Portfolio-Driven Reform: Vermont Teachers' Understanding of Mathematical Problem Solving and Related Changes in Classroom Practice

CSE Technical Report 400

Brian M. Stecher and Karen J. Mitchell
RAND Institute on Education and Training

April 1995

National Center for Research on Evaluation,
Standards, and Student Testing (CRESST)
Graduate School of Education & Information Studies
University of California, Los Angeles
Los Angeles, CA 90095-1522
(310) 206-1532

The work reported herein was supported under the Educational Research and Development Center Program cooperative agreement R117G10027 and CFDA catalog number 84.117G as administered by the Office of Educational Research and Improvement, U.S. Department of Education.

The findings and opinions expressed in this report do not reflect the position or policies of the Office of Educational Research and Improvement or the U.S. Department of Education.

PREFACE

This report is intended for practitioners, researchers, and policy makers concerned about current performance assessment efforts and their effects on instructional quality. It describes the impact of Vermont's statewide portfolio assessment initiative on teachers' understanding of mathematical problem solving and their instructional practices. The report also discusses the implications of these findings for the validity of portfolio scores. The conclusions should be of interest to educational policy makers responsible for assessment reforms in other jurisdictions.

The project was conducted by RAND under the auspices of the Center for Research on Evaluation, Standards, and Student Testing (CRESST). This report supplements findings from previous evaluations of the Vermont portfolio assessment program conducted by the RAND study team.

CONTENTS

LIST OF FIGURES AND TABLES

SUMMARY

The Vermont portfolio assessment program has had substantial positive effects on fourth-grade teachers' perceptions and practices in mathematics. Vermont teachers report that the program has taught them a great deal about mathematical problem solving and that they have changed their instructional practices in important ways. Many can define mathematical problem solving and describe the problem-solving skills they seek to teach. Vermont teachers say the program has taught them that mathematics is more than computation, that there are many everyday applications of mathematics, and that mathematical communication is valuable. They have incorporated problem solving into their curriculum, they routinely assign problem-solving tasks, and they regularly teach problem-solving skills. Vermont teachers appear to have learned much of what has been communicated via the state training materials and network meetings.

However, teachers have had difficulty understanding certain aspects of the reform and making appropriate changes in classroom practice. Vermont teachers do not share a common understanding of mathematical problem solving. Further, though they readily list the problem-solving strategies they strive to teach, they do not agree about which are the most essential skills. Neither do they agree which skills are required by particular tasks. Teachers do not have a rich vocabulary with respect to problem solving, although their discussions of the assessment aspects of problem solving are less spare than their discussions of the instructional aspects. Indeed, teachers often use the language of the state scoring rubrics to describe problem solving, problem-solving strategies, and the characteristics of good problems.

Teachers also focus on the scoring rubrics for practical guidance in understanding the problem-solvingdomain, structuring instruction, and guiding student efforts. Such "rubric-driven instruction" has strengths and weaknesses. On the positive side, the rubrics help teachers focus on some of the important and observable aspects of students' problem solving; for teachers with tentative conceptions of mathematical problem solving, they provide needed instructional clarification. On the negative side, emphasis on the scoring rubrics may cause teachers to neglect important problem-solving skills not addressed by the scoring rubrics and reject tasks not well aligned to the criteria. As one teacher told us, "What's in the rubrics gets done, and what isn't doesn't."

Differences in understanding lead to differences in practice—which ultimately affect the meaning of portfolio scores. Some Vermont teachers "preteach" portfolio tasks by assigning similar, simpler problems prior to student work on portfolio pieces so that assessment problems are not overly novel or difficult for their students. In addition, most teachers provide differential help—scribing, reading, and providing manipulative aids to students who need

assistance. The differential conditions under which students prepare for and complete their pieces threaten the validity of portfolio scores for comparisons of students, classrooms, or schools.

Ultimately, these results speak to the fundamental conflict between good instruction and good accountability-oriented assessment.[1] While good instruction should be responsive to the individual needs and capabilities of students, good assessment for accountability purposes should yield data that are comparable across units of interest. When instruction and assessment are intertwined—as they are with portfolios and other forms of embedded assessment—these two principles are in conflict. At present, Vermont teachers appear to place greater value on instruction than accountability assessment, and Vermont policy makers seem to place greater value on local flexibility than on comparability. The resulting data may be acceptable for the current purposes of the Vermont assessment; however, stakeholders in high-stakes testing programs for which similar reforms are proposed should understand the need for greater controls if they want directly comparable scores. This study suggests that, without common understandings among teachers and adequate controls on student preparation and administrative conditions, scores from nonstandardized, instruction-embedded assessments may not support proposed uses involving comparisons or standards applied to students, classrooms, schools or systems.

