City Schools

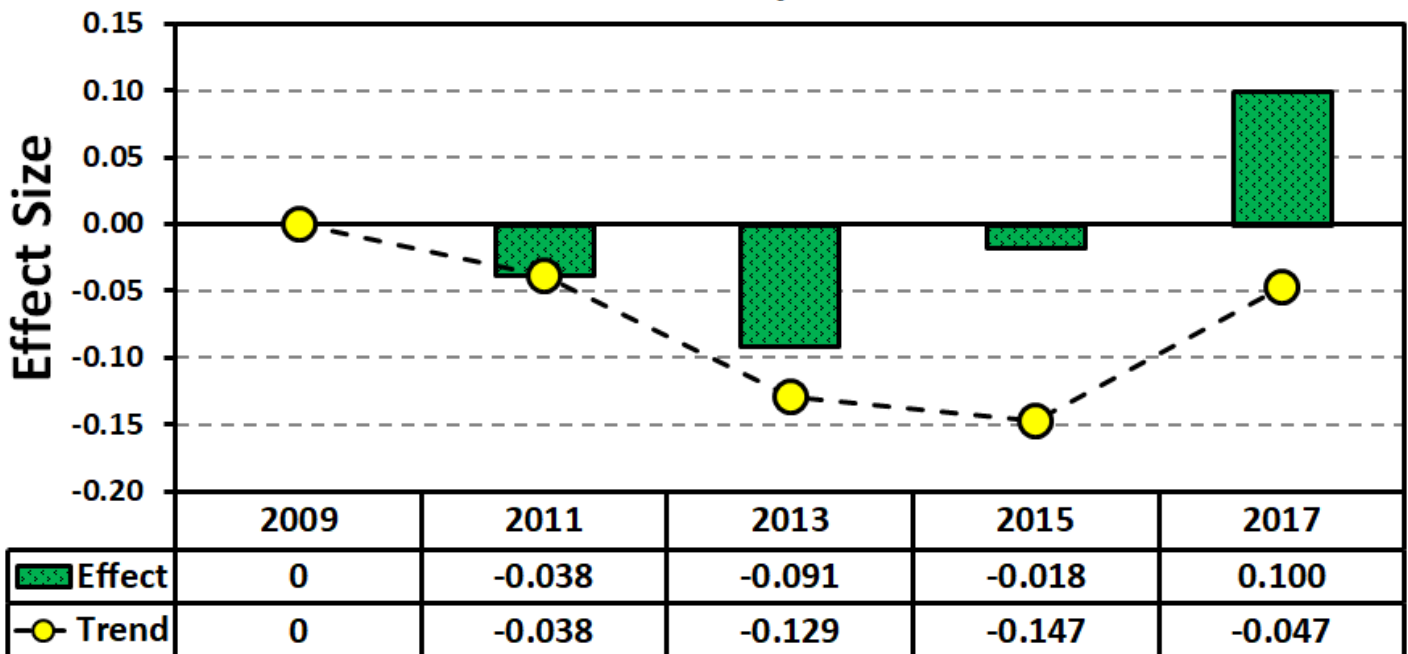


Figure A-24. Idaho effect size results on NAEP mathematics assessments from 2009 through 2017: Grade 8, City schools.

