DOCUMENT RESUME

ED 384 639 TM 023 805

| | |
|---|---|
| AUTHOR | O'Sullivan, Christine |
| TITLE | The 1993 NAEP Science Field Test: Hands-On Tasks and Test Specifications. |
| INSTITUTION | Educational Testing Service, Princeton, N.J. |
| SPONS AGENCY | National Center for Education Statistics (ED), Washington, DC. |
| PUB DATE | Apr 95 |
| NOTE | 8p.; Paper presented at the Annual Meeting of the National Council on Measurement in Education (San Francisco, CA, April 19-21, 1995). |
| PUB TYPE | Reports - Evaluative/Feasibility (142) -- Speeches/Conference Papers (150) |
| EDRS PRICE | MF01/PC01 Plus Postage. |
| DESCRIPTORS | Constructed Response; *Educational Assessment; Elementary Secondary Education; *Field Tests; Grade 4; Grade 8; Grade 12; *Models; Multiple Choice Tests; National Surveys; Problem Solving; Sciences; *Science Tests; *Test Construction |
| IDENTIFIERS | *National Assessment of Educational Progress; Performance Based Evaluation; *Test Specifications |

ABSTRACT

The framework and specifications for the National Assessment of Educational Progress (NAEP) science assessment were developed in 1991-92 and field tested in 1993. The assessment was postponed, however, and will be administered in 1996. The framework calls for performance-based tasks that probe students' abilities to use materials to make observations, perform investigations, evaluate experimental results, and apply problem-solving skills, as well as constructed response and multiple choice items that explore student abilities. The core of the science framework consists of a three-by-three matrix that describes earth, physical, and life science and conceptual understanding, scientific investigation, and practical reasoning. The hands-on tasks were designed and developed in prepackaged kits for administration to large numbers of students. The field tests demonstrated that, given the constraints and challenges of task development, those that were selected for the national assessment did meet the test specifications. Three tables illustrate test characteristics. (SLD)

# The 1993 NAEP Science Field Test: Hands-On Tasks and Test Specifications

Christine O'Sullivan
Science Coordinator
National Assessment of Educational Progress
Educational Testing Service, Princeton, NJ 08541



Paper presented as part of the symposium "Large-Scale Science Performance Assessment Results: Informing Task and Score development" at the 1995 annual meeting of the National council of Measurement in Education in San Francisco, California, April, 1995. This work is funded by the National Center for Education Statistics.

## Introduction

The National Assessment of Educational Progress (NAEP) is the government's primary indicator of what students at three grade/age levels, grade 4/age 9, grade 8/age 13, and grade 12/age 17, know and can do in different scholastic areas. In addition to questions in specific academic areas, NAEP also gathers background data on student, teacher, and school background variables that may be related to subject-area performance.

The framework and specifications for the NAEP science assessment were developed in 1991-1992 under the auspices of the National Assessment Governing Board through a consensus process managed by the Council of Chief State School Officers. The questions for the assessment were field tested in 1993 and based on the results, the science assessment was assembled for administration in 1994. The assessment, however, was postponed and will be administered in 1996.

## The Science Framework and Specifications

The science framework is based on a twofold view: that scientific knowledge should be organized in a structure that connects discrete pieces of information in a meaningful way, and that science proficiency depends on a student's ability to know and integrate facts into larger concepts and themes using the tools, procedures, and reasoning processes of science. As an outgrowth of this view, the framework calls for NAEP's science assessment to include:

- performance-based tasks that probe students' abilities to use materials to make observations, perform investigations, evaluate experimental results, and apply problem-solving skills (no less than 30 percent of the assessment time), and

- constructed-response and multiple-choice questions that explore students' abilities to explain, integrate, apply, reason, plan, design, evaluate, and communicate (no more than 70 percent of the assessment time)

The core of the science framework consists of a three-by-three matrix that describes three major fields of science; earth, physical, and life, and three elements of knowing and doing science; conceptual understanding, scientific investigation, and practical reasoning. In addition to these main dimensions, the framework includes two additional categories that describe science - the nature of science (which includes technology) and the organizing themes of science (models, systems, and patterns of change). The framework can be summarized as shown in Table 1.

**Table 1 - SCIENCE FRAMEWORK**

| | FIELDS OF SCIENCE | | |
|---|---|---|---|
| **COGNITIVE DOMAINS** | Earth | Physical | Life |
| Conceptual Understanding | | | |
| Scientific Investigation | | | |
| Practical Reasoning | | | |
| | Nature of Science | | |
| | **Themes**<br>Models, Systems, Patterns of Change | | |

Table 2 shows the framework in terms of recommended assessment time.