Finally, we note that Vermont teachers appear to need continuing professional development that elaborates and expands their disciplinary and practical knowledge. They also need additional materials to guide pedagogy and classroom activities. The Vermont Department of Education has offered training each year since the initiation of the portfolio assessments, and they have established regional networks to address teachers' staff development needs. These efforts should continue, and the Department should look for ways to help teachers enhance their basic understanding of mathematical problem solving and related pedagogy.

[1] This conflict is much less severe in the case of assessments used for instructional planning and other classroom-level decision making.

ACKNOWLEDGMENTS

This study could not have been completed without the cooperation and assistance of Vermont educators, including Marge Petit, Jill Rosenbaum, and Sue Rigney who reviewed the interview protocols and offered helpful suggestions. Appreciation also is due to twenty Vermont fourth-grade teachers, who responded candidly and thoughtfully to the surveys and interviews. In addition, we received valuable assistance from RAND colleagues Devah Pager, who helped with data collection and coding, and Daniel Koretz and Sheila Barron, who helped frame the study design and offered valuable comments on an earlier draft of this report.

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 provide meaningful accountability information at the school, district, and Supervisory Union levels, and toserve as an impetus for curriculum reform. 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, Klein, McCaffrey, & Stecher, 1993; Koretz, Stecher, Klein, & McCaffrey, 1994a, 1994b; Koretz, Stecher, Klein, McCaffrey, & Deibert, 1993; Stecher & 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.

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 4-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. (The Vermont mathematics scoring rubric appears in Appendix A.) 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, in terms of both 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 teachers' 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 selection and evaluation of 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 10 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, 2 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 5 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 10 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 Figure 2.2 shows the distribution of portfolio experience.

Though the averageexperience of teachers in this sample is slightly lower than that reported in a statewide survey of fourth-grade teachers conducted previously (Stecher & Hamilton, 1994), we consider the study sample to be reasonably representative of fourth-grade teachers in Vermont. All of the

[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.

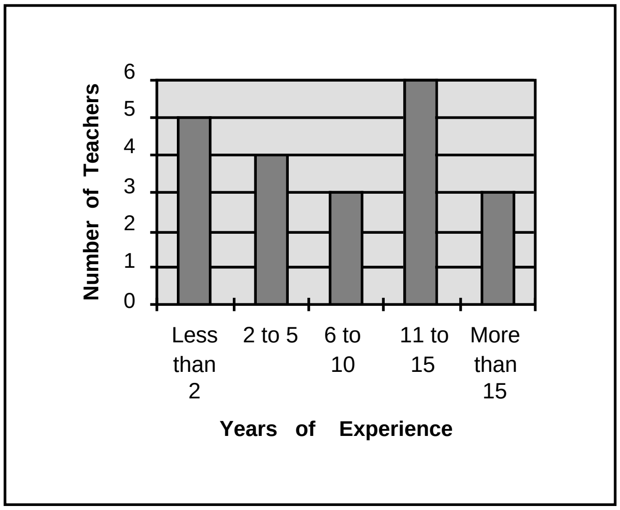

Figure 2.1. Teaching experience.

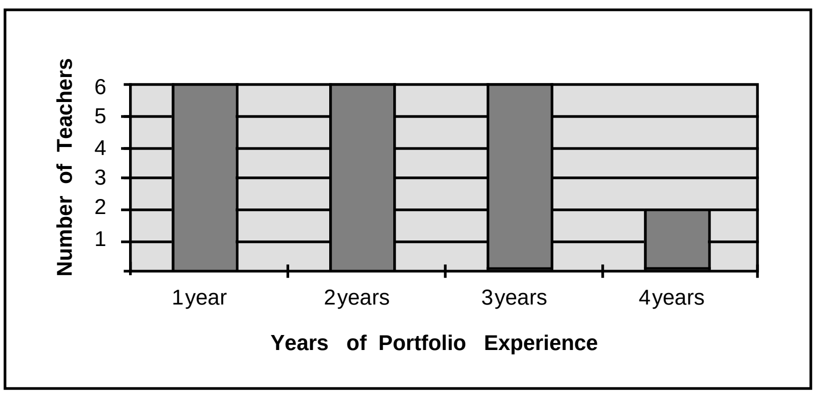

Figure 2.2. Portfolio experience.

