DOCUMENT RESUME

ED 431 796 TM 029 882

AUTHOR	Ruiz-Primo, Maria Araceli; Li, Min; Ayala, Carlos; Shavelson, Richard J.
TITLE	Student Science Journals and the Evidence They Provide: Classroom Learning and Opportunity To Learn.
SPONS AGENCY	National Science Foundation, Arlington, VA.
PUB DATE	1999-03-30
NOTE	29p.; Paper presented at the Annual Meeting of the National Association for Research in Science Teaching (72nd, Boston, MA, March 28-31, 1999).
CONTRACT	SPA-8751511; TEP-9055443
PUB TYPE	Reports - Research (143) -- Speeches/Meeting Papers (150)
EDRS PRICE	MF01/PC02 Plus Postage.
DESCRIPTORS	*Elementary School Students; Evaluation Methods; Grade 5; Intermediate Grades; Performance Factors; *Science Instruction; *Scoring; Student Evaluation; *Student Journals
IDENTIFIERS	*Opportunity to Learn

ABSTRACT

Student science journals were evaluated as an assessment tool to demonstrate student performance throughout the course and the opportunities students have to learn science in their classrooms. The study was conducted with 163 fifth graders from 7 classrooms, although 1 teacher did not collect student journals, reducing the sample size. Close and proximal assessments were administered before and after instruction in each of two units. Student journals within a class were selected based on performance on the posttest as high, medium, or low. Eighteen journals were scored for one unit, and 14 for the other. Each journal was scored by two scorers. Preliminary results indicate that student journals can be scored reliably. Unit implementation and student performance scores were highly consistent across scorers and units. Teacher feedback scores were less reliable, but show potential for use. Inferences about unit implementation using the journals were justified, and inferences about student performance were also encouraging. These results reveal the potential usefulness of assessment through student science journals. (Contains 8 tables, 2 figures, and 16 references.) (SLD)

STUDENT SCIENCE JOURNALS AND THE EVIDENCE THEY PROVIDE: CLASSROOM LEARNING AND OPPORTUNITY TO LEARN*

Maria Araceli Ruiz-Primo, Min Li, Carlos Ayala, and Richard. J. Shavelson**
Stanford University

Paper presented at the NARST Annual Meeting

Boston, MA

March 30, 1999

* The report herein was supported by the National Science Foundation (No. SPA-8751511 and TEP-9055443). The opinions expressed, however, are solely those of the authors.
** The first author wants to thank Heather Lange, then a fourth-grade teacher, and Gabrielle Smith for trying out, with patience, the first iterations of the journal scoring form.

Student Science Journals and The Evidence They Provide: Classroom Learning and Opportunity to Learn

Science journals are a written account of what scientists do in their everyday practice. Consistent with professional practice students use science journals, for example, to describe observations made or procedures followed, and interpret data collected in doing an experiment on a particular day. Students use them to communicate their ideas and findings albeit with varying fidelity and clarity. Sometimes journals contain students' reflections on what they are learning. Because of these characteristics, science journals may be viewed as a potential assessment tool (e.g., Dana, Lorsbach, Hook, Briscoe, 1991; Hewitt, 1974; Shepardson & Britsch, 1997). These journals may provide an unobtrusive indicator of class experiences, an account of what students do in their science class and, possibly, what they learn. In this paper we evaluate science journals as an assessment tool producing scores that bear on students' performance over the course of instruction and on the opportunities students have to learn science in their classroom. At the outset of the paper we present the context of the study in which the science journals were collected. Then we describe the approach and provide technical evidence about the instrument. Finally, we present the information about classroom learning and opportunity to learn of six science classrooms based on the information collected from the science journals.

The Study Context

The evaluation of science journals as an assessment tool is part of a larger study in which the sensitivity of achievement assessments at different proximities to the enacted curriculum was examined. The purpose of the larger study was to provide NSF with an approach to evaluate the impact of inquiry science curricula reform (Ruiz-Primo, Wiley, Rosenquist, Shultz, Shavelson, Hamilton, & Klein, 1998). Information on student achievement in light of large NSF monetary expenditures are currently of considerable interest to Congress.

The "multilevel achievement assessment approach" is based on the idea that if science education reform is having an impact on student achievement, this impact should be located at different levels, the greatest impact should be at the local classroom curriculum level, and then, hopefully, transfer to cross-school curriculum levels as measured by statewide assessments. This approach uses different assessments based on their proximity to the central characteristics of the curriculum. Evidence about impact on student learning, then, is collected at different distances from the enactment of the curriculum: close -- assessments are close to the content and activities of the unit/curriculum; proximal -- assessments tap knowledge and skills relevant to the curriculum, but specific topics can be different from the ones studied in the unit; distal -- assessments are based on state/national standards in a particular knowledge domain; and remote -- assessments focus on general cross-state measures of science achievement.

Two units from the Full Option Science System (FOSS) curriculum were selected for this study, the Variables Unit and the Mixtures and Solutions Unit. For each unit close, proximal, and distal performance assessments were administered to fifth-graders to evaluate the impact of instruction on students' performance (for details see Ruiz-Primo, Wiley, Rosenquist, Shultz, Shavelson, Hamilton, & Klein, 1998). Proximity of the assessments to the central characteristics of the curriculum was defined using three categories: Purpose, Content, and Implementation. Table 1 presents an example of the questions we asked in each category to establish assessment proximity.

To provide an idea of what *close, proximal,* and *distal* assessments are, we describe one of the units and the three most proximal assessments used to evaluate the impact of instruction. In the Variables unit (FOSS, 1993), students are expected to design and conduct experiments; describe the relationship between variables discovered through experimentation; record, graph and interpret data; and use these data to make predictions. During the unit, students identify and control variables, and conduct experiments using four multivariable systems (e.g., Swingers and Lifeboats).

Table 1. Categories Used to Establish Proximity of Assessments

Category	Questions Asked	Aspects Used for Comparison
Purpose	• What is the assessment's purpose based on?	• Instructional activity goals, unit goals, curriculum goals or national/state standards.
Content	• What is the assessment task's content based on?	• Content domain, topic, concepts and principles learned in the curriculum unit and the ones used in the assessment task.
Implementation	• What is the assessment task based on?	• Characteristics of the problems and procedures implemented in the unit versus the ones needed to solve the assessment task.
	• Is the level of structure of the assessment task the same as the instructional activities in the unit?	• Level of structuredness (e.g., students only follow directions to conduct an experiment or they design their own) in the instructional activities versus the structuredness of the assessment task.
	• How similar are the materials used in the assessment task compared to the ones used in the unit?	• Characteristics of the materials students used during the instructional activities compared to the ones used in the assessment task.
	• How similar are the assessment methods used in the assessment task to those used in the unit?	• Characteristics of the measurement methods (e.g., variables measured, instruments, procedures) students learned in the unit versus the one used in the assessment task.

