Project 2061
AAAS

Assessment with Precision

Project 2061 Building a Collection of Test Items Aligned to Standards

As testing gains prominence as a tool for evaluating the progress of students and schools, educators and the public are grappling with the consequences of decisions based on assessment data—from denial of graduation to cuts in school funding. Yet such decisions are commonly made based on data from assessments that are poorly aligned with the content standards for which students, teachers, and schools are being held accountable. While the standards movement in science, mathematics, and technology education has defined and implemented new goals for what students should know and be able to do, today's assessment instruments and practices have not yet adapted to the demands of standards-based reform.

To better serve students and schools, assessments and content standards need to be carefully coordinated so that the tests students take—whether high-stakes state tests or classroom tests and quizzes—actually assess the ideas and skills specified in the standards that shape curricula and instruction. With the federal *No Child Left Behind Act* of 2001 requiring that statewide assessments be based on each state's content standards, the need for test items that are aligned to those standards is more urgent than ever.

A new effort by AAAS Project 2061 aims to help meet that need. With funding from the National Science Foundation's Instructional Materials Development Assessment Program, Project 2061 has begun a five-year, $4.1 million project to develop an online collection of assessment resources that will help users to understand the intent of specific content standards and to select assessment items that will yield the most useful information about student learning. Central to the collection will be a bank of more than 300 high-quality assessment items for middle and early high school science and mathematics that will be electronically linked to national and state content standards.

Evidence of Student Learning

As one of the first organizations to focus on content standards and their role in curriculum, instruction, and assessment, Project 2061 has been studying the alignment and effectiveness of science and mathematics test items. Working with teams of educators and assessment specialists, Project 2061 has developed a set of criteria and a procedure for analyzing the alignment of assessment items to standards. Analyses of items drawn from a wide variety of sources have revealed that while many existing items cover topics—such as cells or fractions—that are identified in content standards, few items are aligned to the precise ideas and skills targeted by standards.

In contrast, the items to be included in the new collection will be specially designed to provide explicit evidence that a student has—or has not—learned a specific idea or skill. "We absolutely must have better tools for finding out if students are learning what we expect them to learn," says Dr. George DeBoer, deputy director of Project 2061 and principal investigator for the assessment project. "With these new items, educational researchers will be able to answer important questions about the impact of various curriculum materials and instructional strategies, and teachers will be able to find out what their students know and can do and to pinpoint

continued on page 2

2061 today

Science Literacy for a Changing Future

Mathematics
Natural Sciences
Social Sciences
Technology

Summer/Fall 2004
Volume 14, Number 2

Assessment with Precision *from page 1*

areas where they need more help."

To build the collection, Project 2061 will screen hundreds of existing middle and early high school science and mathematics assessment items from as many sources as possible, including released items from the Third International Mathematics and Science Study, the National Assessment of Educational Progress, and a variety of state tests. Following the initial screening and sorting, items will undergo a more rigorous analysis to describe their precise alignment to the ideas being targeted by the standards. Revision of items based on this analysis will work to correct deficiencies that hinder alignment. (See next page for the analysis criteria, an example of their application to a test item, and an invitation to analyze an item yourself.) Items will be field tested with students throughout the development process.

Meaningful assessment depends on clear and specific learning goals and on curricula and instruction that are carefully aligned to those goals.

Much-Needed Tools

Because the collection will include assessment items of all types—from selected-response items such as multiple choice questions to more involved performance tasks—and items designed for both diagnostic and evaluation purposes, it will meet a variety of user needs. Dr. Paul Ache, a mathematics professor at Kutztown University who reviewed 5th and 8th grade test items in a study Project 2061 conducted for the Pennsylvania Department of Education, anticipates several ways that teachers will use the aligned items. "I see this bank being used first and foremost as a set of practice items. Teachers have told me that they need practice items to use with students and that it is difficult to find items, let alone good ones," says Ache. "I also see these items being used as diagnostic tools. If teachers know that items from this bank can help them determine what their students know and don't know, they will use them. Finally, I see some teachers using them as a portion of a summative assessment process."

