Assessment in Science

Practical Experiences and Education Research

Edited by Maureen McMahon,
Patricia Simmons, Randy Sommers,
Diane DeBaets, and Frank Crawley
National Association for Research in Science Teaching

NSTApress
NATIONAL SCIENCE TEACHERS ASSOCIATION
Arlington, Virginia

NATIONAL SCIENCE TEACHERS ASSOCIATION

Claire Reinburg, Director
Judy Cusick, Senior Editor
Andrew Cocke, Associate Editor
Betty Smith, Associate Editor
Robin Allan, Book Acquisitions Coordinator

Will Thomas, Jr., Art Director
Tracey Shipley, Cover and Inside Design

PRINTING AND PRODUCTION Catherine Lorrain, Director
Nguyet Tran, Assistant Production Manager
Jack Parker, Electronic Prepress Technician

NATIONAL SCIENCE TEACHERS ASSOCIATION
Gerald F. Wheeler, Executive Director
David Beacom, Publisher

08 07 06 4 3 2 1

LIBRARY OF CONGRESS CATALOGING-IN-PUBLICATION DATA
Assessment in science : practical experiences and education research / Maureen McMahon ... [et al.], editors.
p. cm.
ISBN-13: 978-1-933531-00-7
ISBN-10: 1-933531-00-2
1. Science--Study and teaching (Elementary)--Evaluation. 2. Science--Study and teaching (Secondary)--Evaluation. 3. Science--Ability testing. I. McMahon, Maureen.
LB1585.A785 2006
507'.1--dc22
2006002237

NSTA is committed to publishing material that promotes the best in inquiry-based science education. However, conditions of actual use may vary, and the safety procedures and practices described in this book are intended to serve only as a guide. Additional precautionary measures may be required. NSTA and the authors do not warrant or represent that the procedures and practices in this book meet any safety code or standard of federal, state, or local regulations. NSTA and the authors disclaim any liability for personal injury or damage to property arising out of or relating to the use of this book, including any of the recommendations, instructions, or materials contained therein.

contents

introduction

The National Science Teachers Association and the National Association for Research in Science Teaching have teamed up to create *Assessment in Science: Practical Experiences and Education Research*. This book, intended as a resource for teachers of science, teacher-researchers, and science education researchers, shares methods, stories, and findings about assessment, one of the most pressing issues in today's K–16 science classrooms. *Assessment in Science* links "best-practice" ideas to sound science education research.

Assessment, as defined in the National Science Education Standards (National Research Council 1996), is "a systematic, multi-step process involving the collection and interpretation of educational data" (p. 76). Planning appropriate assessment and evaluation is challenging—and even more difficult in the current climate of high-stakes testing.

The authors whose chapters appear in *Assessment in Science* are practicing K–16 classroom teachers or university-based educators and researchers. The following questions framed the call for papers to which these authors responded:

* Has your analysis of your students' work generated new strategies for your teaching?
* What are the toughest challenges in student assessment, and what are your best solutions?
* Has your research led to insights for identifying evidence of student understanding?
* What practical implications do research findings have for classroom, district, and national reform?

Each chapter is organized into an introduction or background section of relevant research, research questions or issues investigated, methods used to generate and analyze data, results from data analysis, and conclusions and implications for classroom practices. Most authors included a list of resources for further reading and summaries explaining how their studies aligned with the National Science Education Standards.

Authors of the first 13 chapters discuss how they: merged research findings from the cognitive sciences and assessment research base, conducted research on assessment, synthesized a new approach toward assessment for practitioners, and extrapolated research findings on assessment into highly practical information for teachers to employ in the classroom (such as constructing test items). Research studies were conducted at early elementary, elementary, middle, high school, and university levels. The authors employed qualitative and/or quantitative research methods to investigate their respective questions about assessment. Some studies were conducted on a large scale, while others focused on a select small number of individuals. The last 5 chapters contain reports of K–16 educators who were engaged actively in teacher-as-researcher roles. The results reported by all authors have implications for further research and for classroom practices to enhance the teaching and learning of science.

The authors present a variety of perspectives, from bridging theoretical backgrounds to practical tried-in-the-classroom tools:

- **Vitale**, **Romance**, and **Dolan** offer an overall knowledge-based framework for assessment and instruction. Synthesizing research from the cognitive sciences, they outlined a typology for assessing students' science understanding. Using the typology, they describe how teachers can generate student assessment activities and also meet assessment principles of test validity and reliability.
- **Henriques**, **Colburn**, and **Ritz** have organized their article as a helpful how-to guide, outlining tips and strategies for developing test items (multiple choice, true/false, matching, and essay) that align with instructional goals. They also included examples of ineffective and effective test items.
- **King** describes the results of a study with young children in early elementary grades where she functioned as a participant-observer. The students generally responded positively and enthusiastically to open-ended assessment opportunities.
- **Farenga**, **Joyce**, and **Ness** discuss how a state-mandated assessment required teachers to use strategies that are opposite to those required for science inquiry. Adaptive inquiry was the strategy they synthesized to explain how a middle school teacher dealt with this conundrum.
- **McWaters** and **Good** delve into how middle school teachers used national, state, and local science standards to guide their assessment practices. They recommend that teachers, administrators, and researchers collaborate more closely on projects involving classroom instruction, science content standards, and classroom assessment.
- **Harris**, **McNeill**, **Lizotte**, **Marx**, and **Krajcik** report on how they created assessments aligned with their curricular design process (learning-goals-driven

design) and the National Science Education Standards. They describe how specific rubrics were developed, implemented, and evaluated in a chemistry unit for middle grades students and teachers.