The close assessment used to evaluate the Variables Unit was a modified version of the *Pendulum Assessment* in which students were asked to identify the variable that affects the time it takes a pendulum to complete 10 cycles (Stecher & Klein, 1995). Differences between the instructional and the assessment tasks are: (1) the materials used to construct the pendulum and to manipulate the suspended weight, and (2) the way the dependent variable is measured. The proximal assessment was the *Bottles Assessments* in which students were asked to explain what makes bottles float or sink (Solano-Flores, Shavelson, Ruiz-Primo, Schultz, & Wiley, 1997; Solano-Flores, & Shavelson, 1997). Differences between the instructional and the assessment tasks are: (1) the materials used in the assessment are totally different; (2) the procedure used to manipulate the variables is different; and (3) the procedure used in the instructional unit to learn about sinkers and floaters is totally different to the procedure used on the assessment task. Still, the assessment requires knowledge about variables, levels of variables, and how to interpret results. Finally, the distal assessment was the *Trash Performance Assessment* administered by the California Systemic Initiative Assessment Collaborative--CSIAC.[1] The instructional and assessment tasks differ in multiple ways: (1) the focus of the assessment task is on a different domain, physical science; (2) none of the topics learned in the unit (e.g., variables, systems, controlled experiment) are part of the assessment tasks, and (3) the problem, procedures, materials, and measurement methods were different than the ones used as instructional activities.

The study was conducted with 163 fifth-graders from seven classrooms in a medium size school district in the Bay Area. The Variables unit was taught in 3 classes (70 students), and the Mixtures and Solutions in four (93 students). The close and proximal assessments were administered before and after instruction of each unit. Students within each classroom were randomly assigned to take the pretest and the posttest in different sequences (e.g., close-close or proximal-proximal). The distal assessment, part of a different study, was administered after instruction only. Students' science journals were collected at the end of the school year.

Results indicated that instruction had an impact on students' performance. More specifically, significant differences were observed between the pretest and the posttest scores when close assessments were administered, but not with proximal assessments. Moreover, assessment scores were in the predicted direction: close assessments were more sensitive to the changes in students' pre-to-posttest performance (Variables mean effect size = .32; Mixtures and Solutions mean effect size = 1.44) whereas proximal assessments did not show as much impact of instruction (Variables mean effect size = .12; Mixtures and Solutions mean effect size = .07).

High between-class variation in effect sizes for both the close and proximal assessments, and across the two units, suggested the need to examine closely the opportunities students had to learn the units' content. Students' science journals were thought of as a possible source of information that could help to explain, at least in part, the differences between classrooms.

[1] The CSIAC assessment is developed based on the standards proposed on the National Science Education Standards (NRC, 1996) and the Benchmark for Science Literacy (AAAS, 1993) and supports the learning goals of different systemic initiatives funded by NSF.

Science Journals As Assessment Tools

Science journals are seen primarily as a log of what students do in their science class. These journals encourage students to write as a natural part of their daily science class experience. Students may describe the problems they are trying to solve, the procedures they are using, observations they are making, and report their conclusions and reflections. Variations in this basic idea can be easily found (see Shepardson & Britsch, 1997). For example, students may also write what they think about the investigation they will conduct making explicit their questions, ideas, and understandings for later reflection. The main characteristic of science journals, however, is that they are a written account, in more or less detail and with diverse quality, of what students do and, hopefully, learn in their science class.

Indeed, there is a general agreement that science journals can be a formative assessment tool for teachers. Science journals allow teachers to assess students' conceptual and procedural understanding and provide the feedback students need for improving their performance (e.g., Dana, Lorsbach, Hook, & Briscoe, 1991; Hewitt, 1974; McColskey & O'Sullivan, 1993; Shepardson & Britsch, 1997). In what follows we describe another perspective and function of science journals as an assessment tool. We think of science journals as an assessment that can be used also by authorities, external to the classroom (e.g., at the district or state level), for obtaining information not only about students' learning, but also about the opportunities students had to learn science as well as some aspects of the quality of instruction students received.

Assessment Approach

In the context of the multilevel approach, science journals are seen as an immediate assessment – the closest proximity any assessment has to the curriculum. Since journal entries are generated during the process of instruction, the purpose, content, and forms of implementation of the assessment tasks (see Table 1) match those of the instructional activities.

We view science journals as assessment tools at two levels: (1) at the individual level, they may be considered a source of evidence bearing on a student's performance over the course of instruction; (2) at the classroom level, they may be a source of evidence of opportunities students had to learn science.

Our focus on opportunities to learn, and not only student performance, is based on the idea that students cannot be held accountable for achievement unless they are given the adequate opportunity to learn science. Therefore, both students' performance and opportunity to learn science should be assessed (see National Science Education Standards/NRC, 1996). We propose two indicators to evaluate opportunity to learn using science journals: (1) exposure to the science content students have to learn as specified in the curriculum/program adopted, and (2) quality of teachers' feedback to the students' performance as observed in their science journals.[2]

The assessment approach, then, focuses on three aspects of students' science journals: (1) Unit Implementation -- What intended instructional activities were implemented as reflected in the student's journals? Were any other additional activities implemented that were appropriate to achieve the unit goal? (2) Student Performance -- Were students' communications in the journal complete, focused, organized? Did students' communications indicate conceptual and procedural understanding of the content presented? and (3) Teacher's feedback to student performance -- Did the teacher provide helpful feedback on students' performance? Did the teacher encourage students to reflect on their work?

Unit Implementation. As mentioned before, one aspect of opportunity to learn can be defined as students' exposure to the science content. Inferences about opportunities to learn based on students' journals are based on the assumption that science journals are an account of what students do in their science classroom. If this assumption is accepted, it should be possible to map instructional activities implemented in a science classroom when information from individual science journals is aggregated at the classroom level. If none of students' journals for a class has any evidence that an activity was carried out, it is unlikely that the activity was implemented. Furthermore, if science journals allow teachers

[2] We acknowledge that there are many indicators of opportunity to learn at the classroom level (e.g., teacher's content and pedagogical knowledge, and understanding of students). Science journals are seen as one source of evidence, among others, that can be used as an indicator of opportunity to learn.

to assess students' understanding, we think some evidence of this should be found in the students' journals in the form of teacher's comments -- the second indicator we propose to evaluate opportunity to learn.

In this study, the science content to be implemented was specified in two FOSS units, Variables and Mixtures and Solutions. Two questions guided the evaluation of opportunity: (1) What intended instructional activities, as specified by the FOSS units, were implemented as reflected in the student's journals? And, (2) were any other additional activities implemented that were appropriate to achieve the unit goal?

Evidence of the implementation of an instructional activity can be found in different forms in a student's journal: description of a procedure, hands-on activity report, interpretation of results, and the like. Variation in these forms is expected across activities and students' journals. For example, the characteristics of a journal entry vary since each entry may ask students to complete different tasks depending on the instructional activity implemented on a particular day (e.g., write a procedure or explain a concept). Furthermore, journal entries may vary from one student to the next within the same classroom for a number of reasons (e.g., student was absent when a particular instructional activity was implemented). The variety of journal entries can be even wider when students' science journals are compared across different classrooms. To tap the variation in journal entries within- and between-classes, the approach identifies all the different tasks reported in the journals and links them to the intended instructional activities specified in the FOSS units.