To help teachers and researchers make the most of the online collection, Project 2061 is developing the following resources to accompany the assessment items:

- **Item profiles.** For each test item, a profile will describe its match to specific science or mathematics standards, the knowledge needed to answer the item correctly, and features such as whether the item is likely to be approached differently by diverse learners. Users will be able to search for and retrieve items based on the features described in the profiles.
- **Assessment maps.** For each of 16 science and mathematics topic areas covered by the collection, an assessment map will display connections among ideas related to the relevant content standards. These maps, adapted from Project 2061's *Atlas of Science Literacy*, will serve as the main interface for the collection and will give test developers a conceptual framework for ideas they might want to test.
- **Clarifications of standards.** The exact meaning of content standards is not always immediately evident. Yet teachers, curriculum and test developers, and researchers need a clear sense of what students are expected to know and what constitutes evidence of that knowledge in order to assess that knowledge precisely. For each standard referenced in the collection, a clarification statement will suggest what students can be asked to do to demonstrate their knowledge and will describe assessment task contexts that are appropriate and engaging to students at a particular age.

In developing these coordinated resources, Project 2061 and its collaborators recognize that meaningful assessment depends on a commitment to clear and specific learning goals and on curricula and instruction that are carefully aligned to those goals. "Some people claim that teaching to a test is wrong," says Dr. Ache. "But, in my opinion, if the test truly measures understanding of a good set of targeted goals, then we must teach to the test. Otherwise, we are not teaching to the learning goals."

Project 2061 is seeking test items that are aligned with particular learning goals in *Benchmarks for Science Literacy* or the *National Science Education Standards*. Please forward such test items to us for consideration. If you are interested in field testing items in your classrooms, we welcome your participation. For more information, contact Dr. George DeBoer at 202-326-6624 or gdeboer@aaas.org. Further details about Project 2061's assessment research are available at **www.project2061.org/assessment**.

Analyzing an Assessment Item

The following two selected-response items help illustrate how Project 2061's assessment analysis criteria (see table below) can be applied to determine an item's alignment to the targeted learning goal and to judge how effective an item is likely to be in finding out what students actually know about the ideas in the goal. For open-ended items, an additional criterion considers whether scoring rubrics are accurate, clear, complete, and specific.

EXAMPLE 1: PROPERTIES OF SUBSTANCES

Target learning goal: *A substance has characteristic properties, such as density, a boiling point, and solubility, all of which are independent of the amount of the sample.* (National Research Council [NRC], 1996, p. 154; grades 5–8)

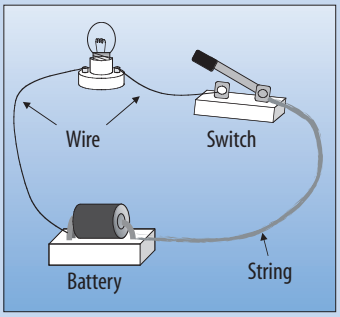
A student found 2 green powders that look the same in the chemistry classroom. The student wants to find out if the 2 powders are the same or different. Which of the following is the best method to use?

a. Mix the powders and then find the density, melting point, and solubility of the mixture.
b. Mix the powders and then find the mass, location, and temperature of the mixture.
c. Find the masses, locations, and temperatures of each powder.
d. Find the density, melting point, and solubility of each powder.*

*Pre/post assessment item from IQWST chemistry curriculum (**http://hice.org/iqwst/**)*

EXAMPLE 2: ELECTRIC CIRCUIT

Target learning goal: *Electricity in circuits can produce light, heat, sound, and magnetic effects. Electrical circuits require a complete loop through which an electrical current can pass.* (NRC, 1996, p. 127; grades K–4)



Javier made an electrical circuit. He decided to replace one of the wires with a piece of string. Choose below what will happen when he closes the switch.

a. The light bulb will light.
b. The light bulb will not light.*
c. The light bulb will blink on and off.
d. The light bulb will blink only once.

