:ies NATIONAL CENTER FOR EDUCATION STATISTICS
Institute of Education Sciences

NAEP
NATIONAL ASSESSMENT OF EDUCATIONAL PROGRESS

U.S. Department of Education
Institute of Education Sciences
NCES 2007–466

The Nation's Report Card™

Problem Solving in Technology-Rich Environments

A Report From the NAEP Technology-Based Assessment Project, Research and Development Series

The National Assessment of Educational Progress

What Is The Nation's Report Card™?

The Nation's Report Card™ informs the public about the academic achievement of elementary and secondary students in the United States. Report cards communicate the findings of the National Assessment of Educational Progress (NAEP), the only continuing and nationally representative measure of achievement in various subjects over time. The Nation's Report Card™ compares performance among states, urban districts, public and private schools, and student demographic groups.

For over three decades, NAEP assessments have been conducted periodically in reading, mathematics, science, writing, history, geography, and other subjects. By making objective information available on student performance at the national, state, and local levels, NAEP is an integral part of our nation's evaluation of the condition and progress of education. Only information related to academic achievement and relevant variables is collected. The privacy of individual students is protected, and the identities of participating schools are not released.

NAEP is a congressionally mandated project of the National Center for Education Statistics within the Institute of Education Sciences of the U.S. Department of Education. By law, the Commissioner of Education Statistics is responsible for carrying out the NAEP project. The National Assessment Governing Board (The Governing Board) oversees and sets policy for NAEP. The Governing Board is an independent, bipartisan group composed of 26 representatives from throughout the United States, including state and local officials, educators, business leaders, and members of the general public.

The National Assessment Governing Board

IES NATIONAL CENTER FOR EDUCATION STATISTICS
Institute of Education Sciences

NAEP
NATIONAL ASSESSMENT OF EDUCATIONAL PROGRESS

Problem Solving in Technology-Rich Environments

A Report From the NAEP Technology-Based Assessment Project, Research and Development Series

U.S. Department of Education
Institute of Education Sciences
NCES 2007–466

August 2007

Randy Elliot Bennett
Hilary Persky
Andrew R. Weiss
Frank Jenkins
Educational Testing Service

In collaboration with
Douglas Forer
Bruce Kaplan
Michael Wagner
Lou Mang
Educational Testing Service

William Tirre
Project Officer
National Center for Education Statistics

U.S. Department of Education
Margaret Spellings
Secretary

Institute of Education Sciences
Grover J. Whitehurst
Director

National Center for Education Statistics
Mark Schneider
Commissioner

The National Center for Education Statistics (NCES) is the primary federal entity for collecting, analyzing, and reporting data related to education in the United States and other nations. It fulfills a congressional mandate to collect, collate, analyze, and report full and complete statistics on the condition of education in the United States; conduct and publish reports and specialized analyses of the meaning and significance of such statistics; assist state and local education agencies in improving their statistical systems; and review and report on education activities in foreign countries.

NCES activities are designed to address high-priority education data needs; provide consistent, reliable, complete, and accurate indicators of education status and trends; and report timely, useful, and high-quality data to the U.S. Department of Education, the Congress, the states, other education policymakers, practitioners, data users, and the general public. Unless specifically noted, all information contained herein is in the public domain.

We strive to make our products available in a variety of formats and in language that is appropriate to a variety of audiences. You, as our customer, are the best judge of our success in communicating information effectively. If you have any comments or suggestions about this or any other NCES product or report, we would like to hear from you. Please direct your comments to:

National Center for Education Statistics
Institute of Education Sciences
U.S. Department of Education
1990 K Street NW
Washington, DC 20006-5651

August 2007

The NCES World Wide Web Home Page address is http://nces.ed.gov.
The NCES World Wide Web Electronic Catalog address is http://nces.ed.gov/pubsearch.

Suggested Citation:

Bennett, R.E., Persky, H., Weiss, A.R., and Jenkins, F. (2007). *Problem Solving in Technology-Rich Environments: A Report From the NAEP Technology-Based Assessment Project* (NCES 2007–466). U.S. Department of Education. Washington, DC: National Center for Education Statistics.