To answer the first question -- *What intended instructional activities, as specified by the FOSS units, were implemented as reflected in the student's journals?* -- we defined the instructional tasks to be considered as evidence that the unit was implemented. The specification of these tasks was based on the description of the implementation presented in the teacher guide for each FOSS unit. For example, each unit (e.g., Variables) had four activities (e.g., swingers, lifeboats, planes, and flippers). On each activity (e.g., swingers), the teacher guide defines: (1) the concepts to be reviewed (e.g., variable and controlled experiment), (2) the different instructional tasks to be implemented (e.g., construct the swinger system, test different variables of the swinger system, predict outcomes and compare results), and (3) the products expected (e.g., swinger picture graph).

Based on this description we created a verification list for the basic instructional tasks for each activity in the unit. Two criteria were used to include an instructional task in the verification list: (a) the teacher guide explicitly called for the implementation of that task (e.g., "introduce the concept of variable" or "review the concept of variable"), and (b) the implementation of the task could not be inferred using another, more relevant, instructional task (e.g., if a variable is tested, say weight of the swinger, it can be inferred that the swinger was constructed, therefore, "constructing a swinger" was not included in the verification list). These criteria helped us reduce to the minimum the instructional tasks used as evidence for the unit implementation and, therefore, to estimate the number of instructional activities implemented and to identify extra activities.

The verification list followed the units' organization: one list for each activity and one for assessments suggested (i.e., hands-on assessments, pictorial assessments, reflective questions). Each activity-verification list contained different Parts (P) that corresponded to the description of the activity (see Table 2). Each unit, then, has four activity-verification lists and one assessment-verification list.

To answer the second question -- *Were any other additional activities implemented that were appropriate to achieve the unit goal?* -- we classified any instructional task not specified in the verification list as: (1) definition of a concept, (2) description of a procedure, (3) inquiry activity (e.g., prediction, observation, recording data, interpreting data), (4) content question not addressed in the unit, (5) quick writes (e.g., what did you learned with this activity?), and (6) unrelated activity (i.e., task not directly related to the unit goal). The verification list allows us to identify "extra-instructional tasks" within each part of an activity. Given the context of the instructional task it is easy to define in which part (P) of the activity the extra task was implemented.

Table 2. Example of the Journal Verification list for the Activity 1, Swingers, of the Variables Unit.

Variables Unit—Activity 1--Swingers	1 Basic/Extra Implemented? 0-1	2 Complete Report/Activity? 0-2	3 Type of Activity 1-6
P.1 Making Swingers			
Defining Pendulum			
Defining Cycle			
Swinger Test: How many times swinger will swing in 15 seconds?			
Replication of swinger test			
Defining Variable			
Extra Activity			
Extra Activity			
P.2 Testing New Variables			
Activity Sheet: Swinger Pendulum Graph			
Review: What is a variable?			
Standard Pendulum System			
Defining Controlled Experiment			
Experiment 1: Release position			
...			
Experiment 2: Weight			
...			
Extra Activity			
Extra Activity			
Reflections on the Activity—Questions at the end of the Act.			
Recall: What variables did we experiment with?			
...			

The shaded boxes (Table 2) in the verification list mean that the criteria do not apply to the instructional task at hand. For example, for the basic-instructional tasks specified in the unit, there is no need to know the type of activity. This criterion only applies to any extra-instructional task. Another example is the criterion, "Completeness of Report," that only applies to the basic-instructional task "Activity Sheet". Activity sheets are provided by FOSS for students to fill out for each activity. They are considered an essential piece of the implementation of any unit activity.

For each basic- or extra-instructional task identified (First Column: 1 = Yes, 0 = No), two sets of criteria are applied: one to evaluate the appropriateness and accuracy of the student's communications according to the requirements of the task —student performance, and another to determine the quality of the feedback provided by the teacher to help students improve their performance —teacher's feedback.

Student's Performance. According to the National Science Education Standards (NRC, 1996), inferences about students' understanding can be based on an analysis of their classroom performances and work products. Communication is considered in the Standards as fundamental for both, the performance and product-based assessments. If science journals are considered as one of the possible product of a student's work, evidence about her performance can be collected from the written/schematic/pictorial accounts of what she does in her everyday science class.

Inferences at the individual level about a student's performance is based on an analysis of the student's communications provided in her science journal. Student's notes, written reports, diagrams, data sets, explanation of procedures or results reported in the science journal can be seen as evidence not only of unit implementation, but also of a student's conceptual and procedural understanding, as well as evidence of her scientific communication skills (e.g., Dana, Lorsbach, Hook, & Briscoe, 1991; Hewitt, 1974; McColskey & O'Sullivan, 1993; Shepardson & Britsch, 1997).

A student's performance can be evaluated for each instructional task represented in her journal. The evidence is provided in different forms of communication (e.g., diagrams, data sets, notes, activity sheets), and each form of communication -- a written/text communication (e.g., explanatory, descriptive, inferential statements), a schematic communication (e.g., tables, lists, graphs showing for example data), or a pictorial communication (e.g., drawing of apparatus) – can be evaluated.

We evaluated each communication along four dimensions (Table 3). The first three focus on the quality of the communication – clarity, completeness, and organization, and the fourth on the level of conceptual or procedural understanding reflected in the communication (e.g., Does a student's explanation apply the concepts learned in the unit correctly? Does the student's description provide examples of a concept that are correct? Is a student's inference justified based on relevant evidence?).

Table 3. Criteria for Assessing Student's Communications.

Aspect	Student's Communication Is Scored As:
Communication	***Complete if*:** • Explanation or description does not lack sentences or paragraphs that make the communication not interpretable. • Table/list/drawing does not lack their main requirements and are filled out completely (e.g., a table should have rows and/or columns with list of items, facts; a list should be a series of names, materials, equipment).
	***Coherent, Clear, and Focused if*:** • Reader can easily identify in the explanation/description the main issue addressed (e.g., a definition, procedure, or interpretation of results). • Reader can easily identify the topic in the table/list/drawing (e.g., data collected, materials used, observations made).
	***Organized if*:** • Explanation/description is arranged in an orderly and systematic way (e.g., communication has subtitles or is arranged in steps). • Table/list/drawing has all the appropriate titles/labels.
Conceptual/Procedural Understanding	***Conceptual if*:** • Communication refers to defining, exemplifying, relating, comparing, or contrasting unit-based concepts. ***Procedural if*:** • Communication refers to a procedure carried out during an activity/experiment, observations/results/outcomes, interpretation of results, conclusions, and investigation plan.

Completeness and *Coherence, Clarity, and Focus* were scored as 0 (No) or 1 (Yes). *Organization of Communication* was evaluated using a three-level score: 0--No Organization (i.e., no sign of organization); 1--Minimal Organization (e.g., student uses only dates to separate information or only lists information); and 2—Strong Organization (e.g., students uses titles, subtitles, labels appropriately). *Conceptual* and *procedural* communications were evaluated on a four-point scale: (NA)—Not applicable (i.e., instructional task does not require any conceptual or procedural understanding); 0—No Understanding (e.g., examples or procedures described are completely incorrect); 1—Partial Understanding (e.g., relationships between concepts or descriptions of observations are only partially accurate or incomplete); 2—Adequate Understanding (e.g., comparisons between concepts or descriptions of a plan of investigation are appropriate, accurate and complete); and 3—Full Understanding (e.g., communication focuses on justifying student's responses/choices/decisions based on the concepts learned or the communication provides relevant data/evidence to formulate the interpretation). If a student's communication is scored on the completeness and clarity scales as 0 no further attempt is made to score the remaining dimensions (i.e., organization and conceptual/procedural understanding).