* correct answer

ANALYSIS CRITERIA

Necessity: Is the knowledge or skill specified in the learning goal *needed* to answer the item correctly?	Knowledge is needed. Must know that properties can be used to distinguish substances from one another and that density, boiling (melting) point, and solubility are properties.
Sufficiency: Is the knowledge or skill specified in the learning goal *enough* by itself to answer the item correctly?	Knowledge is enough. Item only requires knowing that properties can be used to distinguish substances from one another and knowing examples of properties.
Comprehensibility: Are students likely to understand the task statement, diagrams, symbols, etc.?	Students may have difficulty comprehending all aspects of the task. There are some grammar/sentence structure issues (e.g., multiple uses of the word find/found).
Clarity of Expectations: Are students likely to understand what they are expected to do and what sort of response is considered satisfactory?	The directions are clear.
Appropriateness of Context: Is the task context appropriately familiar, engaging, and realistic to students?	The task context is realistic and plausible although perhaps not very engaging.
Resistance to Test-Wiseness: Could students respond satisfactorily to the task by guessing or employing other general test-taking strategies?	Guessing is possible. Response choices (a) and (b) involve mixing the powders, which may seem counterproductive to many students. Also, "location" may be ruled out in (b) and (c) by most students because it is not a plausible option. This leaves the correct answer (d) as the remaining choice.

APPLICATION OF CRITERIA TO EXAMPLE 2

Try an analysis of your own on Example 2:

- Is the knowledge specified in the target learning goal both needed and enough by itself to answer the item correctly?
- Which of the analysis criteria reveal aspects of the item that point to the need for revision?

Reference

National Research Council. (1996). *National science education standards*. Washington, DC: National Academy Press.

Deputy Director Brings Key Questions to the Fore

Dr. George E. DeBoer, Deputy Director of Project 2061

Since becoming deputy director in 2002, Dr. George E. DeBoer has helped lead Project 2061's research into goals-based curriculum, instruction, and assessment. He came to Project 2061 from the National Science Foundation (NSF), where he was a program director in the Division of Elementary, Secondary, and Informal Science. A former high school science teacher, Dr. DeBoer enjoyed a long career as a professor and administrator at Colgate University, where he holds an appointment as Professor of Educational Studies. *2061 Today* recently interviewed Dr. DeBoer about Project 2061's current research, including its work with the Center for Curriculum Materials in Science (CCMS) and his role as principal investigator for a new NSF-funded grant to create a collection of high-quality science and mathematics assessment items.

2061 Today: A small sign on your office door reads: "Should we test *what* they know or *that* they know?" What are you referring to?

GD: I'm referring to whether or not assessment is trying to find out something in particular, for instance knowledge of a particular idea—an approach consistent with the whole standards movement. Because a standard targets a particular idea that you expect students to understand, this approach looks to test *that* they know *that* idea.

Now the *what* students know is much more open-ended. Some view assessment as trying to find out what people know, so you could ask students a variety of things, which would not necessarily be aligned with the content standards but would give you information about what students know. People know a lot more, for example, than one can find referenced in *Benchmarks for Science Literacy*. So you could create a number of open-ended questions or other types of questions that will test *what* they know—about anything. This *that/what* distinction is not really a judgment; instead, it is something to keep in mind when thinking about assessments.

2061 Today: You wrote in your 1991 book *A History of Ideas in Science Education* that curriculum developers face a difficult challenge in deciding what to include because there is simply too much material. Have the subsequent publications of *Benchmarks for Science Literacy* and the *National Science Education Standards* (*NSES*) made that challenge less formidable?

GD: Yes, because it is not as open-ended as it was. But even with the standards documents we have today there is still an enormous amount of material, and that means there are a lot of hard decisions to be made, and I think everyone knows that. So *Benchmarks* and *NSES* have certainly helped but they haven't completely solved the problem. One difficulty is that each idea listed in the standards documents can expand in many different ways through the instances that are chosen to teach it. So, the question still remains: How many and which relevant instances of an idea is one going to teach? And there is still the question of whether all of the benchmarks and standards can be taught well in the time available and, if not, which ones are more important than the others.

2061 Today: One of your key research interests is student achievement. Is your work at Project 2061 building upon previous investigations into this area?

GD: Yes, a great deal of the work I have done involved trying to answer two questions: What specific learning goals are educators trying to achieve at any point in time and what are the most important or ultimate aims of education? So Project 2061 has been a great place to continue to think about these questions, because of its focus on having clear statements of what students should know and be able to do and clear justifications for why those things are important. Much of the new work we are doing involves clarifying these goal statements even further.

