DOCUMENT RESUME

ED 383 731 TM 023 171

AUTHOR Stecher, Brian M.; Mitchell, Karen J.
TITLE Vermont Teachers' Understanding of Mathematical Problem Solving and "Good" Math Problems.
INSTITUTION National Center for Research on Evaluation, Standards, and Student Testing, Los Angeles, CA.; Rand Corp., Santa Monica, CA. Inst. on Education and Training.
SPONS AGENCY Office of Educational Research and Improvement (ED), Washington, DC.
PUB DATE 21 Apr 95
CONTRACT R117G10037
NOTE 30p.; Paper presented at the Annual Meeting of the American Educational Research Association (San Francisco, CA, April 18-22, 1995).
PUB TYPE Reports - Research/Technical (143) -- Speeches/Conference Papers (150)

EDRS PRICE MF01/PC02 Plus Postage.
DESCRIPTORS Educational Assessment; Educational Change; Elementary Education; Elementary School Teachers; Knowledge Level; Mathematics Curriculum; *Mathematics Instruction; Mathematics Tests; *Portfolios (Background Materials); *Problem Solving; State Programs; *Student Evaluation; *Teacher Attitudes; Teaching Methods; Test Construction; Testing Programs; *Word Problems (Mathematics); Writing Evaluation
IDENTIFIERS Performance Based Evaluation; *Vermont

ABSTRACT

For the past five years, Vermont has been developing an innovative statewide assessment system in which portfolios of student work in mathematics and writing are a key element. The use of mathematics portfolios, particularly in elementary school, is a novel aspect of the assessment system. This study examines the elementary school mathematics portfolio assessment and its instructional impact by exploring the conceptions of problem solving, the knowledge of problem-solving strategies, the evaluation of problem-solving tasks, and the instructional practices of 20 fourth-grade teachers. Teachers indicated that the portfolio assessment program has enhanced their understanding of mathematical problem solving and broadened their instructional practices, but that they have encountered difficulty in understanding certain components of the reform and making the relevant changes. Teachers did not share a common understanding of mathematical problem solving or agree on skills that students should master. Eight figures and three tables present study findings. (Contains 13 references.) (SLD)

VERMONT TEACHERS' UNDERSTANDING OF MATHEMATICAL PROBLEM SOLVING AND "GOOD" MATH PROBLEMS

Brian M. Stecher and Karen J. Mitchell

RAND Institute on Education and Training

National Center for Research on Evaluation, Standards and Student Testing (CRESST)

Paper presented as part of the symposium "The Impact of Alternative Assessments on Teachers' Knowledge and Practice" at the annual meeting of the American Educational Research Association, San Francisco, CA, April 1995.

The research on which this paper is based was funded by cooperative agreement number R117G10037 and CFDA catalog number 84.117G as administered by the Office of Educational Research and Improvement, U.S. Department of Education. The views expressed here do not reflect the position or policies of the Office of Educational Research and Improvement, U.S. Department of Education.

1. INTRODUCTION

THE VERMONT PORTFOLIO ASSESSMENT PROGRAM

For the past five years, Vermont has been developing an innovative statewide assessment system in which portfolios of student work in mathematics and writing are a key element. The Vermont program has two purposes: to serve as an impetus for curriculum reform and to provide meaningful accountability information at the school, district, and Supervisory Union levels. Since 1990, RAND has examined the reliability and validity of student and aggregate scores as well as the impact of the assessment program on selected classroom practices (Koretz, et al., 1994a; Koretz, et al., 1994b; Koretz, et al., 1993a; Koretz, et al., 1993b; Stecher and Hamilton, 1994).

Perhaps the most novel aspect of the Vermont assessment is the use of portfolios in mathematics, particularly elementary school mathematics. Writing portfolios are common instructional tools, and a number of jurisdictions include writing samples in their formal assessments. However, we know of no other operational large-scale assessment which is primarily based on collections of open-ended student responses to extended mathematical problems. For that reason, we undertook this focused study of the elementary mathematics portfolio assessment and of its instructional impact.

In Vermont, mathematics portfolios are supposed to contain the best mathematics work students have produced during the school year. Students collect their mathematics assignments throughout the year, and, at the end of the year, they cull from these collections five to seven entries they regard as their "best pieces." These best pieces make up the final portfolios that are submitted for scoring. Each piece is scored on seven dimensions using four-point scales. Four dimensions relate to mathematical problem solving; they are understanding the problem, approaches to solving it, decisions along the way, and outcomes. Three dimensions reflect mathematical communication, including the use of mathematical language, mathematical representations, and the overall presentation. Dimension subtotal scores and an overall total score are computed. Scoring is done by teachers other than the students' own in a single statewide scoring session. The portfolios are supplemented by standardized Uniform Tests in mathematics and writing. The mathematics Uniform Test is primarily but not entirely multiple-choice.

The dual foci of the Vermont mathematics portfolio assessment are problem solving and mathematical communication, neither of which held a prominent place in the instructional program previously. As a result, the implementation of the mathematics portfolio assessment program has required Vermont teachers, particularly fourth and eighth grade teachers, to learn new concepts, teach new content, and apply new methods of instruction. Indeed, teachers have been asked to adopt fundamentally different ways of thinking about mathematics and mathematics instruction. Consequently, teachers' abilities to acquire new knowledge and translate it into practice will determine the success of the portfolio initiative, both in terms of the quality of mathematics instruction presented to students and the quality of information provided for assessment purposes.

PURPOSE OF THIS STUDY

This study explores fourth grade teacher's understanding of mathematical problem solving, an aspect of the reform largely unexamined in previous RAND research. Specifically, we explore teachers' conceptions of problem solving, their knowledge of problem-solving strategies, their evaluation of specific problem-solving tasks, and their instructional practices related to problem solving.

2. PROCEDURES

SAMPLING

A two-stage process was used to select a representative sample of 20 fourth grade teachers who were using math portfolios during the 1993–94 school year. In the first stage, the population of Vermont schools was divided into four groups based on mean 1992–93 uniform test scores; a stratified random sample of 20 elementary schools was then drawn.[1] Two schools with fewer than ten fourth grade students were replaced by others drawn at random from the same strata.