Forms of communications (i.e., written, schematic, or pictorial) are not thought of as been fixed for a particular instructional task (e.g., communicating how the swinger system is constructed may be in writing, a picture or both). Because some forms of communication may be more suitable for certain instructional tasks than others, certain types of instructional tasks (e.g., experimental procedures) may lead to fixed communication forms (e.g. written communications). Also, some instructional tasks may have the same form of communication across students because teachers may have required the form (e.g., "Describe in writing how the swinger was built."). Written communications (i.e., "text" communications) are not assessed by looking at single sentences, but rather by analyzing the entire communication represented for each instructional task. The written communication may be just a paragraph, or a two-page description, but in both cases, a score is assigned to the whole communication.

Teacher Feedback. According to the National Science Education Standards (NRC, 1996), one aspect of opportunity to learn is teacher quality. We acknowledge that systematic observation of teaching performance by qualified observers is probably the best indicator of teacher quality. However, this method is expensive (e.g., large numbers of observations by qualified observers are needed to capture a wide range of teacher performances). Alternative methods have been proposed (e.g., portfolios as those used for certification purposes), each with advantages and disadvantages. We think that students' science journals can be used as a source of evidence about one aspect of teaching, the use of feedback.

Indeed, Black and Wiliams (1998) provide strong evidence on the relation of the nature of feedback and student achievement. Black (1993) has shown that formative evaluation of student work (e.g., feedback) can produce improvements in science learning. However, teachers' effective use of formative evaluation is hard to find (e.g., Black, 1995; Black & Wiliam, 1998). Furthermore, classroom teachers are rarely good at providing useful feedback (e.g., Wiggins, 1993). Most of the time feedback is considered as a comment in the margin that involves praise and/or blame or a code phrases for mistakes (e.g., "seg. sentence!"). Research has found that quality of feedback (i.e., comments, comments with grade or grade only) affects its effectiveness for improving students' performance (e.g., Butler, 1988). If a teacher's feedback is just a grade (e.g., B-) or a code phrase (e.g., "incomplete!" or a happy face sticker), such information can hardly help students redirect their efforts to meet the needs revealed on the journal entries.

If science journals allow teachers to assess students' understanding, we would expect to see some evidence of feedback in the students' journals. If teachers do not respond, probe, challenge, or ask for elaborations of journal entries, the benefit of the journals as a learning tool and as an instrument to inform students about their performance may be lost.

We assessed the quality of teacher feedback for each instructional task identified in the verification list. We used a six-level score: –2—feedback provided, but incorrect (e.g., teacher provides an A+ for an incorrect journal entry); -1—no feedback, but it was needed (e.g., teacher should point out errors/misconceptions/inaccuracies in student's communication); 0—no feedback; 1—grade or code phrase comment only; 2—comments that provide student with direct, usable information about current performance against expected performance (e.g., comment is based on tangible differences between current and hoped performance, "Don't forget to label your diagrams!"); and 3—comments that provide student with information that can help to reflect/construct scientific knowledge (e.g., "Why do you think this is important for selecting the method of separation so as to know whether the material is soluble?). Rules were created for those cases in which one instructional task had more than one type of feedback. All rules follow the idea of providing teachers with the highest possible score.

Method

Students' Journals. Five of the 75 elementary schools in a medium sized urban school district in the Bay Area participated in this study with seven classrooms/teachers and 163 fifth graders (Ruiz-Primo, Wiley, Rosenquist, Shultz, Shavelson, Hamilton, & Klein, 1998). As mentioned before, the Variables Unit was taught in 3 classrooms (70 students) and the Mixtures and Solutions in four (93 students). Unfortunately, one teacher did not collect her students' journals, reducing to six the classes that participated in this part of the study. Information about students' reading and mathematics scores was provided by the school district. Performance assessment scores (close, proximal and distal) were available for each student.

Science journals were collected at the end of the school year. Students' journals within a classroom were selected according to students' performance level on the posttest, high, medium, or low. Journals were randomly selected from two top-, two middle-, and two low-groups within each class. In two of the three classrooms in which Mixtures and Solutions was implemented, only four journals were provided by the teachers, reducing the number of journals scored for that unit. Total number of journals scored were 18 for Variables and 14 for Mixtures and Solutions.

Other Sources of Evidence. At face value, journals would seem to reflect what happened in classes. Nevertheless, the question of corroborative evidence arises. Two independent sources of evidence for unit implementation were also collected, teachers' unit logs and teachers' verification lists. While implementing the unit, teachers kept a Unit Log for each unit activity. The log focused on: (1) time spent on each activity; (2) type of group work used during instruction (i.e., individual work, pair/small group, and large group); (3) type of instructional activity (i.e., teacher presentation, student reading/writing, hands-on investigation, discussion, and other); (4) FOSS materials used (i.e., videos, think sheets); and (5) other non-FOSS activity related to the unit. For each teacher the unit log was collected.

Two teachers from the Mixtures and Solutions classes routinely used their own Verification Lists to score students' journals. The teachers' verification lists included all the activities they did in class. Each student had a verification list with a checkmark for each activity reported in her journal. Unfortunately, only seven students' teacher-verification lists could be collected.

Instrumentation. To score students' science journals two verification lists (see Table 2), one per unit, were developed following our approach. We call the verification lists, "Journal Scoring Forms". A Scoring Criteria Table and Scoring Rules were also developed. The table was designed to provide scorers with criteria, codes and examples of students' performances to use during scoring. Two independent scorers evaluated each student's journal. Scorers were experts in the unit content and activities. Students' journals within units were mixed and randomly assigned an order of scoring. Scorers were unaware of the class or level of student performance.

Results

Preliminary analyses focused on two main issues: (1) Information about the technical quality of the journal assessment -- Can two raters reliably score student's science journals? Do students' science journals provide similar information about the unit implementation when compared with independent sources? And (2) whether information collected through science journals help, in any way, to explain differences observed across classrooms and units in the posttest performance assessments scores. Before describing the information related to the technical quality of journal scores, we present information about performance in the sample of students who participated in the study.

Describing the Sample

We first compare the complete sample of the study with the sample of students used for the journal study. Table 4 presents posttest mean scores and standard deviations for the complete sample (n = 163) and for those students whose science journals were collected.

Table 4. Mean Scores and Standard Deviations for the Complete Sample and the Sample Used for the Journal Study

Unit	Type of Assessment	Max Score	All Sample			Journals Sample		
			n	Mean	SD	*n*	Mean	SD
Variables	Close	16	34	10.40	3.61	9	10.61	3.21
	Proximal	29	36	16.39	5.75	9	15.38	6.92
	Distal	62	57	28.33	13.01	15	30.20	10.38
Mixtures	Close	20	43	8.23	4.40	9	8.06	3.09
	Proximal	18	50	8.85	4.16	5	6.40	4.39
	Distal	62	75	39.71	12.40	8	35.75	10.33

In general, students who participated in the journal study have similar means and standard deviations than those found in the complete sample. Means for those students who took the proximal and distal assessment for the Mixtures and Solutions unit are lower, but not very far away from the original sample. This information suggests that the sample of students whose journals were scored can be considered an appropriate sample of the classrooms that participated in the study.

