

The SEPUP Assessment System: An Overview

Lily Roberts, Mark Wilson and Karen Draney

SEPUP Assessment Project

University of California, Berkeley

June 1997

The SEPUP Assessment System: An Overview

The Science Education for Public Understanding Program (SEPUP) at the Lawrence Hall of Science at the University of California, Berkeley has developed a year-long, middle school science curriculum: *Issues, Evidence and You* (IEY; SEPUP, 1995a). IEY is unique in many ways (which we describe below), but particularly because its *embedded assessment system* was designed in tandem with the curriculum. Assessment developers from the Graduate School of Education at U.C. Berkeley worked cooperatively with the SEPUP curriculum developers and IEY field test teachers to design a comprehensive, integrated system for assessing, interpreting and monitoring student performance. This overview describes the underlying principles as well as the components of the SEPUP Assessment System, the benefits to teachers and students based on field test evaluation results, and next steps being taken to assist schools in implementing this classroom-based assessment system. See Figure 1 for an overview of the important characteristics of the SEPUP Assessment System.

Figure 1. Unique Qualities of the SEPUP Assessment System

- The Assessment System, and the whole curriculum, is structured by five *variables*.
- Assessments are embedded throughout the course.
- Student progress on SEPUP Variables can be tracked throughout the entire year.
- The assessment system measures student learning in a way that teachers and students can understand.
- The system can provide developmentally meaningful feedback to teachers, students, parents and administrators.
- The system contains scoring guides that have been field tested nationally.
- The system provides the tools to put teachers in charge of assessment.

IEY is a four-part course that facilitates student learning of science by providing hands-on, issue-oriented activities in the following topic areas:

- Water Usage and Safety;
- Materials Science;
- Energy; and
- Environmental Impact.

The IEY activities are designed to help students develop an approach to decision making that begins with a careful investigation of related concepts and evidence -- a key element is weighing the evidence and making tradeoffs with respect to various issues that affect our society. The SEPUP approach to decision making is congruent with major reform efforts in science education, including *Benchmarks for Science Literacy* (AAAS, 1993) and the *National Science Education Standards* (National Research Council, 1996). The goals of SEPUP are summarized in Figure 2 (SEPUP, 1995b).

Figure 2. Goals of SEPUP

To provide educational experiences focusing on science and technology and their interaction with people and the environment;

To promote the use of scientific principles, processes, and evidence in public decision making;

To contribute to improving the quality of science education in America; and

To enhance the role of science teachers as educational leaders in the schools and in the community.

The SEPUP Assessment System focuses on the progressive development of the student throughout the year-long curriculum. Teachers are provided “tools” to effectively implement the assessment system, including assessment tasks, scoring guides and exemplars of student work. The assessment tasks come in a variety of formats (e.g., lab reports; town meetings) and are embedded in the curriculum. Teachers are encouraged to work in groups to determine their local standards of performance through consensus moderation. The assessment system contains mechanisms for quality control and feedback to students, teachers, parents, administrators and other school officials or interested individuals.

SOCIETAL DECISION MAKING: A CONCEPTUAL FRAMEWORK FOR IEY

Issues, Evidence and You approaches middle school science education from an issue-oriented, hands-on learning perspective. Students are not expected to memorize a body of facts about water usage, materials science or energy, but rather are expected to build their scientific literacy and to become informed citizens who can make decisions by weighing the evidence and making tradeoffs. For middle school students to be able to address a problem or issue (i.e., to engage in evidence-based, societal decision making), they need a set of tools. In the SEPUP Assessment System, these tools are identified as SEPUP Variables which help teachers “map” student progress over the year on the central instructional themes of the course. The SEPUP Variables are the areas in which students are assessed.

The variables that form the heart of IEY are: *Designing and Conducting Investigations (DCI)*; *Evidence and Tradeoffs (ET)*; *Understanding Concepts (UC)*; *Communicating Scientific Information (CM)*; and *Group Interaction (GI)*. These are further defined in the next section on components of the system (also see Figure 5). Being able to make judgments about evidence requires an understanding of scientific concepts as well as the ability to design and conduct investigations, to work cooperatively and to communicate effectively. The knowledge, skills and abilities that undergird the IEY curriculum are displayed in Figure 3. This figure presents a schematic overview of the conceptual framework for IEY, showing how the SEPUP Variables are related in the societal decision making process.

These variables are the basic building blocks on which the structure of the SEPUP Assessment System is formed. The variables represent aspects of the scientific process as well as scientific concepts and skills.

“Taken as a whole, the SEPUP Variables define the process of societal decision making, which is central to and, in many ways, distinguishes *Issues, Evidence and You*. The process of societal decision making can be summarized as follows: When confronted with an important societal issue, students learn (a) to draw upon their understanding of the relevant scientific concepts and upon the results of sound scientific experiments to evaluate the available scientific evidence and (b) to consider this scientific evidence as well as other important issues in weighing the advantages and disadvantages of various possible solutions in making a decision about courses of action. Students also learn to work collaboratively in collecting and evaluating data and in generating possible solutions to a problem, and to communicate the results of their decision making clearly and accurately” (SEPUP, 1995a).

The implementation of this classroom-based, assessment system relies on the expertise of teachers as the “front-line” professionals who:

- (a) assess student performance on the central concepts and skills of the curriculum;
- (b) set standards of student performance;
- (c) track student progress over the year for both instructional planning and grading purposes; and
- (d) provide feedback on student progress and classroom instruction (to themselves as self-assessors as well as to students, parents, and administrators).

Next, we turn to more detail on the various components of the system.