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] (See Appendix B.) 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.

Data Analysis

With the teachers' permission, the interviews were recorded on audio tape, and the recordings were used to ensure 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

[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.

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

Combined survey and interview results are presented in the following three sections. The initial section deals with teachers' understanding of mathematical problem solving and the knowledge they gained from participating in the portfolio assessment. The latter two sections explore how this knowledge was translated into practice. We focus first on problem-solving tasks, then we examine instructional practices related to problem solving.

Two general comments are appropriate prior to the presentation of specific findings. 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.

Finally, 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—for instance, their relevance to real life, their open-endedness, and their promotion of communication. For example,

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

> 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.

> 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* (National Council of Teachers of Mathematics, 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.

Delineations 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 10 different problem-solving strategies they convey to students. The average number of problem-solving skills listed by respondents was 5, 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.

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

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%

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

problem-solving demands of tasks or the problem-solving approaches of students. Second, teachers did not 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. (See Appendix A.)

Teachers' Assessment 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.

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.

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.

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.

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.

No one knows why it happened, but on Tuesday almost all the students 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.

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 & 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 10 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. (See Appendix C for relevant excerpts from the Vermont training materials.) 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

Practical considerations seem to play a greater role in task selection than the theoretical notions of quality described above. Seventy percent of the teachers mentioned only operational reasons for picking their most recently assigned task: using it to introduce a new math topic, to prepare students for an upcoming test, to promote the use of manipulatives, to speak to other school and class themes, to try out cooperative group problem solving, and so on. The most commonly noted reason was that the task was related to the current math lesson (40%); one-quarter of the teachers said that tasks were selected because they were relevant to students' lives or they were related to other class lessons or themes.

Teachers also indicated that the scoring criteria exert a strong influence on task selection. Seventy-five percent of teachers said they do not assign tasks unless they are likely to address most or all of the seven criteria. Teachers said that they look for tasks that require students to make decisions along the way, to use math language and to devise math representations—key features of the scoring guides. One teacher described the impact of the criteria in this way:

> Regardless of what you call this, kids and parents and administrators put a grade on this stuff. What's in the rubrics gets done, and what isn't doesn't.

Teachers also suggested that some useful mathematics problems are rejected because they would not lead to high-scoring responses on all criteria.

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 1 and 2 reflect aspects of the task that were reported in positive terms and the percent of teachers describing the task as such. Columns 3 and 4 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

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

and quality of problem-solving tasks further illustrates that teachers' practical understanding of problem-solving requirements lags behind their theoretical knowledge.

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* in Figures 3.3 and 3.4.)

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 One-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 its 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.)

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*.

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.

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.

Sources of Problem-Solving Tasks

Tasks distributed by the Vermont Department of Education and network resources were the most common sources of tasks for more than one-half of teachers (60%). In addition to the network materials, teachers relied on two supplemental resources: The *Problem Solver* series (Goodnow & Hoogeboom, 1987) was mentioned by 35% of the teachers, and *About Teaching Mathematics* (Burns, 1992) and other books by the same author were discussed by 30% of the teachers. Several other publications were cited by individual teachers. Teachers extemporaneously voiced a need for additional tasks, portfolio resources, and support. Statements like the following were common:

> Teachers don't have the time or expertise to make up good tasks. We need benchmarks; we need resources.

> I feel like I'm hurting for tasks sometimes; I want a big book that I can use as a resource.

> No one is out there showing us great tasks and answering questions when we get stuck.

Finally, teachers said they tend to use tasks developed by others rather than authoring their own or adapting problems obtained from other sources. On average, about one-half of the tasks (55%) used during the academic year were taken intact from books, workshop and other print materials, and colleagues' materials—although the percent of tasks used "as is" varied widely across teachers (from 10% to 100%). In comparison, an average of 30% of the tasks came initially from other sources but were adapted in some way by teachers (the range was from 0% to 80%), and about 15% of tasks were authored by the teachers themselves (the range was from 0% to 60%). Teachers also indicated that they built upon their experience from prior years by re-using tasks that were successful in the past. Forty percent of tasks used by respondents in their second or subsequent years of teaching were used in previous years.