Journal Scores. According to the approach we proposed, three general scores were obtained for each student's journal: Unit Implementation (UI), Student Performance (SP), and Teacher Feedback (TF).[3] Table 5 provides the descriptive information for each score. Maximum scores for UI across the two units were based on the basic-instructional tasks; no extra-instructional tasks were considered in the preliminary analyzses. For the Variables Unit, no evidence of implementation of the two lasts instructional activities (i.e., planes and flippers) was found in any journal, consequently, maximum scores for SP and TF were calculated considering only the first two activities (i.e., Swingers and Lifeboats).

Table 5. Means and Standard Deviations for Each Type of Score Across Units and Classrooms

Type of Score	Variables (n=18)			Mixtures and Solutions (n = 14)		
	Max	Mean	SD	Max	Mean	SD
Unit Implementation	27	6.19	1.88	46	20.32	9.83
Student Performance	58*	9.64	4.89	201	49.39	26.87
Teacher's Feedback	48*	-1.44	1.19	138	13.42	15.02

* Maximum score based only on two instructional activities, Swingers and Lifeboats.

In general, mean scores were lower for those classrooms in which Variables was taught than in those in which Mixtures and Solutions was implemented, based on the maximum possible score. According to the information provided in the students' journals, around 44 percent of the basic-instructional tasks suggested by the FOSS teacher's guide were implemented in the classrooms where Mixtures and Solutions was taught, whereas only 22 percent of the instructional activities for the Variables unit were implemented. Low performance across the two units revealed that students' communication skills and understanding were far away from the maximum score. Teachers who taught the Mixtures and Solutions unit provided, in general, higher quality feedback than those teachers who taught Variables. In the Variables unit, the mean score was negative. This means that teachers tended not to provide feedback to students despite the fact that errors or misconceptions were evident in the students' communications. Unfortunately, since

[3] Notice that other sub-scores can be obtained within each dimension (e.g., a sub-score for conceptual understanding).

teachers' feedback scores were not reliable (see next section), no final conclusions can be drawn about these scores. In the next sections we discuss these findings in more detail.

Reliability

Each science journal was scored by two scorers. Interrater reliability was calculated for each score across units (Table 6).

Table 6. Interrater Reliability Coefficients Across Units

Type of Score	Unit	
	Variables	Mixtures and Solutions
Unit Implementation	.92	.99
Student Performance	.90	.95
Teacher Feedback	.49	.93

In general, the magnitude of the coefficients are very high across the three types of scores, except for the teacher feedback score for Variables. This means that despite the variability in the students' journal entries and the diversity of the forms of students' communications (written, schematic or pictorial), raters can consistently identify whether or not an instructional task was implemented. Furthermore, raters can consistently score student performance and teacher feedback for each instructional task, at least for the Mixtures and Solution Unit. Unfortunately, this was not the case for the Variables unit. Two points may explain this result. First, although the interrater reliability for UI for Variables was very high, there were so few instructional tasks implemented for this a missed instructional task by a scorer would have a big impact on the total TF score. Second, there was one scoring rule misapplied by one of the raters. Unfortunately, time constraints for producing this paper did not permit a second round of scoring using different raters. However, we are confident that teacher feedback can be consistently scored. Improvement in the scoring rules will help to avoid this inconsistency.

Although TF scores for the Variables unit were not reliable, it is important to note two issues related to the scores in this group: (1) The percent of agreement between raters when individual scores were compared across students was .71. And (2) none of the raters scored any teacher feedback as a "2" or a "3". This means that teachers who taught this unit did not provide any helpful feedback to students.

Validity

To examine whether journals can serve as a trustworthy source of information about experiences students had in their classrooms, we qualitatively compared the information provided by independent sources of information: Teachers' Unit Logs, Teachers' Verification Lists, and Journals Scoring Form.

Teachers' unit logs were collected for all six classrooms. Unit logs did not provide detailed information about the different instructional tasks implemented for each activity. Rather, they only included Activity Parts (e.g., P.1, Making swingers and P.2, Testing new variables). Agreement between teachers' unit logs and the science journals, then, was calculated by part within each activity. We defined an agreement as when in the unit log there was evidence that an activity part was implemented and in any of the students' scoring forms there was evidence that at least one basic-instructional task for that part was implemented.

For two of the three classrooms in which Mixtures and Solutions was taught, information on teachers' verification lists were also available. All teachers' verification lists included a detailed list of the activities implemented, however, they varied in which activities were included. Agreement was calculated by student and averaged. Agreement was defined as when in the teacher's verification list an activity was checked as implemented and our scoring form also identified the same activity. Table 7 provides the results of these two qualitative analyses.

Table 7. Percent of Agreement About Unit Implementation Between Teachers' Units Logs and Students' Science Journals Across Classrooms

	Variables	Mixtures and Solutions	
	Teachers' Unit Logs	Teachers' Unit Logs	Verification List*
Activity 1	100	100	81.63
Activity 2	89	89	76.14
Activity 3	83	100	90.48
Activity 4	100	100	80.97

* Only for Classrooms 1 and 3

Percentages of agreement between teachers' unit logs and information on unit implementation were high. On average, 93 and 97 percent of agreement was found across activities in the Variables and Mixtures and Solution respectively.[4] Percent of agreement with teachers' verification lists was not as high, but still adequate, 82.30, on average, across activities. Agreement using both sources varied according to the class. It is important to note that the main reason for disagreements with teachers' verification lists was that teachers did not provide a check mark for activities that students did have in their journals but were identified by us. We concluded that information gleaned from journals about the opportunity students had to learn the unit content was trustworthy.

To examine whether the journal scores bearing on student performance behaved as an achievement indicator, journal scores were correlated with scores students obtained on the multilevel performance assessment. Table 8 shows the correlation obtained across units according to the proximity of the assessments: close, proximal, and distal. Correlations with reading and math scores are also provided.

Table 8. Correlations Between Different-Proximity Assessment Scores and Reading and Mathematics Scores

Unit		Proximity of Assessments						Other Measures	
		Immediate			Close	Proxi-mal	Distal	Read	Math
		UI	SP	TF					
Variables	Unit Implementation (UI)	-	-	-	.43 (n=9)	.18 (n=9)	.01 (n=15)	.23 (n=16)	.67** (n=16)
	Student Performance (SP)	.40 (n=18)	-	-	.37 (n=9)	.05 (n=9)	.59* (n=15)	.42 (n=15)	.67** (n=15)
	Teacher Feedback (TF)	NA	NA	-	NA	NA	NA	NA	NA
Mixtures	Unit Implementation	-	-	-	.84** (n=8)	.61 (n=5)	.76** (n=8)	.45 (n=13)	.68* (n=12)
	Student Performance	.97** (n=14)	-	-	.85** (n=8)	.62 (n=5)	.80** (n=8)	.52 (n=13)	.71** (n=12)
	Teacher Feedback	.86** (n=14)	.87** (n=14)	-	.62* (n=8)	.55 (n=5)	.72** (n=8)	.15 (n=13)	.44 (n=12)

** Correlation is significant at.01 level
* Correlation is significant at.05 level
NA Not applicable since scores were unreliable

[4] Note that the agreement in Activity 4 in Variables indicates that both teachers' unit logs and our scoring forms did not provide any evidence that the activity was implemented. In Activity 3, one teacher's unit log indicated that two Activity 3-Parts were implemented, but any of students' journals in that class provided any evidence of the implementation.

