VINCENT R. ROGERS AND CHRIS STEVENSON

How Do We Know What Kids Are Learning in School?

Two exploratory studies suggest that to determine what students have learned, we must "dig deeper," using assessment modes that allow us to probe their understanding.

Sam teaches science to 6th graders who have been in the school's advanced track for four years. The class moves steadily through large chunks of content in order to complete the 7th grade text by the end of the year. Units last from one to three weeks, and conventional tests are the major evaluation of students' learning. Sam prides himself on the fact that almost every student averages at least 90 percent on these unit tests.

Not long ago, three months after a typically successful unit about electricity, an ice storm caused power outages in the community. Schools were closed for two days. When the students returned, conversation centered around the unexpected holiday and how families coped with inconveniences. When Sam directed the discussion to their recent study of electricity, the youngsters were unable to relate their earlier study to these immediate events, much to his surprise. When questions went unanswered, he would explain. Occasionally his explanations were met with "Oh yes, I remember that." More frequently, however, the children either withdrew from the discussion or looked away in embarrassment.

Once his 12-year-olds were removed from the immediate context of schoolwork, they didn't know much about content they had "mastered" on tests given a few weeks earlier. Sam pondered, "Just what are my kids really learning here?"

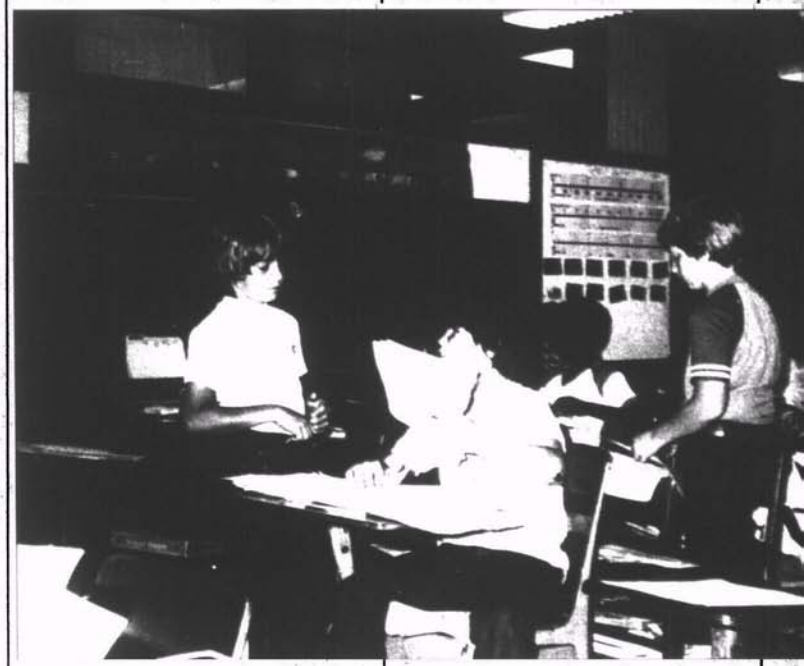
Alternative Assessment Methods

On the surface, the question is deceptively simple. If you want to know what kids are learning, you test them. Mostly you test them with paper and pencil, multiple choice, matching, and fill-in items that can be machine scored to yield numerical results. Yet Sam's experience suggests that the problem is far more complicated than that.

Most of us, of course, are aware of the difficulty of measuring the enduring effects of teaching and of the overall "messiness" of the assessment problem. Finding out what students have learned implies probing the recollections, associations, interpretations, values, and beliefs that they derive from school experiences. It means investigating alternative assessment modes: ways of digging deeper, more thoroughly, more subjectively. So where do we find the techniques to help us?

Anthropologists in particular and qualitative researchers in general have a great deal to offer. They seek the meaning of events as they are embodied in the perceptions of those being studied: their attitudes, values, beliefs, and underlying assumptions about events. Anthropologists' search for "thick description"—the richest, fullest, most comprehensive description possible—has led them to develop techniques that we can adapt to education.

Our mutual interest in these issues led us to develop two exploratory projects. Both were carried out in suburban middle school classrooms, one in a 5th grade social studies class in Connecticut and the other in an 8th grade English class in Vermont. In both schools the units of study we assessed are parts of established curriculums.



The Connecticut Project

We identified a 5th grade class about to begin a one-week social studies unit on the Bill of Rights, the Constitution, and the branches of the U.S. Government. Their teacher described the 23 heterogeneously grouped children as an average class with a wide range of abilities and interests.