Aspects of Practice: Problem-Solving Instruction

Teachers said the portfolio assessment led them to change the way they teach mathematics, and they voiced considerable enthusiasm for these changes. Thirty-five percent of the teachers punctuated their statements about the impact of the program on instructional practice with effusive statements such as "It's totally changed the way I teach" and "I've learned so much." This study provides information about three aspects of instruction: teaching methods that stress student communication and construction, teachers' attempts to adjust instruction so that assessment problems are not overly novel or challenging, and the provision of differential support to assure students do their best work.

Methods for Problem-Solving Instruction

The portfolio assessment program prompted teachers to include problem-solving instruction as an integral, routine part of their instructional program. Teachers said that problem-solving instruction is no longer "something on the side." They gave math problem solving more time and more emphasis than they did before the portfolio assessment, and they incorporated it into varied classroom activities, including those in other disciplines. (Science and social studies were mentioned most frequently.)

Teachers used a number of different strategies for teaching problem-solving skills. Most relied on frequent and, they hoped, meaningful classroom discussion. About one-half of the teachers had students pose alternate strategies for the solution of a particular problem and then collectively explored the effectiveness of each. Teachers stressed the value of having "kids listen to others and realize that there is no one right way to solve a problem." This student-centered approach was described as follows:

> [I] give them a problem and ask how they would handle it. They may think about drawing. I'll divide them into groups to try different strategies. Then they discuss and compare how they got their answers, if they were the same, and which seemed to be more efficient.

> I have students solve problems and then we look at the different ways that they have solved it. We are approaching problem solving, but I am not teaching it. Given the opportunity, these 26 kids will come up with it.

About one-quarter of the teachers said they model desired skills for their students and use guided exploration to prompt student attempts:

> I work with them to identify the problem by going through the problem verbally with them . . . To teach what you need to solve a problem, we go through and underline parts of the problem and then discuss why we need and don't need things. We try it and talk about how we used the skills to solve it. They do one on their own and talk about how they did it.

> We do a think-aloud where we read the problem and make sure we know all the words and what they mean, and we see whether we need to get anything into the classroom to solve the problem. I graphically organize it for them using an overhead. There is a lot of discussion and modeling.

Another popular approach that teachers described (40%) is focusing on one strategy at a time over the course of the year and presenting problems that illustrate each approach. Teachers said,

> I use the *Problem Solver* books from Creative Publications. They have nicely outlined problems that teach the youngsters how to do the strategies. I build from there.

> They are taught as strategies. I do problem-solving tasks once a week; the textbook skills and use of manipulatives take place the other four days . . . I use Creative Publication's *Problem Solver* book, Marilyn Burns' book, and the Heath Math Series to teach individual skills.

One-quarter of the teachers mentioned that they let problem solving carry over into other fields and situations.

> I see a big correlation between this and the scientific process of problem solving. . . . One of the things that we have been looking at as far as how to solve a problem is—once we understand what question is being asked, what can we do to come up with a solution. What resources can we go to? What can we do besides using our own

heads to get at answers? How can we compile information for anyone that wants to look at our stuff? . . . This applies to anything—to science and any problem.

I try to find a problem-solving unit for each Language Arts, Science and Social Studies unit I do. If we are doing myths in Language Arts, for example, I try to come up with a math activity to go along with it.

Finally, two other themes were prominent in teachers' descriptions of their instructional approach: the importance of challenge in mathematical problem solving, and the benefits of encouraging students to think about and discuss their attempts. The former is illustrated by the following comment:

When kids didn't get it before, we rushed right in there because we thought we were helping them. . . . Now we give them the luxury of thinking about it, a long time. . . . Pondering and thinking is learning. It is almost sadistic, but I feel good when I see them struggling.

The latter position was described by another teacher:

Kids like to say, "I just got the answer." It happens in their head and they don't know how it got there. The biggest piece is that they're [Vermont] getting kids to think about the thinking they do. They say, "First of all, I did this and then I did that, and, oh yeah, in between I did this." . . . In my classroom . . . we make visible what is invisible.