Correlations of student-level unit implementation scores, one aspect we propose as an indicator of opportunity to learn, with the other measures were all positive. Although not all the correlations were significant (small Ns), all were in the right direction, indicating that the more opportunities students had to learn science content, the better their performance across different measures. We expected the correlations to be higher with the proximal than the distal assessment; however, for Mixtures and Solutions, the correlation was higher and significant for the distal assessment.

Correlations of students' journal performance with the proximity assessments are of special interest if journal scores are to be used as an achievement indicator. Although correlations were not in the desired pattern, we expected the pattern of correlations to vary according to the proximity of the assessment, it is important to note that correlations with the distal assessment (i.e., state/national assessments) were in all cases significant and high.

For the Mixtures and Solutions unit, correlations of teachers' feedback with the other measures were positive, high, and, except for the proximal assessment, significant. Notice that the correlation with the distal assessment was higher than those for the close and proximal assessments. Correlations of teacher feedback with ability measures, reading and math, were also positive but not significant. These results are consistent with previous research that indicates that feedback, as a form of formative evaluation for students, has positive impact on students' performance (e.g., Black, 1993). Unit implementation and student performance scores correlated positively with reading and math measures. However, only the correlations with math were significant across types of scores.

Based on these correlations, we concluded that students' science journals can provide reliable and valid information about students' performance and opportunity to learn. In the next section we use the information provided by the journals as a way to explain differences in students' performance observed across classrooms.

Using Journal Scores

As mentioned before, results from the multilevel evaluation indicated high between-class variation in effect sizes for both the close and proximal assessments, and across the two units. This suggested the need to examine closely the opportunities students had to learn the units' content. In this section we focus on the use of journal scores as a possible source of information that can help to explain, at least in part, these differences. We use the unit implementation and the teacher feedback scores as indicators of the opportunity to learn. Figure 1 provides information about unit implementation and students performance on the close posttest across classrooms.

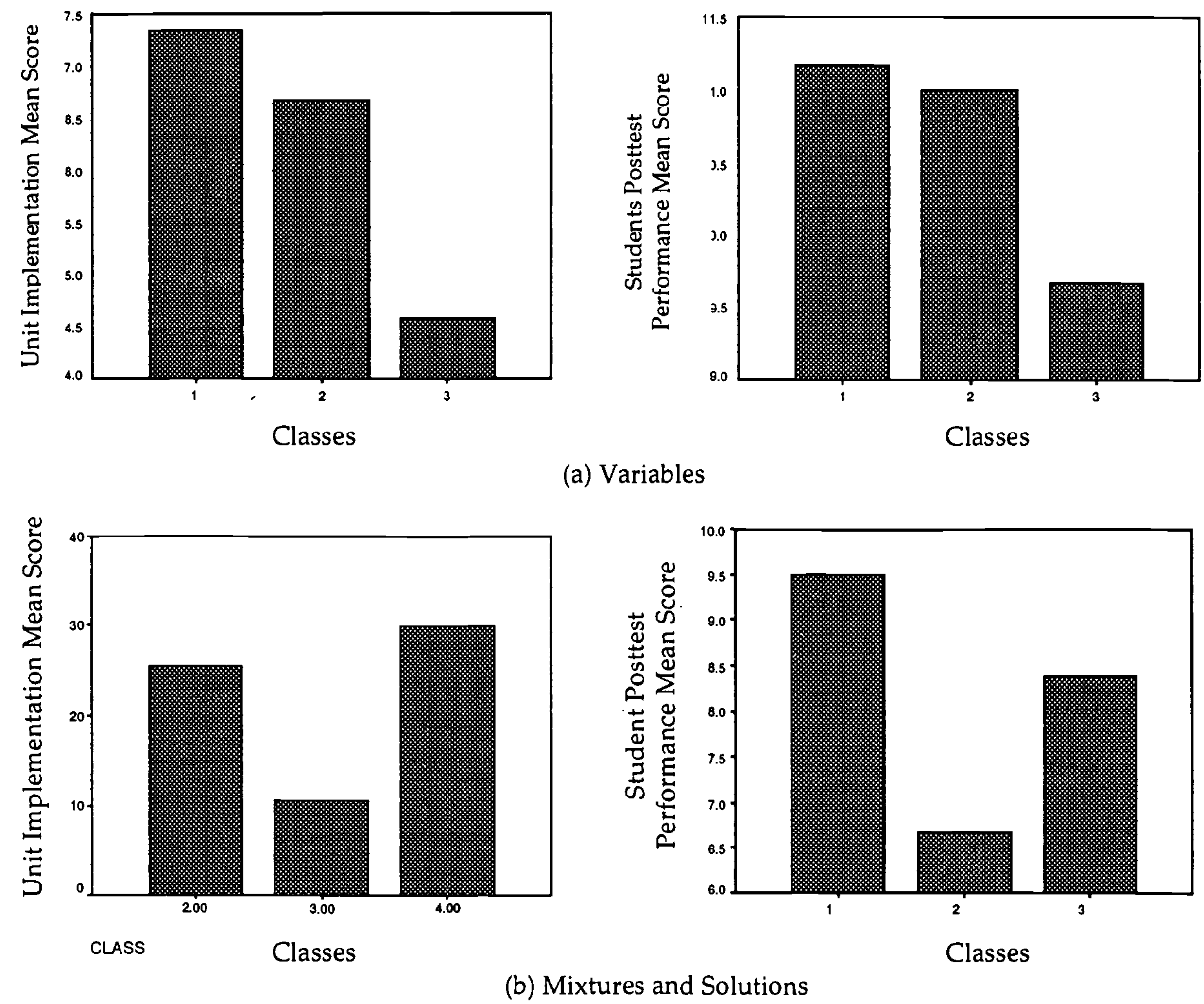

Figure 1. Histograms comparing unit implementation and student performance on the close assessment across units.

Differences in the implementation of instructional tasks, the first aspect of opportunity to learn, across classrooms are evident in both units. The same pattern of differences can also be observed in students' performances on the posttest. We concluded that those students who had less opportunity to learn the unit content as evident in their journals, performed more poorly when compared with classrooms in which more instructional tasks were taught. Furthermore, unit-implementation mean scores across units indicated that more instructional tasks were implemented for the Mixtures and Solutions unit than for the Variables unit, which is also reflected in the magnitude of the effects sizes found in the complete sample (Variables mean effect size = .32 and Mixtures and Solution mean effect size 1.44).