What I am finding intriguing is how assessment plays a role in this thinking about learning goals. Thinking about assessment makes you consider what is reasonable or fair to ask students to do to demonstrate their knowledge or skill. This forces you to look closely again at the learning goal statements to see if they are clearly stated and what exactly they imply about what students should be able

to do with that knowledge. This is one of the most interesting aspects of my work here: seeing the interaction between a learning goal statement and an assessment item.

2061 Today: One of Project 2061's goals is to have an impact on the development of curriculum materials. Will the new Center for Curriculum Materials in Science bring that goal closer to reality?

GD: Oh sure, but predicting the timing of the impact is difficult. There is an immediate impact on the leadership of the Center as they learn from each other, and on the Center's students and postdocs, who will in turn have an impact when they leave the Center. In the same way, the teachers who are involved with the Center will have an immediate impact on their students. Although they will not be developing curriculum materials themselves, they are learning ways to make better decisions about selecting and implementing curriculum materials.

As far as the longer term impact, our CCMS partners have materials development grants and their development efforts are influenced by the Center's work—by the conversations we have and by the tools and resources that we share. They are using a goals-based approach and embedding pedagogical supports into their materials, and we are helping them in that process.

2061 Today: How does the current research into curriculum materials at CCMS compare to previous research efforts?

GD: To put that question in context, the first thing to say is that in the larger world of developing and publishing curriculum materials, there is very little research that has been done. Now, over the last decade or more, largely with the support of the NSF, research-based development efforts have been taking place, where enough resources are made available to allow the developers to do research concurrent with their development work.

This is resulting in much more thoughtful work, and I expect the Center will become a model for looking even deeper into the questions that need to be answered about curriculum materials. To do that, we at the Center share tools and resources with each other for curriculum development research, and we are developing assessment items that can be used to help us see whether or not students do in fact learn from this or that particular approach. So there is a much greater depth of analysis going on in the Center than is typical. We also are beginning to share these tools and resources with developers outside the Center and working with them in the creation of a research-based model that can be used in the larger development community.

2061 Today: Project 2061 has begun to develop an online collection of science and math test items aligned to content standards (see this issue's lead article). When do you envision this work becoming available to educators?

GD: We are already supplying a number of research and development projects with items for their current research. Because the development of test items is being done in conjunction with that work, participants in those projects are having almost immediate access to the assessment work we are involved in.

The larger public, whether it be teachers, test developers, district personnel, or the public in general, will become familiar with this work after we have gone through a very careful process of examining the items. We will go through a process not only of finding items that are aligned to the content standards, but also of creating a utility and a Web-based interface where the items can be accessed along with additional tools like assessment maps. After two years we will have a reverse site visit with NSF, and at that time we should be able to demonstrate some significant progress, and we should have built a prototype of the interface. Whether we actually go public at that point is something we still need to decide. In the meantime, we'd like to have a dialogue about the project with the larger educational community and we invite educators to send us sample test items that are aligned to particular benchmarks or standards, or to consider field testing items in their classrooms as part of our research.

Thinking about assessment makes you consider what is reasonable or fair to ask students to do to demonstrate their knowledge or skill. This forces you to look closely again at the learning goal statements.

About AAAS and Project 2061

Publisher of the peer-reviewed journal *Science*, the American Association for the Advancement of Science (AAAS) is the largest general scientific organization in the world. Its education initiative, Project 2061, has been at the forefront of the K–12 reform movement

- Defining science literacy and promoting it as a goal for all Americans;
- Developing K–12 benchmarks for student learning in science, mathematics, and technology;
- Producing a wide range of innovative tools for educators—books, CD-ROMs, and online resources—to guide their reform efforts; and
- Conducting research on the design and use of curriculum materials, assessment, professional development, and other areas of science teaching and learning.

AAAS gratefully acknowledges the following Project 2061 supporters: Carnegie Corporation of New York, Hewlett-Packard Company, John D. and Catherine T. MacArthur Foundation, Noyce Foundation, David and Lucile Packard Foundation, Siemens Foundation, and the National Science Foundation.

For more information:
AAAS Project 2061
1200 New York Avenue, NW
Washington, DC 20005

Phone: 202-326-6666
Fax: 202-842-5196
E-mail: project2061@aaas.org
Web site: www.project2061.org

To order Project 2061 products:
Call 1-888-737-2061 or visit www.project2061.org/order.