In the second stage, one fourth grade teacher was drawn at random from each school and invited by letter to participate in the study. Neither district administrators nor school principals were notified about the study or of the teachers who were participating. During initial conversations, two of the 20 teachers were removed from the sample because they had not been teaching for the full school year or were not using portfolios in mathematics. Three other teachers declined to participate. Each of the five was replaced at random by another fourth grade teacher from the same school (if there was one) or by sampling another school from the same stratum (if there was no other fourth grade teacher in the school).[2]

Participating teachers had between one and four years of experience with portfolios (four years was possible if a teacher participated in the portfolio pilot in 1990–91) and between one and 40 years of teaching experience. On average teachers in the study had ten years of teaching experience, spending about one-half of this time at the fourth grade level. Figure 2.1 shows the distribution of teaching experience in the sample, and Figur 2.2 shows the distribution of portfolio experience.

---

[1] For the purposes of this study we use the term, "elementary school," to mean any school that included the fourth grade. This population contains mostly K-6 schools, but there are some K-8 and some K-12 schools.

[2] One respondent declined to participate because of the pressures of being a first year teacher; two others declined for personal reasons.

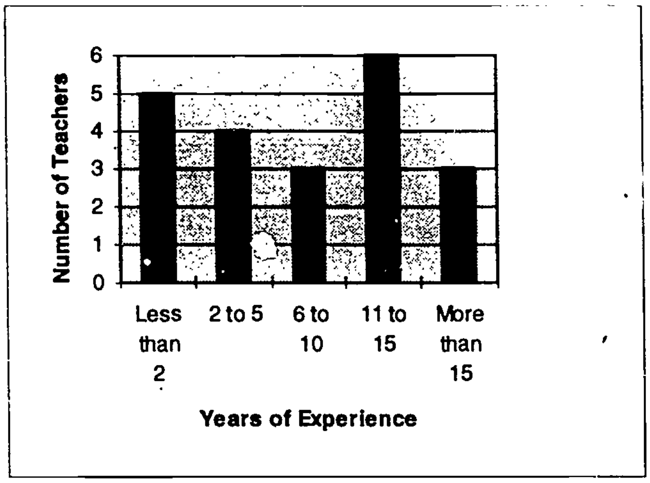

Figure 2.1—Teaching Experience

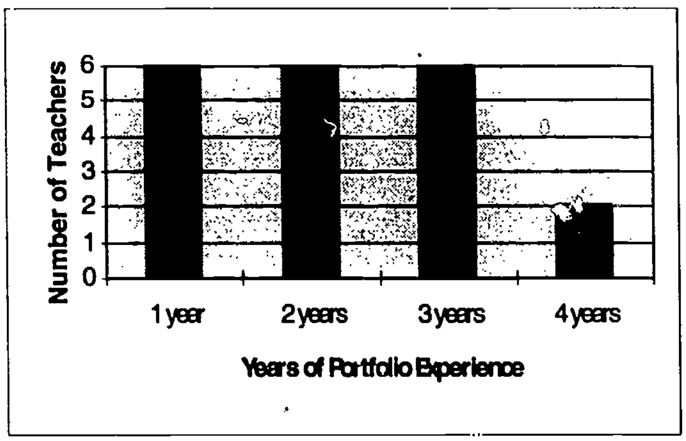

Figure 2.2—Portfolio Experience

Though the average experience of teachers in this sample is slightly lower than that reported in a statewide survey of fourth grade teachers conducted previously (Stecher and Hamilton, 1994), we consider the study sample to be reasonably representative of fourth grade teachers in Vermont. All of the teachers taught multiple subjects in self-contained classrooms with students who were heterogeneous in terms of achievement.

The Vermont Department of Education offered a summer training institute and four subsequent training workshops during the school year. All of the teachers in the sample attended at least one training session during the 1993–94 school year. They also attended training sessions the previous year. Over the past two years, one-quarter of the teachers attended some of the training sessions offered, 40% attended most of the sessions, and 35% attended all training sessions.

DATA COLLECTION

Data were collected using a written survey and an hour-long structured telephone interview.[3] Both the survey and the interview guide were developed by RAND with input from Vermont mathematics educators and a representative of the Vermont Department of Education. The written surveys were mailed two weeks before the scheduled interviews. Teachers were asked to complete the surveys prior to the interview, to retain them for reference during the interview, and to return them to RAND thereafter. All but one teacher completed the survey in advance and had it available for the interview.

In addition to background information on experience, classroom conditions and training, the written survey elicited teachers' judgments about a small number of specific problem-solving tasks. These common tasks provided a basis for quantifying variations among teachers on a number of dimensions. For example, teachers were asked to evaluate and suggest improvements to two specific tasks taken from (or patterned after) those in Vermont portfolios.[4] Teachers also were shown two pairs of portfolio tasks addressing similar mathematical topics. They were asked to explain which member of each pair would be better as an instructional activity and which would be more likely to produce high-scoring best pieces.

The subsequent telephone interviews focused on teachers' understanding of problem solving, task selection, and portfolio-related instruction. Questions also were asked about the portfolio tasks that appeared in the written survey. These questions examined the problem-solving skills students would use in response to the tasks and teachers' judgments about the merits of the tasks.

---

[3] The interviews were scheduled at the teachers' convenience, and almost all were conducted in the late afternoon or early evening.

[4] We used a collection of student portfolios from 1992-93 as a source of tasks. Unfortunately, many of the tasks we initially selected had been reviewed explicitly during network training sessions or appeared in the widely disseminated Resource Guide. In these cases, we generated new tasks of the same style and substance with different settings, values or characters.

DATA ANALYSIS

With the teachers' permission, the interviews were recorded on audio tape, and the recordings were used to insure fidelity to the ideas expressed by the teachers. Survey and interview results were summarized on a question by question basis, retaining key information in the teachers' own words. The two authors conducted all the interviews and the data analysis, each completing the initial summary on interviews he or she conducted.