It is important to mention that even though we found a significant increase from pretest to posttest in the complete sample (n=163; see Ruiz-Primo, et al., 1998), low mean scores across the two units using the close assessment suggested to us that knowledge exhibited by students on the posttest was partial and far from the maximum score (see Table 4).

We acknowledge that other factors may be involved in the trends of effect sizes observed across classrooms, such as class composition and the characteristics of the unit. For example, classroom 2 in the Mixtures and Solutions group had a significantly lower reading mean score when compared to the other classrooms, although no significant difference was found on math. We also believe that the nature of the unit is important to consider. The Variables unit seems to be a more difficult unit to teach than Mixtures and Solutions. For example, when developing the journal scoring form for this unit, our group found that some of the activities proposed by FOSS were not only difficult to implement, but results were hard to replicate across trials (e.g., planes). It seems that the unit requires from teachers more depth of knowledge than does Mixtures and Solutions.

The other aspect of the opportunity to learn we consider in our approach is teachers' feedback. We present information about teachers' feedback scores only for the Mixtures and Solutions unit, since for the Variables unit this type of score was unrealiable. Figure 2 provides the percentage of teacher's type of feedback by classroom.

Across the three classrooms, the type of feedback with the highest percentage is the "Missing" feedback, which represents those instructional tasks not found in the students' journals and, therefore, students cannot receive feedback. The next highest percentage is for Type 1 feedback—teachers provide only a grade or code phrase comment. The teacher in Class 3 provided more Type 2 feedback (comments that provide students with direct, usable information about current performance against expected performance), and Type 3 feedback (comments that provide students with information that can help them reflect/construct scientific knowledge). Although both Type –2 (incorrect feedback) and Type –1 (no feedback when needed) were present across the three classrooms, their percentages were not high (but it would be desirable to have a 0 percentage in these two categories). These two negative categories may reflect teachers' content knowledge, often considered as an indicator of opportunity to learn. Type 9 (incongruent feedback) represents the feedback in which both a positive type of feedback (1, 2 or 3), and a negative type (-2 or –1) were found for one instructional task.

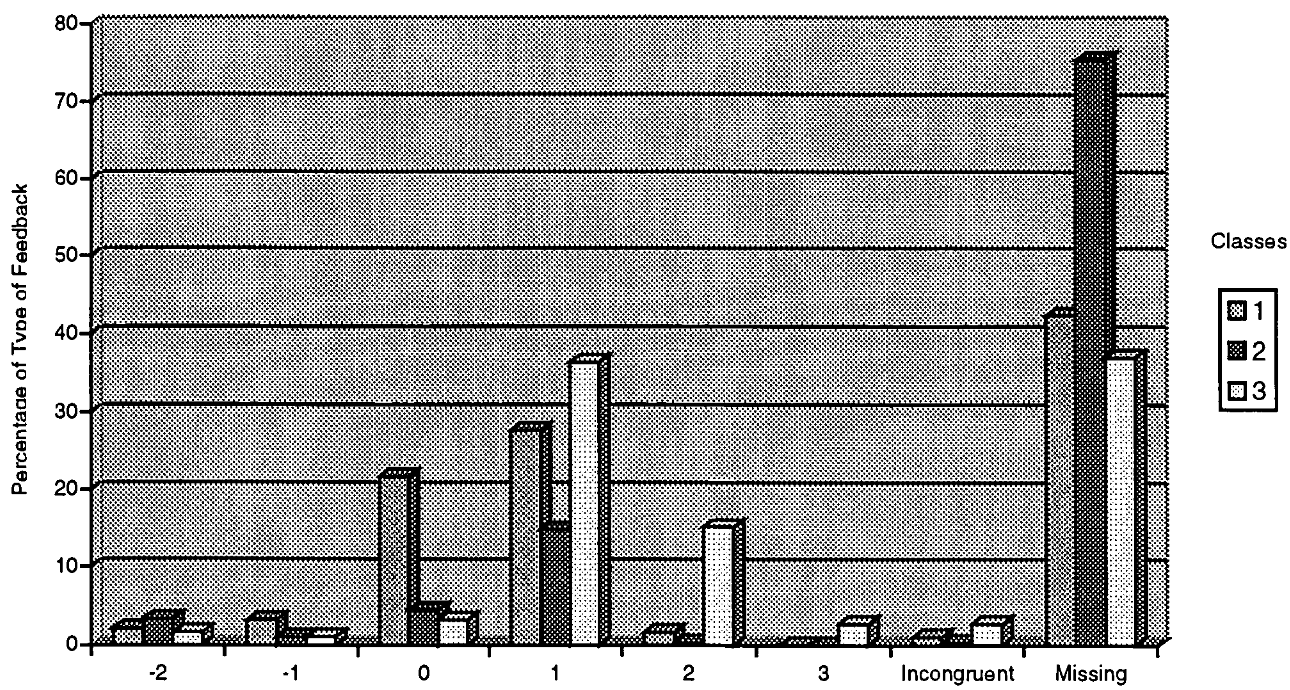

Figure 2. Percentage of teacher feedback by type and classroom.

BEST COPY AVAILABLE

Notice that Classroom 2 is the class in which students' science journals received, on average, less feedback from the teacher, and the class with the most Type –2 feedback. Classroom 2 was also the class with the lowest performance on the posttest (see Figure 1).

We believe that the fact that teachers used more "grades" or "short comments" (involving a praise and/or blame for mistakes) as feedback reflects a limited understanding of what feedback really means. Feedback is information that provides the performer with direct, clear, usable insights into current performance, based on the differences between the current and the expected performance (Wiggins, 1995). This means that for providing feedback, teachers need to have a clear idea of the hoped-for performance. If this is not clearly specified, it is difficult to determine the "differences between the current and the expected performance" and therefore difficult to provide comments that help the students to know how they are doing and how they can improve their performance. We found, for example, that teachers tend to write "great" for written descriptions of procedures which vary in quality. This may be because there is not a clear criterion of what a good description of a procedure is (e.g., the description should allow other students to replicate the procedure described).

Another important characteristic of feedback is that it should be descriptive, not evaluative or comparative. Focusing on labeling student's performance over-emphasizes grading and under-emphasizes learning (Black, 1993). In fact, it has been found that "giving of praise" as a feedback can have a negative impact on low achieving students (Butler, 1988)!

Conclusions

In this study we explored the use of students' science journals as an assessment tool that provides evidence bearing on their performance over the course of instruction and on the opportunities they have to learn science. We examined whether students' journals could be considered a reliable and valid form of assessment and whether they could be used to explain, at least partially, between-class variation in performance.

Our preliminary results indicate that: (1) Students' science journals can be reliably scored. *Unit implementation* and *student performance scores* both were highly consistent across scorers and units. Teacher feedback scores, however, proved to be consistent across raters only for the Mixtures and Solutions unit. Nevertheless, we believe that *teacher feedback* should be considered a reliable score once the criteria and rules for scoring this aspect of the Variables-Unit journals are improved. (2) Inferences about unit implementation using journal scores were justified. A high percent of agreement with independent sources of information on the instructional activities implemented indicated that the *unit implementation score* was valid for this inference. (3) Inferences about students' performance are also very encouraging. High and positive correlations with other performance assessment scores indicate that the *student performance score* can be considered as an achievement indicator. Although the pattern of correlations were not the same across the two units, in general, correlations were in the right direction. (4) The unit implementation score helped to explain differences in the performance across classrooms. Those classrooms in which journals showed that more instructional activities were implemented, were associated with higher performance means. (5) Low student performance mean scores across the two units revealed that students' communication skills and understanding are far away from the maximum score. And (5) teacher feedback scores helped to identify teacher feedback practices across classrooms.