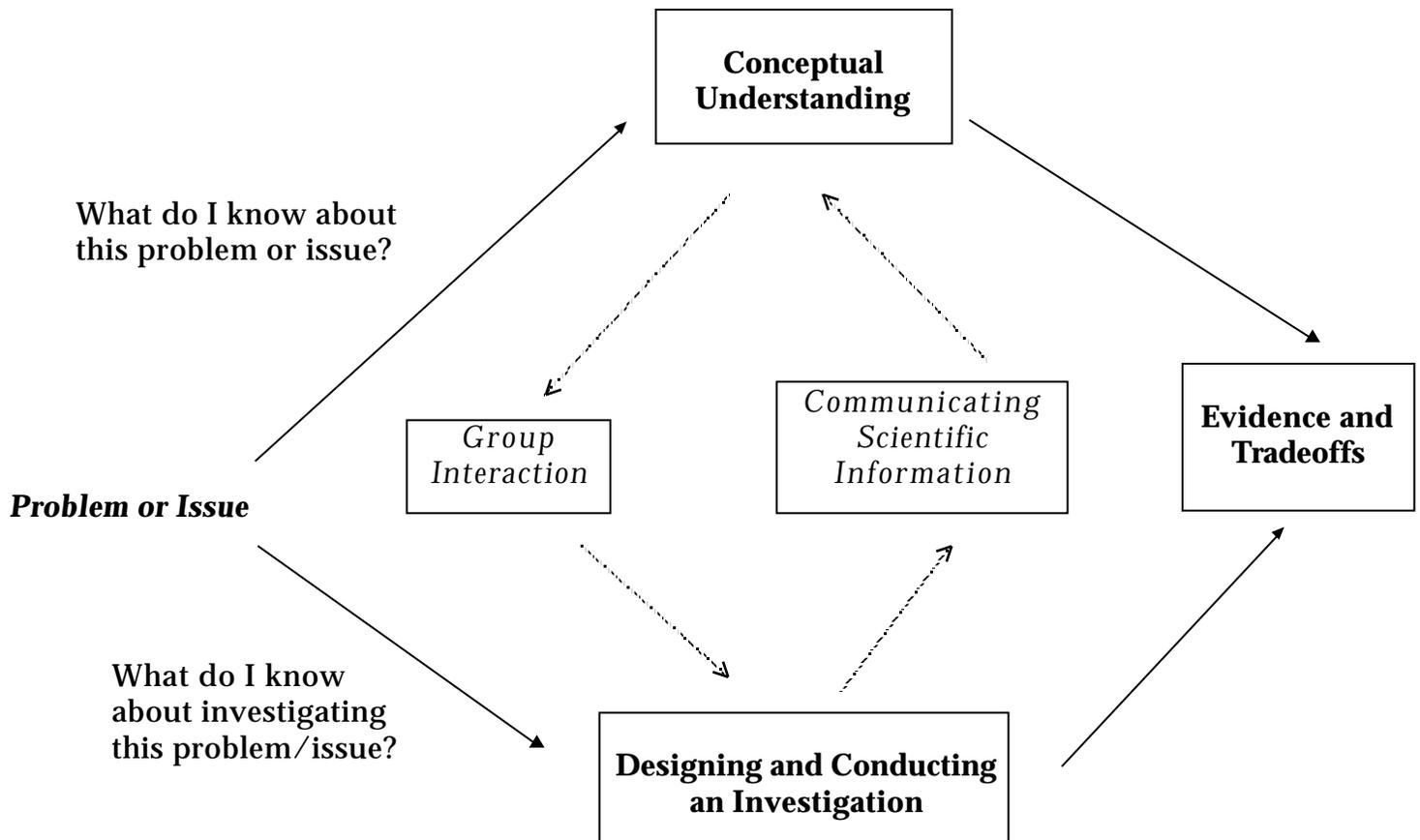


Figure 3. Evidence-based, Societal Decision Making

COMPONENTS OF THE SEPUP ASSESSMENT SYSTEM

There are eight components in the SEPUP Assessment System. Figure 4 provides a summary of all components. The discussion that follows describes the components and provides examples.

Figure 4. SEPUP Assessment Components at a Glance

1. SEPUP Variables - Five variables which represent student learning in terms of the core concepts of IEY.
2. Assessment Tasks - Activities that are an integral part of regular instruction.
3. Scoring Guides - Rubrics that establish baseline criteria for assessing levels of student performance.
4. Assessment Blueprints - Sequential list of IEY activities and opportune points in instruction for assessing student learning.
5. Exemplars - Examples of student work that have been scored and moderated for each variable and score level.
6. Assessment Moderation - A process by which teachers discuss and reach consensus on local standards for scoring student work.
7. Performance Maps - Graphical representations of student development on the SEPUP Variables.
8. Link Tests - Additional assessment activities for teachers' use at major course transitions that are also based on the SEPUP Variables.

The SEPUP Assessment System is guided by three central components. First, the SEPUP Assessment System defines the instructional content using the *SEPUP Variables* and embedded *Assessment Tasks*. Further, the SEPUP Assessment System aids teachers in setting standards for assessing student performance through variable-specific *Scoring Guides*. These three components are integral parts of the curriculum.

Three Central Components

1 - SEPUP Variables. These five variables (see Figure 5 for an overview and description) represent student learning in terms of the core concepts of IEY that are based on the model of evidence-based societal decision making. The five SEPUP Variables form the conceptual framework for IEY instruction and assessment and provide a ***developmental perspective*** from which to view student progress. More information on SEPUP Variables is available in Section II of the *SEPUP Assessment Resources Handbook*.

Figure 5. The SEPUP 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 a solution to a problem, effectively, and 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.

2 - Assessment Tasks. Assessment activities are an integral part of instruction and were developed in the same way as the other instructional activities. These tasks serve the same purpose as regular instruction with the addition of assessment characteristics. However, these activities do not diminish instructional time as do end-of-chapter tests. Two types of tasks are embodied in IEY: formal *Assessments* and *Quick Checks* (see Figure 6). These tasks represent a variety of assessment formats (e.g., graphical representations; lab reports; oral presentations; short essays).

Figure 6. Types of Assessment Tasks Defined

Assessment 	Quick Check 
These activities are distributed throughout IEY at critical junctures where formal assessments of individual student work are appropriate. Assessments are directly related to the SEPUP Variables and can be scored to assess each student's progress.	These activities provide teachers with opportunities to check for student understanding of important concepts, skills, or procedures that are being introduced or practiced in the activity. A sample of these might be scored to see how the class is doing as a whole.

3 - Scoring Guides. The scoring guides, one specific to each SEPUP Variable, establish baseline criteria for assessing levels of student performance and interpreting student work. Figure 7 describes the general scoring levels that are embodied in the SEPUP Scoring Guides, ranging from off-task remarks (scored as 0) to an advanced response (scored as 4) that goes beyond "correct and complete."

Figure 7: General Scoring Guide for the SEPUP Variables

Score level	Description
4 "Advanced"	Goes beyond expected answer (Level 3) in some significant way.
3 "Correct "	Correct and complete answer.
2 "Incomplete"	Shows understanding but something important is missing.
1 "Incorrect"	Includes some appropriate terms or concepts but response is incorrect.
0 "Off task"	No response, or response is irrelevant.
X	Student had no opportunity to respond.