Preteaching to Assessment Tasks

Some Vermont teachers structured instruction so that assessment problems were not overly novel or difficult for their students. One-quarter of the teachers said they would prepare students in advance for a particular portfolio task by "preteaching" to it. These interviewees described a process in which they precede portfolio tasks with similar, but simpler, problems so that the assessment tasks do not present too great a challenge. One teacher described this approach in relation to the *Raisins* task, saying,

Sometimes before a piece, I may do some sort of warm-up. If I was going to give this on Friday, maybe on Monday we might generate a little class data—how long is everyone's pencil or something like that—and recall how we could find out what's the most common length. We might do a small scale something like it, a similar thing on a much smaller scale. If we hadn't done anything like it, I probably wouldn't give this problem.

Another teacher said,

> Originally, I would read problems with kids and stress a couple of points . . . but that wasn't enough and I didn't get the results that I needed. Now, I choose a task and I reword it, change the numbers and do one that is similar, and then I hand [them] the problem with different figures and different wording and have them do it.

Differential Assignment of Tasks and Support for Instruction

Most teachers reported that they assign the same portfolio task to all students in class; however, they often vary the kind of support they provide to accommodate the needs of individual students. For special education students, 75% of teachers provided individualized support based on the learning needs of students. Such support included adapting manipulatives for students with motor difficulties, providing scribes for students who cannot write, and reading tasks to those with severe reading difficulties.

More importantly, 70% of teachers said they individualize their interactions with any students who may need help to understand and to perform their best on the portfolio tasks. Sixty-five percent of the teachers said they record information for students with writing difficulties, one-quarter said they give additional attention to students who need it, and 20% mentioned offering manipulative aids to students who would benefit from them.

There are other variations in teachers' descriptions of their task assignment policies. A few teachers (15%) give different tasks to groups of students based on math or reading ability. One teacher said she always gives students a choice from a set of tasks, and another varies the procedure during the year—sometimes giving the same problem to all and sometimes permitting students to choose. Some teachers reported experimenting with group work as well, permitting students to collaborate on tasks.

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 and produce meaningful assessment data for accountability purposes. 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 implications of these findings for score validity.

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 corresponded 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 notwidely shared. Teachers' lists of problem-solving skills were dissimilar; only three of the 20 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 interrelated 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 interrelationships. 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, Lio, Stone, & Ankenmann, 1993). Vermont teachers might use this type of framework to organize the discrete, concrete strategies they talk about—for example, 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. Similarly, they have students spend much more time working on problem-solving tasks. However, teachers differ at the level of curriculum specifics. They do not emphasize the same problem-solving skills, and they do not appear to select the same tasks for students to solve.[7] Overall, the portfolio assessment seems to have pushed teachers in a common direction with respect to curriculum, but they have varied along this path.

Similarly, most of the reported changes in teaching methods are consistent with the goals of the portfolio assessment, but there are differences in the approaches adopted by Vermont teachers. Respondents described a few methods for teaching mathematical problem solving, ranging from a student-centered

[7] A previous study that examined the contents of Vermont portfolios reached the same conclusion (Stecher & Hamilton, 1994).

approach that calls for students to nominate and collectively evaluate alternate problem-solving strategies to a more teacher-centered sequential presentation of discrete strategies. Most of the practices teachers described emphasize student construction and communication, which are consistent with the themes of the portfolio program. The more didactic presentation of discrete skills appears to be slightly at odds with the intent of the reform, but we do not have enough information about the nature of student participation in these classes to judge its appropriateness. Again, most instructional changes appear to have a common theme, but differ in the specifics.

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. (See Appendix A.) 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.

In addition, the rubrics affect task selection. Vermont teachers said tasks are desirable if they are rich with respect to the scoring rubrics; that is, they permit students to produce work that can be scored on all seven criteria.

Teachers also said they reject otherwise useful tasks that cannot be scored on all criteria.

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 undesirable consequences similar to those observed when teachers focus on multiple-choice tests (Shepard & Dougherty, 1991). These negative consequences include increased instructional time and emphasis given to tested knowledge/skills over nontested 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 is less of a problem with portfolios than in the multiple-choice context, where the test relies on a specialized, abbreviated format that is different from normal instructional presentations. However, some aspects of this narrow test preparation phenomenon are relevant to portfolios. Teachers may emphasize some problem types or response formats over others because they fit the rubrics, or they may discard otherwise appropriate problems that only permit high scores on four or five of the scoring criteria. To the extent the rubrics oversimplify problem solving and fail to represent useful problem-solving skills, teachers may do students a disservice by overemphasizing the rubrics in curricular and instructional planning.

