National Assessment of Educational Progress
The Nation's Report Card: Science 2009

Jack Buckley
Commissioner
National Center for Education Statistics
January 25, 2011

Today I am releasing the 2009 National Assessment of Educational Progress science results.

Students were assessed at the fourth, eighth, and twelfth grades. Over 156,000 students at grade 4, 151,000 at grade 8, and 11,000 at grade 12 took the assessment. We have national results for public and private school students at all three grades. At grades 4 and 8, we also have results for public school students in 46 states and the Department of Defense schools. The state samples were combined and augmented with sampled students from the four non-participating states plus the District of Columbia, along with a national sample of private school students, to create the full national samples for grades 4 and 8. The twelfth-grade sample is smaller because there are no state-representative samples at that grade.

The 2009 assessment was based on a new framework that emphasized four science practices—identifying science principles, using science principles, using scientific inquiry, and using technological design—that describe *how* students use their scientific knowledge by measuring what they *are able to do* with the science content. The framework also provided an increased focus on the understanding of science principles and asked students to answer questions that cut across the three science content areas: life science, physical science, and Earth and space sciences. At grades 8 and 12, there was a shift in emphasis in the content areas from the previous NAEP science framework.

- At grade 4, the assessment placed equal emphases on all three content areas, as it did in 2005.
- At grade 8, we increased the emphasis on Earth and space sciences and decreased the emphasis on life science.
- At grade 12, we increased the emphasis on physical and life sciences and decreased the emphasis on Earth and space sciences.

Results are reported in two ways: as average scale scores on a 0–300 scale, with a separate scale for each grade; and as percentages of students at or above three achievement levels—*Basic*, *Proficient*, and *Advanced*.

The achievement levels were developed by the National Assessment Governing Board. They set standards for what students should know and be able to do. For each subject and for each grade, the Governing Board has established standards for *Basic*, *Proficient*, and *Advanced* performance. Ultimately, the goal is to have all students performing at or above the *Proficient* level.

Because the 2009 framework represents a substantial change from the previous science framework, results from the 2009 science assessment cannot be compared to those from previous assessments.

Grade 4 Results

Let's begin by examining student performance at grade 4.

Seventy-two percent of fourth-graders performed at or above *Basic*, which includes students at the *Basic*, *Proficient*, and *Advanced* levels. Thirty-four percent scored at or above *Proficient*. One percent scored in the *Advanced* range.

The average scale score for fourth-graders was set at 150 on a 0 – 300 point scale to serve as a reference point for current and future assessments. The average score of 163 for White students was higher than the average for any of the other racial/ethnic groups.

The average score of 151 for male fourth-graders in 2009 was 2 points higher than the score for female fourth-graders.

We have national results for public school students at grade 4, as well as separate results for all private school students, and for private school students attending Catholic schools. Private school students accounted for 9 percent of fourth-graders of which roughly one-half attended Catholic schools. The average scores for all private school students and for Catholic school students were higher than the average score for public school students.

NAEP categorizes schools according to location—city, suburb, town, and rural—based on information from the Census Bureau. In 2009, students attending city schools had an average score of 142, which was lower than the average score of any of the other three groups.

NAEP uses student eligibility for the National School Lunch Program as a measure of family income. Students whose families have an income that is less than 130 percent of the federal poverty level are eligible for free lunches. Those whose families have an income that is between 130 and 185 percent are eligible for reduced-price lunches, while those whose families are above 185 percent of the level are not eligible.

Students who are eligible for free lunches scored lower than the other two groups. Students eligible for reduced-price lunches scored lower than those who are not eligible. About 44 percent of students are eligible for free or reduced-price lunches.

When comparing state results to the nation's, 23 states and the Department of Defense schools had higher average scores, 10 states had lower scores, and 13 states had scores that were not significantly different from the nation.

We examined the scores, and the score gaps, between White and Black students and between White and Hispanic students at the state level. I'll discuss the results for White and Hispanic students at grade 4 and describe the results and score gaps for White and Black students at grade 8.

The White-Hispanic gap for fourth-grade public school students in the nation was 32 points. White students had an average score of 162, while the average score for Hispanic students was 130.