Our University of Connecticut research team obtained copies of the school's curriculum guide, lists of goals and objectives, textbook, other instructional materials, and teacher-developed tests. The researchers developed the following assessment devices to supplement the conventional tests in use.

1. *Small-group interviews.* Each team member conducted loosely structured interviews with groups of four to five children about fundamental rights such as freedom of the press, freedom of religion, and so on. The interviewers emphasized clarification

"Teachers teach and children learn, but what they learn, if we probe deeply enough, often bears little resemblance to the 'taught' curriculum."

and elaboration ("Can you tell me what you mean by that? Can you give me an example?"). All responses were considered acceptable, and students were encouraged to talk to each other as well.

2. *Situational pictures.* The researchers showed the children simple sketches of situations in which basic American rights would come into play; for example, a sketch of a nativity scene on public property being taken down two weeks before Christmas. Children were asked to explain why they thought this was happening and whether or not they thought it was justified. Another sketch showed two policemen with a criminal suspect. One policeman says to the other, "Give me 10 minutes alone with him, and I'll make him talk." Children were asked to discuss whether the second policeman should consent or object to the "grilling" of a suspect.

3. *Card sorts.* Children were given packs of three-by-five cards, each with the name of a key person from the various branches of the federal government, including the President, Supreme Court Justices, and various cabinet members. Each child was asked to sort the cards into "most important" and "least important" piles and to explain his or her reasoning for sorting the cards that way.

4. *Learning logs.* At the beginning of the unit, each student received a blank notebook. At the end of each day's lesson the children spent a few minutes writing responses to these questions:

- What was the most important thing you learned today? Why do you think so?
- What are you still confused or uncertain about?

5. *Leader snapshots.* The researchers showed pictures of key government figures—senators, cabinet members, and so on—to small groups of children and asked them to identify the individuals and then tell what they do. For example, the children were shown a picture of a U.S. Congresswoman and asked, "What do you think she does in the Congress each day?"

6. *Open-ended versions of conventional tests.* The original test on the Bill of Rights asked students to agree or disagree with 10 statements. For example, "The police or FBI may sometimes be right in 'grilling' a person to make him or her give them information." The open-ended version added two parts to each question: (1) "If you agree with this statement, list examples of when police would be justified in grilling a suspect; and (2) "If you disagree or are uncertain, explain why."

"A large majority of the responses confirm the impressionistic nature of these students' understandings—misconceptions, overlapping fragments of concepts, mistaken connections of events, misconstructions of relationships."

These six assessment techniques, the results of most of which were taped and later transcribed, produced a mass of data. We used some devices (the small-group interviews, for example) both before and after the unit. We would like to share some of the more provocative responses gathered during our post-teaching assessment.

For instance, the children's written responses to the open-ended version of the conventional test reveal an interesting theme pervading their thinking.

- In response to "In some cases the police should be allowed to search one's home without a warrant," students who agreed gave these examples of situations:

Police should search homes if they think you are selling drugs or killed someone.

. . . if a really bad crook did something like rob two banks.

- In response to, "Some criminals are so bad they should not be allowed to have a lawyer":

Yes—criminals who murder, steal, or rape.

If a criminal is so bad, he should just be put in jail.

- In response to "The police or FBI may sometimes be right in grilling a person to make them give information":

Yes, if they kill somebody or rob a bank.

Yes, if someone shot somebody or kidnapped someone.

- In response to "In criminal cases, the government should have the right to hold secret trials":

Yes, if you shot a policeman or stole something from a military base.

Yes, if you murdered someone.

These sample responses suggest that, regardless of the "right" in question, many children seemed quite willing to dispense with it if the crime was serious enough as they perceived it; that is, if murder, drugs, or stealing something valuable were involved. They believe in basic rights if the crime is minor. If the crime is "really bad," however, rights may be ignored.

Consider also the varying conceptions of "freedom of the press" that emerged as we examined sample responses from the small-group interview transcripts.

I think that freedom of the press means someone from a newspaper can come in and ask you questions without you saying, "Get out of here; I don't want you here." They should have the right to come in on certain things. I mean that if it is a real private thing, you don't want them in it. So, you might want to keep some things private. Like if somebody from the *Courant* came in and said, "I want to know all your life story," and you don't want to say it, I mean you don't have to say it. You have the right to keep that secret.

I think that freedom of the press might be like newspaper press—maybe like their right to go places and do things. Like to go to South America and come back—you should be able to go places and do things—be able to follow people—some things that normal people won't be able to do.