Although we do not have much information about the extent of such undesirable consequence in Vermont, there is some evidence that the rubrics may have driven instruction in inappropriate directions. A few teachers believe that the scoring rubrics ask students to respond in unnatural or inappropriate ways. During unstructured conversation, three teachers (15%) raised concern about the effects of the scoring criteria on students' efforts. For example, *PS4*, the *So*

What—Outcomes of Activities criterion, asks students to extend their solutions to more complicated situations. A couple of teachers described this as an unnatural and developmentally inappropriate activity. One teacher said, "When you solve a problem, you don't say, 'Well how can I apply this to other things in my life?' " Similarly, the *PS1* criterion, *Understanding the Problem*, was described as contrived: It asks students to identify special factors that would influence the student's approach before starting the problem.

One Vermont teacher told us, "What's in the rubrics gets done, and what isn't doesn't." This is cause for concern, and the Vermont Department of Education should continue to monitor the possible effects of rubric-driven instruction as the portfolio assessment program matures. We think it wise to be cautious at this stage of the portfolio assessment program and to be alert to any potential narrowing of the problem-solving domain.

Sustaining Teacher Professional Development

Flexer, Cumbo, Borko, Mayfield, and Marion (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 & Ball, 1990).

Conditions Affecting Score Validity

Lack of a common understanding of problem solving and problem-solving pedagogy contributed to variations in practice—including task evaluation, task selection, and skills instruction—which affect the meaning of scores assigned to individual work and the meaning of comparisons between classrooms. We want to discuss two of these differences in practice because they provide an opportunity to examine the question of validity in the context of an operational portfolio assessment program and because they point to a fundamental conflict when this type of assessment is used for accountability purposes.

Two practices that threaten score validity at the individual and classroom levels are the provision of individualized assistance on problem-solving tasks and an instructional practice referred to as "preteaching."[8] Ironically, both are completely justified as instructional activities, but they alter the meaning of pieces as assessment products. They are analogous to "score pollutants" arising from differential test administration activities or conditions, which logically compromise score comparisons across students, schools and systems (Haladyna, Nolen, & Haas, 1991; Messick, 1984).

One of the undesirable consequences of high-stakes testing is the provision of differential assistance during assessment to students who teachers believe need extra help (Shepard & Dougherty, 1991). Similar differential assistance was widely reported in Vermont. Seventy percent of Vermont teachers said they provide individual assistance—including scribing, reading, and providing manipulative aids—to help students do their best work. Additionally, a minority of teachers said they assign different problems to students of differing ability levels, further complicating score interpretation. Although the cause was somewhat different than that seen in the high-stakes testing programs—Vermont teachers believe it is appropriate to offer assistance to students based on their individual needs because portfolios are embedded in the instructional program—the results are just as troubling from the perspective of assessment. Individualized assistance changes the meaning of a student's performance. The product no longer represents the student's independent response to the task, and the score cannot be interpreted as an independent and comparable indication of the

[8] We do not know whether the errors in student and classroom scores introduced by these instructional practices are greater or lesser than the errors introduced by differences in the selection of problem-solving tasks or differences in procedures for compiling portfolios.

student's individual capability. With personalized assistance one can never know "Whose work is it?" (Gearhart, Herman, Baker, & Whittaker, 1993). Such teacher intervention confounds comparisons between students or groups of students. In addition, local or state pressure to raise scores may lead to more help and to further decrements in score interpretability.

A substantial minority of Vermont teachers reported that they "preteach" to a task by assigning similar, but simpler, problems before students are given the target task. The purpose of preteaching is to ensure that the target task is not too novel, but represents just the right level of challenge for the students. Teachers understand that if there is nothing novel about the task, then it is not a problem but an exercise. However, if it is too novel, students will perform poorly and not produce pieces that will score well on all criteria. To overcome this dilemma, preliminary tasks are administered to assure that students have access to the knowledge and skills addressed by the upcoming assessment task.[9] However, this practice clouds the meaning of scores assigned to students' work. Without knowing exactly what experience preceded a best piece, a reader cannot know how difficult it was for the student or what level of problem-solving skill was demonstrated. In fact, with too much preteaching on similar problems, a task may no longer contain any novel elements; rather than posing a problem for students, it becomes merely a routine exercise. Different pre-assessment practices among teachers probably result in inappropriate comparisons among portfolio scores—at the class, school, district, and Supervisory Union levels.