* **Siegel**, **Hynds**, **Siciliano**, and **Nagle** describe the use of authentic-embedded assessments (part of the SEPUP modules) that assisted students in creating a clear sense of expectations for learning and helped them focus on learning rather than grades. The authors provide examples for setting up rubrics, detail the thinking behind the generation of criteria for scoring guides, and report on teacher responses from field testing in middle schools.
* **Kubitsky**, **Fishman**, **Margerum-Leys**, **Fogleman**, **Brunvand**, and **Marx** discuss two urban, middle school teachers' implementation of concept mapping as one way to assess their students' learning.
* **Barton**, **Cartier**, and **Collins** describe a long-term collaboration between university researchers and high school science teachers that focused on curriculum design. They report on how teachers used multiple and varied assessments (in a ninth-grade astronomy unit) to determine gains in student knowledge, inquiry skills, and how knowledge is judged in science.
* **Passmore** and **Stewart** write about assessment instruments developed for a high school course in evolution. Investigating students' understanding and ability to use models to explain phenomena, they document the changes in students' ideas through specially designed writing assessment assignments.
* **Hofstein**, **Mamlok**, and **Rosenberg** describe the influence of a new high school chemistry project on changing teaching practices, teachers' beliefs, and students' views toward the new assessments. Both groups perceived that the project established healthy learning environments in their chemistry classrooms.
* **Knight** and **Sullivan** discuss the lessons learned from a K–12 university engineering outreach initiative. Their comprehensive assessment plan provided the analytical structure to guide project activities, measure success, and inform future planning on the initiative.
* **Akerson**, **McDuffie**, and **Morrison** describe the partnerships between preservice and inservice teachers during a science methods course that resulted in mutual learning benefits. Their collaboration documented greater understanding by both groups about the role of performance assessments linked to the National Science Education Standards.
* **Buxton** writes about how conducting research in urban middle schools led him to question if new approaches to teaching assessment (contextually authentic inquiry models) can be reconciled with the demands of high-stakes testing.
* **Flick** and **Tomlinson** describe how their professional collaboration (university/elementary educators) led to the implementation of four cognitive strategies in class. They recount how they synthesized the Cognitive Assessment Inventory, merging research on assessment from the reading comprehension and cognitive science literature bases.

* **Sato**, **Baker**, **Fong**, **Gilbertson**, **Liebig**, and **Schwartzfarb** explain how they (all but Sata are middle school teachers) moved away from the role of teacher as giver of grades to teacher as conductor of learning, through their participation in a university-school research collaboration (CAPITAL).
* **Cheung**, **Cody**, **Hirota**, **Rubin**, **Slavin**, **Coffey**, and **Moorthy** focus their article on five middle school science teachers (part of a larger research-practice-collaboration) who changed their beliefs and actions about assessment. The teachers' reflections form the central component of this chapter.
* **Goodnough** and **Long** describe a study in which mind mapping (a visual assessment tool) was used to examine middle school students' understanding of science and understanding of team work.

Readers are also referred to the websites listed below for additional, current information on K–16 assessment.

American Association for Higher Education
www.aahe.org/assessment/web.htm

Annenberg/CPB Project's Website for Learners and Educators: Assessment in Math and Science
www.learner.org

Buros Center for Testing
www.unl.edu/buros/

National Center for Research on Evaluation, Standards, and Student Testing
http://cresst96.cse.ucla.edu/index.htm

National Science Education Standards
www.nap.edu/readingroom/books/nses/html/

National Science Teachers Association (NSTA)—Position Statement on Assessment in Science Teaching
www.nsta.org/positionstatement&psid=40

No Child Left Behind: Regulations, Information, Implications
www.ed.gov/nclb/landing.jhtml?src=pb

Northwest Regional Educational Laboratory—Assessment
www.nwrel.org/assessment/

U.S. Departments of Education: State Comprehensive Testing, Accountability and Assessment—All 50 States
www.eduhound.com/k12statetesting.cfm

We thank the authors and Claire Reinburg and David Beacom at the National Science Teachers Association for their patience during the assembly of this volume. We extend our appreciation to the reviewers for their timely and helpful commentary and feedback. In addition, we thank Connie Quinlan (University of Missouri-St. Louis) for her assistance with the preparation of the articles.

about the editors

Maureen McMahon teaches elementary science methods at California State University, Long Beach, where she chairs the Department of Science Education.

Patricia Simmons holds the Orthwein Endowed Professorship in Life-Long Learning in the Sciences at the University of Missouri-St. Louis (UMSL) and is director of the Institute for Mathematics and Science Education and Learning Technologies.

Randy Sommers is completing his doctorate in educational leadership at UMSL. He is a former elementary school teacher in Missouri.

Diane DeBaets is a research assistant at the Institute for Mathematics and Science Education and Learning Technologies at UMSL.

Frank Crawley, chair of the Department of Science Education at East Carolina University, teaches courses in physical science, collaborative action research, and learning and teaching science.

CHAPTER 7

using rubrics to foster meaningful learning

Marcelle A. Siegel, Paul Hynds, Marc Siciliano, and Barbara Nagle

Introduction

When the school day ended and the last student complaining about a grade finally left the classroom, the middle school science teacher slumped into her chair. How many—if any—students, she wondered, had learned something significant during the unit on household materials? She had posed some short essay questions that asked them to analyze the chemical information on several cleansers and to discuss potential toxic hazards. Students compared the benefits and risks of the use of the materials. When reading the essays, she saw that the students recalled the labels and some of the hazards of certain household chemicals, but could not weigh the advantages and disadvantages of using them. Their responses revealed a lack of critical analysis. She could still hear the student's voice: "What's wrong with my answer? I told you all about the ingredients in bleach and other cleaners and that some of them are dangerous. I should at least get a B!"

Unfortunately, similar scenes are played out in many classrooms. The teacher's expectations are not met by the student's performance, and yet the teacher is not certain what to do next. Students often assume that if they give any kind of detailed answer it should be sufficient. Teachers sometimes find themselves talking with students more about what grades they are getting than about what the students are learning. What kinds of tools would help teachers clarify expectations and focus attention more on learning and producing quality work than on grades?

Background

Since its inception in the mid-1980s, the Science Education for Public Understanding Program (SEPUP) at the Lawrence Hall of Science (University of California-Berkeley) has developed an array of "issue-oriented" instructional and assessment materials that meet the recommendations of major reform efforts and the expectations of schools. Students gather scientific evidence during guided investigations and apply the evidence during discussions, debates, role plays, and other activities related to societal or personal issues (Thier and Nagle 1994).