NAEP Reading, 2009 thru 2017 Idaho Grade 4, Suburban Schools

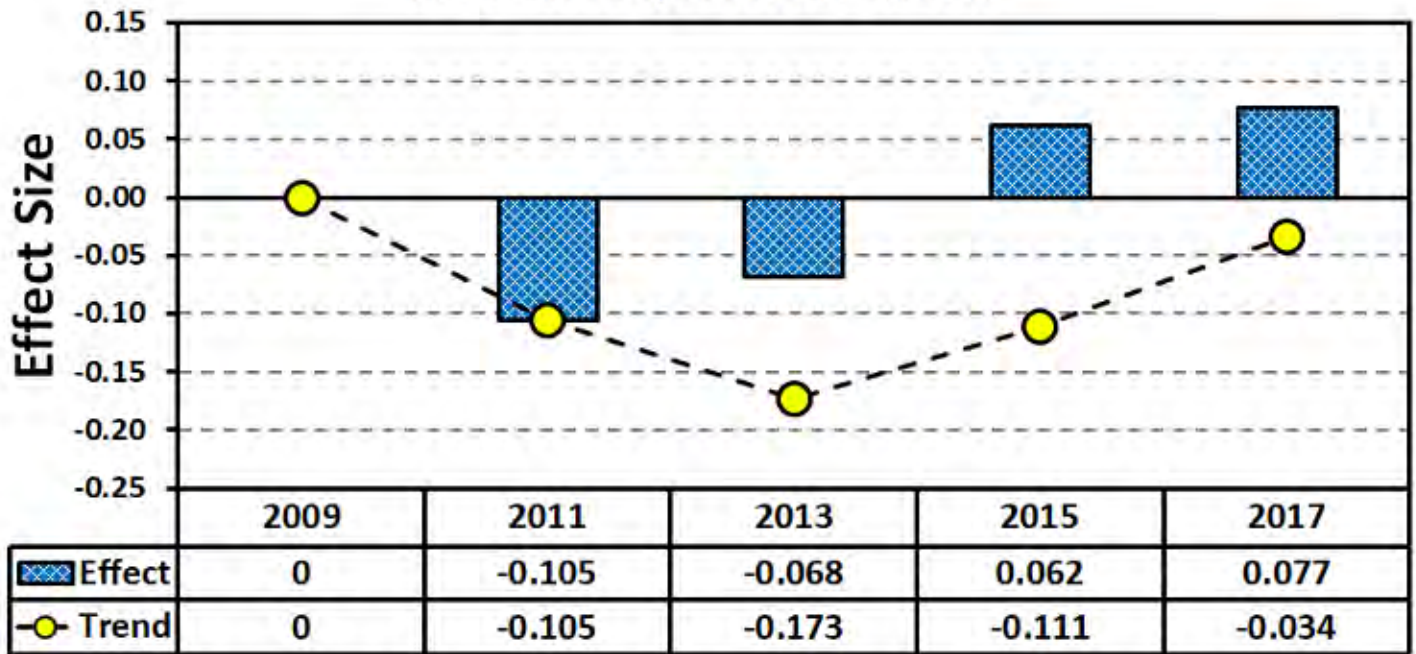


Figure A-25. Idaho effect size results on NAEP reading assessments from 2009 through 2017: Grade 4, Suburban schools.

NAEP Mathematics, 2009 thru 2017 Idaho Grade 4, Suburban Schools

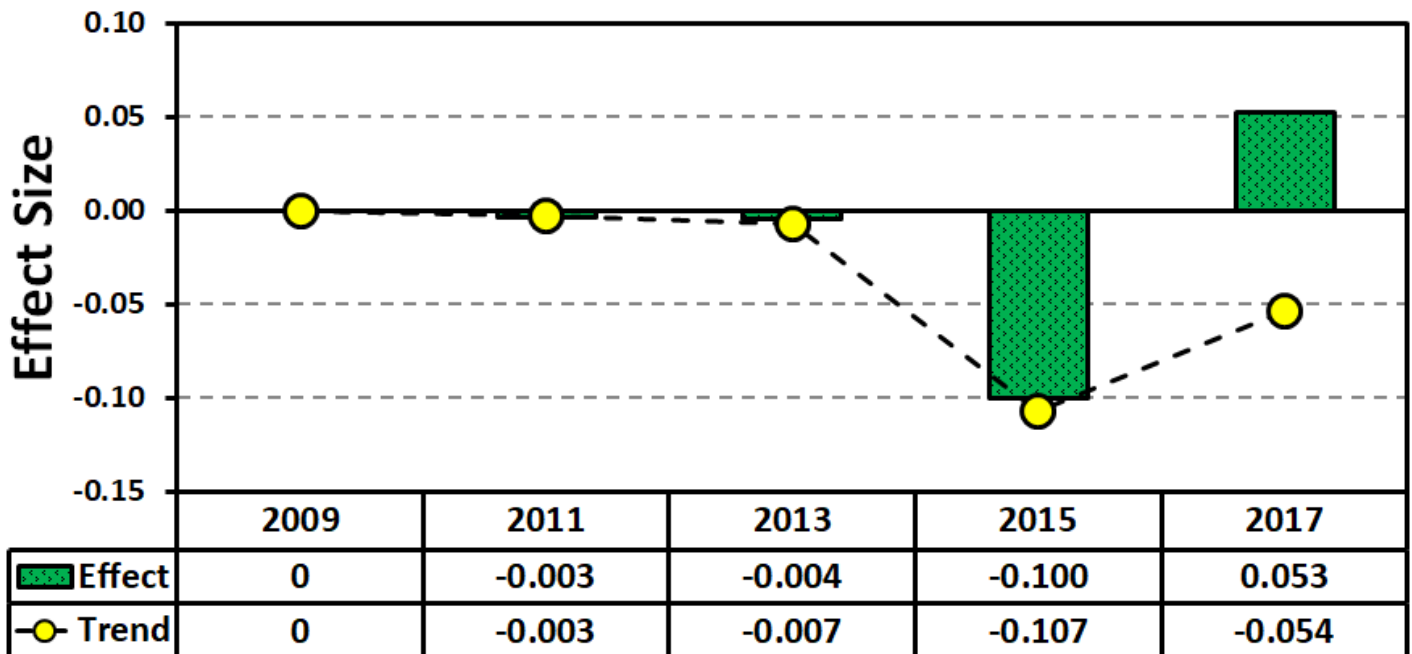


Figure A-26. Idaho effect size results on NAEP mathematics assessments from 2009 through 2017: Grade 4, Suburban schools.

