Exploring and Explaining Trends in NAEP Mathematics Performance: A Discussion

Peter Kloosterman, Indiana University
Crystal Walcott, Indiana University Purdue University Columbus
Arnulfo Pérez, Ohio State University
Michael Roach, Indiana University
Michael Daiga, Indiana University
Ayfer Eker, Indiana University
Mark Creager, Indiana University
Erik Winarski, St. John Fisher College

We will briefly report results from the Main and LTT mathematics NAEP assessments and then move into discussion of why there are such varying interpretations of the results. Participants will be encouraged to speculate on why we see the trends that we do and what the trends mean for curriculum, teaching, and policy.

These slides and additional publications from the What Mathematics Do Students Know Project are available at http://ceep.indiana.edu/ImplicationsFromNAEP/ (or just Google “What Mathematics Do Students Know”)

This research is supported by the REESE Program of the National Science Foundation (grant number DRL-1008438). Opinions, findings, conclusions, and recommendations are those of the presenters and do not necessarily reflect the views of the National Science Foundation.
Session Format

• Brief description of trends in NAEP results over time (20 minutes)
• Small group discussion of the results and what they might imply for curriculum, teaching, teacher education, and policy (30 minutes)
• Sharing of ideas from each group, overall comments (25 minutes)
National Assessment of Educational Progress (NAEP) Mathematics Programs

- **Main NAEP (1990 to Present)** – grades 4, 8, and 12; results are representative of the entire US population (results available for each state for grades 4 and 8, grade 12 results are only available for a handful of states)

- **Long-Term Trend NAEP (1973 to Present)** – uses different items and testing schedule, national sample only, ages 9, 13, 17
Not good enough: Math, reading scores up slightly

Report on US 4th-, 8th-graders released

WASHINGTON — Sometimes the best isn't good enough: Most American fourth- and eighth-graders still lack basic skills in math and reading despite record-high scores on a national exam.

Yes, today's students are doing better than those who came before them. But the improvements have come at a snail's pace.

The 2013 National Report Card released Thursday finds that the vast majority of the students still are not demonstrating solid academic performance in either math or reading. Stubborn gaps persist between the performances of white children and their Hispanic and black counterparts, who scored much lower.

Overall, just 42 percent of fourth-graders and 35 percent of eighth-graders scored at or above the proficient level in math. In reading, 35 percent of fourth-graders and 35 percent of eighth-graders hit that mark.

Still, as state and federal policies evolve in the post-No Child Left Behind era, the nation's schoolkids are doing better today on the test than they did in the early 1990s, when such tracking started, with more improvement in math than in reading. Students of all races have shown improvement over the years.

The results come from the National Assessment of Educational Progress, or NAEP, which is given every two years to a sample of fourth- and eighth-graders.

This year's results, compared with results in 2011, show average performance gains of about one or two points on a 500-point scale in math and reading in both grades, although the one-point gain in fourth-grade reading was not considered statistically significant.
Main NAEP Mathematics Performance: 1990 to 2013
The media react with alarm every time the NAEP scores appear because only about one-third or so of students are rated “proficient.” This is supposed to be something akin to a national tragedy because presumably almost every child should be “proficient.” Remember, under No Child Left Behind, ALL students are supposed to be proficient in reading and math by the year 2014. (Diane Ravitch Blog, May 14, 2012)
Grade 4 Students Reaching the Proficient Level
Grade 8 Students Reaching the Proficient Level
• Proficient is akin to a solid A…. In math, the proportion in fourth grade who were proficient rose from 18% to 40% in the past twenty years, an absolutely astonishing improvement. In eighth grade, the proportion proficient in math went from 21% in 1992 to an amazing 35% in 2011.

• Below basic is where we really need to worry. These are the students who really don’t understand math or read well at all. The proportion who are below basic has dropped steadily in both reading and math in fourth and eighth grades since 1992. *(Diane Ravitch Blog, May 14, 2012)*
When the scores are broken out by race, you can really see dramatic progress, especially in math. In 1992, 80% of black students in fourth grade were below basic. By 2011, that proportion had dropped to 49%. Among white students in fourth grade math, the proportion below basic fell in that time period from 40% to only 16%. (Diane Ravitch Blog, May 14, 2012)
Grade 4 Students Reaching the Basic Level
Grade 8 Students Reaching the Basic Level
LTT NAEP Mathematics
Performance: 1973 to 2013

* Significantly different ($p < .05$) from 2012.