In order to assess students' higher-level thinking, decision-making, and process skills, a new type of assessment system was necessary. In collaboration with the Berkeley Evaluation and Assessment Research (BEAR) group at the University of California at Berkeley Graduate School of Education, SEPUP developed an embedded, authentic assessment system as part of the SEPUP full-year course, Issues, Evidence and You (Roberts, Wilson, and Draney 1997; Wilson and Sloane 2000). The basic assessment system had three components:

* Five variables (see Figure 1) that defined the key domains in which students were expected to make progress during the year
* Actual assessment tasks
* Rubrics used to evaluate student performance on the tasks

Each variable had an associated scoring guide (also called a *rubric*, the term we will be using throughout this chapter), which provided criteria for different levels of student performances. These criteria set the standards of performance for different levels of responses to a task. These levels, or standards of performance, not only informed students of what they were doing, but pointed the way to improving performance by setting clear expectations.

In this chapter, we discuss the use and adaptation of rubrics in many classrooms nationally as teachers (and students) moved toward meeting their goals. One SEPUP teacher wrote about the power of using rubrics: "Students knew what was expected of them and it showed in some of the answers that I received. They thought more thoroughly about what they were going to answer."

Frequent and systematic use of rubrics in the classroom can fundamentally change the dynamics of the teacher-learner interaction. This change can be seen in greater student motivation, improved learning and metacognitive skills, and greater teacher understanding of students' learning. The use of rubrics can also mirror scientific habits of mind and reinforce the importance of evidence.

The SEPUP assessment system was cited by the National Research Council as an exemplary model of measurement (NRC 2001). It was originally developed for a specific middle school course, Issues, Evidence and You. Results from a series of studies indicated that the assessment system was psychometrically sound, that SEPUP students out-performed control students, and that SEPUP users of the

FIGURE 1.
The SEPUP Assessment System Variables

Scientific Process

Designing and Conducting Investigations (DCI)—Designing a scientific experiment, performing laboratory procedures to collect data, recording and organizing data, and analyzing and interpreting the results of an experiment.

Evidence and Tradeoffs (ET)—Identifying objective scientific evidence as well as evaluating the advantages and disadvantages of different possible solutions to a problem based on the available evidence.

Scientific Concepts

Understanding Concepts (UC)—Understanding scientific concepts (such as properties and interactions of materials, energy, or thresholds) in order to apply the relevant scientific concepts to the solution of problems.

Scientific Skills

Communicating Scientific Information (CM)—Organizing and presenting results of an experiment, or explaining the process of gathering evidence and weighing tradeoffs in selecting an effective solution to a problem that is free of technical errors.

Group Interaction (GI)—Developing skill in collaborating with teammates to complete a task (such as a lab experiment), sharing the work of the activity, and contributing ideas to generate solutions to a given problem.

Source: Science Education for Public Understanding Program (SEPUP), Lawrence Hall of Science (University of California-Berkeley).

assessment system out-performed those who used the course alone (Wilson and Sloane 2000). The SEPUP developers collaborated with field-test teachers to adapt the core components of the assessment system for use in modules and two additional full-year courses—Science and Life Issues (middle school life science) and Science and Sustainability (integrated high school science).

Research Questions

The most recent research focused on (1) how teachers used the SEPUP assessment system, (2) how the system was modified to retain its power while making it easier to use, and (3) the additional tools that were incorporated into the system.

The first author of this chapter and SEPUP staff engaged in several ongoing studies to refine the variables and associated rubrics for the Science and Life Issues course, the Science and Sustainability course, and 12 shorter modules. The purpose was to refine the rubrics for different instructional goals, develop additional reliable items, and increase the ease of use for students and teachers. This process included discussion and testing among SEPUP developers, two years of field testing in nearly 50 classrooms, and analysis of data. After two rounds of field-test conferences and collection of written feedback from participants, teacher representatives from the field-test centers joined the SEPUP staff for a summer conference and participated in final revision of the embedded items and the rubrics. Research on the life science course involved (a) piloting new multiple-choice and extended items with 600 students, (b) analyzing data (using Rasch modeling with Conquest software and traditional techniques using SPSS software), and (c) refining items for a national field test in 2002–2003. Based on these studies, SEPUP developed more focused variables, simpler language for the rubrics, approximately 100 additional reliable items, and a feedback form.

DEVELOPING AN EMBEDDED ASSESSMENT SYSTEM

The SEPUP assessment system was based on four principles (Wilson and Sloane 2000). First, it was based on a developmental perspective of student learning, with a focus on student progress over the course of a year. The use of the same five variables and rubrics throughout the year reflected this developmental perspective.

Second, the system matched instructional goals. It was focused on understanding important scientific concepts, the processes of scientific investigation and analysis of information, and evidence-based decision making. To ensure a match between curriculum, instruction, and assessment, the instructional materials and assessment tasks were developed at the same time, and the assessment system was part of all field-testing activities.

Third, the system was designed to generate quality evidence. The assessment tasks, methods of measurement, analysis, and reporting needed to be of high technical quality, which meant maintaining standards of fairness, ensuring that results were compared across time and context, and performing traditional studies of validity, reliability, and equity.

Fourth, the system was built on the principles of teacher management and responsibility. The teacher used the assessment evidence to guide the learning process. SEPUP and BEAR involved teachers in all aspects of the development of the assessment system, including (1) developing the tasks and rubrics, (2) collecting and scoring student work, and (3) interpreting the results.

Alternative assessment created new challenges for teachers, such as finding time to score open-ended responses, translating rubric scores to letter grades, and helping students understand that the new form of assessment was intended to guide learning rather than judge student performance. To help teachers manage the new assessment system, SEPUP staff created tools (see p. 96 for the tools) to help them use the system effectively.