Subsequent data analysis proceeded in stages. For each question, both authors first read one-quarter of the responses and independently developed coding schemes. Second, the two coding approaches were compared and reconciled. Third, the reconciled coding schemes were used to code the remainder of the responses. Any further additions or modifications to the coding scheme at this stage were made by mutual consent. Fourth, an analysis spreadsheet was developed for each question, including the teacher identification numbers and background information, response summary codes, and key direct quotes from the teachers. Fifth, data were summarized further through careful inspection and tabulation of information in the spreadsheets. Because of the small size of the sample, we adopted a conservative analytic approach demanding that differences be quite apparent before we were willing to report them. Finally, to examine the possible effects of experience and training on teachers' responses, the spreadsheets were sorted in rank-order on each of the experience/training variables and the coded and narrative teachers' data were examined for possible patterns of responses.

3. RESULTS

Results from the survey and interview were combined to investigate teachers' understanding of mathematical problem solving and how this understanding was translated into practice in terms of specific problem-solving tasks and activities. Two general comments are appropriate prior to the presentation of specific findings. First, the reader is reminded that all results are based on teachers' self-reports. We have no reason to suspect that respondents purposefully misstated their opinions or observations, but we know that memory is selective and there is unknown bias in these uncorroborated data. Although it complicates the text, we will periodically interject phrases such as "teachers reported" or "teachers said" to remind the reader of the origin of the information. Second, we found almost no differences in teachers' responses that were related to teaching experience, portfolio experience, or current year portfolio training. This finding ran counter to our expectations, and may be attributable partly to the small size of the sample.

TEACHERS' UNDERSTANDING OF PROBLEM SOLVING

Vermont fourth grade teachers asserted that the portfolio assessment program increased their knowledge of mathematical problem solving. These reports were supported by the facts that many teachers could define problem solving, describe the problem-solving skills they sought to teach, and analyze the problem-solving demands of particular tasks. However, teachers did not yet appear to share a common understanding of problem solving, and their conceptions of problem solving seemed somewhat vague and fragmented. There was variation in the problem-solving skills teachers said they address, and teachers did not agree about the problem-solving demands of specific tasks.

Definitions of Problem Solving

The portfolio program broadened fourth grade teachers' understanding of the domain of mathematics. Most importantly, teachers learned the importance of problem solving as an element of mathematics. Typically, teachers said they learned that mathematics is more than just computation and that there are many everyday applications of mathematical problem solving. Further, when we asked teachers to define problem solving, over half (60%) gave a reasonable definition, one that was consistent with the notion of responding to situations for which correct solutions are not immediately evident. The following statements are typical of the problem-solving definitions offered by Vermont teachers:

> I believe it to be facing some situation in which the answer is not immediately evident and it requires that the person face that challenge, recognize that it is a problem, that you have to give some thought to it, and they are willing to continue with it.
>
> You have a question that has to be answered. You have to look for a way to deal with the question. We will have different means to solve the problem . . . You have to analyze information and translate it into something that can be evaluated by another person. Then you learn if you're right or wrong.
>
> I look at problem solving as realistic. . . . As a good problem solver, you should be able to break down what's going on, figure out what your alternatives are, decide on the best alternative and try it. You can apply that to areas other than math.

Although teachers said they know far more about problem solving than they did prior to the advent of the portfolio assessment program and most teachers were able to provide a reasonable definition of the concept, their knowledge appeared tentative. Quite a number of respondents struggled to answer the question, "What is problem solving?", which was not typical of their responses to other questions.[5] For example, several interviewees initiated their responses, apparently became dissatisfied with their comments, and began their explanations again. Teachers appeared to have difficulty framing a description of problem solving.

The efforts of the teachers who struggled to define problem solving were not unproductive, however. Several teachers referred to relevant concepts from the Vermont training materials, such as the characteristics of good problems—e.g., their relevance to real life, their open-endedness, and their promotion of communication. For example,

> A problem is something that is relevant and meaningful to students. They can't arrive at answers off the top of their heads. There has to be some decision making.

[5] We asked a broadly-worded question so that teachers could speak to any aspects of problem solving that were salient to them.

I would say it is a task that you are asked to understand and to find a possible solution to. There may not be one right answer.

Problem solving to me is the ability to communicate to another person something you have done or are attempting to do in solutions. These are real-life. Decisions, decision making.

Over one-third of the teachers offered extremely broad definitions of problem solving that stretched beyond the domain of mathematics. They drew connections between mathematical problem solving and critical thinking in other disciplines and situations. For instance,

The word, problem solving, is kind of self explanatory. You have a problem with multiple factors and you are looking for a viable solution that fits your needs. It fits in all different subject matters and genres. Not only math. We talked about Nancy Kerrigan and Tonya Harding and whether the Olympic commission should let her skate. . . . Problem solving in general is thinking ahead. You think, "If I throw that spit ball across the room, what will happen?". . . . In social studies we are talking about bringing lumber across a river to a mill 100 years ago. This is problem solving to me.

A challenge. Any problem. They need problem-solving strategies across the curriculum. They need to decode in reading; they need technical skills in science. Problem solving encompasses life. I look at problem solving more broadly than mathematics. I also look at it as conflict resolution and problem-solving skills. How are they dealing with each other socially? . . You know what I'm talking about is critical thinking skills. I guess that's what I'm thinking of as problem solving.

Problem solving relates to real life—like when your clothes dryer is broken and you have to figure out what to do about it.

Although less common, this broad view of problem solving as an essential part of life may have important curricular and instructional consequences. However, the present study did not compare the effects on students or scores of this perspective to one that focuses more narrowly on mathematical problem solving.

To put the teachers' responses in context, it should be noted that we ourselves found it difficult to define problem solving. The Vermont training materials do not provide a definition of problem solving, although they do discuss the features of good problems. Even the *NCTM Curriculum and Evaluation Standards* (1989) beg the question. "Mathematics as Problem Solving" is the first NCTM standard, but one has to dig 75 pages into the text to find this indirect definition of problem solving:

> To solve a problem is to find a way where no way is known off-hand, to find a way out of a difficulty, to find a way around an obstacle, to attain a desired end that is not immediately attainable, by appropriate means (Polya in Krulik, 1980, p. 1, quoted in the *NCTM Curriculum and Evaluation Standards*, p. 75).