**Table 2 - NAEP SCIENCE ASSESSMENT FRAMEWORK SUMMARY**

| | Grade 4 | Grade 8 | Grade 12 |
|---|---|---|---|
| Life Sciences | 33% | 40% | 33% |
| Physical Sciences | 33% | 30% | 33% |
| Earth Sciences | 33% | 30% | 33% |
| Conceptual Understanding | 45% | 45% | 45% |
| Scientific Investigation | 45% | 30% | 30% |
| Practical Reasoning | 10% | 25% | 25% |
| Nature of Science/Technology | 15% | 15% | 15% |
| Themes | 30% | 30% | 30% |
| Paper-and-pencil | Not more than 70% | Not more than 70% | Not more than 70% |
| Performance-based | Not less than 30% | Not less than 30% | Not less than 30% |
| Multiple-choice | Not more than 50% | Not more than 50% | Not more than 50% |
| Short constructed-response | Not less than 33% | Not less than 33% | Not less than 33% |
| Extended constructed-response | Not less than 17% | Not less than 17% | Not less than 17% |

**Development of Hands-on Tasks**

A number of challenges and constraints faced test developers, science specialists, scoring specialists, and measurement specialists during the development of hands-on tasks. Briefly the challenges included

- the large number of students who could be assessed at one time
- the location of the assessment
- cost of materials
- standardization of equipment and chemicals
- shelf-life of chemicals
- reading burden (appropriate reading level)
- questions that were independent of one another
- questions that could be scored reliably
- scoring guides that were comprehensive allowing for non-standard answers and
- tasks that met the objectives of the science framework

The constraints included

- safety regulations of each state
- no toxic or corrosive chemicals
- no live organisms (except dormant) and
- no equipment requiring an electrical outlet

The assessment is often administered to fairly large groups of students at one time (30-100 students) in settings such as school cafeterias with only one or two assessment supervisors. Thus, the types of materials and equipment that can be used by students in their tasks is somewhat limited. All materials and equipment for each task had to be included in a pre-packaged kit of manageable size. The cost of the kit for each student had to be no more than $10, and preferably a lot lower. In addition, because materials had to be shipped to assessment supervisors well in advance of the assessment, and to areas of different temperatures (the assessment is administered in January and February), no task could involve materials that had a limited shelf-life, or were influenced by temperature changes. Also no task could include live organisms (with the exception of dormant material), toxic or corrosive chemicals, flames or other heat source, or equipment requiring an electrical outlet (although batteries could be used).

In addition to these considerations, the tasks themselves had to contain enough information for students to do them without them asking questions of the administrator, who was only allowed to ascertain that students had the correct materials and equipment. A general framework for all tasks was developed and task developers worked within this framework. Specifically, each task has introductory information explaining briefly what is to be accomplished in the task; a diagram of the materials and equipment in each packet; directions for accomplishing the task written in language that is developmentally appropriate; and questions, both multiple choice and constructed response. Whilst this shell is standard, there are differences in the way the directions and questions are scaffolded. Some tasks are carefully scaffolded, students being prompted to write down responses as they proceed through the task; in others the students have to complete the task and then answer the questions; a third type presents a problem that students have to solve given certain materials and equipment.

Each task that was eventually selected for field testing had undergone rigorous pre-testing. The tasks were first administered on a one-to-one basis and then piloted in various schools. Students were questioned at each stage about the appropriateness and clarity of language and also monitored closely whilst the tasks were performed.

## The 1993 Science Field Test

The purpose of the 1993 field test was to administer a large set of items and tasks so that those with the best statistical properties could be selected for the 1994 assessment. To obtain the exercises for the operational science assessment, approximately twice as many exercises as were needed were field tested at grade four. At grades eight and twelve, approximately 40 percent more exercises than were needed were field tested. In total, the 1993 science field test contained 673 items, of which 450 were constructed-response items.

Table 3 shows the number of blocks of items field tested. Each block at grade 4 consists of set of questions that take 20 minutes to complete and each block at grades 8 and 12 consists of a set of questions that take 30 minutes to complete.

**Table 3 - NUMBER OF BLOCKS IN THE SCIENCE FIELD TEST**

| | Paper/Pencil | Performance | Total Blocks |
|---|---|---|---|
| 4 only | 10 | 5 | 15 |
| 4/8 overlap | 4 | 2 | 6 |
| 8 only | 6 | 3 | 9 |
| 8/12 overlap | 4 | 2 | 6 |
| 12 only | 10 | 5 | 15 |
| Total | 34 | 17 | 51 |

## Challenges of Scoring

Four hundred and fifty constructed-response questions were scored at National Computer Systems (NCS) in Iowa City. Five hundred students answered each question. Scoring guides for each question had been prepared during the development of the questions and tasks, however, these did have to be revised when student responses were seen. The scoring guides were analytical in nature and usually consisted of 3 or 4 points. Prior to the scoring, training packets made up of anchor papers, practice papers, qualification papers, and calibration papers were assembled from student responses. Four tables of ten scorers took seven weeks to score the questions. Each table had a trainer who was responsible for training the scorers and a table leader who was responsible for back reading and making certain that the scorers were scoring according to the scoring guide. The constructed-response questions in each block of items were trained as a unit, then the scorers scored each students block of questions. The questions in the hands-on tasks were mostly constructed-response and presented their own special scoring challenges, for example, when conclusions based on results were asked for and the results were not the expected ones, judgements had to be made about the plausibility of the conclusions based on the discrepant results. The scorers always did the tasks prior to scoring them. This enabled them to see that

- the equipment and materials did not always behave as expected
- there were not always right answers

Despite the challenges, interrater reliability was for the most part excellent - above 80% and in the majority of questions above 90%.