NAEP Reading, 2009 thru 2017 Idaho Grade 8, Suburban Schools

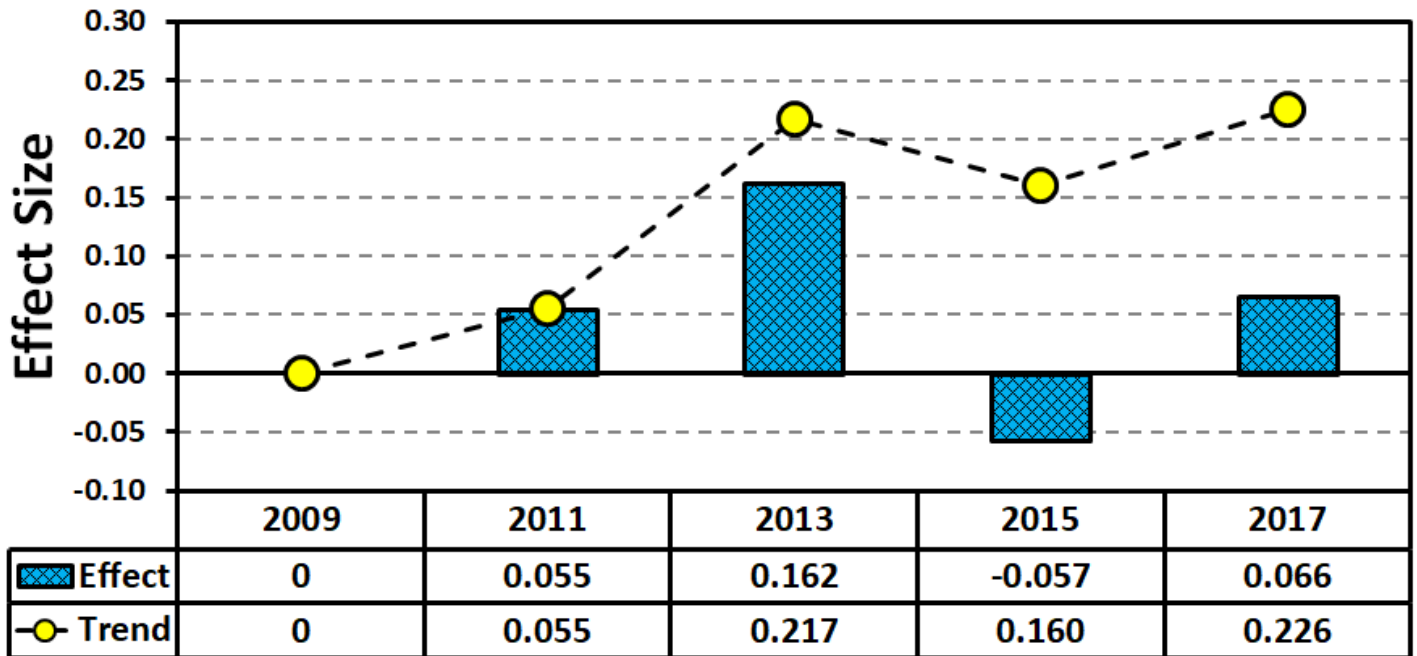


Figure A-27. Idaho effect size results on NAEP reading assessments from 2009 through 2017: Grade 8, Suburban schools.

NAEP Mathematics, 2009 thru 2017 Idaho Grade 8, Suburban Schools

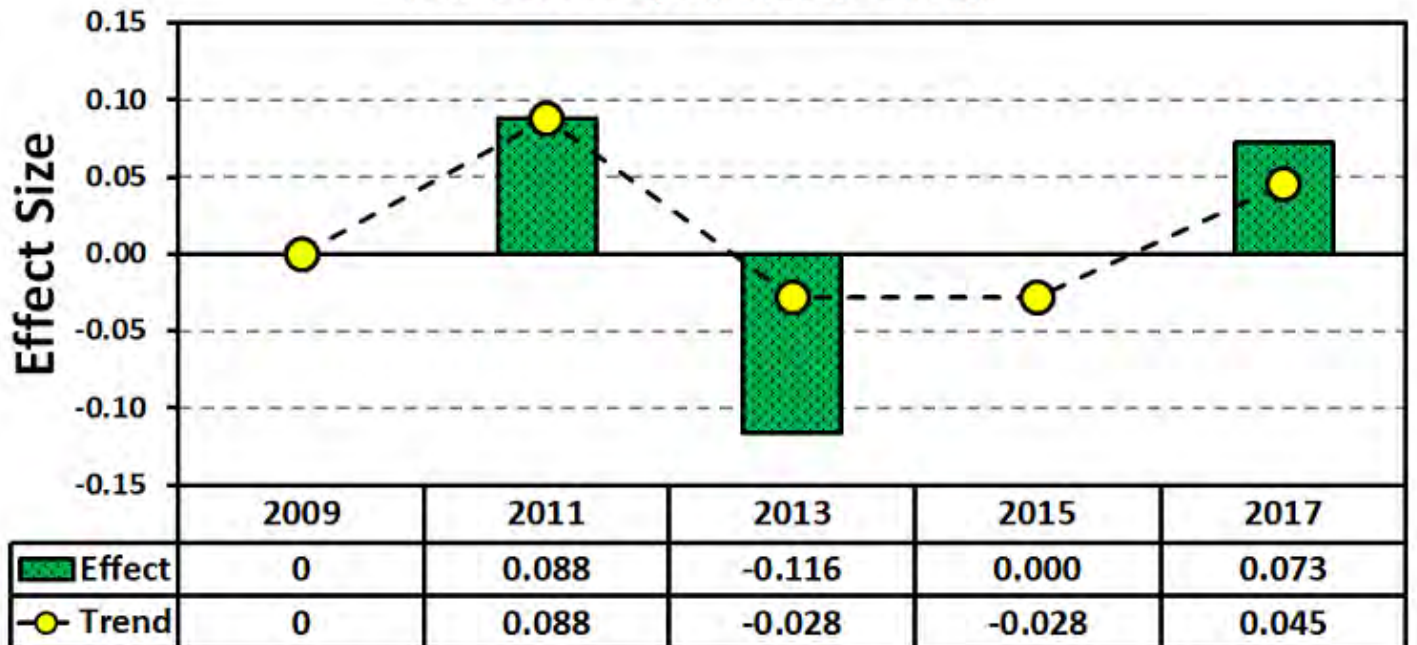


Figure A-28. Idaho effect size results on NAEP mathematics assessments from 2009 through 2017: Grade 8, Suburban schools.