In light of the difficulty of defining problem solving, it is encouraging that more than one half of the Vermont teachers (60%) gave a definition consistent with Polya's; that is, they described a process in which one is confronted with a question for which the answer is not obvious and then takes steps to resolve it.

Finally, we note two instances in which the portfolio scoring criteria affected teachers' conceptions of problem solving. First, a few teachers framed their view of problem solving with words that reflect the scoring rubrics. For example,

> I think problem solving is being confronted with a question and you have to figure out what is being asked, how to solve it, and explain how you reached the decision.

Second, three teachers (15%) intimated that the scoring demands of the portfolio assessment program may neglect some meaningful aspects of problem solving. Unfortunately, the teachers could not identify these omissions. They could only describe neglected constructs in vague terms, such as,

> There are certain elements that have to be in portfolios so I am focusing to make sure I hit those . . . (but) . . . some of the kids have ideas and insights about problems that are extraordinary . . . they have such problem-solving strategies inside themselves that they are not getting scored on.

Delineation of Problem-Solving Strategies

Most teachers (85%) said they changed their instructional program to include discrete, specific problem-solving skills as a result of the portfolio assessment. Further, they could describe the skills they teach. Those skills discussed by at least one-quarter of the teachers are listed in Table 3.1. A few teachers were able to relate as many as ten different problem-solving strategies they convey to students. The average number of problem-solving skills listed by respondents was five, and over 20 different problem-solving skills were mentioned in all.[6]

It is difficult to interpret the quality of the information contained in Table 3.1 because there is no widely-held set of mathematical problem-solving skills requisite for fourth grade learners. Neither is there a prevalent taxonomy of problem solving that can be used to judge the completeness of the list or the relative importance of the skills on it.

Table 3.1

Commonly Taught Problem-Solving Strategies

Strategy	Percent of Teachers (N=20)
Make a table or organized list	60%
Represent/communicate to others	55%
Read/understand the problem	55%
Make a picture or diagram	50%
Look for alternative approaches	50%
Relate to real world	40%
Pick out important information	35%
Work backwards	30%
Guess and check	25%
Use manipulatives	25%
Use other information resources	25%

However, it is possible to use these data to make some statements about fourth grade teachers' understanding of problem-solving skills. First, only three skills were mentioned by more than one-half of the interviewees: making a table or list, representing/communicating information to others, and reading and understanding the problem. The fact that no two teachers provided the same or even highly similar lists suggests they do not have a common perception of the problem-solving demands of tasks or the problem-solving approaches of students. Second, teachers did not

[6] This includes a few approaches--such as, "taking risks" and "persevering"--that might be considered dispositions, not strategies.

appear to have a conceptual framework for analyzing, structuring, and recalling problem-solving skills; nor did they have a common vocabulary for describing this domain. The vast majority of teachers discussed discrete skills in no particular order, without mentioning their relative importance or their relationships to each other. Third, many of the problem-solving skills teachers were familiar with are direct translations of the portfolio scoring rubrics; the first seven on the list, in fact, are skills addressed by the rubrics either in dimension descriptions or in score-level annotations for dimensions.

Teachers' Assessments of Specific Tasks

Further insight into Vermont fourth grade teachers' understanding of mathematical problem solving can be derived from their assessments of the problem-solving demands of specific tasks. Teachers agreed about the problem-solving skills that would be elicited by a traditional word problem of the type discouraged in Vermont, and they agreed about its strengths and weaknesses. However, there was some disagreement in teachers' evaluations of a richer investigative task involving data representation and analysis. These differences in judgment about the demands of specific problem-solving tasks reflect the variation in Vermont teachers' understanding of mathematical problem solving.

Stickers and Brushes. The first task we asked teachers to evaluate was called *Stickers and Brushes*. (See Figure 3.1.) Teachers were asked to judge a number of different aspects of this task. They were unanimous in their judgment that this relatively simple, traditional word problem would be very easy for their students. Forty-five percent agreed with the following interviewee, "I'd have students who could yell out the answer right away. It's too simplistic. . . . It's an open-and-shut case." Over half (55%) noted that the task relies exclusively on arithmetic computation. One teacher said, "It's not a problem; it's an exercise." Ninety percent of the teachers agreed, as well, that no special preparation would be necessary for their students to respond well to this problem.

> You want to buy a package of stickers for 79 cents and a pair of paintbrushes that cost 29 cents each. You have $1.50. Can you buy them? How do you know?

Figure 3.1—Stickers and Brushes

The most common criticisms of *Stickers and Brushes* were that the problem is too basic or simple, is closed or single-answered rather than open-ended, and does not relate to the scoring criteria. Each of these points was made by 40%–50% of the teachers. Teachers often described the weaknesses of the problem in terms of the scoring criteria.

> The task is very limited. It does not lend itself to the criteria by which students are assessed. There is no way a student could get beyond a 1 or a 2 on each criterion because of the task.

The major strength any teachers saw in the problem was its ease—students could understand the problem and solve it. A number of teachers also liked the phrase, "How do you know?", because it encouraged students to elaborate on their thought processes and extend their discussions.

Raisins. The second task teachers reviewed was called *Raisins*. (See Figure 3.2.) Although most teachers said they believed their students would do well on this richer, more complex, exploratory task, there was a moderate amount of disagreement about its skill requirements and quality.

> No one knows why it happened, but on Tuesday almost all the student in Mr. Bain's class had small boxes of raisins in their lunch. One student asked, "How many raisins do you think are in a box?" Students counted their raisins and found the following numbers:
>
> 30 33 28 34 36 31 30 27 29 32 33 35 33
>
> 30 28 31 32 37 36 29
>
> What is the best answer to the question, "How many raisins are in a box?" Explain why you think this is the best answer.