For More Information

Content contact:
William Tirre
202-502-7361
William.Tirre@ed.gov

To obtain single copies of this report (limited number of copies available), or for ordering information on other U.S. Department of Education products, call toll free 1-877-4ED-PUBS (877-433-7827), or write:

Education Publications Center (ED Pubs)
U.S. Department of Education
P.O. Box 1398
Jessup, MD 20794-1398
TTY/TDD 1-877-576-7734
FAX 1-301-470-1244

The work upon which this publication is based was performed for
the National Center for Education Statistics, Institute of Education Sciences,
by Educational Testing Service and Westat.

Executive Summary

The Problem Solving in Technology-Rich Environments (TRE) study is the last of three field investigations in the National Assessment of Educational Progress (NAEP) Technology-Based Assessment Project, which explores the use of new technology in administering NAEP. The TRE study was designed to demonstrate and explore an innovative use of computers for developing, administering, scoring, and analyzing the results of NAEP assessments. The prior two studies, Mathematics Online (MOL) and Writing Online (WOL), compared online and paper testing in terms of issues related to measurement, equity, efficiency, and operations.

In the TRE study, two extended scenarios were created for measuring problem solving with technology. These scenarios were then administered to nationally representative samples of students. The resulting data were used to describe the measurement characteristics of the scenarios and the performance of students.

The context for the problem-solving scenarios was the domain of physical science. The TRE Search scenario required students to locate and synthesize information about scientific helium balloons from a simulated World Wide Web environment. The TRE Simulation scenario required students to experiment to solve problems of increasing complexity about relationships among buoyancy, mass, and volume; students viewed animated displays after manipulating the mass carried by a scientific helium balloon and the amount of helium contained in the balloon. Both scenarios targeted grade 8 students who were assumed to have basic computer skills; basic exposure to scientific inquiry and to concepts of buoyancy, mass, and volume; and the ability to read scientifically oriented material at a sixth-grade level or higher.

In the TRE study, data were collected from a nationally representative sample of grade 8 students in the spring of 2003. Over 2,000 public school students participated, with approximately 1,000 students taking each assessment scenario. (See appendix B for detailed information about the TRE sample selection.) Students were assigned randomly within each school to one of the scenarios—Search or Simulation. Students took the scenarios on school computers via the World Wide Web or on laptop computers taken into the schools. For both scenarios, data were collected about student demographics; students' access to computers, use of computers, and attitudes toward them; and students' science coursetaking and activities in school.

Methodology

The TRE study used Evidence-Centered Design (ECD) (Mislevy, Almond, and Lukas 2003) to develop the interpretive framework for translating the multiplicity of actions captured from each student into inferences about what populations of students know and can do. In ECD, the key components of the interpretive framework are student and evidence models. The student model represents a set of hypotheses about the components of proficiency in a domain and their organization. The evidence model shows how relevant student actions are connected to those components of proficiency, including how each relevant action affects belief in student standing on each proficiency component. The structure provided by ECD is particularly important for complex assessments like TRE, for which meaningful inferences must be drawn based on hundreds of actions captured for each student.

For the purposes of TRE, the student model represented the components of student proficiency in the domain of problem solving in technology-rich environments. Two primary components were postulated: scientific inquiry and computer skills. Scientific inquiry was defined as the ability to find information about a given topic, judge what information is relevant, plan and conduct experiments, monitor efforts, organize and interpret results, and communicate a coherent interpretation. Computer skills were defined as the ability to carry out the largely mechanical operations of using a computer to find information, run simulated experiments, get information from dynamic visual displays, construct a table or graph, sort data, and enter text.

Evidence of these skills consisted of student actions called "observables." Observables were captured by computer and judged for their correctness using scoring criteria called "evaluation rules," and summary scores were created using a modeling procedure that incorporated Bayesian networks (Mislevy et al. 2000). Bayesian models belong to a class of methods particularly suited to the TRE scenarios because these methods account for multidimensionality and local dependency, neither of which is explicitly handled by the measurement models typically used in NAEP assessments.