FIVE TYPES OF LEARNING ASSESSED IN SEPUP

The SEPUP assessment system measured five types of content and process learning, called "variables," that were central to the instructional materials. The variables were Designing and Conducting Investigations, Evidence and Tradeoffs, Understanding Concepts, Communicating Scientific Information, and Group Interaction (Figure 1). These five variables represented student learning in terms of the core concepts of SEPUP courses based on decision making about societal issues and clarified the conceptual framework for instruction and assessment.

Two SEPUP rubrics

Each SEPUP variable had an associated rubric that set forth the expected levels of performances for students. For example, the rubrics began with a 1, a minimal response teachers often described as "on your way." A more thoughtful response earned a 2, or "almost there." The goal for students was a 3 for a "complete and correct" response. This included scientific and conceptual understanding, the use of evidence in communicating that understanding, and evaluating alternatives. The criteria for 4, or "going beyond," described answers in which students displayed that they were going further in their thinking, such as connecting the specifics of their responses to other ideas. The five rubrics had these four levels of criteria in common, but each also included specific, unique criteria. For example, the Evidence and Tradeoffs (ET) Rubric (Table 1) measured a core goal of SEPUP courses, the ability to use scientific evidence to analyze the advantages and disadvantages of a real-world decision. At Level 3 in that rubric, a student provided the major reasons for or against a decision and supported each with relevant and accurate scientific evidence. Often, students gathered the evidence in the form of data from a hands-on activity; at other times, they obtained evidence from simulations and readings. One SEPUP field-test teacher wrote that using the ET rubric helped middle school students make decisions: "First of all, the questions [i.e., the assessment questions in the curriculum] ask the students to take a stand, make a decision. Typical middle school youngsters like to ride the fence; noncommittal is safe. So all of a sudden they are expected to come up with at least two options. This threw them quite a bit." Eventually, students were able to clarify their viewpoints based on evidence.

Many teachers also found that secondary students could generate a conclusion, but could not explain their reasoning regarding, and judgment of, multiple sources of evidence. Using the ET rubric helped students improve at this type of sophisticated thinking. One new SEPUP teacher commented, "Students actually realize they must provide evidence for their opinions!"

The Communicating Scientific Information (CM) Rubric (Table 2) measured how well students expressed their arguments and/or ideas. This rubric was designed to measure written (e.g., lab report), oral (e.g., presentation), and visual (e.g., poster) student reports. One element of the variable referred to the organization of the response, and the second element involved technical aspects, such as grammar and eye contact.

TABLE 1.
Evidence and Tradeoffs (ET) Rubric for Use With Middle School Science Students

Score	Using Evidence Response uses objective reason(s) based on relevant evidence to argue for or against a choice.	Using Evidence to Make Tradeoffs Recognizes multiple perspectives of issue and explains each perspective using objective reasons, supported by evidence, in order to make a choice.
4	Accomplishes Level 3 AND goes beyond in some significant way, e.g., questioning or justifying the source, validity, and/or quantity of evidence.	Accomplishes Level 3 AND goes beyond in some significant way, e.g., suggesting additional evidence beyond the activity that would further influence choices in specific ways, OR questioning the source, validity, and/or quantity of evidence and explaining how it influences choice.
3	Provides major objective reasons AND supports each with relevant and accurate scientific evidence.	Uses relevant and accurate scientific evidence to weigh the advantages and disadvantages of multiple options, and makes a choice supported by the evidence.
2	Provides some objective reasons AND some supporting evidence, BUT at least one reason is missing and/or part of the evidence is incomplete.	States at least two options AND provides some objective reasons using some relevant evidence BUT reasons or choices are incomplete and/or part of the evidence is missing; OR provides only one complete and accurate perspective.
1	Provides only subjective reasons (opinions) for choice; uses unsupported statements; OR uses inaccurate or irrelevant evidence from the activity.	States at least one perspective BUT only provides subjective reasons and/or uses inaccurate or irrelevant evidence.
0	Missing, illegible, or offers no reasons AND no evidence to support choice made.	Misses information, illegible, or completely lacks reasons and evidence.
X	Student had no opportunity to respond.	

Source: Science Education for Public Understanding Program (SEPUP), Lawrence Hall of Science (University of California-Berkeley).

TABLE 2.
Communicating Scientific Information (CM) Rubric for Use With Middle School Science Students

Score	Organization	Technical Aspects
	Response logically organizes arguments, evidence and/or ideas related to a problem or issue. Ideas are frequently, but not always, organized in the following way: * Introduction * Explanation of procedures * Presentation of relevant evidence * Consideration/ interpretation of the evidence * Conclusion	Response conveys a concept or idea clearly by using the assigned medium appropriately. Possible forms of communication and areas to examine are * *written (e.g., report):* sentence structure, grammar, spelling, and neatness * *oral (e.g., presentation)*: enunciation, projection, and eye contact * *visual (e.g., poster)*: balance of light, color, size of lettering, and clarity of image
4	Accomplishes Level 3 AND goes beyond in some significant way.	Accomplishes Level 3 AND enhances communication in some significant way. No technical errors.
3	All parts present and logically organized.	Presents response that is clear and easy to understand, with few minor errors.
2	Shows logical order BUT part is missing.	Provides understandable response BUT clarity is missing in places; technical errors may exist BUT do not prevent audience from understanding the message.
1	Lacks logical order OR is missing multiple parts.	Detracts audience from understanding the message with unclear and technical errors.
0	Missing, illegible, or contains no evidence or ideas related to the task.	Misses evidence, illegible, incoherent, or contains no evidence or ideas related to the task.
X	Student had no opportunity to respond.	

Source: Science Education for Public Understanding Program (SEPUP), Lawrence Hall of Science (University of California-Berkeley).

TEACHER TOOLS IN THE ASSESSMENT SYSTEM

Additional components of the SEPUP assessment system were designed to help teachers use the assessments effectively. Assessment blueprints provided a chronological list of course activities with potential assessment opportunities. Link tests and item banks were sets of additional assessment tasks and questions, some closely tied to the curriculum and some transferring items that were used by the teacher to monitor understanding or conduct a summative assessment. Assessment moderation was a process for teachers to compare ideas for scoring student work, deepen their understanding of students' responses, and come to an agreed-upon standard. The SEPUP developers and BEAR researchers used the moderation process with teachers to select exemplars of student work at each scoring level for a variety of tasks. Exemplars provided teachers with a model of how to score student work and a tool for modeling expectations to students.