Figure 8: Scoring Guide for the Evidence and Tradeoffs Variable

Score	<i>Using Evidence:</i> Response uses objective reason(s) based on relevant evidence to argue for or against an option.	<i>Using Evidence to Make Tradeoffs:</i> Response recognizes multiple perspectives of issue and explains each perspective using objective reasons, supported by evidence, in order to make choice.
4	Accomplishes Level 3 AND goes beyond in some significant way, such as questioning or justifying the source, validity, and/or quantity of evidence.	Accomplishes Level 3 AND goes beyond in some significant way, such as suggesting additional evidence beyond the activity that would further influence choices in specific ways, OR questioning the source, validity, and/or quantity of evidence & explaining how it influences choice.
3	Provides major objective reasons AND supports each with relevant & accurate evidence.	Makes a tradeoff using <u>at least two</u> perspectives or options AND provides objective reasons, supported by relevant & accurate evidence, for each option.
2	Provides <u>some</u> 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 perspectives or options AND provides some objective reasons using some relevant evidence BUT reasons are incomplete and/or part of the evidence is missing; OR only one complete & accurate option has been provided.
1	Provides only subjective reasons (opinions) for choice and/or uses inaccurate or irrelevant evidence from the activity.	States at least one perspective/option BUT only provides subjective reasons (opinions) and/or uses inaccurate or irrelevant evidence.
0	Response missing, illegible or lacks reasons AND offers no evidence to support decision made	Response missing, illegible or lacks reasons AND offers no evidence to support decision made.
X	Student had no opportunity to respond.	

The SEPUP Variables have subparts called *elements* in the SEPUP Assessment System vocabulary. For example (see Figure 8), the Evidence and Tradeoffs Variable has two elements: *Using Evidence* and *Using Evidence to Make Tradeoffs*. Student work may be scored on one or more elements of a variable depending on the assessment task, and in some cases, on more than one variable. The Evidence and Tradeoffs Variable will be used for illustrative purposes throughout this overview because it is a unique contribution of IEY. Note that the number of elements varies from two to four for the different variables (see Figure 9 for the elements of other SEPUP Variables).

The Scoring Guides provide more variable-specific descriptions of each scoring level for every element of each SEPUP Variable. For example, a correct response (i.e., a “3”) on the *Using Evidence* element of the Evidence and Tradeoffs Variable is described as an answer that *provides major objective reasons AND supports each with relevant and accurate evidence*. Scoring Guides for all SEPUP Variables are included as resources in Section II of the *SEPUP Assessment Resources Handbook*.

Five Support Components

The next five components of the SEPUP Assessment System assist teachers in their implementation of the IEY curriculum as well as their use of the three central components.

4 - Assessment Blueprints. Assessment Tasks are distributed throughout the IEY curriculum at opportune points for checking and monitoring student performance on the SEPUP Variables. These points are indicated on a table called an Assessment Blueprint. The Assessment Blueprints (one for each of the four parts of the course) list all IEY activities sequentially and identify the points in instruction where a formal *Assessment* or *Quick Check* could be applied to indicate student progress on one or more of the SEPUP Variables (see Figure 9; note that the Group Interaction Variable is not included here, because it is not yet fully implemented). Assessment Blueprints are included as a resource in Section II of the *SEPUP Assessment Resources Handbook*.

5 - Exemplars. Exemplars are samples of student work that have been scored and moderated (see Section IV of the *SEPUP Assessment Resources Handbook* to learn more about moderating student work) by SEPUP field test teachers who used the IEY curriculum and SEPUP Assessment System. Exemplars offer ideas as to what a teacher might expect from students at varying levels of development throughout the course. Examples of student work at each scoring level for each variable and its elements are included in the *SEPUP Assessment Resources Handbook* in Section VII. Figure 10 presents an exemplar of actual student work (spelling errors, etc., are from the original student paper) for a complete and correct (level 3) response to Activity 12.

Figure 9: Partial Assessment Blueprint

Variables and Elements					
Activity	Designing and Conducting Investigations (DCI)	Evidence and Tradeoffs (ET)	Understanding Concepts (UC)	Communicating Scientific Information (CM)	Group Interaction (GI)
	<ul style="list-style-type: none"> * Designing Investigation * Selecting & Recording Procedures * Organizing Data * Analyzing and Interpreting Data 	<ul style="list-style-type: none"> * Using Evidence * Using Evidence to Make Tradeoffs 	<ul style="list-style-type: none"> * Recognizing Relevant Content * Applying Relevant Content 	<ul style="list-style-type: none"> * Organization * Technical Aspects 	<ul style="list-style-type: none"> * Time Management * Role Performance/ Participation * Shared Opportunity
1 - Water Quality					
2 - Exploring Sensory Thresholds			: Both Elements <i>(Measurement and Scale)</i>		
3 - Concentration			: Applying Relevant Content		
4 - Mapping Death					: Time Management; Shared Opportunity
5 - John Snow and Search for Evidence		A: Using Evidence		A: Both Elements	
6 - Contaminated Water	: Designing Investigations				
7 - Chlorination	A: All Elements				

In Activity 12, students receive the following assignment.

Assignment

You are a public health official who works in the Water Department. Your supervisor has asked you to respond to the public's concern about water chlorination at the next City Council meeting. Prepare a written response explaining the issues raised in the newspaper articles. Be sure to discuss the advantages and disadvantages of chlorinating drinking water in your response, and then explain your recommendation about whether the water should be chlorinated.