Like if the President sent a missile out in space, and they want to put it in the newspaper, and he doesn't want it, it can't be kept top secret.

To most children, however, freedom of the press seemed to focus on the theme, "If you didn't have it, you wouldn't get any warning about possible dangerous events about to occur." The notion of freedom of the press as intellectual license, freedom to express ideas and opinions, rarely occurred. The following small-group interview responses were typical.

People like to know what's going on. If there was a tiger on the loose and you would be walking on the street, you wouldn't know.

If you didn't have freedom of the press, this is what I think—let's just say Russia was going to attack, and we didn't have any newspapers to be informed at all.

Because there would be nothing to tell you what's going on. Russia could send over one of their nuclear missiles and blow this whole school away.

The children's attitudes toward the Soviet Union as they discussed freedom of religion are reflected in the following two sample responses.

In Russia, if you're outside by ten o'clock you would be persecuted, but you had to be inside by ten o'clock. In a church? No, in your home. If anybody is in the streets after ten o'clock they'd be shot.

You have the right to a religion if you want, but like in Russia, you couldn't do that sort of thing, they just give you a religion, and that's it.

In addition, children almost always described the meaning of freedom of religion in the following terms.

You can be any religion you want, and no one can treat you differently. You can have the same job as Catholics; no one can bother you. You can go to any church you want, and you can study different religions.

If I were Jewish and everybody else was Christian and they said, "Hey, you can't be Jewish, so we'll come over with a machine gun and kill you." But here nobody cares like if that guy is Catholic and next door there is a guy who's Jewish; he's not going to come out with a machine gun and go into your house and hold you prisoner or give you a choice: either you turn Christian, or I'm going to kill you, or hang you, or crucify your children.

It's important because if there's just one religion people might not be satisifed with it, and they really wouldn't want it.

" . . . our experience also suggests that the best assessment devices are often the most effective teaching devices. . . ."

Most of the children have internalized the view that, in this country, you can go to any church you want, and that freedom of religion means *choice*. Only two children mentioned the equally important right *not* to believe.

For some children, freedom is seen to exist only in the United States, as this transcript excerpt reveals.

Interviewer: When I say Bill of Rights, what do you think about?
Mary: Freedom. Our country has freedom, more freedom than other countries.
Interviewer: How do you know that we have more freedom than other countries?
Mary: Well, if you watch the news or read the paper . . .
Interviewer: What other countries are you thinking about?
Robert: Arabia and Africa.
Interviewer: Are there other countries that have as much freedom as the United States does?
Jeffrey: Not that I know of.
Mary and others: I don't know.

And, finally, what can one say about Chuck's response to an item dealing with cruel and unusual punishment?

Well, robbing the bank or something you'd usually get about 20 years in prison, but if you raped somebody and then had an abortion and all of that, and you kept doing it for nine months or so and they catch you then you're going to have to get the death punishment because you're going along raping all these ladies.

Interviewer: What are you saying? Is that cruel and unusual punishment?

Chuck: Yeah, it is. Because all they're doing is having their fun in their way and they're not hurting anybody. They're just doing what they feel they should do. They didn't kill anybody. They're jut having their fun.

These responses defy quantification. Some were the responses of individual children—personal, idiosyncratic—yet vitally important to the education of those single learners. Others revealed themes that dominated the thinking of the class as a group, although there were always exceptions to these trends.

A large majority of the responses confirm the impressionistic nature of these students' understandings—misconceptions, overlapping fragments of concepts, mistaken connections of events, misconstructions of relationships.

Obviously, this kind of information suggests the need for more discussion, elaboration, clarification, and for reteaching some of the unit's most important concepts. It also suggests that unintended learning was taking place, much of it outside of school. How does the committed teacher respond to Chuck's comments on rape, for example? Ignore them because they are not part of the curriculum? Or act on the belief that "I must begin wherever my children are?"

The Vermont Project

The Vermont project followed a different tack. Chris Stevenson taught a two-week unit about Robert Frost to students in one section of 8th grade English. He explained at the outset that he was teaching the unit because he wanted to learn more about how 8th graders learn. Stevenson observed the regular English teacher to ensure that his students moved at the same pace and covered the same material as those in the other two English sections. His students used identical materials and did the same work as the other classes.

The assessment plan for this unit consisted of five techniques.