Legend:
- Extrapolated data
- Original assessment format
- Revised assessment format
LTT NAEP Reading Performance: 1971 to 2012

* Significantly different (p < .05) from 2012.
In a statement released Thursday, Indiana Gov. Mike Pence noted that Indiana was the second fastest improving state in the nation when it came to NAEP test scores. Pence credited the sweeping and controversial education reforms put into place under previous Gov. Mitch Daniels for the rise in test scores. (Maureen Hayden, *New Albany-Jeffersonville News and Tribune*, November 7, 2013)
• NAEP notes that over the past four decades, the demographic makeup of American students has changed considerably, with Hispanic students accounting for a larger proportion and whites accounting for a lower proportion than in the 1970s. At age 13, for example, the proportion of Hispanic students taking NAEP’s math assessment more than tripled between 1978 and 2012, from 6 percent to 21 percent, while the proportion of white students taking the assessment decreased from 80 percent to 56 percent. Black students accounted for 13 percent of those taking the math assessment in 1978, and 15 percent in 2012. (Susan Jones, CNSNews.com, July 2, 2013)
• The results from the 2012 long-term trend National Assessment of Educational Progress show that over the last four decades, our nation has made very real progress for all groups of students. *(Education Trust Press Release, June 27, 2013)*
Since closing the achievement gap was the goal of No Child Left Behind, which went into effect a decade ago under the administration of former president George W. Bush, the NAEP results tell us that it was an abject failure. (Valarie Straus, *Washington Post*, November 1, 2011)
### Grade 4 Black-White Gap

<table>
<thead>
<tr>
<th>Year</th>
<th>Scale Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>'90</td>
<td>188*</td>
</tr>
<tr>
<td>'92</td>
<td>193*</td>
</tr>
<tr>
<td>'96</td>
<td>199*</td>
</tr>
<tr>
<td>'00</td>
<td>203*</td>
</tr>
<tr>
<td>'03</td>
<td>216*</td>
</tr>
<tr>
<td>'05</td>
<td>220*</td>
</tr>
<tr>
<td>'07</td>
<td>222*</td>
</tr>
<tr>
<td>'09</td>
<td>222*</td>
</tr>
<tr>
<td>'11</td>
<td>224</td>
</tr>
<tr>
<td>'13</td>
<td>224</td>
</tr>
</tbody>
</table>

**Score Gap**

- **Accommodations not permitted**
- **Accommodations permitted**

*Note: The asterisk (*) indicates accommodations not permitted.*
Grade 4 Hispanic-White Gap

![Graph showing the score gap between Hispanic and White students from 1990 to 2013. The graph indicates a widening gap with accommodations not permitted.](image)
Grade 4 Performance by Eligibility for Free or Reduced-Price Lunch
Grade 4 Gender Gap

The graph illustrates the score gap between male and female students over the years from 1990 to 2013. The lines represent different score levels, with accommodations not permitted and accommodations permitted, showing a general trend of increasing scores over time for both genders.
Grade 4 Performance by Percentile
Grade 8 Black-White Gap
Grade 8 Hispanic-White Gap

[Graph showing the score gap between Hispanic and White students from 1990 to 2013, with accommodations not permitted and permitted.]
Grade 8 Performance by Eligibility for Free or Reduced-Price Lunch
Grade 8 Gender Gap

[Graph showing the gender gap in scale scores from '90 to '13, with accommodations not permitted and permitted lines, and male and female markers.]
Grade 8 Performance by Percentile
Trends in PISA

<table>
<thead>
<tr>
<th>Year</th>
<th>US Score</th>
<th>OECD Average</th>
</tr>
</thead>
<tbody>
<tr>
<td>2000</td>
<td>493</td>
<td>492</td>
</tr>
<tr>
<td>2003</td>
<td>483</td>
<td>500</td>
</tr>
<tr>
<td>2006</td>
<td>474</td>
<td>494</td>
</tr>
<tr>
<td>2009</td>
<td>487</td>
<td>496</td>
</tr>
<tr>
<td>2012</td>
<td>481</td>
<td>494</td>
</tr>
</tbody>
</table>

Like TIMSS, the PISA scoring system is designed so the that average is dependent on the participating countries and is about 500 for each administration.
2012 PISA Problem-Solving Assessment

• First time PISA has assessed general problem solving (not necessarily mathematics problem solving)
• Results released April 1, 2014
• In contrast to 2012 PISA, where the U.S. was significantly below 19 of the other 33 OECD countries, the U.S. was significantly below only 5 of the 27 OECD countries participating in the problem-solving assessment.
## Trends in U.S. TIMMS Scores