- In 15 states and the Department of Defense schools, the White-Hispanic gap was smaller than the national gap. However, in some cases, the comparatively small gap was the result of higher-performing Hispanic students. In others, it was the result of lower-performing White students.
- In four states (Delaware, Montana, Virginia, and Wisconsin), and the Department of Defense schools, both White and Hispanic students had scores that were higher than their peers nationally, but scores for Hispanic students were high enough to create relatively small gaps. In Virginia, for example, White students had an average score of 172—10 points above the national average for White students—but Hispanic students had an average of 152—22 points higher than the national average for Hispanic students, resulting in a 19-point gap.
- In Arkansas and Nevada, the relatively low performance on the part of White students in comparison to the nation contributed to the narrower gaps. Here, scores for Hispanic students were comparable to the national average for their peers, while the scores for White students were lower.
- In nine states (Florida, Iowa, Kentucky, Louisiana, Maryland, Michigan, Mississippi, South Dakota, and Wyoming), the average scores for Hispanic students were above the national average and scores for White students were either comparable to the national average or below it. The relatively high scores for Hispanic students in these states contributed to the narrower gaps.
- In Connecticut, the score gap between White and Hispanic students was larger than the gap at the national level. While White students in Connecticut scored above the national average for their peers, the score for Hispanic students was comparable to the national average, resulting in the 39-point gap. In the remaining 30 states, the White-Hispanic score gaps were comparable to the gap nationally.

In the report card you will find an "item map" in each grade section with questions from the three science content areas ranked according to difficulty on the 0– 300 science scale.

An item map gives us an idea of the kinds of questions students at different levels of performance can answer. As we move up the scale in the item map, increasing the content difficulty, fewer students are able to answer these types of questions.

For example, one question falling below the *Basic* cut point on the scale asks students to identify the data on a chart. A question in the *Basic* range asks students to recognize an example of a change in a physical state. A question in the *Proficient* range asks students to explain an example of heat transfer. A question in the *Advanced* range asks students to determine the source of sound during an investigation about the pitch of sounds.

A sample grade 4 question in the physical science content area shown in the report, asks students how to determine if two different kinds of cups hold the same volume of water. Thirty-five percent of students chose the correct answer, "Pour all of the water from Cup 1 into Cup 2 to see if the water completely fills Cup 2 without spilling over." Twenty-three percent of students in the below *Basic* range answered the question correctly, compared to 88 percent in the *Advanced* range.

Grade 8 Results

Sixty-three percent of eighth-graders performed at or above the *Basic* level, while 30 percent scored at or above *Proficient*. Two percent scored in the *Advanced* range. As at grade 4, the average score was set at 150.

When we make comparisons of student performance at grade 8, we see many of the same patterns among student demographic groups that we saw at grade 4. For example, male students scored 4 points higher than female students. At grade 8, we ask students the level of education attained by each of their parents, and we then classify the students according to the highest level reached by either parent. Seven percent of students reported that neither parent completed high school, while 49 percent reported that at least one parent had completed college. Higher levels of parental education were associated with higher scores.

Twenty-four states and the Department of Defense schools had higher average scores than the nation, 15 states had lower scores, and 7 states had scores that were not significantly different.

- The score gap between White and Black public school students in the nation was 36 points. The average score for White students was 161 compared with 125 for Black students. In the Department of Defense schools, the average for Black students was high

enough to create a relatively smaller gap compared to the nation. In these schools, scores for both groups were higher than the national averages for White and Black students.

- In Hawaii, Nevada, and West Virginia, relatively low scores for White students contributed to the small gaps.
- In Delaware, Kentucky, and Oregon, relatively high scores for Black students contributed to the small gaps.
- In Arkansas, Illinois, and Wisconsin, relatively low scores for Black students contributed to large gaps of 43 or 44 points.
- In the remaining 37 states for which we have data, the state White-Black score gaps were comparable to the gap nationally.

The item map for grade 8, also found in the report card, gives an indication of what students can and cannot do, depending on where students score on the scale.

The following examples are all drawn from the Earth and space sciences content area. A question shown in the below *Basic* range asked students to describe part of a valid experiment to compare heating rates of different materials. This question ranked at 119, below the cut point for the *Basic* achievement level. Students scoring at or above *Basic* would be likely to answer this question correctly.