NAEP Reading, 2009 thru 2017 Idaho Grade 4, Town Schools

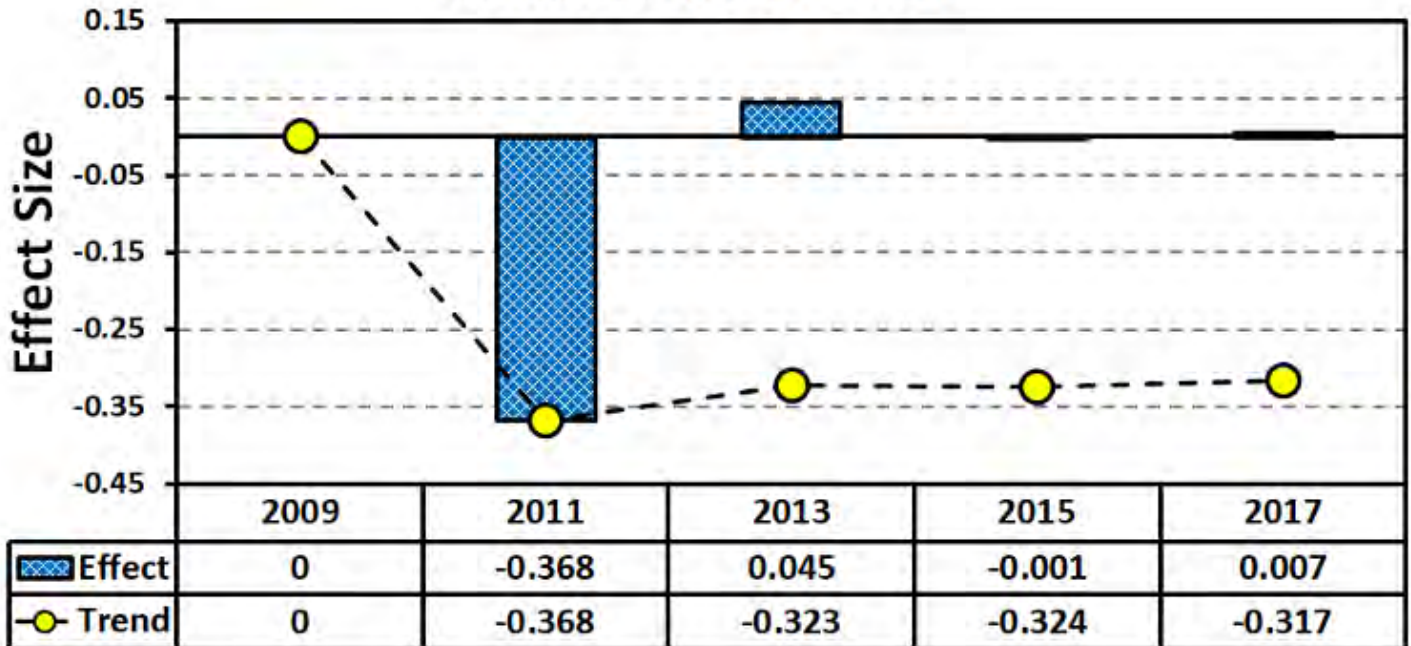


Figure A-29. Idaho effect size results on NAEP reading assessments from 2009 through 2017: Grade 4, Town schools.