The TRE Scenario Scales and Results

Because the TRE study used measures that are experimental, data were analyzed to explore how well the TRE scenario scales captured the skills they were intended to summarize. For each scenario, the following measures were obtained: internal consistency; the relations of student scores to students' prior knowledge; the TRE scale intercorrelations; the correlations of each observable with each subscale; the locations of the observables on the scales; the response probabilities for prototypic students (i.e., hypothetical students with low, medium, and high levels of proficiency); and the relations of relevant student background information to performance. Results were considered to be statistically significant if the probability of obtaining them by chance alone did not exceed the .05 level.

Readers are reminded that the TRE project was intended as an exploratory study of how NAEP can use technology to measure skills that cannot be easily measured by conventional paper-and-pencil means. This report will discuss the ability of a nationally representative student sample to solve problems using technology in the TRE context. However, the results pertain to student performance in only two scenarios employing a limited set of technology tools and a range of science content sufficient only for demonstration purposes. Therefore, results cannot be generalized more broadly to problem-solving in technology-rich environments for the nation's eighth-graders.

The Search Scales and Results

TRE Search consisted of 11 items (or observables) and produced a total score and two subscores, scientific inquiry and computer skills.

- The internal consistency of the three TRE Search scores (total, scientific inquiry, and computer skills) ranged from .65 to .74, as compared to .62 for the typical main NAEP science assessment hands-on task block, which, although measuring skills different from TRE, also includes extended, problem-solving tasks.
- The Search scores provided overlapping but not redundant information; the (disattenuated) intercorrelation of the subscores was .57. This value contrasts with intercorrelations of .90 to .93 for the main NAEP science assessment scales.
- The scientific inquiry skill scale score was most related in the student sample to the following scale observables: the relevance of the World Wide Web pages visited or bookmarked, the quality of the constructed response to a question designed to motivate students to search for and synthesize information from the Web, and the degree of use of relevant search terms (r range between performance on the observable and scale score = .51 to .71).
- The computer skills scale score was related in the student sample primarily to the following scale observables: the use of hyperlinks, the use of the Back button, the number of searches needed to get relevant hits (an efficiency measure), and the use of bookmarking (r range = .60 to .69).
- Statistically significant differences in performance were found on one or more TRE Search scales for NAEP reporting groups categorized by race/ethnicity, parents' highest education level, students' eligibility for free or reduced-price school lunch, and school location. No significant differences were found, however, for reporting groups categorized by gender.

The TRE Simulation Scenario Scales and Results

The TRE Simulation scenario consisted of 28 observables and produced a total score and three subscores: scientific exploration, scientific synthesis, and computer skills.

- The internal consistency of the four scales ranged from .73 to .89, as compared to .62 for the typical main NAEP science assessment hands-on task block, which, although measuring skills different from TRE, also includes extended, problem-solving tasks.
- The Simulation scores provided overlapping but not redundant information; the (disattenuated) intercorrelations of the subscores ranged from .73 to .74. These values contrast with intercorrelations of .90 to .93 for the main NAEP science assessment scales.
- The scientific exploration skill scale score was most related in the student sample to three scale observables: which experiments students chose to run to solve the Simulation problems, whether students constructed tables and graphs that included relevant variables for solving the problems, and the degree to which experiments controlled for one variable in the one problem demanding controlled experimentation.
- The scientific synthesis scale score was primarily related in the student sample to the degree of correctness and completeness of conclusions drawn for each Simulation problem.
- Performance on the computer skills scale was related in the student sample mainly to the number of characters in the written responses students gave for each of the three Simulation problems.
- Statistically significant differences in performance were found on one or more TRE Simulation scales for NAEP reporting groups categorized by race/ethnicity, parents' highest education level, and students' eligibility for free or reduced-price school lunch. No significant differences were found, however, for reporting groups categorized by gender or school location.