Findings

Teachers found that even though the assessment system offered new challenges, it was also rewarding to use (Roberts and Wilson 1998). One SEPUP field-test teacher commented that the assessment system gave her and her students "specifics to look and aim for. We had common language for our discussions. The application of the concepts gave both of us a positive way to judge growth." Another field-test teacher remarked on the usefulness of the exemplars: "I was totally impressed with the quality of writing that I received from students.... Students were empowered by the rubric and the exact criteria for which they would be measured. The Level 3 examples given in the teaching materials also enabled students to have a model for excellence." Another teacher said she spent a lot of time giving students feedback on why they received a score and that it paid off in more focused answers: "[I highlighted] in different colors each of the items I was looking for and scored them. (This was very time consuming and I [used] lots of highlighters.) It was worth it, though! The temptation for [students] to just put down words, write for weeks, was somewhat halted since they knew the criteria. (Quality, not quantity is our slogan—along with evidence, not emotion)."

MODIFICATION AND FIELD TESTING

After the collaborative project to develop the assessment system for the Issues, Evidence and You course, we continued to study and refine the variables and associated rubrics for two more courses—Science and Life Issues and Science and Sustainability. This resulted in more focused variables, with fewer sub-variables and somewhat simpler language in the rubrics. Recently, we refined and field-tested three rubrics (Designing Investigations, Analyzing Data, and Evidence and Tradeoffs) for use with 12 supplementary science modules for grades 6–12. Only one rubric was emphasized in each module.

Our process of adapting the rubrics for the modules included changes in content, clarity, and usability. We first discussed the criteria for each level of the rubric and attempted to achieve clear cut-offs between the levels. We scored additional student work to see if the new rubric was effective, and selected exemplars of student work for each level that illustrated the criteria for that level. Based on this evidence, we revised the rubric and analyzed it in light of additional student work.

During this process, we devised a new tool related to rubrics for teachers called a feedback form (Table 3). Teachers often told us that the rubrics were text-heavy and, at first, overwhelming for students. The feedback form was a concise version of a rubric that only listed the criteria for Levels 3 and 4. It was intended to help the teacher introduce rubrics and to show them how to provide feedback to students. It contained space for the teacher to offer written comments to the student about why his or her answer received a particular score. The feedback form was also designed to help teachers organize their comments to students.

TABLE 3.

Feedback Form for Revised Evidence and Tradeoffs Rubric for Modules

Complete and Correct Response	Yes	Almost	No
You use evidence to support a logical interpretation of the data. You evaluate the source, quality, and/or quantity of evidence.			
Comments:			

You accomplish the above and go beyond in some significant way, such as:	Yes	Almost	No
You present a thorough examination of evidence. You connect your ideas with the science concepts learned. You provide an explanation for why alternative ideas were discarded. You provide suggestions for further relevant investigations. You include a diagram or visual to clarify your ideas. Other:			
Comments:			

Source: Science Education for Public Understanding Program (SEPUP), Lawrence Hall of Science (University of California-Berkeley).

Each module and its embedded assessments and associated rubric were field-tested by approximately 25 teachers in 10 centers. Selected teachers from each field-test site joined SEPUP staff during a three-day conference and communicated with project staff regularly during the nine months of testing. At the conclusion of the school year, teachers responded to questions about each activity in the module as well as about the module as a whole. This information was collected in an online database that gave SEPUP staff an efficient way to sort and analyze all of the feedback. Another useful data source was the adaptations to the materials and rubrics that teachers made.

As teachers used rubrics to shift student attention from grades to learning, improvements in performance were observed. For example, at the Young Women's Leadership Charter School in Chicago, students were evaluated on individual outcomes each trimester with task-specific rubrics with three common levels: "Not Yet," "Proficient," or "High Performance." Steve Torres, an eighth-grade physical science teacher, elected not to use rubrics or related scoring rubrics during the first trimester, except for the final assessment. In the second and third trimester, he began using rubrics and consistent assessment language related to the rubrics for all assignments. Torres wrote phrases such as "incomplete" at the top of students' assignments, along with sidebar comments on how to improve work. As one might expect, student outcomes appeared to decrease in the second trimester (possibly because students were held accountable to a higher standard than the first trimester). The number of trimester 1 scores for "Not Yet," "Proficient," or "High Performance" were 206, 449, and 104, respectively. Trimester 2 scores were 292, 366, and 73, respectively. With further use of the criteria and rubrics in the third trimester, the average performance increased dramatically (172, 345, and 240, respectively). While no conclusions can be drawn from a single case during one year, the results suggested what many teachers have observed: The more times that students were exposed to rubrics, the more students performed at the highest levels.

SEPUP middle school students in classrooms using the assessment system achieved statistically and educationally significant gains across the five SEPUP variables, including the ability to use evidence and make tradeoffs (Wilson et al. 1995). Wilson and Sloane (2000) reported that gains for students with teachers who received professional development for fully using and implementing the assessment system during the year were 3.46 times greater than those of the control non-SEPUP group and 2.14 times greater than the SEPUP group that received nonassessment professional development (differences between average gains were significant at the .05 level).

Implications for Teaching: How to Use Rubrics in the Classroom

Based on extensive field-test research and ongoing communication with teachers after they implemented the revised, commercial versions of SEPUP materials, we have learned how teachers used the assessment system, and especially the rubrics, to enhance learning. Following are six recommendations that teachers can apply.

1. Create An Effective Learning Environment

All teachers want their classrooms and schools to provide effective learning environments. Most teachers have in mind a picture of a classroom where the focus of attention is on the exchange of ideas, not on performance rankings (grades). Yet students often do not perceive the classroom in this way. Throughout most students' lives, they have been trained to measure their successes by the grades that they receive. Getting the "right answer" is what counts. How can a teacher create an environment that will help students move beyond the comfort zone of the easy and the familiar into the more challenging, and sometimes threatening, arena of testing their ideas and thinking for themselves?