Figure 10: Exemplar for a Level 3 (“Complete and Correct”) Response on Activity 12, The Peru Story (Element Scored: Using Evidence to Make Tradeoffs)

Level 3	Uses relevant and accurate evidence to weigh the advantages and disadvantages of multiple options, and makes a choice supported by the evidence.
<p>“As an educated employee of the Grizzelyville water company, I am well aware of the controversy surrounding the topic of the chlorination of our drinking water. I have read the two articles regarding the pro’s and cons of chlorinated water. I have made an informed decision based on the evidence presented the articles entitled “The Peru Story” and “700 Extra People May bet Cancer in the US.” It is my recommendation that our towns water be chlorin treated. The risks of infecting our citizens with a bacterial disease such as cholera would be inevitable if we drink nontreated water. Our town should learn from the country of Peru. The article “The Peru Story” reads thousands of innocent people die of cholera epidemic. In just months 3,500 people were killed and more infected with the disease. On the other hand if we do in fact chlorine treat our drinking water a risk is posed. An increase in bladder and rectal cancer is directly related to drinking chlorinated water. Specifically 700 more people in the US may get cancer. However, the cholera risk far outweighs the cancer risk for 2 very important reasons. Many more people will be effected by cholera where as the chance of one of our citizens getting cancer due to the water would be very minimal. Also cholera is a spreading disease where as cancer is not. If our town was infected with cholera we could pass it on to millions of others. And so, after careful consideration it is my opinion that the citizens of Grizzelyville drink chlorine treated water.”</p>	<p><u>Comment</u></p> <p>Both sides of the chlorinating issue have been presented and supported. The choice to chlorinate was made.</p>

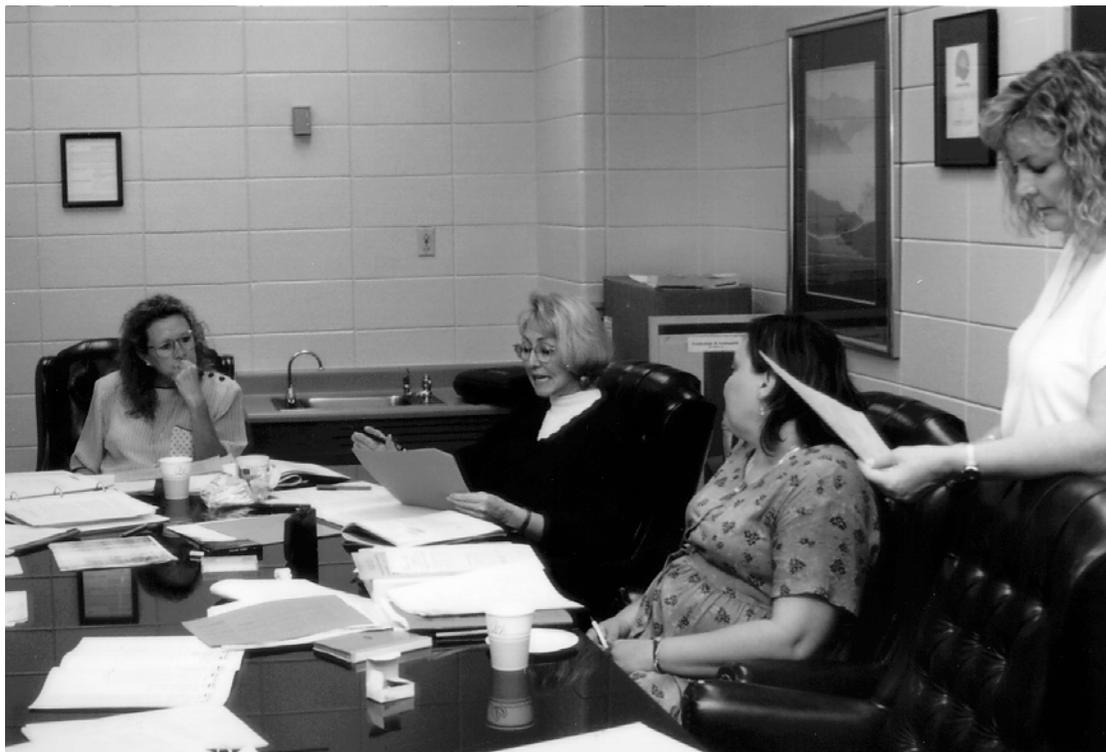
Exemplars are included as a resource for teachers as they initially begin using the SEPUP Assessment System, applying the Scoring Guides to assess their own students’ work, and engaging in assessment moderation. Section VII of the *SEPUP*

Assessment Resources Handbook includes all available exemplars by activity.

6 - Assessment Moderation. Assessment moderation is the process in which a local group of teachers come together to score and discuss student work. This process serves two purposes. First, it ensures technical quality by allowing for agreement among teachers as to the appropriate score level for a particular student response (i.e., local consensus building to set standards of student performance). Second, and more importantly, it provides for *teacher professional development* in assessment. One-shot teacher professional development has failed time and again to serve teachers' needs in our dynamic educational system; to be truly effective teacher enhancement must also be in-depth and long-term (Fullan, 1990; Loucks-Horsley, et al., 1989; St. John, 1991). Changing assessment practices is no less difficult to promote and support, but we believe local assessment moderation provides teachers with ongoing support to this end (Roberts, 1996, 1997; Roberts, Sloane, & Wilson, 1996).

In local assessment moderation sessions, teachers discuss the scoring, interpretation, and use of student work, and make decisions regarding standards of performance and methods for reliably judging student work relative to those standards (see Figure 11). Further, this process provides a forum in which teachers may discuss common mistakes or difficult concepts that can then be addressed in subsequent instruction. Collegiality is a hallmark of successful local assessment moderation (Roberts, 1996) and also fulfills one of the basic tenets of high quality professional development (Little, 1993; McLaughlin, 1991). More information is available in Section IV of the *SEPUP Assessment Resources Handbook*.

Figure 11. Picture of Teachers Engaged in Local Assessment Moderation



7 - Performance Maps. Performance or variable maps are graphical representations of student development on the SEPUP Variables. Each map shows how a variable unfolds or evolves over the year in terms of an individual student's performance on the assessment tasks. Maps of student performance on the tasks can be used by teachers for their own planning and to show students, administrators, and parents how students are developing on the SEPUP Variables over the year. For more information about using Performance Maps refer to Section V of the *SEPUP Assessment Resources Handbook*.

As a result of teachers managing and using the SEPUP Assessment System, maps can be produced which allow them to assess both individual and class progress. A teacher could plot the entire class to gain insights about student progress overall, thus informing her or his instructional planning. For instance, if the class has not performed well on a section of the course, then the teacher might feel the need to go back and review concepts or issues reflected in the assessment results. In this way, performance maps make scores more meaningful, while also providing a basis for evaluation by teachers and others.