Foreword

The Research and Development series of reports has been initiated for the following goals:

1. To share studies and research that are developmental in nature. The results of such studies may be revised as the work continues and additional data become available.

2. To share results of studies that are, to some extent, on the cutting edge of methodological developments. Emerging analytical approaches and new computer software development often permit new, and sometimes controversial, analysis to be done. By participating in "frontier research," we hope to contribute to the resolution of issues and improved analysis.

3. To participate in discussions of emerging issues of interest to educational researchers, statisticians, and the federal statistical community in general. Such reports may document workshops and symposiums sponsored by the National Center for Education Statistics (NCES) that address methodological and analytical issues or may share and discuss issues regarding NCES practice, procedures, and standards.

The common theme in all three goals is that these reports present results or discussions that do not reach definitive conclusions at this point in time, either because the data are tentative, the methodology is new and developing, or the topic is one on which there are divergent views. Therefore, the techniques and inferences made from the data are tentative and are subject to revision. To facilitate the process of closure on the issues, we invite comment, criticism, and alternatives to what we have done. Such responses should be directed to:

Marilyn M. Seastrom
Chief Statistician
Statistical Standards Program
National Center for Education Statistics
1900 K Street NW, Suite 9000
Washington, DC 20006

Acknowledgments

The NAEP Problem Solving in Technology-Rich Environments (TRE) study was part of the Technology-Based Assessment (TBA) project, a collaborative effort led by the National Center for Education Statistics (NCES) and the National Assessment Governing Board, and carried out by Educational Testing Service (ETS) and Westat. The project was funded through NCES, in the Institute of Education Sciences of the U.S. Department of Education. We appreciate the support and guidance of Acting Commissioner of Education Statistics Grover J. Whitehurst, NAEP project director Suzanne Triplett, TBA project directors Holly Spurlock and William Tirre, and NCES consultants Vonda Kiplinger and Bob Evans.

NAEP is grateful to the students and school staff who participated in the study, and to the Westat staff who administered the study.

NAEP activities at ETS were directed by Stephen Lazer and John Mazzeo, with assistance from John Barone. The ETS management for the TBA project included Randy Bennett, James S. Braswell, Beth Durkin, Christine O'Sullivan, and Clyde Reese.

The TRE study was managed and coordinated by Hilary Persky of ETS. Development and scoring activities were conducted at ETS with contributions from Malcolm Bauer, Dwight Bawcom, Mary Lauko, Claudia Leacock, Christine O'Sullivan, and Andrew Weiss. Janet Stumper designed the Simulation scenario interface with the assistance of Debbie Pisacreta.

Data analysis activities were carried out by Kevin Bentley, Frank Jenkins, Gerry Kokolis, and John Willey. Katharine Pashley conducted the TRE database work.

The design and production of this report were overseen by Loretta Casalaina with assistance from Rick Hasney, Parul Kiran, and Susan Mills. Ming Kuang coordinated the documentation and data-checking procedures with assistance from Kit Chan and Janice Goodis. Carmen Payton and Joanne O'Brien reviewed tabular presentations for consistency with NCES standards. Arlene Weiner coordinated the editorial and proofreading procedures with assistance from Mary Daane, Alice Kass, Janice Lukas, and Linda Myers. The Web version of this report was coordinated by Rick Hasney.

This project could not have been completed without Westat, which conducted student sampling, administration, field support, and weighting. Westat's activities were managed by Dianne Walsh, Dward A. Moore, Jr., Brice Hart, and Brenda Ennis, with the assistance of Nia Davis. Sampling and weighting were conducted by Louis Rizzo and Tom Krenzke. Weighting systems work was completed by Bill Wall. The day-to-day management of Westat's technical and software systems was accomplished by Brice Hart with assistance from Fran Cohen and Karen Dennis.