Figure 3.2—Raisins

Eighty percent of the teachers believed their students would respond well to the task—which indicates a common sense of the difficulty it poses for students. Although teachers were able to describe one or more problem-solving strategies they thought their students would use to solve the task, most of these descriptions did not mention the same skills. (See Table 3.2.) Although almost all the strategies listed in the table could be used to solve the problem, the fact that no one strategy was mentioned by even one-half of the teachers says something about the lack of common terminology for describing problem-solving skills, as well as, something about the level of agreement concerning the skills needed to solve this task. Again, we note that at least two teachers answered in terms of the scoring criteria (using mathematics language) rather than solution strategies.

Table 3.2

Problem-Solving Strategies Evoked by *Raisins*

Problem-Solving Strategies	Percent of Teachers (N=20)
Using manipulatives/raisins	40%
Graphing/charting	35%
Tabulating/listing	35%
Averaging	30%
Counting	30%
Finding the range and frequency of numbers	25%
Guessing and estimating	15%
Using math language	10%
Discussing	10%
Adding and subtracting	10%

NOTE: Table includes responses mentioned by two or more of the respondents. Eight other strategies were mentioned by single respondents.

ASPECTS OF PRACTICE: PROBLEM-SOLVING TASKS

This section examines a more concrete component of the Vermont mathematics portfolio assessment program: the tasks teachers use to elicit student problem solving. In contrast to the difficulty teachers had defining problem solving, they spoke easily and at length about the ideal qualities of problem-solving tasks in the abstract. Further, there was broad agreement among teachers on a common core of desirable problem features, which were consistent with the state's training materials. However, in practice, teachers had difficulty applying their abstract notions of task quality to specific tasks. As earlier noted, there was considerable variability in their judgments of the merits and demerits of typical tasks. Similarly, Vermont teachers made a distinction between tasks that are suitable for assessment purposes and those that are better for instruction, but when shown pairs of tasks, there was only moderate agreement on which tasks fell into each group. Teachers appear to base their day-to-day task selection on the practical demands of instruction as much as on their theoretical notions of task quality. As in previous years, teachers reported having difficulty finding appropriate problems (Koretz, et al. 1994b; Stecher and Hamilton, 1994).

Key Features of Problem-Solving Tasks

Teachers spoke easily and at length about the desirable characteristics of good and bad problems, and the majority agreed on a number of key features. The typical teacher mentioned ten different features, and there were 35 different characteristics mentioned in all. Seven features of good problems were mentioned by more than one-half of the teachers: (Features that relate directly to the scoring criteria are italicized.)

- Relate to math studied in class (95%)
- Admit multiple approaches, multiple solutions (70%)
- *Lead to the use of mathematical representations* (70%)*
- Are open-ended, not overly structured (65%)
- Require critical thinking, reasoning (65%)
- Are at an appropriate level of difficulty (65%)
- *Lead to use of mathematical language* (60%)*

Seven additional features were mentioned by between 35% and 45% of the teachers:

- Relate to other school lessons, subjects or themes (45%)
- *Lead to evidence that students understood the problem* (45%)*
- *Speak to most or all seven scoring criteria* (45%)*
- *Lead to effective presentation of results* (40%)*
- Have meaning for or relevance to students (40%)
- Lead to evidence about the decisions students made while solving the problem* (35%)
- Interest or engage students (35%)

Comparing these lists to the training materials prepared by the Vermont Department of Education suggests that teachers have learned many of the relevant concepts in the abstract, although the following section suggests they do not always apply them in specific situations. Teachers are cognizant of many of the characteristics of good tasks identified by the state. The teachers' descriptions appear somewhat more practical than the formulations in the training notebooks, but all elements of the formal definitions were mentioned more than once during the interviews. Generally speaking, Vermont teachers said that good problems should relate to what is going on in regular mathematics instruction, be of appropriate difficulty, eschew too much structure, admit multiple solutions, and demand critical thinking or reasoning skills. These views largely are consistent with the standards of the NCTM as well. In addition, Vermont teachers said they believe good tasks are rich with respect to the scoring rubrics; that is, good tasks permit students to produce work that will address all seven criteria.

Selection of Problem-Solving Tasks

Teachers' evaluations of the *Raisins* task (described earlier) illustrate the difficulty with which the tenets of task quality provided by the training materials are applied during task selection. There was considerable disagreement among teachers about the quality of the *Raisins* task as a problem-solving activity. Teachers' judgments about the task's features are presented in Table 3.3. The table is arranged to highlight the contradictions between teachers' judgments about features of the task. The table shows the percent of teachers (out of 20) extemporaneously citing a particular trait as a strength or weakness of the *Raisins* task. Columns one and two reflect aspects of the task that were reported in positive terms and the percent of teachers describing the task as such. Columns three and four indicate the percent of the sample attributing the same underlying characteristics to the task, but in negative terms. The table contains only those features mentioned in either positive or negative terms by at least 15% of the respondents.

Table 3.3 suggests that teachers disagree about the characteristics of specific tasks. For example, while a majority of teachers praise the *Raisins* task for being open, some condemn it for being too structured. Similarly, some said it would elicit good math language; others disagreed. Twenty percent of the teachers said *Raisins* is a good task because it is understandable, while 30% said it is a poor task because it is confusing. It would appear that a contemporary version of an old adage applies to problem-solving materials in Vermont: "Good problems are in the eye of the beholder." This lack of common judgment about the characteristics and quality of problem-solving tasks, further illustrates that teachers' practical understanding of problem-solving requirements lags behind their theoretical knowledge.

Table 3.3

Teachers' Evaluations of *Raisins*

Positive Aspect	Percent of Teachers (N=20)	Percent of Teachers (N=20)	Negative Aspect
Is open ended in approach or solution	60%	15%	Is too structured
Is realistic, relevant, engaging, interesting	55%	5%	Is not relevant, personal, or exciting
Elicits good math language	40%	20%	Does not elicit good math language
Calls for estimation, averaging	40%	15%	Focuses too specifically on math operations
Elicits good mathematical representations	35%	10%	Does not elicit good mathematical representations
Encourages manipulatives, is a hands-on activity	30%		
Requires documentation of approach and decisions	30%	10%	Not good for math presentation
Encourages extensions	20%	10%	Hard to generate general rules
Is understandable	20%	30%	Is confusing, hard to get started
Requires problem solving or reasoning skills	15%	5%	Not much problem solving involved

Distinctions between Instructional and Assessment Tasks

The interviews suggested that Vermont teachers assess the merits of tasks differently when they are thinking about instruction than when they are thinking in terms of assessment. However, teachers did not always agree whether a task was better for instruction or for assessment. Agreement rates on four specific tasks ranged from 45% to 75%. It also appeared that teachers had at their disposal a sparser vocabulary for discussing the instructional aspects of problem solving than for describing the scoring aspects.