A question falling near the bottom of the *Basic* range asked students to identify the mechanism of a weather pattern. Students in the *Basic*, *Proficient,* and *Advanced* ranges, would be likely to answer it correctly.

A question asking students to list soils in the order of their permeability ranked near the upper boundary of the *Proficient* range. At the *Advanced* level, students would be likely to correctly answer a question that asked them to predict the Sun's position in the sky.

In the question about soil permeability, students were asked to identify the relative rates at which water would flow through funnels filled with pebbles, fine sand, or coarse sand. Forty-five percent chose the correct answer, recognizing that water flows more easily through coarse materials.

Grade 12 Results

At grade 12, we have national results only. Sixty percent of students performed at or above the *Basic* level, and 21 percent performed at or above *Proficient*. One percent performed at *Advanced*. Many of the comparisons of student groups at grade 12 are similar to those at grade 8.

For example, White and Asian/Pacific Islander students had higher average scores than students in the other three racial/ethnic groups.

We asked grade 12 students what science courses they had completed or were taking. Thirty-four percent said they were taking or had taken three courses—biology, chemistry, and physics; 38 percent reported taking biology and chemistry; and 28 percent reported taking biology only or other science courses.

On average, students who took more advanced science courses had higher scores. The average score for all students who took all three courses was 166, compared to the overall average of 150. (However, this doesn't necessarily mean that increasing course requirements will guarantee higher scores. Students may take more science courses because they do well in the subject, rather than doing well because they take more courses.)

Fifty-eight percent of Asian/Pacific Islander students reported taking biology, chemistry, and physics, which was higher than the percentages for White, Black, or Hispanic students. When comparing performance by race/ethnicity among students who took all three science courses, White and Asian/Pacific Islander students had higher scores than Black or Hispanic students.

For grade 12, I will highlight individual questions in each achievement-level range from the life science content area. For example, a question in the below *Basic* range asked students to draw a conclusion about population growth based on data. A question asking students to determine relationships between species based on an evolutionary tree fell just above the *Basic* cut point. Again, students scoring at or above *Basic* were likely to answer this question correctly.

Students scoring in the *Proficient* range could correctly answer a question that asked them to "evaluate two methods to help control an invasive species."

The final example, asking students to critique a conclusion about photosynthesis, was quite difficult. Only students scoring at *Advanced* were likely to receive full credit on a question of this difficulty. One question, from the life science content area, asked students to consider an experiment involving the growth patterns of aerobic bacteria in the vicinity of a strand of algae capable of photosynthesis after light rays of varying colors were passed through the algae. To answer this question correctly, students had to understand the significance of this phenomenon, to know that aerobic bacteria require oxygen to multiply, and that plant cells, when using energy from light rays to engage in photosynthesis, release oxygen as a by-product. Where there are no bacteria, there is no oxygen.

Students were asked whether they thought the results of the experiment justified a conclusion that the algae used in the experiment were green. They were also asked to explain their conclusion, using results from the experiment to support their answer.

To receive the highest rating of "Complete," students had to agree with the conclusion that the algae were green. In their answer, they had to explain that green light is not used or is less

effective for photosynthesis than other forms of visible light, that few bacteria were found in the region where the algae was exposed to green light, and that this indicated that green light was reflected or not absorbed by the algae.

One percent of students received a “Complete” rating on this item. Three percent received the next highest rating of “Essential,” while 19 percent received a “Partial” rating and 71 percent received an “Unsatisfactory/Incorrect” rating. Six percent omitted the question.

There is much more information on student performance in the 2009 Science Report Card.

In addition, the NAEP website gives you extensive information on the performance of students at both the national and state levels for grades 4 and 8 and at the national level for grade 12, along with access to released questions through NAEP’s Questions Center and access to the NAEP Data Explorer, our online data analysis tool.

Later this year, I will be releasing a second science report based on this assessment. In 2009, some students participated in an extended assessment, performing hands-on tasks, which require students to use scientific equipment and materials, or interactive computer tasks, which exploit the resources of modern computers to allow students to engage in a variety of forms of scientific investigation and analysis. This second report will give the results for student performance on these special tasks at all three grades.

In conclusion, I would like to thank all the students and schools who participated in the assessment.