Because rubrics define the expected levels of student performance, they focus attention on criteria for improvement. For example, in some SEPUP rubrics, the criteria for Level 4 ("going beyond") clarified ways to improve critical thinking, such as questioning the quantity or quality of evidence or making specific suggestions for further investigations. When students internalized this way of analyzing their own work, they began asking, "How can I make this better? What else do I know that is relevant to this problem?" Thus, the process of teaching students to use the rubrics as guides for their work moved the classroom focus to the quality of work—exactly where teachers want to go.

Rubrics foster some of the same habits of mind that educators wish to promote in science students. Accepting a proposition as valid requires the evaluation of information and evidence, rather than individual authority. SEPUP teachers often used the Evidence and Tradeoffs Rubric to help students understand that scientific evidence can provide information for personal and societal decisions and that the interpretation of evidence is an integral part of many human endeavors. Another aspect of the scientific approach fostered by rubrics is that the information available may not yield a "perfect" scientific answer or decision—one must develop an ability to make progress despite uncertainty and to be willing to revise a solution in the light of evidence. In the real world, in science, and in the rubric-based classroom, answers are subject to revision as more evidence is collected. Using the Evidence and Tradeoffs Rubric, students learned about weighing the quality and quantity of evidence on different sides of an issue. Classroom conversations revolved around topics of reliability and reproducibility. Another study also found that SEPUP high school students' attitudes about the relevance of science in their lives were enhanced (Siegel and Ranney 2003).

2. Introduce Rubrics Right Away

All teachers understand the importance of the critical first weeks of a school year. It is during this time that classroom structure is established, goals are set, and expectations declared. In order to implement rubrics successfully, one teacher developed a strategy that focused on building observation skills while introducing a rubric-based assessment system at the same time. Mike Lach, a National Board Certified teacher, designed a unit on observations and interpretations for an introductory unit to physical science courses. One objective of this unit was to help students distinguish between observation and inference. He and Marc Siciliano implemented this unit while teaching ninth grade at Lake View High School in Chicago.

The unit consisted of a series of activities that took students through progressive stages of learning about observations and creating practice rubrics. First, students learned the importance of making detailed observations in science. Students were presented with open-ended stories, ambiguously written, so that multiple interpretations were possible. Students shared their interpretations and saw that confusion existed when stories lacked detail and facts. As the notion of details and facts was expanded, students learned the basic definition of a scientific observation.

Second, the goal was to have students see the value in rubrics by doing the same task twice—once "blind" and the second time with a rubric—and ending with a more coherent outcome. Without a rubric, students were asked to classify a set of observations of an object (a common chalkboard works nicely) to test their understandings of "good" observations. The students initially placed the observations into two groups, one representing "better" observations and the other indicating "poorer" ones. These two groups were then further divided so that there were four groups representing a range from best to worst. Students then named each category and listed the defining characteristics of each. They suggested characteristics such as complete sentences, multiple adjectives, spelling errors, exactness, inclusion of interpretations, and so forth. When the class then scored their observations based on these insufficient categories, they saw that they scored many observations differently. They saw that some observations could be placed in multiple categories and began to recognize the notion of a continuum. These realizations sparked interesting questions about the process and revealed a practical need for rubrics to set criteria for better responses.

As students shared their categories and the rationales for the categories, they saw the limitation of the activity: The task was vague. Thus, the third step was to have students come to consensus by deciding on clear characteristics for each category. They created a rubric and used it to score observations of a different object. This time, the task only took a half period, and students said, "Let's do it this way again." They bought into the concept of a rubric, because they experienced its power and efficiency.

Siciliano and Lach spent approximately two weeks introducing rubrics. They recommend that teachers spend ample time introducing rubrics. One field-test teacher commented about a module, "At first it was very difficult but it got much better as the students grew in their understanding." It should be noted that it took longer than two weeks for students to completely understand the process.

Once students understood the usefulness of basic rubrics, they were primed for more advanced rubrics. Lach and Siciliano introduced the SEPUP rubrics at this point, and showed students that they would be assessed on five variables (see Figure 1, p. 91) throughout the school year. The rubrics were placed over the walls of the classroom. One of Siciliano's students remarked, "Oh, to get a 3, I have to show that I did the experiment, I have the data, I've synthesized the data, and this is how it can be applied to my conclusion." The expectations were established and the focus became a process of learning to meet those expectations.

3. Customize the Rubrics

Because the SEPUP rubrics set the learning goals in the class, teachers must make sure that they match their own goals. SEPUP teachers sometimes added another rubric to the mix or adapted an existing rubric to better reflect their goals. Many teachers also recommended rewriting the rubrics in students' own words so students could see the connections between the rubric language and their own words. Teachers then posted the student rubrics around the classroom and referred to them. Donna Parker, a high school integrated science teacher in Columbus, Ohio, rewrote the Group Interaction Rubric using student language (Table 4).

Another SEPUP teacher adapted the Designing and Conducting Investigations Rubric to emphasize her goal regarding student predictions. The Designing Investigation element of this goal stated that a Level 2 response "[i]ncompletely states a problem or the design of an experiment." She adapted the criteria at each level to emphasize the importance of a prediction in the design of an experiment or investigation. Her Level 2 criteria were as follows: "Stated purpose is appropriate to the assignment, but may be incomplete. A prediction is mentioned, but not clearly stated or not related to the stated purpose."

4. Shift Students' Attention From Grades to Learning

In one classroom in Chicago, Doug Goodwin made comments to his students such as, "That looks like a 2 to me. How can you make it a 3?" His colleague, Marc Siciliano, wrote feedback questions that referred to getting a Level 4 for an answer and not settling for a 3. For instance, he stated that three pieces of evidence were needed in order to get a 4, even though students may have only learned two. Such expectations stretched students' thinking and encouraged connections to prior knowledge or extensions in order to achieve at high levels.