Fractions tasks. Teachers were asked to compare two pairs of tasks in terms of their instructional merits and their capacity for generating best pieces. The first pair of tasks relates to fractions. (See *Fractions Close to One-Half* and *Building Rectangles* at Figures 3.3 and 3.4 below.)

For each situation, decide whether the best estimate is more or less than 1/2. Record your conclusions and reasoning.

1. When pitching, Joe struck out 7 of 17 batters.
2. Sally made 8 baskets out of 11 free throws.
3. Bill made 5 field goals out of 9 attempts.
4. Maria couldn't collect at 4 of the 35 homes on her paper route.
5. Diane made 8 hits in 15 times at bat.

Make up three situations and exchange papers with a classmate.

Figure 3.3—Fractions Close to One-Half

You need: color tiles, squared paper, markers or crayons.

Use tiles to build a rectangle that is 1/2 red, 1/4 yellow and 1/4 green. Record and label it on squared paper. Find at least one other rectangle that also works. Build and record.

Now use the tiles to build each of the rectangles below. Build and record each in at least two ways.

1/3 green, 2/3 blue
1/6 red, 1/6 green, 1/3 blue, 1/3 yellow
1/2 red, 1/4 green, 1/8 yellow, 1/8 red
1/5 red, 4/5 yellow

Figure 3.4—Building Rectangles

When asked which of these tasks would be better for the purpose of instruction, teachers were in broad agreement. By roughly a three-to-one margin, teachers said *Building Rectangles* was a better instructional activity than *Fractions Close to One-Half*. Many were enthusiastic about the instructional merits of the task because of the use of manipulative aids as learning tools.

> The rectangle task [is better] because of the use of manipulatives. They could see the fractions themselves. Building rectangles is a good introductory or exploratory activity.

Those who thought *Fractions Close to e-Half* was a better instructional piece noted that students "have more experience using halves than other fractions" and might find *Building Rectangles* confusing.

On the other hand, teachers' opinions were evenly divided when asked which task would lead to better scores on the portfolio criteria. Those who thought *Fractions Close to One-Half* would lead to higher-scoring pieces usually noted the opportunities it created to use math language. Other attributes mentioned were the task's relevance to the real world, its utility as a starting point for creating students' own situations, and its usefulness as a foundation for general rules. Teachers who thought *Building Rectangles* was richer from the perspective of the scoring criteria usually cited the opportunity it provides for creating mathematical representations. Teachers also mentioned it's open-endedness, the ease with which students could extend it to other questions, and the fact that it provides a good basis for writing about the approach they took.

Teachers' language about the instructional merits of the fractions tasks was less specific and less extended than their narratives about assessment quality. Teachers on average made two directed statements when describing the instructional value of these tasks, including assertions that they are well suited to the use of manipulatives and that they call for the application of knowledge about fractions. In discussing the tasks' scoring merits, the typical teacher made reference to three of the scoring criteria; some addressed their relation to all seven. Length of discourse (as indicated by the number of lines of response text) similarly suggested greater fluency with the assessment qualities. While 55% of the teachers spoke at equal length about the instructional and assessment merits of the two tasks, the other 45% spoke at greater length about the scoring promise of the tasks than about their instructional utility.

Exploration tasks. Teachers also compared two "explorations"—*Weather* and *What Shows with 100 Throws?*—on instructional value and utility for generating best pieces likely to receive higher scores. (See Figures 3.5 and 3.6 below.)

Find the average high and low temperatures of a U.S. city over a 10 day period.

Figure 3.5—Weather

You need a pair of dice. Roll both dice and add the two numbers. The sums you can get are 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, and 12. Throw the dice 100 times. Keep a chart, and tally the sums each time they appear.

Sums	Tally
2	/
3	/
4	
5	///

and so on

What kind of pattern can you see? Write about it.

Figure 3.6—What Shows with 100 Throws

Teachers were divided in their assessments of the instructional merits of the two tasks. Thirty percent said that *Weather* is a better instructional activity, 45% said that *What Shows with 100 Throws* is better for instruction, and 25% of the teachers either were undecided or said it would depend on their instructional objectives and/or their students' interests.

In contrast, there was a strong preference for *What Shows with 100 Throws* as an assessment task. Most teachers (75%) indicated that *What Shows with 100 Throws* is the task most likely to lead to high-scoring best pieces. The teachers' discourse about this task's suitability for scoring was targeted and specific. One teacher said it "hits a lot of the criteria bullets." Ninety-five percent of the teachers responded to this question about best pieces by referencing the scoring rubrics; the typical teacher discussed the tasks in relation to three or four of the scoring criteria. The two criteria mentioned most frequently were *PS4, So What–Outcomes of Activities* and C2, *Math Representation*.

As with the previous pair of tasks, the teachers' language about the instructional merits of these tasks was less specific and prolific than their language about assessment merits. Sixty-five percent of the teachers described the instructional value of the tasks in relation to specific instructional objectives; of those who used targeted language about instruction, most made two or three specific statements, including assertions that one task or the other reinforces patterning, or provides practice computing averages, or promotes information-gathering skills. This compares to the 95% of teachers who described the scoring characteristics of the tasks using the language of the state rubrics. Examining the length of teachers' discourse (as indicated by number of lines of response text) reveals that more teachers (65%) spoke at greater length about the scoring promise of the tasks than about their instructional utility.