Many thanks are due to the numerous reviewers, both internal and external to NCES and ETS. The comments and critical feedback of the following reviewers are reflected in the final version of this report: Tajuana Bates, Deven Carlson, Young Chun, Mary Crovo, Aaron Douglas, Ray Fields, Hilary Forster, Stephanie Germeraad, Arnold Goldstein, Steve Gorman, Fraser Ireland, Andrew Kolstad, Susan Loomis, Teresa Neidorf, Taslima Rahman, Linda Shafer, William Tirre, Alan Vanneman, Melissa Vrana, David Williamson, Matthias Von Davier, and Rima Zobayan.

Contents

List of Tables

Page

List of Figures

Page

Introduction

For more than 30 years, the National Assessment of Educational Progress (NAEP) has regularly collected, analyzed, and reported valid and reliable information about what American students know and can do in a range of subject areas. As authorized by the U.S. Congress, NAEP typically assesses nationally representative samples of students in grades 4, 8, and 12. Since 1990, NAEP has also assessed representative samples of students at grades 4 and 8 in states and other jurisdictions that participate in the NAEP state-by-state assessments. In 1988, Congress established the National Assessment Governing Board to oversee and set policy for NAEP.

In response to the ever-increasing importance of technology in educational and workplace settings, and to maintain its leadership role in the area of large-scale assessment, NAEP initiated the Technology-Based Assessment (TBA) Project in 1999. The TBA Project was intended to explore the many uses of new technology in NAEP, among them specific NAEP processes (e.g., item creation, test delivery), assessment of specific content domains, and assessment of technology skills.

The TBA Project focused on several key questions:

1. *What are the measurement implications of using technology-based assessment in NAEP?* Technology-based assessment may change the meaning of our measures in unknown ways. It may allow assessment of skills that could not be measured using paper and pencil or preclude measuring skills that could be tested by conventional means. It may allow the assessment of emerging skills, particularly those requiring students to employ new technology in learning and problem solving.
2. *What are the implications for equity?* If not carefully designed, technology-based assessment could inaccurately reflect the skills of some groups of students, especially those with differing degrees of access to computers. At the same time, it could increase participation of students with disabilities. It may also better reflect the skills of students who routinely use the computer to perform academic tasks like writing.
3. *What are the efficiency implications of using technology-based assessment compared with paper and pencil?* Along with other new technologies, the Internet may afford significant time and cost savings for large-scale assessments.
4. *What are the operational implications of technology-based assessment?* Moving from a paper-based program to an electronic one raises significant issues concerning school facilities, equipment functioning, administrator responsibilities, and school cooperation.

To answer these questions, the NAEP program undertook three empirical studies with students: Math Online (MOL; Sandene at al. 2005), Writing Online (WOL; Horkay et al. 2005), and Problem Solving in Technology-Rich Environments (TRE).

The MOL and WOL studies were designed to investigate the effects of delivering existing paper tests via computer. In contrast, the TRE study was designed to demonstrate and explore innovative uses of computers in NAEP by developing two sample extended problem-solving scenarios. This report describes the methodology, technology, and results of the TRE study.

The TRE Project was guided by several principles:

1. *TRE should use the computer to do what cannot easily be done on paper.* The TRE scenarios allow students to answer questions by searching electronic databases and by using a simulation tool to conduct experiments. All student actions are captured by computer for later scoring, allowing for evaluation of the processes used in problem solving. These capabilities could not be easily achieved with conventional paper-and-pencil testing. Chapter 1 of this report describes in detail the two grade 8 TRE problem-solving scenarios—the Search scenario and the Simulation scenario.
2. *TRE should represent the type of problem solving done with computers in educational and work environments.* TRE attempts to capture the multidimensionality characteristic of problem solving with technology by requiring students to demonstrate both science skills and basic facility with the computer. Further, technology in TRE is used as a means of solving substantive problems, rather than as an end in itself.
3. *To the degree possible, TRE should allow the disentangling of component skills.* The two TRE scenarios were intended to measure both basic computer skills and science skills in an integrated way; that is, students would need to use both skill sets simultaneously to solve the problems in the scenarios. For example, students were required to demonstrate mastery of searching for information in a World Wide Web environment, but this skill was to be used in a specific scientific domain that demanded the ability to select and synthesize relevant scientific material.