Goodwin and Siciliano constantly discussed terms like *almost there* and *complete and correct* with each other and their students, making sure that the message was consistent. In order for students to associate numbers with achievement, the teachers also named each level of the class rubric ("awesome," "cool," "embarrassing," and "bogus"). After all of the names were compiled, the class agreed on the following names for each level: "Above and Beyond" = 4; "Complete and

TABLE 4.
Customized Group Interaction (GI) Rubric in Student Language

Score	Task Management What to look for: Group stayed on task, managing time efficiently.	Group Participation What to look for: Group members worked together as a team and the ideas of all members were valued and weighed in working toward the common goal.
4 (A 50 pts)	Accomplishes Level 3 AND goes beyond in some significant way, e.g., group defines own approach to more effectively manage activity, group members actively help each other accomplish the task, group uses extra time productively. **(Help other groups; assign roles; different people try things different ways; rework; extra work.)**	Accomplishes Level 3 AND goes beyond in some significant way, e.g., group members actively <u>ask questions</u> about each other's ideas, group members <u>compromise</u> if there are disagreements, group members actively help <u>each other accomplish</u> the task.
3 (B 40 pts)	Group <u>managed time well</u> and <u>stayed</u> on task <u>throughout</u> the activity.	All group members participated and respectfully considered each others' ideas. **(All working, contributing, listening. No arguing. No horseplay.)**
2+ (C 35 pts) 2 (D 30 pts)	**(Group works only when told.)** Group stayed on task <u>most</u> of the time.	Unequal **(one person does not do his or her part)** group participation OR group respectfully considered some **(only listen to people you like)**, but not all, ideas.
1 (F 20 pts) 0	Group was off-task majority of the time, but task completed. **(Talking in groups or talking to other groups.) (Gossip. Not doing anything. Horseplay. Sleeping. Working on other subjects.)** Group did not stay on task, which caused task not to be completed.	Significantly unequal group participation **(one person does all work)** OR group totally disregarded some members' comments and ideas. **(Ignore group member. "Shut up." "That's stupid!" Talking while other members talk.)** Single individual does entire task.
X	Group was not present.	

Note. Underlining was done by the teacher; parentheses and **different type** *indicate new text added based on student input.*

Source: Science Education for Public Understanding Program (SEPUP), Lawrence Hall of Science (University of California-Berkeley).

Correct" = 3; "Almost There" = 2; and "Major Errors" = 1. Developing the names helped students take ownership.

In these classrooms, students later questioned why certain of their responses had received a 2 rather than a 3. One student, examining an answer to an Evidence and Tradeoffs question, thought that the answer included sufficient relevant evidence. Siciliano explained that such dialogue demonstrated a greater level of student concern and understanding than did the students' questions during previous classes. This discussion opened the door for the teacher to focus on the additional evidence needed for a complete answer.

Rubrics also helped students view learning as a process. Siciliano's philosophy was that students start at different places on a continuum and some may take longer than others to reach Level 3 and 4 responses. He found something to build on in each student's work. Many of his students stated that rubrics motivated them in science. Having worked on a response only to have it marked with a big red X was much different than receiving a 2. There was no feedback from an X. A score, based on a rubric, clearly defined what was correct in the response and what could be improved.

5. Help Students Monitor Their Own Progress

In Chapter 12 of *Teaching Problems and the Problems of Teaching*, Lampert (2001) discussed ways to help students at different starting points improve in mathematics. One way is to foster independent learning by having students monitor their own progress (an aspect of metacognition). Students can do this by scoring their own and each other's work, thereby becoming more invested in the results. Another strategy, used by Chicago teacher Siciliano, is to have the teacher score student work and then simply tell students to check the rubric and figure out on their own why they received a particular score. This latter strategy demands practice and repetition, but gradually students do monitor their own learning. Siciliano found that, before he used rubrics, it took a lot of time to go over scored papers with students, because it was necessary to break down each problem. Rubrics help students understand what type of element is unclear in their responses, so they can be more in control of their learning.

An additional strategy is to have students keep a running record of their scores. Periodically, students reflect on their progress using their score records, and they make a note of any science questions or confusions they still have. This helps the student and teacher to monitor understanding and the teacher to adapt instruction.

6. Use Rubric Scores Within a Letter Grade System

Most school districts require letter grades and/or percentages for permanent records; however, this should not discourage teachers from using rubrics. Although Levels 0–4 may correspond with traditional grades, they should not be equated (a

4 should not be equated with an A, and a 2 should not be equated with a C). The rubric is a tool for documenting student development over time. Because Levels 3 and 4 of a SEPUP rubric describe major learning goals, it is unlikely that very many students would achieve a Level 3 at the beginning of the school year. Some students may be able to improve a scoring level over the course of a unit, while others will require a longer period of time to improve their performances. Thus, consistent Level 3s may not be necessary for a student to get an A or a B, especially at the beginning of a unit or course.

We encourage teachers to place rubrics into the contexts of their classes and local standards. The overall grading system is likely to include other criteria, such as completion of assignments and class participation. For example, out of a total of 20 points, a teacher may grade students' investigations and embedded assessments and assign a rubric score of 4, with the remaining 16 points determined by other criteria. Another teacher may decide that students who are able to improve by one scoring level (over a semester or a year) should receive an A or B for this aspect of work. A student who consistently scores 3s has demonstrated substantial competence, but no growth, and would not receive an A. Siciliano used spreadsheet formulas to convert rubric scores to more traditional grades for summative assessments. Adapting grading software also works well. Generally, these types of approaches satisfy the district, the parents, and those students who need to see a letter grade to confirm their successes.

Policy documents have increasingly referred to assessment practices as either summative (at the end of a unit or course to document learning) or formative (during instruction to inform learning). Using assessments formatively with students is an essential part of creating a culture of learning, although it does represent a major change for students and teachers (Bell and Cowie 2001).

Conclusion

Rubrics are tools for teachers and students. Teachers can clarify learning goals, give feedback, and help students build understanding through rubrics. Students can better understand learning goals, focus on learning rather than grades, and monitor their progress through rubrics. The development of an appropriate assessment system and classroom setup enables teachers to maximize the power of rubrics and use them to foster meaningful learning!