Resources for New Textbook Development

Findings from Project 2061's evaluations of widely used middle and high school science textbooks indicate that existing materials have a long way to go in presenting a coherent set of learning goals and providing high-quality support for students and teachers (see evaluation reports at **www.project2061.org/textreview**). To help address the need for better textbooks, Project 2061 is developing an online collection of resources aimed at helping educators, curriculum materials developers, and publishers to take advantage of the growing body of knowledge about how students learn science concepts and skills.

With funding from the National Science Foundation, Project 2061 is working with content area and science education experts to review a broad range of print and digital resources (e.g., activities, diagrams, and sets of questions) for inclusion in the collection. At the same time, we are producing resources that serve to clarify and connect learning goals and to summarize research on student learning. Everything in the collection will be aligned to national and state content standards and based on the most relevant research.

The collection will use 13 strand maps drawn from Project 2061's *Atlas of Science Literacy* as its main interface. Users will be able to click on the text of a learning goal displayed on a map to access resources in the following categories:

- **Clarifications:** Elaborations of the intent of the learning goal that specify the level of sophistication intended and identify peripheral ideas and terms that are not on target.
- **Connections:** Relationships to prerequisite and other benchmark ideas found in *Atlas for Science Literacy*, *Benchmarks for Science Literacy*, and *Science for All Americans*.
- **Ideas Students Have:** Summaries of research that sheds light on students' commonly held ideas related to the learning goal and on the likely sources of these ideas.
- **Diagnostic Questions:** Questions or tasks educators can use to elicit their students' thinking and track their understanding.
- **Phenomena/Activities:** Descriptions of phenomena to help students view abstract scientific ideas as plausible, along with hands-on or demonstration activities to engage students with the phenomena.
- **Representations/Activities:** Illustrations, models, and simulations to help clarify abstract ideas for students.

A prototype of the online interface is now available. Please try the prototype at **http://test.p2061.org/curriculum** and give us your feedback. Send e-mail to Dr. Sofia Kesidou (skesidou@aaas.org), for feedback regarding the resources, and to Dr. Francis Molina (fmolina@aaas.org), for feedback on the interface.

Director's Notes

Tried, but Untested

A few years ago, my colleagues and I spent an evening with the school board of a suburban Pennsylvania community in an attempt to persuade them to allow their teachers to consider and test alternative science curriculum materials. We failed. In thanking us for coming, the superintendent explained that the school board just wanted what was best for their children and didn't want to "experiment" with their children's learning. She seemed justifiably alarmed, however, when I pointed out that their current curriculum materials had in fact never been tested and that Project 2061's careful analysis of them had found a number of weaknesses.

Now, a new National Research Council (NRC) report reminds us just how little we know about the effectiveness of curriculum materials. The NRC took a look at the quality of information about effectiveness provided by evaluation studies of 19 mathematics curricula. Despite nearly 150 studies analyzed, they found that in order to determine with confidence the effectiveness of individual curricula, we need a greater number of studies on each curriculum, a wider array of methods used, and more consistency in the quality of the studies. What is more, nearly 75% of the analyzed studies were conducted on National Science Foundation-funded curriculum materials. That means that commercially developed textbooks, used by the vast majority of U.S. students, remain untested.

Determining Curricular Effectiveness

As the NRC report points out, no single study can unequivocally establish the effectiveness of a curriculum material. Instead, the report recommends that effectiveness be established through the use of multiple, scientifically valid evaluation methods. Such methods include curriculum analyses (such as those carried out by Project 2061), comparative studies using experimental or quasi-experimental designs, case studies investigating the relationships between program components and the quality of their implementation, and a report synthesizing and integrating the findings from these different studies.

The NRC also calls for increased federal support for empirical studies of curriculum materials and the development of new, more precise, methods for evaluating how materials are being used in the classroom, what students are learning, and what kinds of professional development can support teachers most effectively. With our new effort to build an online collection of assessment items aligned to standards (see this issue's lead article), Project 2061 hopes to contribute to this larger effort.