Figure 12 shows a hypothetical plot of three students with varying developmental patterns on the ET Variable during the first part of the course. Criterion zones are defined in the far-right hand column in the figure and represented by the alternating horizontal bands of shading. The Criterion Zones provide a way to interpret student performance that is directly tied to the score levels on the Scoring Guides. Student (S1) has made good progress on this variable. A second student (S2) starts out doing quite well (at Level 3), but drops to incorrect (Level 1) by the end of Segment D of Part 1. This large drop raises a “red flag” for the teacher and the needed intervention takes place, leading to improved performance on Link Test 1. The final example is a student (S3) who starts out on the Pretest at Level 2, a mostly correct response, but then drops during the first three parts of the course to incorrect responses. The teacher intervenes in the case of this student to provide the appropriate support and S3 moves back up to the level where he started at the Pretest by the end of Part 1. By the end of Part 1, all three of these students are working at about the same level, so the teacher feels comfortable moving on to Part 2.

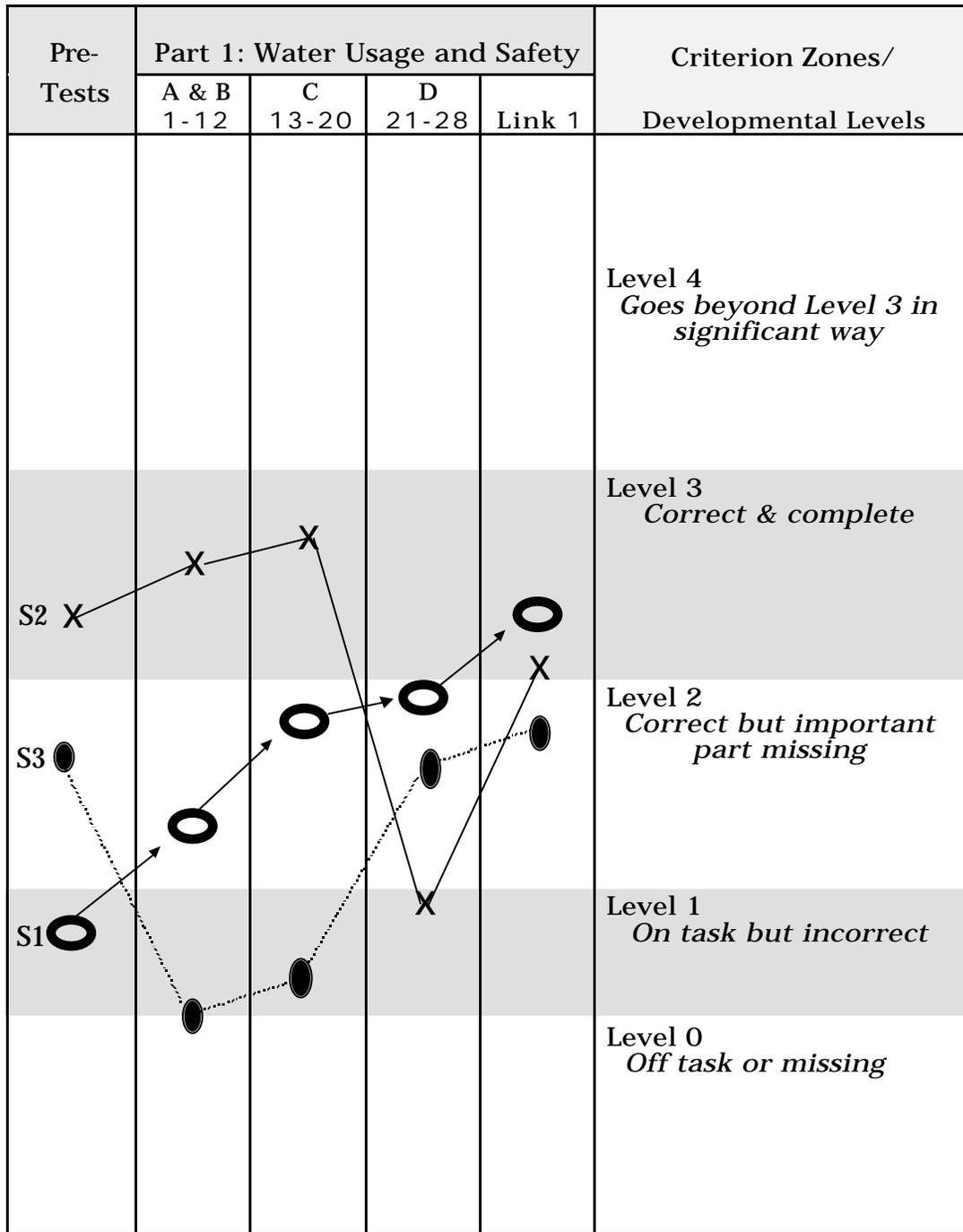
8 - Link Tests. Link tests are a series of tests that are intended to be given at major transition points in the IEY curriculum (e.g., at the end of Part 1, Water Usage and Safety, and so on). Each link test contains open-ended items related to the content of the course which further assess each student's ability on the SEPUP Variables. Since no single assessment can optimize both reliability and instructional validity, link tests serve to scaffold the embedded IEY assessment tasks. Student scores on the link tests also provide additional information for tracking student progress on the variable maps. Link tests provide teachers with an "item bank" from which to draw other assessment tasks when designing their own tests or quizzes. More information is available in Section VI of the *SEPUP Assessment Resources Handbook*.

Summary

In summary, these eight components (refer to Figure 4) comprise the SEPUP Assessment System. The first three define the instructional content and set standards for assessing student performance. The latter five are supportive components that serve the needs of teachers engaged in implementing the SEPUP Assessment System.

In the next section, the benefits of using the SEPUP Assessment System are identified in terms of: (1) teacher change in assessment, collegial and instructional practices; and (2) student change in attitudes toward science and gains in achievement. These benefits were identified during the evaluation of the field test of IEY in 1994-95.