4. DISCUSSION

Teachers indicate that the Vermont portfolio assessment program has enhanced their understanding of mathematical problem solving and broadened their instructional practices in mathematics. However, they have encountered difficulty understanding certain components of the reform and making relevant changes to classroom practice. In this section we discuss the balance between positive changes and teachers' lingering difficulties and reflect on the degree to which these difficulties may have limited the program's ability to improve mathematics teaching. We also comment on the extent to which the lessons learned in Vermont are relevant to assessment reform elsewhere. Three issues are examined: teachers' understanding of mathematical problem solving, changes in instructional practices, and the need for sustained professional development.

TEACHERS' UNDERSTANDING OF KEY CONCEPTS

The portfolio assessment program has introduced new mathematical concepts to Vermont educators in the belief that improved instruction will result from a solid understanding of these concepts and their application to classroom practice. Teachers appear to have learned many of the concepts in the abstract, but they had difficulty applying them to concrete situations. For example, most teachers could define problem solving to a reasonable degree, and almost all could delineate the multiple problem-solving skills they seek to teach. Furthermore, two-thirds or more of the teachers agreed on desirable features of problem-solving tasks in the abstract, and their descriptions correspondea to the Vermont training materials.

However, not all teachers have a complete and clear understanding of mathematical problem solving. Forty percent of interviewees struggled to define problem solving, relying on terminology from the scoring rubrics and on descriptions of specific task characteristics. Furthermore, understanding of problem solving was not widely shared. Teachers' lists of problem-solving skills were dissimilar; only three of the twenty problem-solving skills mentioned by teachers were included on more than one-half of the lists.

There was wide variation in the way teachers' applied their knowledge to concrete situations. Teachers agreed to a much greater degree on the positive features of tasks in the abstract than they did when shown specific tasks. Similarly, although Vermont teachers agreed on the problem-solving demands of a simple task, agreement broke down when more challenging tasks were considered. Teachers also appeared to lack a common vocabulary for talking about the problem-solving demands of specific tasks. This suggests it is not easy to translate theoretical conceptions of important task features into judgments about specific tasks. It may be that different classroom experiences and different student capabilities affect teachers' judgments of task requirements and difficulty. This remains an open question.

Although we did not ask teachers to explain how problem-solving skills were inter-related or to indicate their relative importance, we were surprised that teachers did not appear to have an organizing structure when they talked about problem-solving skills. Teachers described numerous problem-solving skills, but the vast majority of interviewees listed discrete skills in no particular order and made no mention of their relative importance or their inter-relationships. The teachers seemed to lack a useful structure for organizing their knowledge of these skills.

To the extent that teachers lack a structure through which different types of mathematical problems, problem difficulty, and children's problem-solving strategies are related, their progress in implementing the Vermont reforms may be slowed. Research suggests that teachers need a thorough understanding of the topics and issues that define a discipline and a taxonomy to serve as an organizing framework (Shulman, 1986). It is not sufficient for teachers to attend to isolated mathematics concepts and skills, as Vermont teachers appeared to do.

What would such a framework look like? The Quantitative Understanding: Amplifying Student Achievement and Reasoning (QUASAR) project of the Learning Research and Development Center offers teachers a taxonomy of the cognitive processes that underlie problem solving in mathematics. These processes include understanding a mathematical problem, discerning mathematical relationships, organizing information, using mathematical strategies, formulating conjectures, evaluating the reasonableness of answers, generalizing results, justifying answers or procedures, and communicating mathematical ideas (Lane, et al., 1993). Vermont teachers might use this type of framework to organize the discrete, concrete strategies they talk about—e.g. identifying relevant information, making lists, and working backwards—under larger meaningful units. Such a framework might provide a better way for them to analyze, retain, and recall these important constructs.

Vermont educators are not unaware of the value of such structure. The Department of Education earlier attempted to categorize problem-solving tasks into three types: puzzles, applications and explorations. However, this system could not be applied consistently to problems, and nothing has emerged to take its place. We recommend that the state work with teachers to develop a framework that relates mathematical problem solving to problem types and problem-solving strategies.

CHANGES IN TEACHING PRACTICES

Teachers reported substantial changes in their mathematics curricula and instruction, but their comments raised questions about the understandings on which these changes are based and the support they received to implement classroom reforms. After discussing the nature of the changes in teaching practice, we explore the role of the scoring rubrics in shaping changes and the adequacy of the support provided to teachers.

All of the teachers report changing their curriculum in the direction encouraged by the portfolios, but they have not all moved in the same way or at the same pace. In all cases, teachers said they place far more emphasis on problem-solving skills than they did prior to the portfolios. However, teachers differ at the level of curriculum specifics. For example, they do not emphasize the same problem-solving skills. Overall, the portfolio assessment seems to have pushed teachers in a common direction with respect to curriculum, but they have varied along this path.

Rubric-Driven Instruction

We are concerned that gaps in teachers' understanding of problem solving increase their reliance on the scoring rubrics, and this emphasis may have undesirable consequences. The problem arises in part because teachers have been asked to implement a new problem-solving curriculum with somewhat limited assistance and support. Their task has been complicated by the fact that many lack a firm understanding of problem solving and of problem-solving pedagogy. Furthermore, they realize that, in the long run, high stakes may be attached to school-level portfolio scores. The scoring rubrics contain concrete operational definitions of the aspects of problem solving that should be encouraged and of the student behaviors that will be rewarded. Consequently, they are attractive targets for instruction, and almost all teachers indicated that the scoring rubrics played a prominent role in shaping their instructional practices.

For example, there was evidence that the Vermont scoring rubrics affect which problem-solving skills are taught. Almost all teachers described ways in which the rubrics affect their choice of problem-solving skills. Many of the procedures being taught as problem-solving strategies are direct translations of the portfolio scoring rubrics. In fact, the seven most frequently cited skills are strategies addressed by the scoring rubrics, either in the dimension descriptions or in the score-level annotations for dimensions.

This reliance on the scoring rubrics for curricular and instructional guidance has both positive and negative consequences. On the positive side, the rubrics represent teachers' judgments about the observable and important aspects of students' problem solving. Much time and effort went into their creation, and they embody some of the elements Vermont teachers believe to be important components of problem solving. In this way they are "vehicles of instructional clarification" (Popham, 1987) and are helpful to teachers. To the extent that the scoring guide captures the most important and essential elements of problem solving, it is good that teachers focus on these central concepts.