1. *First unit test.* The regular teacher had designed a four-page, four-part test to evaluate students' mastery of selected content from the Frost unit. The first part consisted of short-answer items about the poet's life. The second part required students to define and give examples of literary terms such as "stanza" and "couplet." In the third section, students were asked to identify and analyze selected lines from Frost's poems in terms of literary devices they had been taught such as "alliteration" and "simile." The fourth part of the test asked students to write brief essays about the poet and his work. In addition, students had to write a Frost poem they had memorized during the unit. This test was administered during class on the final day of the two-week unit.

2. *Interviews.* A week after the unit's completion, a team of graduate students from the University of Vermont interviewed the students individually about their reactions to the unit. To assure students' anonymity, the interviewers did not record their names. The use of open-ended questions encouraged candid descriptions of the unit, the teacher's instruction, and students' work and responsibilities. For example, the first question was "Tell me about the Frost unit." When students appeared shy or reluctant, interviewers asked more specific questions such as "What went well for you?" or "What did you enjoy most?" Once the interviews began to flow comfortably, the interviewers asked for concrete examples of specific learnings and opinions.

3. *Second unit test.* Eight weeks after first administering the unit test, Stevenson returned with fresh copies of the test. He encouraged students to do their best on this unexpected examination, which he assured them would not affect their semester grades. After they completed the test, Stevenson asked them to write the poems they had memorized on the back on the final page. Results were scored, analyzed, and compared to the initial tests.

"If one takes this sort of data seriously, one cannot possibly cover as much content; but is there any value in superficial coverage assessed superficially?"

4. *Ten-week assessment.* Ten weeks after the second unit test, Stevenson returned to collect anonymous assessment information about student reactions to the Frost unit. Using rating scales, the students appraised teaching assignments, materials used, difficulty of the material, classmates' reactions to the unit, and so on. Self-evaluation explored how students felt about themselves in regard to poetry. Also included, with lines from Frost poems, were selections by other poets to assess whether or not students could recognize Frost's style. Finally, students wrote advice for 7th graders who would be studying the unit next year.

5. *Final group interview.* One week before the end of the school year, Stevenson returned a final time to conduct a whole-group interview. His questions and statements were drawn from the results of previous testing and assessment experiences. He urged students to think seriously about the unit, the issues he raised, and their classmates' comments. The overarching question was "What are your thoughts about the study of Frost's poetry as part of your education?"

Insights and More Questions

The assorted assessment techniques used in this project produced an abundance of information about what students learned, understood, and believed about the study. Distillation of the data produced both provocative insights and challenging new questions.

1. *Test performance.* On the first test, half of the students scored between 83 and 97 percent; only four scored less than 70 percent. Two unusually low scores placed the class average at a deceptive 77 percent. As unit tests go, a conclusion that students had learned acceptably well is justified. Errors were scattered throughout the four sections, and item analysis did not identify any specific questions as particularly difficult. Students correctly wrote their memorized poems and made only occasional spelling and punctuation errors.

Overall performance on the second test eight weeks later produced interesting results. Scores dropped by an average of 32 percent, but two students actually scored *higher* the second time. While on the first test all but four students had scores 70 percent or higher, only four students scored that high the second time.

Analysis of performance on individual items and sections revealed that content recalled most successfully on the second test was narrative material, such as details of Frost's personal life. The items that students found difficult were those requiring conceptual applications; for example, identifying theme, metaphor, personification, and so on. Although most of the students had accurately performed such tasks during the teaching unit and on the first test, two months later they were unable to apply the same concepts correctly.

Only one student correctly wrote the poem memorized two months earlier. Three students wrote approximately 75 percent of the poem accurately, 11 reconstructed less than half of the poem, and 3 were unable to recall any lines.

2. *Individual interviews.* The overall tone of students' comments about the unit was positive. A few indicated that some of it was boring, but most reported liking it. When pressed for

examples of helpful practices, they most frequently mentioned teacher-led discussions. Several youngsters also liked the memorization, interpreting poems, and the teacher's explanations. A few students said that they sometimes found it difficult to understand the meaning of poems. Everyone spoke favorably of Frost and complimented his poetry.

Student reactions as revealed in the interviews were predictably supportive of the guest teacher, the unit content, and the place of Robert Frost's work in the school curriculum. Interviewers reported that students were polite, agreeable, and pleased to be a part of the project. No one offered even mild criticism, suggesting either that the unit was a great success or that students were unwilling to be completely candid with the unfamiliar interviewers.

3. *Ten-week assessment.* The data generated in the ten-week assessment is too voluminous to describe here, but several noteworthy themes emerged.

