NATIONAL ASSESSMENT OF EDUCATIONAL PROGRESS (NAEP) SCIENCE ASSESSMENT

Summary of Frameworks and Assessment Design

Chronology of NAEP Science assessment administration

- National sample at grade 12 in 1996, 2005, and 2009. (Science was assessed at grade 12 in 2000, but Florida did not participate.) Next assessment will be in 2015.
- National and State samples:
    - Grade 4 in Florida in 1996, 2005, and 2009. (Science was assessed in 2000 but Florida did not participate). Next assessment will be in 2015.
    - Grade 8 in Florida in 2005, 2009, and 2011. (Science was assessed in 2000 but Florida did not participate.) Next assessment will be in 2015.
- TUDA samples:
    - TUDA grades 4 and 8 in 2005, 2009, and 2011. Next assessment will be in 2015.
- NAEP Science is usually administered every four years and participation is at the discretion of individual state education chiefs. Participation in Science is not required under the No Child Left Behind (NCLB) 2001 Federal legislation.

Students chosen to be assessed in NAEP

- NAEP assessments are administered to a stratified random sample of students from grades 4, 8, and 12.
- Both public school and nonpublic school students in grades 4 and 8 are assessed at the national level.
- At the state level, only the results of public school students are reported.
- Forty-seven jurisdictions participated in NAEP 2009 Science—46 states and the Department of Defense Activity schools (Alaska, Kansas, Nebraska, Vermont, and the District of Columbia Public Schools did not participate).
- Accommodations are offered to English language learners (ELLs), students with a section 504 plan, and students with disabilities (SD) who have Individual Education Plans (IEPs). The most typical accommodations include:
    - extra testing time
    - individual or small-group administrations
    - large-print booklets

Science content assessed

- All NAEP frameworks are developed under the guidance of the National Assessment Governing Board (NAGB).

- A new NAEP Science Framework was used for the 2009 NAEP Science assessment. This will start a new trendline as a different science framework was used by NAEP for the 1996, 2000, and 2005 Science assessments.
- A framework specifies what is to be assessed and how it is to be assessed. The new science framework can be accessed at http://www.nagb.org/publications/frameworks/science-09.pdf.
- A NAEP framework does not specify how a subject should be taught nor does it prescribe a particular curricular approach to teaching.
- The NAEP science framework specifies
    - three broad areas of content:
        1. Physical Science – matter, energy, and motion
        2. Life Science – structures and functions of living systems and changes in living systems
        3. Earth and Space Sciences – Earth in space and time, Earth structures, and Earth systems
    - four science practices used to generate student performance expectations:
        1. Identifying Science Principles
            - describe, measure, or classify observations
            - state or recognize correct science principles
            - demonstrate relationships among closely related science principles
            - demonstrate relationships among different representations of principles
        2. Using Science Principles
            - explain and predict observations of phenomena
            - suggest examples of observations that illustrate a science principle
            - propose, analyze, and/or evaluate alternative explanations or predictions
        3. Using Scientific Inquiry
            - design or critique aspects of scientific investigations
            - conduct scientific investigations using appropriate tools and techniques
            - identify patterns in data and/or relate patterns in data to theoretical models
            - use empirical evidence to validate and/or relate patterns in data to theoretical models
        4. Using Technology Design
            - propose or critique solutions to problems, given criteria and scientific constraints

        - identify scientific tradeoffs in design decisions and choose among alternative solutions
        - apply science principles or data to anticipate effects of technological design decisions

Science Assessment Design

- The Science Assessment takes approximately 90 minutes to administer. This includes two separately timed blocks of items, each 25 minutes in length, and background questions.
- Students are given assessment booklets containing three types of items. A small sample of the students receive extra time to participate in the hands-on performance or interactive computer tasks:
    - multiple-choice items
        - students read, reflect, and then select an answer from four alternatives provided
    - constructed-response items
        - short constructed-response – students must generate relevant information
        - extended constructed-response – involve multiple content statements, practices, and/or cognitive demands
        - concept-mapping tasks – measure student ability to make reliable and valid assessments based on ability to make connections among science principles
    - combination items
        - item clusters – provide opportunities to assess students' understanding of a particular key science principle at some depth
        - predict-observe-explain (POE) item sets
            - situation is described
            - student provides a prediction of what will happen
            - student provides an explanation for what appears to be an anomaly
        - hands-on performance tasks
            - student manipulates selected physical objects and tries to solve a scientific problem involving objects
            - student combines science knowledge with the investigative skills that reflect the nature of science and inquiry
        - interactive computer tasks – four possible types
            - information search and analysis
            - empirical investigation
            - simulation
            - concept mapping

- Testing time on NAEP Science is divided evenly among multiple-choice items and the three types of constructed-response items (short-constructed response, extended constructed-response, and concept-mapping tasks).
- Tests at each grade level should contain at least one of each of the combination items.
- Because each block is spiraled with other blocks and is administered to a representative sample of students, the results can be combined to produce average group and subgroup results based on the entire assessment.
- Background questions gather information that is used for analyzing a number of student demographic and instructional factors related to student achievement. Questions such as "How many books are there in your home?" appear with four different number ranges for the student to choose from.