## Conclusion

Given the constraints and challenges that were encountered during the development of the hands-on tasks, those that were selected for the national assessment do meet the requirements stipulated in the science framework. Specifically they "probe students' abilities to use materials to make observations, perform investigations, evaluate experimental results, and apply problem-solving skills."

Reference
National Assessment Governing Board, Science Framework for the 1994 National Assessment of Educational Progress (no date, pre-publication draft; 202-357-6938).

# END

## U.S. Dept. of Education

Office of Educational
Research and Improvement (OERI)

# ERIC

## Date Filmed
## November 17, 1995

National Council on Measurement in Education
April 19-21, 1995

TM023605

**U.S. DEPARTMENT OF EDUCATION**
*Office of Educational Research and Improvement (OERI)*
*Educational Resources Information Center (ERIC)*

ERIC

# REPRODUCTION RELEASE

(Specific Document)

## I. DOCUMENT IDENTIFICATION:

Title: The 1993 NAEP Science Field Test: Hands-On Tasks and Test Specifications

Author(s): Christine O'Sullivan

Corporate Source: Educational Testing Service

Publication Date:

## II. REPRODUCTION RELEASE:

In order to disseminate as widely as possible timely and significant materials of interest to the educational community, documents announced in the monthly abstract journal of the ERIC system, *Resources in Education* (RIE), are usually made available to users in microfiche, reproduced paper copy, and electronic/optical media, and sold through the ERIC Document Reproduction Service (EDRS) or other ERIC vendors. Credit is given to the source of each document, and, if reproduction release is granted, one of the following notices is affixed to the document.

If permission is granted to reproduce the identified document, please CHECK ONE of the following options and sign the release below.

☐ Sample sticker to be affixed to document

Sample sticker to be affixed to document ☐

**Check here**

Permitting microfiche (4" x 6" film), paper copy, electronic, and optical media reproduction

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

__________

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

Level 1

"PERMISSION TO REPRODUCE THIS MATERIAL IN OTHER THAN PAPER COPY HAS BEEN GRANTED BY

__________

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

Level 2

**or here**

Permitting reproduction in other than paper copy

## Sign Here, Please

Documents will be processed as indicated provided reproduction quality permits. If permission to reproduce is granted, but neither box is checked, documents will be processed at Level 1.

"I hereby grant to the Educational Resources Information Center (ERIC) nonexclusive permission to reproduce this document as indicated above. Reproduction from the ERIC microfiche or electronic/optical media by persons other than ERIC employees and its system contractors requires permission from the copyright holder. Exception is made for non-profit reproduction by libraries and other service agencies to satisfy information needs of educators in response to discrete inquiries."

| | |
|---|---|
| Signature: [signature] | Position: Science Coordinator, NAEP |
| Printed Name: Christine O'Sullivan | Organization: Educational Testing Service |
| Address: P.O. Box 6710<br>Princeton, NJ 08541-6710 | Telephone Number: ( ) 609-734-1918<br>Date: May 10, 1995 |

**You can send this form and your document to the ERIC Clearinghouse on Assessment and Evaluation. They will forward your materials to the appropriate ERIC Clearinghouse. ERIC/AERA Acquisitions, ERIC Clearinghouse on Assessment and Evaluation, 210 O'Boyle Hall, The Catholic University of America, Washington, DC 20064, (800) 464-3742**

CUA

THE CATHOLIC UNIVERSITY OF AMERICA
*Department of Education, O'Boyle Hall*
*Washington, DC 20064*
*202 319-5120*

March 1995

Dear NCME Presenter,

Congratualations on being a presenter at NCME. The ERIC Clearinghouse on Assessment and Evaluation would like you to contribute to ERIC by providing us with a written copy of your presentation. Submitting your paper to ERIC ensures a wider audience by making it available to members of the education community who could not attend the session or this year's conference.

Abstracts of papers that are accepted by ERIC appear in RIE and are announced to over 5,000 organizations. The inclusion of your work makes it readily available to other researchers, provides a permanent archive, and enhances the quality of RIE. Your contribution will be accessible through the printed and electronic versions of RIE, through the microfiche collections that are housed at libraries around the country and the world, and through the ERIC Document Reproduction Service.

We are gathering all the papers from the NCME Conference. We will route your paper to the appropriate clearinghouse and you will be notified if your paper meets ERIC's criteria. Documents are reviewed for contribution to education, timeliness, relevance, methodology, effectiveness of presentation, and reproduction quality.

To disseminate your work through ERIC, you need to sign the reproduction release form on the back of this letter and include it with two copies of your paper. You can drop of the copies of your paper and reproduction release form at the ERIC booth (615) or mail to our attention at the address below. Please feel free to copy the form for future or additional submissions.

Mail to: NCME 1995/ERIC Acquisitions
The Catholic University of America
O'Boyle Hall, Room 210
Washington, DC 20064

Sincerely,

Lawrence M. Rudner, Ph.D.
Director, ERIC/AE

ERIC® Clearinghouse on Assessment and Evaluation