NAEP Mathematics, 2009 thru 2017 Idaho Grade 4, Town Schools

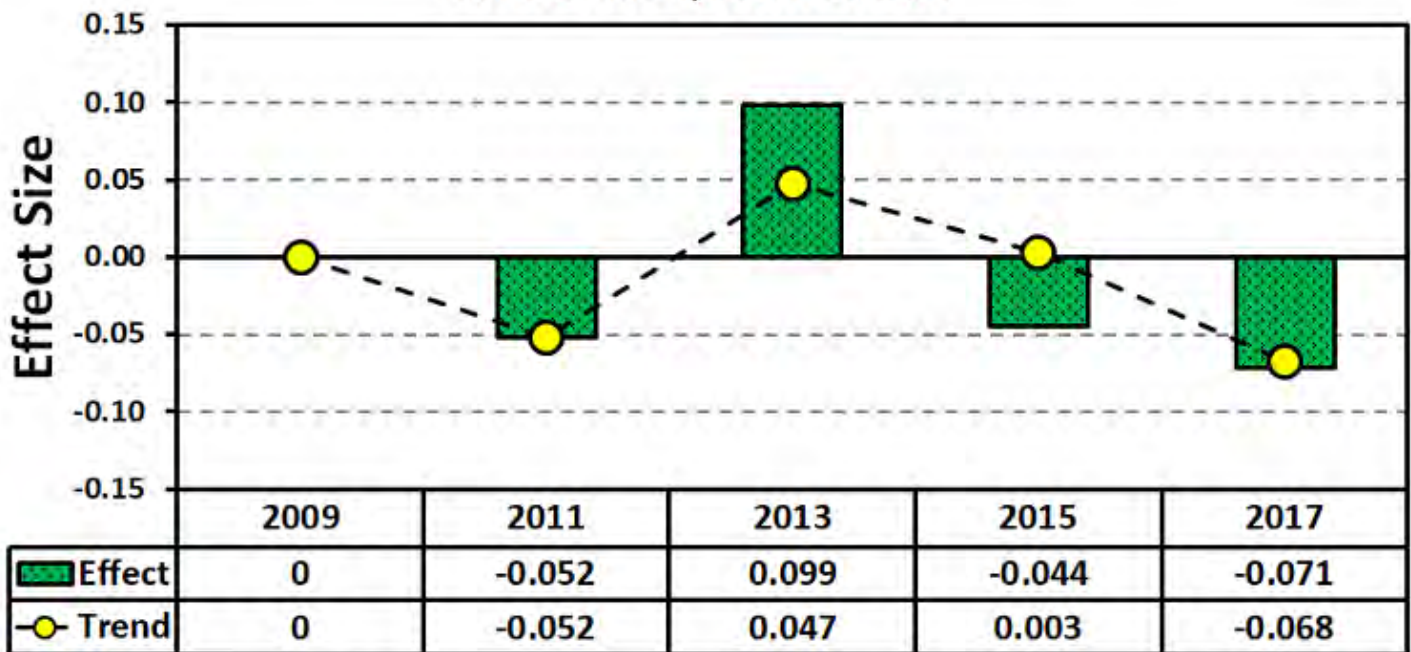


Figure A-30. Idaho effect size results on NAEP mathematics assessments from 2009 through 2017: Grade 4, Town schools.

NAEP Reading, 2009 thru 2017 Idaho Grade 8, Town Schools

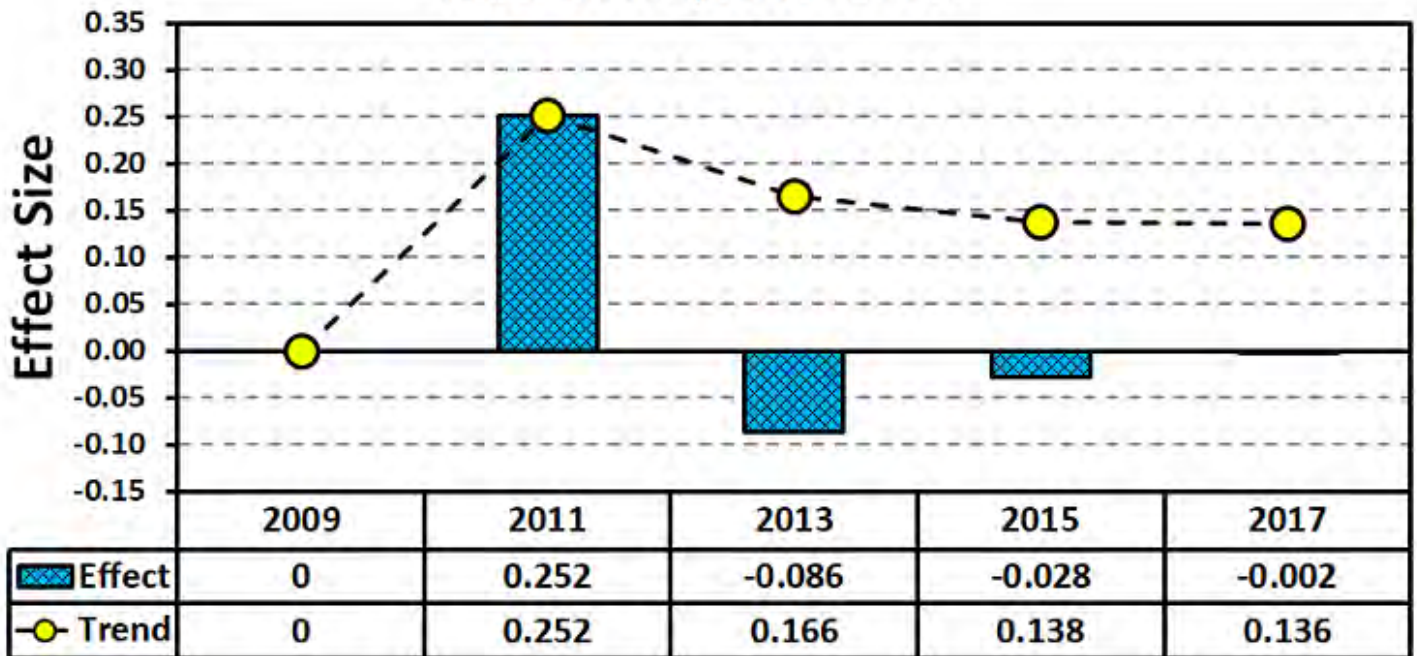


Figure A-31. Idaho effect size results on NAEP reading assessments from 2009 through 2017: Grade 8, Town schools.