First, several students believed that they could handle certain tasks better than they actually did. For example, one item asked students to rate on a scale of 1 to 10 (low to high) the statement, "I can paraphrase poems." The average rating students gave this item was about 8, but only two students correctly handled paraphrasing tasks on the test. The rating for "I can recognize metaphor in poems" was similarly high, but only one student did so correctly on the test. Differences between self-assessment and actual performance on such items invite further inquiry.

A second interesting theme concerned differences between the ratings students gave themselves and the ones they gave their classmates. For example, students consistently rated their own interest in the Frost unit high but their classmates' interest relatively low. Students also rated themselves high and their classmates low in terms of enjoyment of the Frost unit, understanding of poetry, attitude toward future study of poetry, and ability to interpret poems. When these differences were explored in the final group interview 16 weeks after the unit, it became clear that students were not completely certain that their responses were kept anonymous. One youngster expressed the sentiments of others, "I just needed to be sure I had all the bases covered." This issue will be discussed later in detail under *Group Interview*.

"Although most of the students had accurately performed . . . tasks. . . during the teaching unit and on the first test, two months later they were unable to apply the same concepts correctly."

Third, unless these students were directed to look specifically for particular literary devices in poems, they were not inclined to see them. For instance, eight passages of two, three, or four lines that had been studied in class as examples of the use of particular poetic devices were printed on a separate sheet. Each passage included at least one device studied in class which had been included on the test: metaphor, simile, alliteration, monologue, personification, dialogue, couplet, and rhyme pattern. Students were invited to write any thoughts, interpretations, descriptions, or other reactions that occurred to them. Only one student identified a literary device (couplet). One-third of the lines elicited no response at all. Given that so much class time was spent on these very lines as examples of structures in poetry, this result was surprising.

A fourth major theme is that these students equated poetry with analysis or dissection. Several items invited students to make free associations with poetry; invariably the association implied or directly stated that "poems equal analysis." For example, a sentence completion item ("The most important thing about a poem is . . .") resulted in a preponderance of references to "structure, interpretation, meaning, theme." In response to "The feelings I have about Frost's poems are . . .", the typical reaction of the students is captured in ". . . that they mean something I need to figure out."

The evidence suggests that there were at least two dimensions of thought and value about the Frost unit. First, the unit was part of the school program—nothing out of the ordinary for 8th graders. Their responsibility was to do what they were told, try to understand the content, learn as much as they could, and make the best of it. The students fulfilled the instructor's expectations that they do what had been planned for them. Second, we sensed that the youngsters withheld reactions. Experienced teachers know from the expressions on students' faces, from the context of comments, and from things that are not said that another agenda also exists. It was that set of ideas, values, impressions, and understandings that Stevenson sought to explore in the final assessment interview.

4. *Group interview.* This interview was intentionally scheduled a week before the end of the year. Four months had passed since completion of the Frost unit. Students were eager for the festivities of their graduation to high school, and they were in excellent spirits. If they felt connected to the unit in a meaningful way, this would be a good opportunity to collect their candid reactions.

In response to the interviewer's question, "Should the study of poets and poetry be a part of everyone's education?", there was a consensus: no. Why? After some speculative comments, a unifying idea came to the surface: students should have choices about what they learn. When the interviewer sought clarification, ideas flowed easily. These students didn't really know whether or not an educat-

ed person needed to know poetry. It wasn't something over which they had any control; they had to "take it on faith," as one expressed it. A popular suggestion was that students should have an opportunity to sample a course of study and then decide whether or not to study it further. One girl offered, "If you chose it and it didn't work out—that's OK. But if you're forced, it really turns you off." Everyone applauded her comment. Their issue was *choice*, not poetry or Robert Frost.

The next logical question was, "If you had a choice, what would you choose to learn?" The first and most popular response was mathematics. When asked to explain that choice, they related it to being able to use it now and in the future in getting jobs. The extent to which they judged the worth of a topic on strictly utilitarian grounds was unavoidable. One student commented, "You can't use poetry," and others agreed.

In response to, "How did learning how to analyze poems affect your reactions to Frost's poems?", one girl quickly commented that "it helps you analyze what you read." Other students appeared puzzled by her comment. No one else felt that the study had helped them look analytically at other reading material. Next the interviewer posed a hypothetical situation: "If you were just about to begin a study of Carl Sandburg and his poetry, what did you learn in the Frost unit that would be useful in the new one?" Only four students responded. One girl mentioned "monologue and dialogue," another girl mentioned "finding the theme," a boy suggested "paraphrasing," and a student who had joined the class after the Frost unit had been taught offered "similes." While it should not be assumed that these statements represent the only transfer of knowledge that would occur, they constitute scant linkages in these students' perceptions.