On the negative side, focusing on the rubrics may have undesireable consequences similar to those observed when teachers focus on multiple-choice tests (Shepard and Dougherty, 1991). These negative consequences include increased instructional time and emphasis given to tested knowledge/skills over non-tested content, and extensive classroom time devoted to test preparation. Both concerns are relevant for portfolios, although in slightly different guises. Inappropriate instructional emphases could occur if teachers favor some aspects of problem solving over others out of proportion to their relative importance. Narrowly focused test preparation may occur if teachers emphasize some problem types or response formats over others because they fit the rubrics, or if they discard otherwise appropriate problems that only permit high scores on four or five of the scoring criteria. To the extent the rubrics over-simplify problem solving and fail to represent useful problem-solving skills, teachers may do students a disservice by over emphasizing the rubrics in curricular and instructional planning.

Sustaining Teacher Professional Development

Flexer et al., (1994) note that fundamental changes in content and pedagogy require that teachers have access to practice-oriented professional development materials, instructional resources, and ad hoc support. This continues to be a need in Vermont. Teachers turn to supplemental text materials and assistance from network leaders and other colleagues to fill gaps in their understanding and to obtain instructional materials.

Teachers author few of their own tasks. Only 15% of the tasks teachers used in 1993–94 were developed by the teachers themselves. Furthermore, most find that existing curriculum materials do not provide good problem-solving tasks. Under these circumstances, they turn first to the state training materials as a source of portfolio tasks. Next in popularity are supplemental mathematics books such as the *Problem Solver* series. In fact, for about 20% of the teachers interviewed, supplemental books are becoming *de facto* curricula in problem solving, without formal review or adoption.

Many of the teachers also told us they continue to need ongoing support from other professionals. They continue to rely on other teachers to help them with problem-solving curricula and instruction. It seems clear to us that they will need sustained professional development and support to be able to "teach mathematics that they never learned, in ways that they never experienced" (Cohen and Ball, 1990).

CONCLUSIONS

The key elements of the Vermont portfolio assessment program—the dual goals of instructional improvement and accountability, and assessment embedded in instruction—are present in testing reform efforts in other states, so the results of this study of Vermont teachers should be relevant to educators elsewhere. Although we devoted more space to discussing negative findings, we should reiterate that the portfolio assessment had strong positive effects on teachers. Teachers reported that they learned a great deal about mathematical problem solving and problem-solving pedagogy. They also changed their curricular and instructional practices to try to promote problem solving and mathematical communication. Moreover, teachers remained enthusiastic about the reform, despite the demands it placed on their classroom time and their personal time.

However, there are still important gaps in teachers' understanding and in the support they receive to implement the reform that should be addressed in Vermont. For example, teachers do not share a common understanding of mathematical problem solving—a key construct of the reform—nor do they agree on the essential problem-solving skills students should master. As a consequence, teachers have focused on the scoring rubrics for practical guidance. However, such rubric-driven instruction may lead to fragmentation and narrowing of the curriculum. Instead, teachers should receive additional professional development that elaborates and expands their disciplinary and practice knowledge. They also need materials to guide pedagogy and classroom activities.

REFERENCES

Cohen, D. and Ball, D. (1990). Policy and practice: An overview. *Educational Evaluation and Policy Analysis*, (12), 3, 347–353.

Flexer, R., Cumbo, K., Borko, H., Mayfield, V., and Marion, S. (1994). How "messing about" with performance assessment in mathematics affects what happens in classrooms. Presented at the Annual Meeting of the American Educational Research Association and the National Council on Measurement in Education, New Orleans, LA.

Koretz, D., Klein, S., McCaffrey, D., and Stecher, B. (1993a). *Interim report: The reliability of Vermont portfolio scores in the 1992–93 school year* (CSE Technical Report 370). Los Angeles, CA: CRESST/UCLA. (Available from RAND as Reprint No. RP–260.)

Koretz, D., Stecher, B., Klein, S., McCaffrey, D., and Deibert, E. (1993b). *Can portfolios assess student performance and influence instruction? The 1991–92 Vermont experience* (CSE Technical Report 371) Los Angeles, CA: CRESST/UCLA. (Available from RAND as Reprint No. RP–259).

Koretz, D., Stecher, B., Klein, S., and McCaffrey, D. (1994a). *The evolution of a portfolio program: The impact and quality of the Vermont program in its second year (1992–93).* (CSE Technical Report 385). Los Angeles: UCLA Center for the Study of Evaluation.

Koretz, D., Stecher, B., Klein, S., and McCaffrey, D. (1994b). The Vermont portfolio assessment program: Findings and implications. *Educational Measurement: Issues and Practice*, 13, (3), 5–16.

Lane, S., Lio, M., Stone, C., and Ankenmann, R. (1993). Validity evidence for QUASAR'S mathematics performance assessment. Paper presented in the symposium "Assessing Performance Assessments: Do They Withstand Empirical Scrutiny?" at the Annual Meeting of the American Educational Research Association, Atlanta, GA.

Polya, G. (1980). On solving mathematical problems in high school. In *Problem Solving in School Mathematics*, 1980 Yearbook of the National Council of Teachers of Mathematics, edited by Stephen Krulik, pp. 1–2. Reston, VA: NCTM.

National Council of Teachers of Mathematics. (1989). *Curriculum and Evaluation Standards for School Mathematics*. Reston, VA: NCTM.

Popham, W. (1987). The merits of measurement-driven instruction. *Phi Delta Kappan*, (68), 679–682.

Shepard, L. and Dougherty, K. (1991). Effects of high stakes testing on instruction. Paper presented at the Annual Meeting of the American Educational Research Association and the National Council on Measurement in Education, Chicago, IL.

Shulman, L. (1986). Those who understand: Knowledge growth in teaching. *Educational Researcher*, 15, (2), 4–14.

Stecher, B. and Hamilton, E. (1994). Portfolio assessment in Vermont, 1992–93: The teachers' perspective on implementation and impact. Paper presented at the Annual Meeting of the National Council on Measurement in Education, New Orleans, LA.