NAEP Mathematics, 2009 thru 2017 Idaho Grade 8, Town Schools

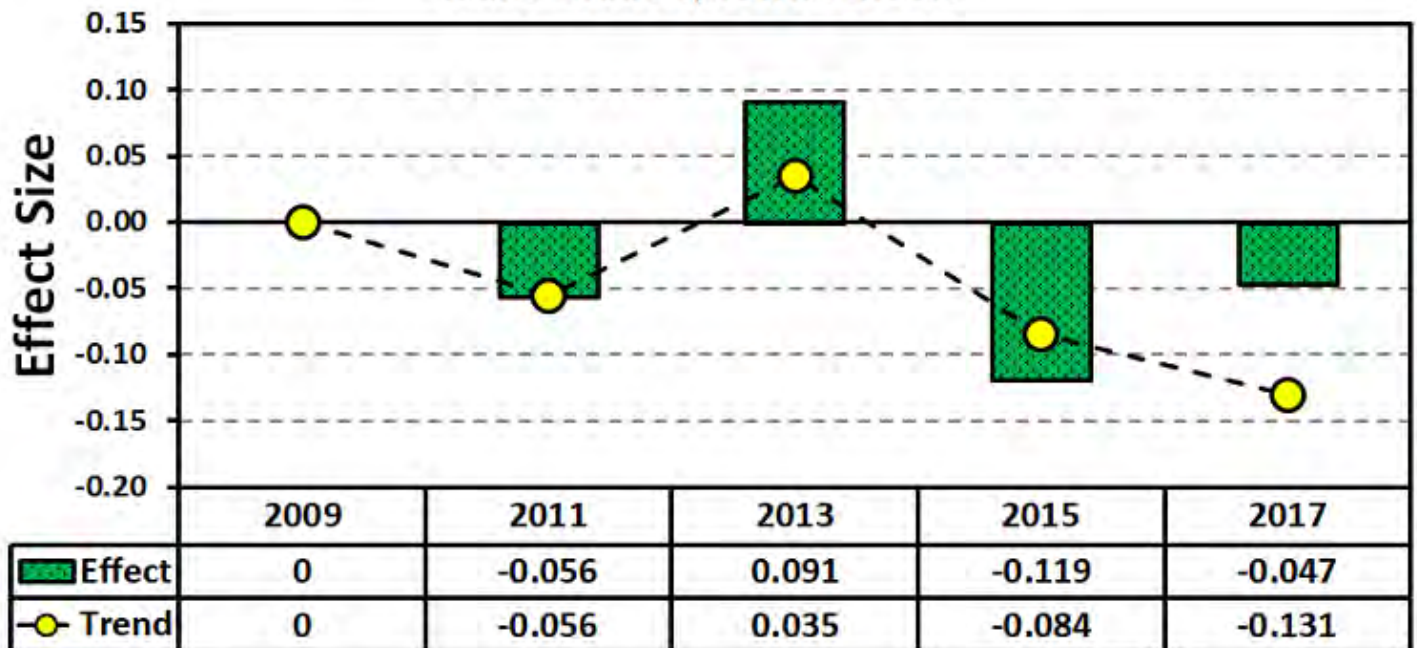


Figure A-32. Idaho effect size results on NAEP mathematics assessments from 2009 through 2017: Grade 8, Town schools.

NAEP Reading, 2009 thru 2017 Idaho Grade 4, Rural Schools

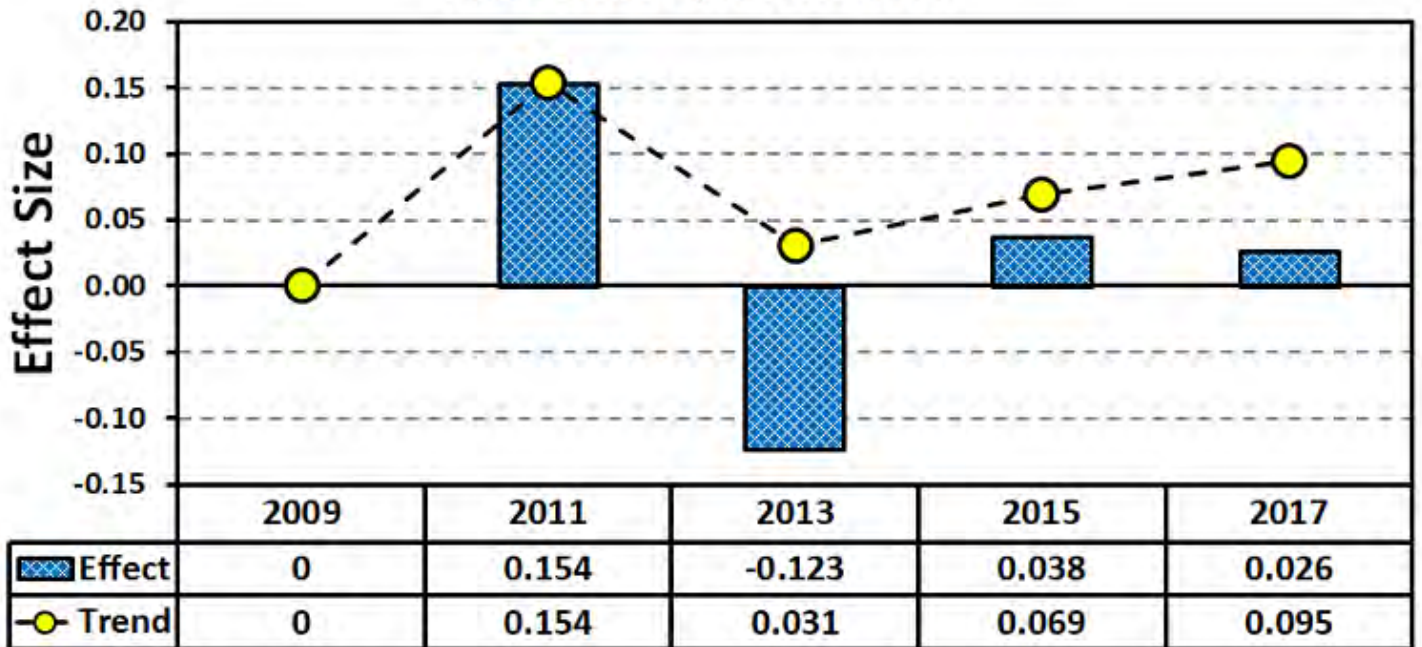


Figure A-33. Idaho effect size results on NAEP reading assessments from 2009 through 2017: Grade 4, Rural schools.