Figure 12. Performance Map for a Hypothetical Class' Growth on the ET Variable from Pre-Test to the End of Part 1, Water Usage and Safety



USING THE SEPUP ASSESSMENT SYSTEM: BENEFITS TO STUDENTS AND TEACHERS

The SEPUP Assessment System was field-tested during the 1994-95 school year at six centers throughout the United States which we called "Assessment Development Centers" (ADCs). In addition to the ADCs, there were seven centers referred to as Professional Development Centers (PDCs), which were required to use the IEY curriculum, but not the assessment components. Each Center selected at least one comparison teacher and class for evaluation purposes. The comparison teachers were chosen to be of equivalent quality as science teachers as were the ADC and PDC teachers. While ADC and PDC teachers were to follow specific SEPUP and IEY guidelines, comparison group teachers were to use their regular science curriculum and assessment program. The evaluation of the field test used a quasi-experimental pre/post, comparison group design. The methods of data collection for teachers included: pre-post surveys administered by mail and personal interviews and focus groups conducted during site visits or inservice activities. Student data was collected using pre-post surveys as well as pre-post test, link tests, and the actual assessment tasks in IEY. The results of the evaluation indicate significant benefits for teachers and students in the Assessment Development Centers.

Benefits to Students

The IEY curriculum affects both students' perceptions of the relevance of science (Roberts & Henke, in press) as well as their science achievement on the SEPUP Variables (Wilson, Sloane, Roberts, & Henke, 1995; Wilson & Draney, 1996, 1997).

Student Achievement. Students in the ADCs received the greatest benefits from the course as evidenced by the significant differences between ADC, PDC and comparison students' gains in achievement on the combined SEPUP Variables (see Figure 13). Note that there was steady growth through April, then a slump in ADC scores by the end of the year. However, the end of year average for the ADC students is still significantly larger than that of either the PDC or comparison students.

The PDC students' achievement was comparable to that of the comparison students who were in traditional science courses. This result indicates that the IEY curriculum alone is not sufficient to bring about changes in student achievement on the SEPUP Variables. Remember that the difference between ADC and PDC teachers was the requirement of the ADC teachers to use the components of the SEPUP Assessment System whereas the PDC teachers were only required to teach the IEY course. Thus, the effect we observed seems to be in the added value that the assessment system brings to the curriculum.

Figure 13
Gains for the ADC, PDC, and comparison groups on the combined SEPUP Variable during the 1994-95 school year

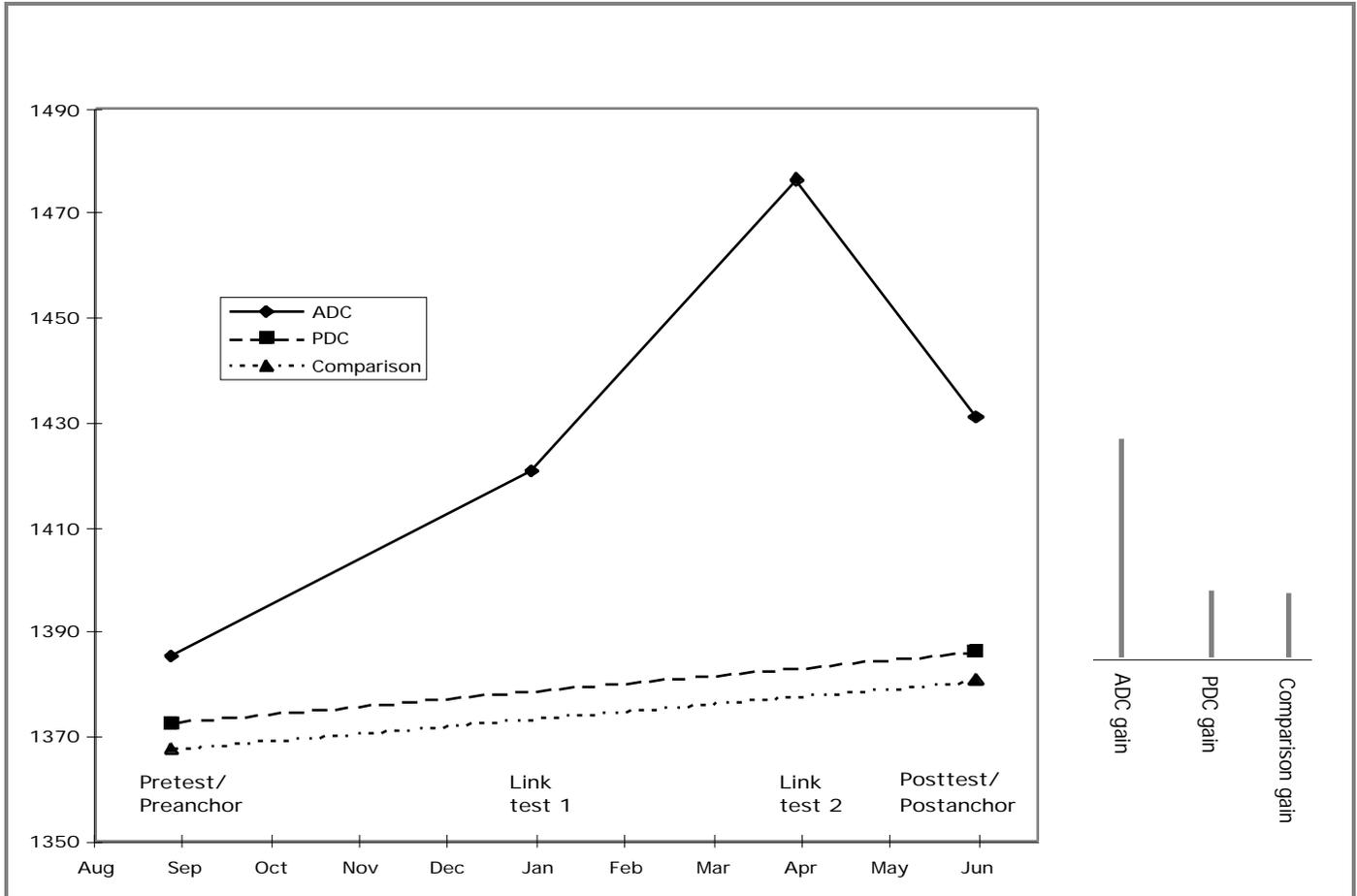


Figure 13

The drop at the end of the year among the ADC students is interesting also. There are several possible explanations for this drop in student proficiency at the time of the posttest (Wilson & Draney, 1997). The first is that the posttests were most often administered during the last week or two of the school year, which tends to be a busy time for both teachers and students. Consequently, it is possible that the administration of some of the posttests was somewhat rushed. Also, as can be seen in Figure 13, much more time was spent on the first sections of the curriculum than on the last section. In particular, most ADC classes spent nearly half of the school year on Part 1 (Water) of the curriculum, and many spent only a month or so on Part 3 (Energy). Thus, one would expect that students would have felt much more secure with material from Part 1 than with material from Part 3. Since questions from Part 3 are included in both Link 3 and the Post-anchor test (refer to Figure 13), this may have adversely affected the proficiency estimates of the ADC students at the time of the posttest. These results illustrate how assessment information can be used to provide important formative evaluation information (Wilson & Draney, 1997).