<table>
<thead>
<tr>
<th>Year</th>
<th>Grade 4</th>
<th>Grade 8</th>
</tr>
</thead>
<tbody>
<tr>
<td>1995</td>
<td>518</td>
<td>492</td>
</tr>
<tr>
<td>1999</td>
<td>NA</td>
<td>502</td>
</tr>
<tr>
<td>2003</td>
<td>518</td>
<td>504</td>
</tr>
<tr>
<td>2007</td>
<td>529</td>
<td>508</td>
</tr>
<tr>
<td>2011</td>
<td>541</td>
<td>509</td>
</tr>
</tbody>
</table>

The TIMSS scoring system is designed so that the average score is about 500 for each administration. Although TIMSS uses an equating method to reduce the impact of changes in countries on the average score, trends over time are dependent on trends within each participating country.
NAEP Item Formats

• **Multiple Choice Format**
  - 4\textsuperscript{th} grade – Four choices
  - 8\textsuperscript{th} grade – Five choices
  - 12\textsuperscript{th} grade – Five choices

• **Short Constructed Response**
  Two types:
  1. Students write their answers in the space provided
  2. Multiple questions or a brief rationale (Main NAEP only)

• **Extended Constructed Response**
  Multi-part, scored with focused holistic rubrics (Main NAEP only)
Most Difficult 2008 Released Age 13 LTT Item (29% correct)

• Which of the following shows perpendicular lines?

1. 

2. 

3. 

4. 
Each figure in the pattern below is made of hexagons that measure 1 centimeter on each side.

If the pattern of adding one hexagon to each figure is continued, what will be the perimeter of the 25th figure in the pattern? (36% correct) Show how you found your answer. (18% correct)
Both items were used in grade 4 Main NAEP. Is one significantly harder?

• $N$ stands for the number of hours of sleep Ken gets each night. Which of the following represents the number of hours of sleep Ken gets in 1 week?
  - A. $N + 7$
  - B. $N - 7$
  - C. $N \times 7$
  - D. $N \div 7$

• Each of the 18 students in Mr. Halls class has $p$ pencils. Which expression represents the total number of pencils that Mr. Halls class has?
  - A. $18 + p$
  - B. $18 - p$
  - C. $18 \times p$
  - D. $18 \div p$
Tools Available on the NAEP Website

NAEP Tools and Applications

The NAEP website features a number of applications designed to give users quick and easy access to questions from previous assessments, performance comparisons, and NAEP assessment data for quick or complex analyses; read a brochure, NAEP Tools on the Web (1.1 MB), describing the tools. See more information about each tool below, and print Quick Reference Guides if you are a new user.

Data Explorer

The NAEP Data Explorer (NDE) creates customizable tables and graphics to display NAEP results. See the results of an assessment across multiple years, and broken down across a variety of student groups. For some assessments, results are available by state or participating urban district. Results can be filtered by content areas.

For in-depth exploration, the NDE provides significance testing, gap analysis, and regression analysis. You can export tables and charts to Word documents, Excel workbooks, and PDFs. Special versions of the NDE focus on the High School Transcript Study (HSTS) and the National Indian Education Study (NIES).

More information:

- Watch a short video about the NDE
- Use the Quick Reference Guide (505 KB)
- Learn about NDE features from the tutorial or access Help
Mathematics Questions Search

NAEP Items Released in Recent Years

• At grade 4, 54 items were released in 2007, 31 in 2009, 51 in 2011, and 46 in 2013
• At grade 8, 53 items were released in 2007, 34 in 2009, 47 in 2012, and 47 in 2013
• At grade 12, 53 items were released in 2005 and 27 in 2009 (more items are likely to be released when the 2013 grade 12 results are released)
• The only data on released items are for the year released although with modest overall gains since 2005, performance on most items has been relatively stable since then.
Jurisdiction Data
NAEP Data Explorer
Discussion Topic

• What can explain the patterns in NAEP results?
• What do the results imply for mathematics curriculum, teaching, teacher education, and public policy?
Factors Suggested as Impacts on NAEP Trends

- NCLB/increased use of high-stakes testing
- NCTM Standards/CCSS/changes in curriculum
- State policies and standards
- Increased accountability for teachers
- Economic factors
- Changes in demographics
- Increased time spent on mathematics (more coursework in high school, more time dedicated to mathematics in elementary school)
- Changes in teacher preparation
- Teaching students to be better test takers