A consequence of this close integration of skills, however, is that a deficiency in one skill can prevent the expression of another. The TRE team sought to limit such occurrences in several ways. For example, to reduce the chances that limited computer skills would keep students from showing their science skills, tutorials were supplied to help students understand the scenario interfaces, common interface conventions were used (e.g., dialog boxes and wizards), and a computer-related help function was made available. To prevent lack of science skills from impeding the demonstration of computer skills, students were supplied with a science help tool to access basic information relevant to both scenarios; the Simulation interface tools were organized to facilitate a structured inquiry process built around designing experiments, running experiments, and interpreting results; certain choices in the Simulation scenario were constrained (e.g., the choice of variables to include on each graph axis); and the Simulation scenario began with a relatively simple problem. Finally, an interpretive framework was used that allowed for the simultaneous estimation of related proficiencies.

4. *TRE should be positioned so it can inform the development of a future assessment of emerging skills or of more traditional subject matter.* It should be possible to incorporate meaningful exercises using a simulation tool or electronic information search into existing NAEP subject-matter assessments; for example, a likeness of the TRE Simulation scenario could find a logical place in the NAEP science assessment to measure skills needed for scientific investigation. It should also be possible to use the TRE scenarios as models for measures of problem solving with technology generally.
5. *TRE should be an assessment, not instruction, but students should be able to learn from it incidentally.* Both scenarios involve discovery; hence, students may learn from working with the TRE scenarios in a way that participation in the typical large-scale assessment does not provide.

Overview of the Study

Educational Testing Service (ETS) assessment development and research staff created the two TRE scenarios with expert input and reviews from a TRE Development Committee. The committee was composed of science and technology educators and curriculum experts. (The membership of this committee can be found in appendix A.) NCES staff provided oversight and guidance as to the appropriate direction and nature of the scenarios. The development of the TRE scenarios was further informed by a variety of sources, among them the NAEP Science Framework (National Assessment Governing Board 2000) and current research in problem solving and scientific inquiry. Also important were various state and national science and technology standards, including the National Science Education Standards (National Academy of Sciences 1996) and the National Educational Technology Standards (International Society for Technology in Education 2002).

The scenarios were created for grade 8 students who were assumed to have basic computer skills; basic exposure to scientific inquiry and to concepts of buoyancy, mass, and volume; and the ability to read scientifically oriented material at between a sixth-grade and an eighth-grade level. NAEP project staff assumed that most grade 8 students have at least basic computer skills because the 2002 NAEP Writing Online data suggest that virtually all students use computers for schoolwork at least to some extent (Horkay et al. 2005). Further, because of the prevalence of experimental methodology and physics content in grade 8 science curricula, NAEP project staff assumed that members of the grade 8 population have had some basic exposure to scientific inquiry and to basic concepts of buoyancy, mass, and volume.[1]

The TRE study tested a nationally representative sample of grade 8 students in the spring of 2003. Over 2,000 public school students participated, with approximately 1,000 students taking each assessment scenario. (See appendix B for detailed information about the TRE sample selection.) Students were assigned randomly within each school to one of the scenarios—Search or Simulation. For both scenarios, data were collected about student demographics; students' access to, use of, and attitudes toward computers; and students' science coursetaking and activities in school. Additionally, before starting each scenario, students answered prior knowledge questions designed to determine the degree to which they had the computer and/or science knowledge and skills being assessed.