NAEP Mathematics, 2009 thru 2017 Idaho Grade 4, Rural Schools

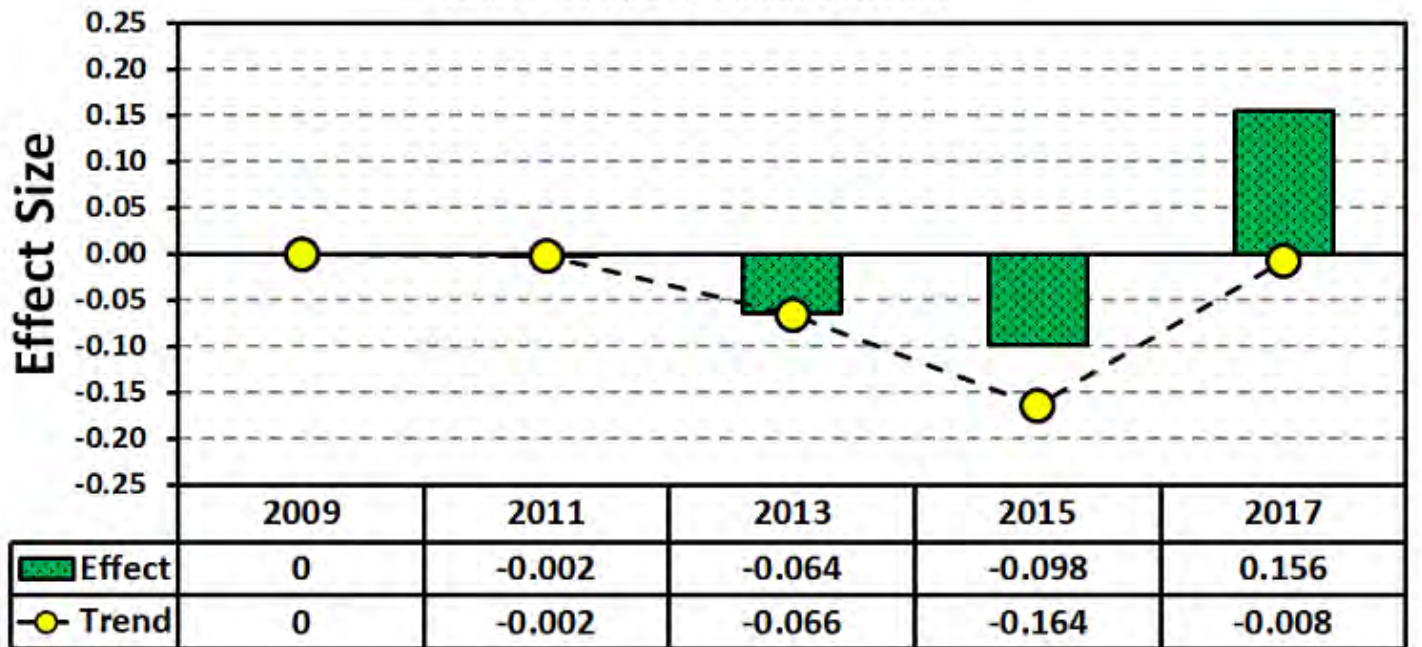


Figure A-34. Idaho effect size results on NAEP mathematics assessments from 2009 through 2017: Grade 4, Rural schools.