The final issue addressed concerned students' aforementioned assignment of relatively high ratings for their own interest in poetry and the Frost unit versus their predicted low levels of interest among their classmates. They appeared puzzled by the questions, so their earlier responses were described in detail. As the description unfolded, they began to smile in recognition. Almost everyone agreed that they "wanted to make a good impression on the teacher." A few students disagreed, but they offered no other explanation. In short, then, we were all engaged in a masquerade, each doing what we thought we were supposed to do, accepting quite a lot of data about each other on faith. Students were reconciled to their roles, and they accepted the teacher's efforts on assumptions of his good intentions.

"Finding out what students have learned implies probing the recollections, associations, interpretations, values, and beliefs that they derive from school experiences."

New Directions for Teaching and Learning

We've gathered an enormous amount of information during these two years of work. The data reported here are illustrative in nature; we have been more concerned with arousing interest and suggesting possibilities than with making definitive statements about teaching effectiveness or how much or how little specific children were learning. Nevertheless, we would like to share a few tentative conclusions resulting from our experiences.

First, we are more convinced than ever of the personal, idiosyncratic nature of the phenomenon we call "learning." Teachers teach and children learn, but what they learn, if we probe deeply enough, often bears little resemblance to the "taught" curriculum.

Second, this suggests, of course, the need for continuing assessment, for formative data, for feedback, for continual curriculum revision and review, and, often, for reteaching. To do so may not be practical; certainly it is time-consuming and inefficient. If one takes this sort of data seriously, one cannot possibly cover as much content; but is there any value in superficial coverage assessed superficially?

Perhaps more important, our experience also suggests that the best assessment devices are often the most effective teaching devices. Indeed, the two are difficult to separate. Having 3rd graders teach 1st graders about apples or insects by organizing and writing books, reading them, and discussing them with younger children is sound pedagogy. But careful attention to and reflection about what the 3rd graders chose to share with the 6-year-olds will reveal a great deal about what mattered most to them in their own learning.

Next, we are convinced of the importance of real-life application as an assessment technique. Our most revealing responses came when we asked children to respond to pictures depicting real-life applications of the principles emphasized in formal teaching (for example, the "nativity scene" sketch).

The data we gathered also raised serious questions about the cognitive match of curriculum to learner. Many of the children's responses suggested that answers beyond the superficial might be too much to expect, given the abstractness and subtlety of the subject matter being taught (for example, the many meanings of freedom of the press).

Also, we became convinced that much valuable information could be

gained by using some form of pre-teaching assessment. Many of our teachers found it instructive to have some idea of children's conceptions and misconceptions prior to formal study.

We were unpleasantly surprised at the difficulty many children had in using language to respond to the questions raised in small groups or one-on-one interviews and in open-ended written assessment devices. Simply put, this kind of task seemed foreign to many children; they found it exceedingly difficult to explain, illustrate, describe, or clarify their ideas in either a written or spoken statement. This finding suggests that we should have developed an even greater variety of assessment modes—techniques that include children's drawing, dramatic play, construction, and so on.

Finally, assessment and evaluation are vastly different processes. As one of the teachers commented, "When you are evaluating, some kids' answers are better than others. When you are assessing, all answers are equal." While the kind of data we gathered might well contribute to a teacher's ability to make a more informed judgment when giving a "grade" to a child, its greatest value lies in the insight it gives us about the quality of the teaching-learning act in a given classroom and the opportunity it provides to revise and reteach intelligently.

The work we have described, while not precise or conclusive, supports the metaphor of education as a continuing journey, with barely discernible beginnings and unknown endings. Perhaps the experiences we've described may help to make the journey more informed and more intelligent for both teacher and learner.□

Authors' note: We wish to acknowledge the contributions of anthropologist Pertti Pelto of the University of Connecticut to our work on qualitative forms of assessment. This work was supported by the Matsushita Foundation.

Vincent R. Rogers is Professor, Curriculum and Instruction, University of Connecticut, School of Education, Box U-33, Room 322, 249 Glenbrook Rd., Storrs, CT 06268.
Chris Stevenson is Associate Professor of Education, University of Vermont, Burlington, VT 05405.