Distribution of Items on Science Assessment

The distribution of items among each content area differs by grade level to reflect the knowledge and skills appropriate for each. The distribution also reflects the relative importance and value given to each content area.

Distribution of Items by Content Area and Grade

Target Percentage of Student Response Time

Content Area	Grade 4 (%)	Grade 8 (%)	Grade 12* (%)
Physical Science	33.3	30.0	37.5
Life Science	33.3	30.0	37.5
Earth and Space Sciences	33.3	40.0	25.0

***** These recommendations are based on NAEP data regarding students' course-taking patterns in 12th grade. If these patterns change substantially after 2009, the recommendations might be reconsidered.

Distribution of Items by Science Practice and Grade

Target Percentage of Student Response Time

Science Practices	Grade 4 (%)	Grade 8 (%)	Grade 12 (%)
Identifying Science Principles	30	25	20
Using Science Principles	30	35	40
Using Scientific Inquiry	30	30	30
Using Technological Design	10	10	10

Performance expectations are derived from the intersection of content statements and science practices. By comparing student responses with the particular science content and practice being assessed, inferences about what students know (about particular science principles) and can do (with respect to particular science practices) are made.

Cognitive Demands

- four cognitive demands are used in item development and analyzing student responses:
    1. knowing that – declarative knowledge
    2. knowing how – procedural knowledge
    3. knowing why – schematic knowledge
    4. knowing when and where to apply knowledge – transfer of current knowledge to new tasks or problems
- used as a lens to facilitate item development
- provides framework to analyze student responses
- one or more cognitive demands associated with each Science Practice
- balance across the Science Practices ensures a range of cognitive demands will be covered in items

Scoring criteria for constructed-response items used on all NAEP assessments

- Unique scoring guides are developed for each question.
- Scoring guides describe the specific criteria for assigning a score level to student responses.
- Scoring process:
    - Expert scorers are extensively trained on how to determine what score level to assign to student responses
    - Scoring is monitored to ensure that scoring standards are being adhered to reliably
    - Monitoring measures the consistency of scoring the same items administered in different assessments and thus ensures consistency of application standards across assessment years
    - The percentage of exact agreement among raters of the same student responses are tracked

Reporting NAEP scores on all NAEP assessments

- Results are used to compile national, state, and selected urban district data. No results are generated for schools or individual students.
- National results reflect the performance of all grades 4, 8, and 12 students in public schools, private schools, Bureau of Indian Education (BIE) schools, and Department of Defense Activity schools.

- National public and state results reflect the performance of students in public schools in grades 4 and 8.
- Overall results are reported for various groups of students: by race/ethnicity, eligibility for free/reduced-price lunch, students with disabilities (SD), English language learners (ELL), and gender.
- NAEP reports scores in two different ways: average scale scores and achievement levels. Both scores are based on the performance of samples of students, not the entire population.
- Science average scale scores indicate how much a student knows and can do based on a 0–300 scale. The scores are reported as:
    - Average scale scores (range from 0–300)
    - Percentiles (10th, 25th, 50th, 75th, and 90th), which show trends in performance for lower-, middle-, and higher-performing students
- Achievement levels
    - offer a means of identifying percentages of students who have demonstrated certain proficiencies
    - are performance standards based on scale scores and define the degree to which student performance meets expectations of what students should know and be able to do
    - are set by NAGB
        - *Advanced* represents superior performance
        - *Proficient* represents solid academic performance--students reaching this level have demonstrated competency over challenging subject matter
        - *Basic* denotes partial mastery of prerequisite knowledge and skills that are fundamental for proficient work at a given grade
    - Below *Basic* is reported, but is not considered to be an achievement level.
    - Achievement-level descriptors for grade 4 science can be found at http://nces.ed.gov/nationsreportcard/science/achieveall.asp#grade4.
    - Achievement-level descriptors for grade 8 science can be found at http://nces.ed.gov/nationsreportcard/science/achieveall.asp#grade8.
    - Achievement-level descriptors for grade 12 science can be found at http://nces.ed.gov/nationsreportcard/science/achieveall.asp#grade12.

Interpreting NAEP scores for all NAEP assessments

- Differences between average scale scores or between achievement-level percentages are discussed in NAEP reports only when they are statistically significant. Statistically significant means it is unlikely the differences in scores occurred by chance. The differences are referred to as “significant differences” or as being “significantly different.”

- NAEP assesses a representative sample of students in each state and in selected urban districts. The number of students tested determines the standard error for a particular jurisdiction. Because of sample design, performance standard error must be considered in reporting NAEP results. Statistical tests that factor in the standard errors are used to determine whether the differences are significant at the 0.05 level.
- Estimates based on smaller groups are likely to have relatively large standard errors. In these cases, some seemingly large differences may not be statistically significant.
- Results data for all NAEP assessments can be found at the NAEP Data Explorer Web site at http://nces.ed.gov/nationsreportcard/nde. Reports of Florida's NAEP results can be found at http://www.fldoe.org/asp/naep/.