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

[9] The New Standards Project provides pre-assessment activities with a similar intent.

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.

One important principle that emerges from these findings, although it was not the primary focus of this study, is the fundamental conflict between good instruction and good assessment for accountability purposes. Most educators would agree that good instruction should be responsive to the individual needs and capabilities of students. Good accountability assessment, by contrast, should provide comparable, interpretable data. When assessment and instruction are intertwined—as they are with portfolios and other forms of embedded assessment—these two principles are in conflict. To the extent that teachers individualize their interaction with students during the preparation of portfolio pieces, the scores assigned to these pieces will not reflect the independent capabilities of the students. Similarly, if students include different pieces in their final portfolios and if teachers from different schools assign different tasks, then neither individual student scores nor classroom aggregate scores will be directly comparable. We do not know the relative size of these sources of error, only that they are additive. As Messick wrote in 1975:

> To judge the value of an outcome or end, one should understand the nature of the processes or means that led to that end, as Dewey emphasized in his principle of the means-end continuum: it's not just that the means are appraised in terms of the ends they lead to, but the ends are appraised in terms of the means that produce them.

At present Vermont teachers appear to place greater value on instruction than on assessment, and Vermont policy makers seem to place greater value on local flexibility than on comparability. Under these circumstances, the scores assigned to Vermont portfolios may be acceptable for the purposes of the Vermont assessment. However, similar data are less likely to be acceptable in the contexts in which current assessment reforms are being proposed. This study suggests that scores from nonstandardized, embedded assessments may not support proposed uses involving comparisons or standards applied to students, classrooms, schools and systems.

APPENDIX A: VERMONT MATHEMATICS SCORING RUBRIC

MATHEMATICS PROBLEM SOLVING CRITERIA

Vermont Portfolio Scoring Guide: Mathematics Rev. 2/5/93	Level 1	Level 2	Level 3	Level 4
PS1 Understanding the Problem	*...didn't understand enough to get started or make progress.*	*...understood enough to solve part of the problem or to get part of a solution.*	*...understood the problem.*	*...identified special factors that influenced the approach **before** starting the problem.*
	Part of a Problem: Some problems are multi-step. If all the parts of the problem are not addressed, then all of the problem was not understood. **Solution**: A solution includes all of the work that was done to complete the problem, an explanation of the decisions made along the way, and an answer. **Special Factors**: Special factors are factors that might affect the outcome of the problem. **Special Considerations**: A Level 4 for this criterion is dependent on identifying factors to be considered **before** starting the problem.			
PS2 How You Solved the Problem	*...approach didn't work.*	*...approach would only lead to solving part of the problem.*	*...approach would work for the problem.*	*...approach was efficient or sophisticated.*
	Approach: The strategy or skill used to solve the problem. **Would**: An approach that would work for a problem even if computation errors or an incomplete response prevented a solution is credited as a Level 3. **Efficient**: Efficiency is determined by the directness of the approach. Use of an algorithm to solve a problem suggests this was just an application of knowledge, not a real problem. If finding the least common multiple, an approach which lists all multiples of a number as compared to the use of prime factorization is not as efficient. The second case might be sophisticated because of its efficiency. **Sophisticated**: A sophisticated approach is one that is not common for students this age. **Special Considerations**: A piece scored at a Level 1 or 2 on PS1 can not score more than a Level 2 on PS2.			
PS3 Why - Decisions Along the Way	*...had no reasons for the decisions made.*	*...may have used reasoning, but it is hard to see from the work.*	*...didn't clearly explain the reasons for decisions, but work suggests reasoning was used.*	*...clearly explained the reasons for the decisions made along the way.*
	Suggests: Work suggests there is reasoning if: there is a change in approach, but no reason given for the change; there is more than one approach to the problem but no comparisons were made to show this was done as verification; work of other approaches is given but without explanation of their part in problem solving. **Clearly Explained**: Clear reasoning can be seen in the following ways: a written explanation of the decisions made along the way; a written justification for why a path was followed; multiple approaches with comparisons and verifications; and commentary on the approach that shows evidence of thinking.			
PS4 So What - Outcomes of Activities	*...solved the problem and stopped.*	*...solved the problem and made comments about something in the solution.*	*...solved the problem and connected the solution to other math OR **described** a use for what was learned in the "real world."*	*...solved the problem and made a general rule about the solution or extended the solution to a more complicated situation.*
	Connections can be: between mathematical ideas; between problems; to other classes or content areas; to other cases. **General Rule**: a rule that can be used no matter what the numbers in the problem are. A general rule need not be an algebraic rule; it can also be a generalization of the problem to a more complicated situation. **Prompted Response**: General prompts (e.g., Can you think when you might use this? Do you notice anything interesting in your solution?) are OK for a teacher to give. Specific prompts (e.g., What does this problem have to do with factors? How is this similar to pricing items at a grocery store?) limit scoring to a Level 2. **Special Consideration**: For this criterion, the Levels are independent of each other. For example, a student could score a 4 without a comment or description of real world use. Score Level 4 if a generalization was made at any point in the problem, whether a requirement of the problem or not, as long as an explanation showing understanding or derivation of the generalization is included.			