In a larger study than the one described here, we collected information using the multilevel achievement assessment in 20 classrooms from 12 schools over the two units, Variables, and Mixtures and Solutions. Information was collected at all assessment levels -- immediate (i.e., students' science journals), close, proximal and distal. We expect that this sample of about 500 students will provide more definite results about the multilevel assessment approach we have proposed and the importance of students' science journals as an immediate assessment tool.

References

American Association for the Advancement of Science (1993). Benchmarks for Science Literacy. Washington, DC: National Acdemy Press.

Black, P. (1993). Formative and summative assessment by teachers. Studies in Science Education, 21, 49-97.

Black, P. & Wiliam, D. (1998). Assessment and classroom learning. Assessment in Education, 5(1), 7-74.

Dana, T. M., Lorsbach, A. W., Hook, K., & Briscoe, C., (1991). Students showing what they know: A look at alternative assessments. In G. Kulm and S. M. Malcom (Eds.), Science assessment in the service of reform (pp. 331-337). Washington, D.C.: American Association for the Advancement of Science.

Full Option Science System (1993). Britannica Science System. Chicago, IL: Encyclopaedia Britannica Educational Corporation.

Hewitt, H. C., (1974). The journal. Science and Children, 11(8), 30-31.

McColskey, W. & O'Sullivan, R. (1993). How to assess student performance in science: Going beyond multiple-choice tests. Tallahassee, FL: Southeastern Regional Vision for Education (SERVE).

National Research Council (1995). National science education standards. Washington DC: Author.

Ruiz-Primo, M. A., & Shavelson, R. J., (1996a). Problems and issues in the use of concept maps in science assessment. Journal of Research in Science Teaching, 33(6), 569-600.

Ruiz-Primo, M. A., & Shavelson, R. J., (1996b). Rhetoric and reality in science performance assessments: An update. Journal of Research in Science Teaching, 33(10), 1045-1063.

Ruiz-Primo, M. A., Wiley, E., Rosenquist, A., Schultz, S., Shavelson, R. J., Hamilton, L., & Klein, S. (1998, April). Performance assessments in the service of evaluating science education reform. Paper presented at the NCME Annual Meeting, San Diego, CA.

Shepardon, D. P. & Britsch, S. J. (1997). Children's science journals: Tool for teaching, learning, and assessing. Science and Children, 34(5), 13-17, 46-47.

Solano-Flores, G., Shavelson, R J., Ruiz-Primo, M.A., Schultz, S. E., & Wiley, E. (1997, March). On the development and scoring of classification and observation science performance assessments. Paper presented at the AERA Annual Meeting. Chicago, IL.

Solano-Flores, G., & Shavelson, R. J., (1997). Development of performance assessments in science: Conceptual, practical, and logistical issues. Educational Measurement: Issues and Practices, 16(3), 16-25.

Stecher, B. M. & Klein, S. P. (1995). Performance assessments in science: Hands-on tasks and scoring guides. (Technical Report DRU-1087-NSF). Santa Monica, CA: RAND Corporation.

Wiggins, G. P. (1993). Assessing student performance. Exploring the purpose and limits of testing. San Francisco: Jossey-Bass Publishers.

U.S. Department of Education
Office of Educational Research and Improvement (OERI)
National Library of Education (NLE)
Educational Resources Information Center (ERIC)

ERIC®

TM029882

REPRODUCTION RELEASE

(Specific Document)

NCME

I. DOCUMENT IDENTIFICATION:

Title: "Student Science Journals and The Evidence They Provide: Classroom Learning and Opportunity to Learn"

Author(s): Maria Araceli Ruiz-Primo, Min Li, Carlos Ayala, Richard Shavelson

Corporate Source: Stanford University

Publication Date:

II. REPRODUCTION RELEASE:

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

If permission is granted to reproduce and disseminate the identified document, please CHECK ONE of the following three options and sign at the bottom of the page.

The sample sticker shown below will be affixed to all Level 1 documents	The sample sticker shown below will be affixed to all Level 2A documents	The sample sticker shown below will be affixed to all Level 2B documents
PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL HAS BEEN GRANTED BY Sample TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC) 1	PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL IN MICROFICHE, AND IN ELECTRONIC MEDIA FOR ERIC COLLECTION SUBSCRIBERS ONLY, HAS BEEN GRANTED BY Sample TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC) 2A	PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL IN MICROFICHE ONLY HAS BEEN GRANTED BY Sample TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC) 2B
Level 1	Level 2A	Level 2B
☑	☐	☐
Check here for Level 1 release, permitting reproduction and dissemination in microfiche or other ERIC archival media (e.g., electronic) *and* paper copy.	Check here for Level 2A release, permitting reproduction and dissemination in microfiche and in electronic media for ERIC archival collection subscribers only	Check here for Level 2B release, permitting reproduction and dissemination in microfiche only

Documents will be processed as indicated provided reproduction quality permits.
If permission to reproduce is granted, but no box is checked, documents will be processed at Level 1.

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

Sign here,→ please

Signature: Maria Araceli Ruiz Primo	Printed Name/Position/Title:	
Organization/Address:	Telephone:	FAX:
	E-Mail Address:	Date:

(over)

III. DOCUMENT AVAILABILITY INFORMATION (FROM NON-ERIC SOURCE):

If permission to reproduce is not granted to ERIC, or, if you wish ERIC to cite the availability of the document from another source, please provide the following information regarding the availability of the document. (ERIC will not announce a document unless it is publicly available, and a dependable source can be specified. Contributors should also be aware that ERIC selection criteria are significantly more stringent for documents that cannot be made available through EDRS.)

Publisher/Distributor:
Address:
Price:

IV. REFERRAL OF ERIC TO COPYRIGHT/REPRODUCTION RIGHTS HOLDER:

If the right to grant this reproduction release is held by someone other than the addressee, please provide the appropriate name and address:

Name:
Address:

V. WHERE TO SEND THIS FORM:

Send this form to the following ERIC Clearinghouse:

THE UNIVERSITY OF MARYLAND
ERIC CLEARINGHOUSE ON ASSESSMENT AND EVALUATION
1129 SHRIVER LAB, CAMPUS DRIVE
COLLEGE PARK, MD 20742-5701
Attn: Acquisitions

However, if solicited by the ERIC Facility, or if making an unsolicited contribution to ERIC, return this form (and the document being contributed) to:

ERIC Processing and Reference Facility
1100 West Street, 2nd Floor
Laurel, Maryland 20707-3598

Telephone: 301-497-4080
Toll Free: 800-799-3742
FAX: 301-953-0263
e-mail: ericfac@inet.ed.gov
WWW: http://ericfac.piccard.csc.com