Staff members employed by Westat, the NAEP data collection contractor, administered the TRE scenarios and proctored all administrations using procedures generally similar to those employed for NAEP assessments. Testing was conducted either on school

[1] A range of state curricula surveyed by the authors included experimental activities and methods as well as mastery of the basic concepts of buoyancy, mass, and volume at the eighth-grade (middle school) level. Two typical examples are state middle school curricula for North Carolina and Massachusetts (North Carolina State Department of Education 2004; Massachusetts Department of Education 2001).

computers connected to the Internet or on laptop computers brought in by NAEP administrators. All computers, whether supplied by the school or by NAEP, had to meet minimum hardware and software specifications to ensure that the test would operate uniformly (see appendix C for these specifications). NAEP staff at ETS conducted the scoring and analysis of results.[2]

Analysis of student responses was conducted for two purposes. The first purpose was to evaluate the functioning of the TRE scenarios. The analyses included internal consistency, the relations of student scores to students' prior knowledge, the TRE scale intercorrelations, the correlations of each observable with the TRE subscales, the locations of the observables on the scales, the response probabilities for prototypic students (i.e., hypothetical students with different levels of proficiency), and the relations of relevant student background information to performance. The second purpose was to describe student performance on the scenarios in quantitative and qualitative terms. For differences in mean scores and for differences from zero of correlation coefficients, .05 was used as the level for deciding that a result was statistically significant, with score differences between group means evaluated for statistical significance using independent *t*-tests.

Chapter 1 of this report describes in detail two grade 8 TRE problem-solving scenarios—the Search scenario and the Simulation scenario. Chapter 2 describes how the TRE team used Evidence-Centered Design (ECD; Mislevy, Almond, and Lukas 2003; Mislevy et al. 2001) to help develop an interpretive framework for translating the multiplicity of actions captured from each student who took TRE into inferences about student proficiency. Chapter 3 describes TRE student responses to background questions concerning computer use, attitudes toward computers, and engagement in school science. Chapter 4 discusses how the evaluation rules, or scoring criteria, developed using ECD were applied to student performances by both machine and human scoring, and chapters 5 and 6 present the results of analyses of student performance. Finally, chapter 7 summarizes the TRE study results.

The appendixes that appear in this report are as follows: appendix A lists the members of the TRE Development Committee; appendix B discusses the TRE assessment sample selection process; appendix C identifies the computer specifications for schools that participated in the TRE assessment; appendix D presents the prior-knowledge computer and science questions students took before each scenario, and the background questions students responded to when they had completed the scenarios; appendix E shows the Simulation scenario tutorial and individual screens from the Computer and Science Help in the Simulation scenario; appendix F discusses the use of Bayesian estimation in the study; appendix G lists the rules used for the ETS automated scoring tool, c-rater, for scoring students' search queries; appendix H presents the Search and Simulation scenario scale scores and percentiles by student reporting groups; appendix I presents summary statistics for prior-knowledge measures and mean scale scores for background-question response options; appendix J shows student performance on observables for the Search and Simulation scenarios; and appendix K presents definitions for each of the TRE student reporting groups.

Limitations of the Study

Readers are reminded that the TRE project results pertain to student performance in only two scenarios. These scenarios employed a limited number of technology tools and a range of content sufficient for demonstration purposes only.

A second limitation is that the TRE study was not based on an existing NAEP content-area framework. As such, the conceptualization of the TRE construct domain used in this study did not involve the broad representation of diverse constituencies typical of NAEP assessment frameworks.

A third limitation is that the TRE assessment instruments and analysis methods were experimental ones drawing upon extended computer-delivered performance tasks and Bayesian modeling methods not previously used in NAEP assessments.

Because of these limitations, TRE study results should not be generalized to problem solving in technology-rich environments for the nation's eighth-graders, nor should they be used to draw general conclusions about the science knowledge or computer skills of those students.

[2] No analysis of performance on laptops vs. school computers was conducted because the meaning of any observed performance differences would be ambiguous. Since the assignment of students to computer type was not done at random but rather according to the fit of school technology infrastructure with the requirements of the test delivery system, performance differences could be caused by differences in other factors related to the quality of school technology (e.g., in socioeconomic status) and not by differences in the suitability of one or the other computer type for online assessment. Further, there were no measures of skill independent of computer type that could have been used to adjust statistically for pre-existing differences between groups. But see Horkay et al. 2005 for an analysis of performance differences on laptops vs. school computers for 8th-grade students.