Change in Students' Attitudes. The IEY curriculum also had a positive effect on students' perceptions of the relevance of science. SEPUP IEY students during the pilot year were significantly more positive about the benefits of science to their daily lives than comparison students (Roberts & Henke, in press). The analysis of the field test year comparisons between ADC, PDC and comparison students is in progress.

Below are selected comments from students involved in the SEPUP pilot and field tests. The students describe the Assessment System as they experienced it and their perceptions of the range of benefits. The quotes are reprinted directly from student comments.

SEPUP is different from traditional science class; look at "both sides" of an issue. "The form of assessment that we've taken the past 2 yrs [is] easier. I think that the 0 - 4 grading system is O.K. I guess I like it better than any normal science class. I feel that this way of testing is better because it teaches kids how to explain themselves. Also now we know how to accept both sides of a situation and not just pick the one we like. We consider both sides."

Better retention of information related to "everyday" life. "I think that the [assessment] system is much better. I learn more because when you finish a topic you have a [assessment] or project. I learn more if we take it in steps, not just plop down and answer multiple choice questions for an hour and a half. 'Cause soon as we leave so does everything we learn. So I really think it is better 'cause all of the problems are put to everyday things and that makes it easier."

Scoring and grading system is clear and shows students their progress over time. “These assessments were different because at the end of each unit, we had to write up what we had learned. I think this was better because it helped me understand the unit more. It also helped me see both sides of an issue. I think the grading system was fair also. A four signifies that we really know what was going on, all the way down to a one which means we didn't understand it at all. I thought that the portfolios were a good idea, because it showed how I progressed throughout the year.”

Scaffolds student learning and shows ability better than a traditional test. “I understand more when we aren't just reading and taking notes. I think most kids will learn more with the experiments and explanations over reading and taking notes. The portfolio and papers instead of a final exam, for me, was much easier ‘cause I do good work, but I choke on test and exams, so I felt this way showed more about my ability than just an exam.”

Promotes thinking and expression of own ideas. “I think [SEPUP] assessments are better, because it allows you to think for yourself. You express your own ideas about something. It allows you to come up with an answer for yourself instead of having so many possible answers in a multiple choice test.”

Benefits to Teachers

Teacher change was assessed in terms of assessment, instructional and collegial practices. The evaluation of SEPUP's impact on teacher change examined the differences between ADC and PDC teachers as well as between IEY and non-IEY (i.e., comparison) teachers (Roberts, 1996). The comparison teachers did not change on any of the measures used. However, there were interesting differences between the ADC and PDC teachers.

Instructional Practices. Importantly, there were no significant differences between ADC and PDC teachers on instructional activity use -- both groups had identical gains as was expected, given they used the same course and materials. The gains reflect the increased use of student-directed, teacher-facilitated methods and decreased use of traditional methods, such as lecturing to students.

Collegiality. Both ADC and PDC teachers gained on the measures of collegiality, but the gain was greater for ADC teachers. Collegiality was more evident in ADCs where teachers were all from the same district, where collaboration occurred between moderation meetings, and where there was strong leadership for the group. *Strong collegiality enhanced an ADC's success with local assessment moderation.*

Assessment Practices. The most interesting findings were for the measures related to assessment practice: use of alternative assessment strategies and perception of the usefulness of alternative assessment strategies. In the beginning, both ADC and

PDC teachers perceived alternative assessment strategies to be useful. However, when faced with the reality of using alternative assessments to assess learning, guide subsequent instruction and grade students, ADC teachers turned to the practical issues and concerns that they faced in their classrooms daily:

How do I score 150 student papers every week that can be several pages long? or even 30 papers?

How do I convert these scores to grades that fit my school's grading system without re-scoring all the papers?

Why is my score different than the other teachers' scores on this student's paper when we all used the same scoring guide?

These questions paraphrase concerns that were voiced by ADC teachers during the field test of the SEPUP Assessment System. The moderation meetings were the main support mechanism for teachers as they worked collaboratively to solve these daily dilemmas.

ADC and PDC teachers were significantly different on measures of assessment strategy use and the perceived usefulness of alternative assessment strategies for assessing learning. ADC teachers increased their use of open-ended questions and significantly decreased their use of closed-ended questions. Meanwhile, PDC teachers increased their use of closed-ended questions from the beginning to the end of the school year (i.e., they were using traditional testing such as multiple choice) and reduced their use of alternative assessment strategies by the end of the field test. This evaluation result indicates that the PDC teachers did not grapple with the same practical issues about assessment as their ADC counterparts. Furthermore, ADC teachers re-assessed the usefulness of alternative assessment strategies through practice and reflection and collegial interaction -- which resulted in small, non-significant decreases rather than anticipated increases on their perceptions of the usefulness of such strategies for assessing learning, guiding instruction and grading. However, the ADC teachers remained committed to alternative assessment strategies (Roberts, 1996, 1997). The PDC teachers retained their rosy perceptions of the usefulness of alternative assessment strategies, but never really used them. *These results clearly underscore the **rhetoric** versus the **reality** in these middle school science teachers' perceptions of alternative assessment.*

Local Assessment Moderation. For ADC teachers to manage the SEPUP Assessment System, ongoing support and professional development was necessary for successful implementation. SEPUP IEY field test teachers told us that “teachers need to internalize the variables”. In other words, teachers need to develop an understanding of what it means, for example, to have their students “using evidence to make tradeoffs.” Simply giving teachers IEY *scoring guides* (see Figure 8, Scoring Guide for the Evidence and Tradeoffs Variable) did not sufficiently

promote an understanding of this variable (Roberts, 1996). Teachers learned about variables by engaging in the local assessment moderation in which they were able to set common standards for judging student performance.

The consensus of the ADC participants was that ongoing support through local assessment moderation was critical to teachers' use of the IEY curriculum as well as the SEPUP Assessment System. As teachers used assessment tasks and the scoring guides on their own, they encountered questions or issues of implementation that they shared during moderation meetings with their colleagues. These meetings provided the ADC teachers the ongoing support necessary to change their instructional and assessment practices.