How Assessment Can Help

Rigorous research studies with replicable results—such as those called for in the NRC's report—require assessment items that are aligned to the specific ideas and skills that students are expected to learn. For example, researchers need items that are aligned to standards but independent of any single materials development project in order to compare the effectiveness of different materials more objectively. And curriculum developers need well-aligned items to test and refine their materials as they are being developed. After all, the long-term success of standards-based reform depends on making sure that curriculum, assessment, and all elements of the education system work together to help students achieve the learning goals set forth by national and state content standards. Without credible evidence that new and innovative materials can help students achieve these goals, school boards and other stakeholders may decide—and understandably so—that the costs of implementing such materials are not justified by the benefits.

Jo Ellen Roseman
Director

As the National Research Council points out, no single study can unequivocally establish the effectiveness of a curriculum material.

FYI

Online Update

- ***2061 Connections***, Project 2061's new electronic newsletter, brings you our latest research and findings. Read the current issue and sign up for a free subscription at **www.project2061.org/2061Connections**.

New Cities Expand Partnership's Reach

Thanks to funding from AAAS, 20 new communities will help spread the "Science. It's Everywhere" message, bringing to 26 the number of sites across the country. The new Partnership for Science Literacy collaborating sites are engaging families in their children's science education through online guides to local science resources like science centers, parks, zoos, and museums. Albuquerque, Atlanta, Boston, and Springfield, Missouri, are among the new cities that have already posted local versions of an English/Spanish *Family Guide to Science* on the Partnership's Web site (**www.ScienceEverywhere.org**) hosted by TryScience.org.

Project 2061 Takes Its Show on the Road!

Project 2061 will share its latest research and resources at the following two conferences this fall:

- **Association of Science-Technology Centers (ASTC) Annual Conference**, September 18–21, San Jose, California. Visit exhibit booth 1024 at the San Jose McEnery Center and speak with representatives from Project 2061 and the AAAS Education and Human Resources (EHR) Directorate. AAAS is also hosting a luncheon at the Tech Museum of Innovation on September 21 to report on the progress of the Partnership for Science Literacy, its national science literacy initiative with a local infrastructure. For more on the Partnership, visit **www.ScienceEverywhere.org**. For more on ASTC, see **www.astc.org**.
- **National Science Teachers Association (NSTA) Eastern Area Convention**, December 2–4, Richmond, Virginia. Visit the AAAS booths at the Greater Richmond Convention Center and meet staff from the EHR, Membership, and Project 2061 offices. For more on the NSTA regional meetings, see **www.nsta.org/conventions**.

Register Now for October *Atlas* Workshop

Project 2061 is pleased to offer another of our popular "Using *Atlas of Science Literacy*" workshops, to be held October 13–15, 2004, at AAAS headquarters in Washington, DC. *Atlas of Science Literacy* is a collection of strand maps that present conceptual connections among the ideas and skills that serve as goals for K–12 student learning. Participants will use *Atlas* and other Project 2061 resources to enhance their understanding of science literacy and to improve curriculum, instruction, and assessment. For workshop details and registration information, visit **www.project2061.org/workshops**.

Introducing…

Marcia L. Triunfol joins the staff as a senior program associate developing curriculum resources that promote science literacy. She comes to Project 2061 from AAAS *Science* Editorial, where she initially served as associate editor of SCOPE (Science Controversies On-line: Partnerships in Education) and later worked with *Science*'s Next Wave, GrantsNet, and *AIDScience*. Her previous work includes creating DNA Goes to School, a nonprofit organization in Brazil that offers hands-on workshops for high school students and science teachers. Triunfol holds a Master's in genetics from Universidade Federal do Rio de Janeiro and a Ph.D. in molecular biology through a collaboration program between Universidade Federal Fluminense (Brazil) and the Kennedy Krieger Institute at the Johns Hopkins University. Intern **Philip Piety** contributed to Project 2061's assessment work this summer. Piety is a doctoral student in education studies at the University of Michigan and holds a Master's in communications, culture and technology from Georgetown University. A former software industry consultant, product designer, and instructor, his primary research interests are the usability and accessibility of instructional materials for diverse students.

2061 *today*

American Association for the Advancement of Science
1200 New York Avenue, NW
Washington, DC 20005

Return Service Requested

Nonprofit Organization
U.S. Postage
PAID
Washington, DC
Permit No. 5676

This could be **your last issue**… *see card inside to renew your FREE subscription!*

www.project2061.org