Acknowledgments

We are grateful to our colleagues on the SEPUP and BEAR staff who contributed to development of the assessment system and to the following dedicated teachers for their input: Doug Goodwin, Michael Lach, Donna Parker, and Steve Torres.

Links to the National Science Education Standards

The work described in this chapter addresses many of the concerns of the National Science Education Standards for Science Content, Teaching, and Assessment (NRC 1996). For example, SEPUP's Evidence and Tradeoffs Rubric provides a way to assess students' application of science in personal and social perspectives. The SEPUP Assessment System was created based on four principles that mirror Assessment Standards A–E, such as technical quality, lack of bias, and clear assumptions about student development. This study also incorporated the principles of Teaching Standards B and C.

References

Bell, B., and B. Cowie. 2001. *Formative assessment and science education*. Dordrecht, The Netherlands: Kluwer Academic.

Lampert, M. 2001. *Teaching problems and the problems of teaching*. New Haven, CT: Yale University Press.

National Research Council (NRC). 2001. *Knowing what students know: The science and design of educational assessment.* Washington, DC: National Academy Press.

Roberts, L., and M. Wilson. 1998. *An integrated assessment system as a medium for teacher change and the organizational factors that mediate science teachers' professional development.* BEAR Report Series SA-98-2. Berkeley, CA: University of California.

Roberts, L., M. Wilson, and K. Draney. 1997. *The SEPUP Assessment System: An overview.* BEAR Report Series SA-97-1. Berkeley, CA: University of California.

Siegel, M. A., and M. A. Ranney. 2003. Developing the Changes in Attitude about the Relevance of Science (CARS) questionnaire and assessing two high school science classes. *Journal of Research in Science Teaching* 40 (8): 757–775.

Thier, H. D., and B. W. Nagle. 1994. Developing a model for issue-oriented science. In *STS education: International perspectives on reform,* eds. J. Soloman and G. Aikenhead. New York: Teachers College Press.

Wilson, M., and K. Sloane. 2000. From principles to practice: An embedded assessment system. *Applied Measurement in Education* 13 (2): 181–208.

Wilson, M., K. Sloane, L. Roberts, and R. Henke. 1995. *SEPUP Course I, Issues, Evidence and You: Achievement evidence from the pilot implementation.* BEAR Report Series SA-95-2. University of California, Berkeley (*www-gse.berkeley.edu/research/BEAR /publications/sepup95.pdf*).

Resources

The Science Education for Public Understanding Program (SEPUP)

SEPUP courses include a student book with investigations and readings; an extensive teacher's guide that provides suggested teaching approaches, scientific background, information about students' ideas and possible responses, and suggestions for modifying or enhancing lessons for different student groups; and a full equipment kit. The modular materials provide student handouts as photocopy masters. More information is available about SEPUP courses and modules at *http://sepuplhs.org/currmat/index.html*, *http://sepuplhs.org/resources/assessment,* and *http://sepuplhs.org/profdev.*

One of SEPUP's professional development programs, Elementary Science Teacher Leadership (ESTL), has developed 10 guides for preservice and inservice teacher educators. Each guide includes activities, rationales, and resources to help teacher educators lead nine hours of professional development sessions. The Learning About Assessment guide is particularly relevant: *http://sepuplhs.org/profdev/index.html#teacher_educators*.

The National Center for Education Statistics (NCES)

NCES, in the U.S. Department of Education, is the primary federal resource for collecting and analyzing data related to education in the United States and other nations (*http://nces.ed.gov/practitioners/*).

The Center for the Assessment and Evaluation of Student Learning (CAESL)

CAESL seeks to improve student learning and understanding in science by focusing on effective assessment. The following link takes users to the resources page to find articles and video-based tools: *www.edgateway.net/cs/caesl/print/docs/310*.

Arter, J., and J. McTighe. 2001. *Scoring rubrics in the classroom: Using performance criteria for assessing and improving student performance.* Thousand Oaks, CA: Corwin.

This is a practical book that defines types of rubrics, provides help for developing new rubrics, and offers ways of using them to enhance learning. The resources section contains many sample rubrics to adopt or adapt.

Siegel, M. A., D. Markey, and S. Swann. 2005. Life science assessments for English learners: A teacher's resource. Unpublished manuscript. Berkeley, CA: University of California. Forthcoming at *www.lmri.ucsb.edu*.

This resource guide provides guidelines for assessments that are equitable and effective for English learners, assessments related to the National Science Education Standards for middle school life science, rubrics, teacher tips, and student work samples.

AUTHOR AFFILIATIONS

Marcelle A. Siegel studies and supports teacher and scientist partnerships at the Science and Health Education Partnership program at the University of California-San Francisco. She is principal investigator of "Investigating and Improving Science Learning and Assessment for Middle School Linguistic Minority Students," funded by the University of California Linguistic Minority Research Institute. She was a developer at the Science Education for Public Understanding Program (SEPUP) at the Lawrence Hall of Science, University of California at Berkeley, from 2001 to 2004 and a fellow at the National Institute for Science Education at the University of Wisconsin-Madison from 1999 to 2001.

Paul Hynds is a freshman biology teacher at Gilroy High School, in Gilroy, California, and was an instructional materials developer with SEPUP during 2000–2001. He has 15 years secondary teaching experience in science, 3 at the high school level, and 12 at the middle school level.

Marc Siciliano is the science director at the Young Women's Leadership Charter School in Chicago, Illinois. He directed the Math, Science, and Technology Academy (MSTA) for three years and taught science for five years at Lake View High School in Chicago. Currently a developer of instructional materials at Northwestern, he was a teacher adviser for SEPUP.

Barbara Nagle is director of SEPUP, and project coordinator for the development of SEPUP's two middle school courses, Issues, Evidence and You (IEY) and Science and Life Issues (SALI). She has a PhD in cell biology and taught high school chemistry in an inner-city Oakland school for six years before joining SEPUP as an instructional materials developer.

index

Page numbers printed in **bold** type reference information contained in figures, illustrations, or tables.

C

D

S

T