Below are a few selected comments from ADC teachers related to some of the key roles local assessment moderation played in providing collegial support for learning about the SEPUP Assessment System (Roberts, Sloane, & Wilson, 1996).

Building an understanding of embedded assessment and using it effectively. One teacher noted that although she receives strong collegial support at her school, she feels that moderation with her ADC colleagues was important in helping her understand the Scoring Guides and how to apply them to specific activities. She said that she feels that the moderation process helped her to use the Scoring Guides effectively, with practice making it easier to score each time.

Feedback on student learning. One teacher reported that she has been able to realize through moderation that she “hasn’t gotten through” to her students about *designing investigations* and that she expects too much of her students sometimes. She indicated that the moderation process allows “time to reflect on what you did right and wrong” and that you have “the opportunity to listen to others and get ideas to use next time.”

Promoting reflective practice. One teacher noted that she feels that the moderation process has contributed to her professional growth because it makes her examine her teaching and students' learning. She adds, “you reflect on your purpose as a teacher: to see students grow and change, and have students see that they’re learning something that’s applicable.” She reported, for example, that she sees that her students are referring back to prior knowledge and applying it to a new task.

Figure 14 presents the ideal benefits of the SEPUP Assessment System once it is fully operational. The final section summarizes the next steps in investigating the implementation of the full SEPUP Assessment System.

Figure 14. Ideal Benefits of the SEPUP Assessment System

Benefits to Students	Benefits to Teachers
<ul style="list-style-type: none"> • Informs students about their own progress. • Enables students to self-assess and adjust learning. • Promotes student learning about scientific process and concepts as well as development of relevant skills. 	<ul style="list-style-type: none"> • Informs teachers about student progress. • Enables teachers to evaluate student progress and to self-evaluate the success of their instruction, and hence, modify instruction as needed. • Promotes teacher change in science instruction and assessment aligned with current reform efforts. • Provides opportunities for professional development through collaboration with colleagues.

Next Steps: Investigating the Implementation of the SEPUP Assessment System

The SEPUP Assessment System is now in its second phase of development with new funding from the National Science Foundation. The first phase ended with the field test evaluation described above and the initial development of variable maps. The second phase will produce materials for training and supporting teachers' use of IEY and the SEPUP Assessment System. Further, technology will be adapted to reduce the workload for teachers in the use and scoring of the alternative assessment activities. After the new support materials have been developed and pilot tested, a final field test of the full SEPUP Assessment System will be evaluated and the results will be disseminated.

Over the next few years and subject to sufficient funding, the following activities will be undertaken:

- refining performance maps and preparing materials to support teachers' use of the maps for instructional improvement as well as tracking student progress.
- developing software so that teachers can easily collect data on student performance, score student work in accordance with the levels of the scoring guides, and produce maps of their own students' progress on the different SEPUP Variables.

- developing training materials and building a cadre of IEY teacher leaders who can provide training and support to others.
- selecting districts that are using IEY and interested in participating in a field test evaluation of the full SEPUP Assessment System.
- evaluating the implementation of the full SEPUP Assessment System once the support materials, revised maps and expanded link tests are completed.
- making final improvements to the SEPUP Assessment System based on the evaluation results.

Reference List

- American Association for the Advancement of Science (AAAS), Project 2061. (1993). *Benchmarks for science literacy*. New York: Oxford University Press.
- Fullan, M.G. (1990). Change processes in secondary schools: Toward a more fundamental agenda. In M.W. McLaughlin, J.E. Talbert, & N. Bascia (Eds.), *The contexts of teaching in secondary schools: Teachers' realities* (pp. 224-255). New York: Teachers College Press.
- Little, J.W. (1993). Teacher professional development in a climate of educational reform. *Educational Evaluation and Policy Analyses*, 15 (2), pp. 129-151.
- Loucks-Horsley, S., Carlson, M.O., Brink, L.H., Horwitz, P., Marsh, D.D., Pratt, H., Roy, K.R., & Worth, K. (1989). *Developing and supporting teachers for elementary school science education*. Washington, D.C.: The National Center for Improving Science Education.
- McLaughlin, M.W. (1991). Enabling professional development: What have we learned? In A. Lieberman & L. Miller (Eds.), *Staff development for education in the '90s: New demands, new realities, new perspectives* (2nd ed.; pp. 61-82). New York: Teachers College Press.
- National Research Council. (1996). *National Science Education Standards*. Washington, D.C.: National Academy Press.
- Roberts, L.L.C. (1996). *Methods of evaluation for a complex treatment and its effects on teacher professional development: A case study of the Science Education for Public Understanding Program*. Unpublished dissertation, University of California at Berkeley.
- Roberts, L. (1997, March). *Using maps to produce meaningful evaluation measures: Evaluating middle school science teacher change in assessment, collegial and instructional practices*. Paper presented at the ninth International Objective Measurement Workshop, March 21, Chicago, IL.
- Roberts, L., & Henke, R. (in press). Mapping middle school students' perceptions of the relevance of science. In Wilson, M., Engelhard, G., & Draney, K. (Eds.), *Objective Measurement: Theory into Practice* (Vol. 4). Norwood, NJ: Ablex.
- Roberts, L., Sloane, K., & Wilson, M. (1996). *Local assessment moderation in SEPUP*. Paper presented at the Annual Meeting of the American Educational Research Association. New York, April.
- SEPUP. (1995a). *Issues, Evidence and You: Teacher's Guide*. Berkeley: Lawrence Hall of Science.

SEPUP. (1995b). *SEPUP Annual Report, 1994-95.* Berkeley: Lawrence Hall of Science.

St. John, M. (1991). *Science education for the 1990's: Strategies for change.* Inverness, CA: Inverness Research Associates.

Wilson, M., & Draney, K. (1997). *Developing maps for student progress in the SEPUP assessment system.* Paper presented at the annual meeting of the American Association for the Advancement of Science, Seattle, WA.

Wilson, M., & Draney, K. (1996). *Mapping student progress with embedded assessments.* Paper presented at the Annual Meeting of the American Educational Research Association. New York, April.

Wilson, M., Sloane, K., Roberts, L., & Henke, R. (1995). *SEPUP Course 1, Issues, Evidence and You: Achievement evidence from the pilot implementation.* Berkeley, CA: Graduate School of Education, University of California.