NAEP Reading, 2009 thru 2017 Idaho Grade 8, Rural Schools

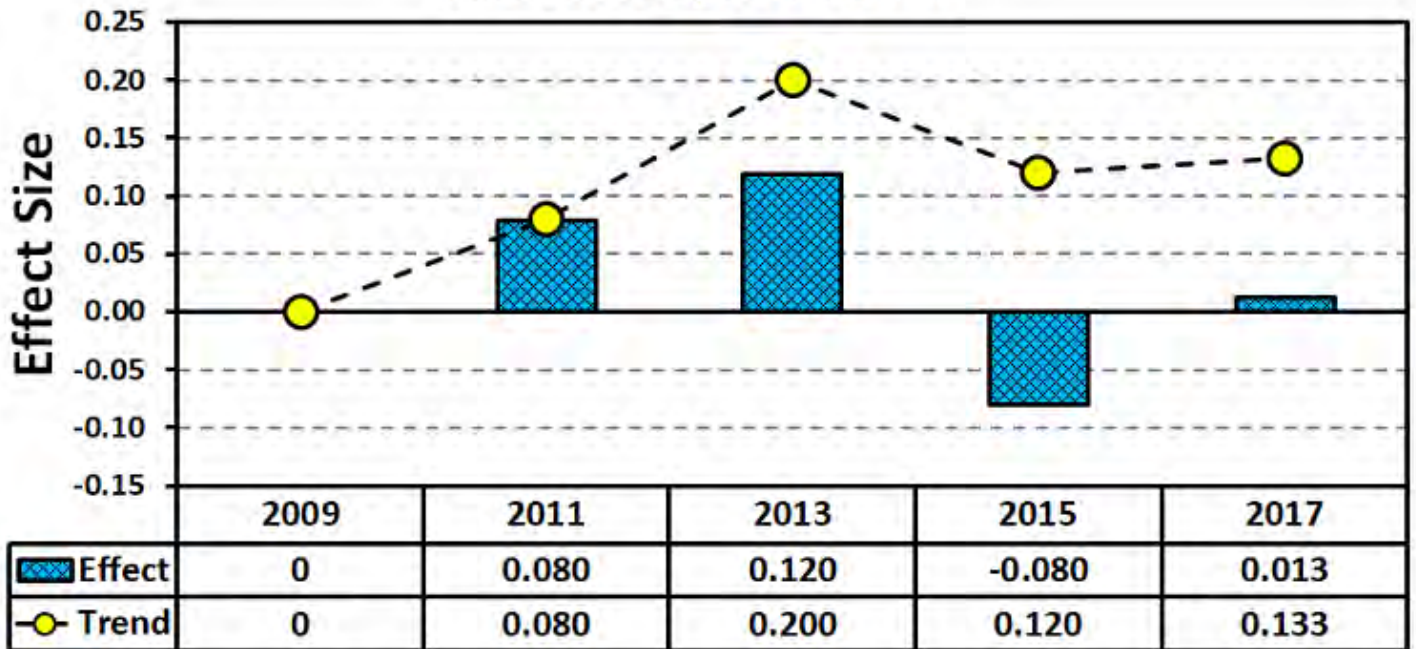


Figure A-35. Idaho effect size results on NAEP reading assessments from 2009 through 2017: Grade 8, Rural schools.

NAEP Mathematics, 2009 thru 2017 Idaho Grade 8, Rural Schools

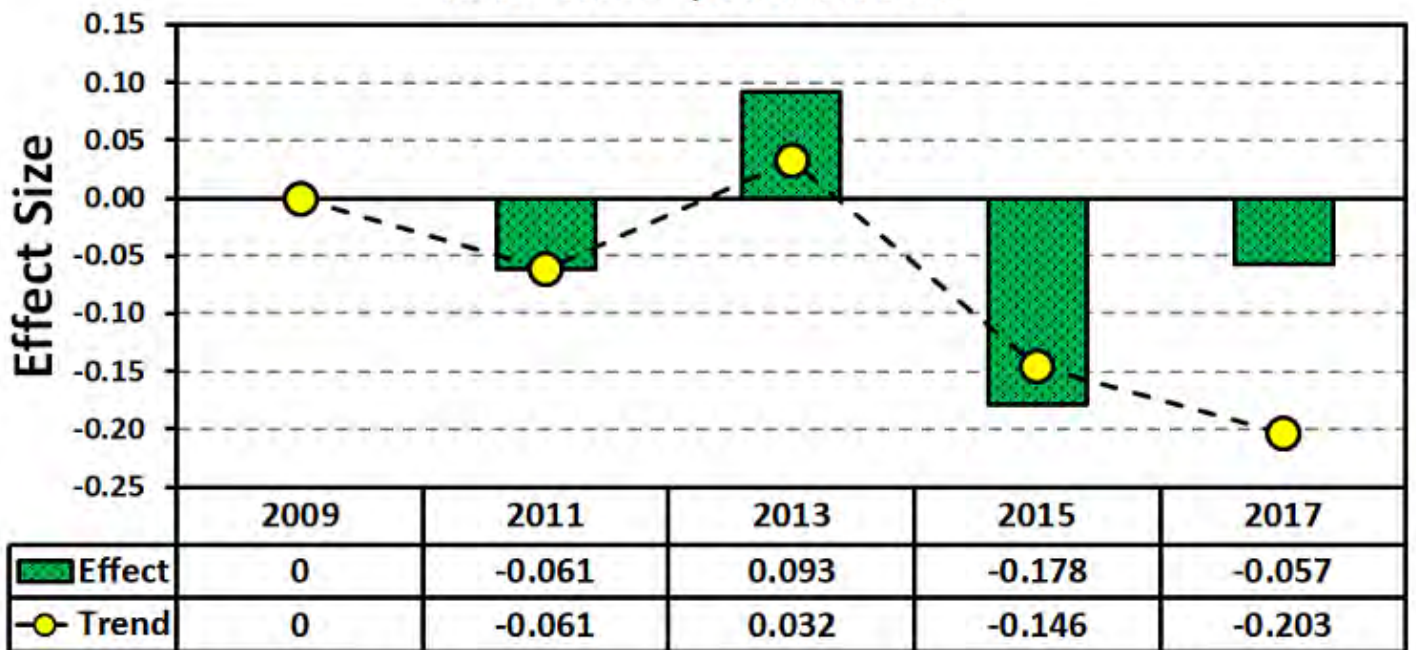


Figure A-36. Idaho effect size results on NAEP mathematics assessments from 2009 through 2017: Grade 8, Rural schools.