MATHEMATICS COMMUNICATION CRITERIA

<table>
<tr><th>Vermont Portfolio Scoring Guide: Mathematics
Rev. 2/5/93</th><th>Level 1</th><th>Level 2</th><th>Level 3</th><th>Level 4</th></tr>
<tr><td rowspan="2">C1 Mathematical Language</td><td>...didn't use any math vocabulary, equations, or notations or used them incorrectly.</td><td>...used basic math words or basic notation accurately.</td><td>...went beyond occasional use of basic math language and used the language correctly.</td><td>...relied heavily on sophisticated math language to communicate the solution.</td></tr>
<tr><td colspan="4">Basic Mathematical Language is limited to the math words and symbols that are commonly used at the student's grade level. For example, words like multiply and subtract, or notation like + and = . Don't consider math language that is included in the problem statement. Simply repeating words in the problem earns only Level 1.
Sophisticated Mathematical Language includes language not commonly used at this student's grade level. For example, words like exponent or sequence, or notation like $x < 6$ or 2^7.
Special Considerations: One mistake in accuracy will not drop a score down to a Level 1. Similarly, use of one math term rarely merits scoring a Level 2.</td></tr>
<tr><td rowspan="2">C2 Mathematical Representation</td><td>...didn't use any graphs, tables, charts, models, diagrams or drawings to communicate the solution.</td><td>...attempted to use appropriate representation.</td><td>...used appropriate math representation accurately and appropriately.</td><td>...used sophisticated graphs, tables, charts, models, diagrams, or drawings to communicate the solution.</td></tr>
<tr><td colspan="4">An Appropriate Representation is one that is related to the problem. Using the representation appropriately is executing the representation properly.
Mathematical Representations include graphs, charts, tables, models, diagrams, and equations that are linked to representations. Completing a structured chart at Grade 4 earns Level 2; at Grade 8 earns Level 1.
Accurate graphs include: 1) labeled and correctly scaled axes, 2) appropriate titles, and 3) keys, when necessary.
Sophisticated representations include perceptive representations; combinations of many graphs, charts, and tables to organize, display, and link data; and representations that were relied upon to obtain a solution.
Special Considerations: For a representation to be related it must reflect the problem.</td></tr>
<tr><td rowspan="2">C3 Presentation</td><td>...response is unclear.</td><td>...response contains some clear parts.</td><td>...if others read this response, they would have to fill in some details to understand the solution.</td><td>...response is well organized and detailed.</td></tr>
<tr><td colspan="4">Unclear suggests the reader has little or no idea what was done to solve the problem.
Some clear parts suggests the reader understood some of the work but has so many questions about what the student did that the reader is uncertain what was done to solve the problem.
Fill in some details means that although most of the solution is organized, it may be missing some details or it is detailed but lacks organization, and the reader is required to fill in.
Well organized pieces of work have all the parts connected to each other (e.g., if there are graphs and tables, there is an explanation of their part in the solution).</td></tr>
</table>

APPENDIX B: WRITTEN SURVEY AND INTERVIEW PROTOCOL

RAND MATHEMATICS PORTFOLIO SURVEY

1993–94

Dear Teacher,

Thank you for agreeing to participate in this study of the effects of the Vermont portfolio assessment program on fourth grade mathematics instruction. The study is voluntary, and we appreciate your willingness to participate. This study, like all of RAND's past research in Vermont, is strictly confidential. No information about individual teachers or students will be shared, and no participants will be identified by name or location, except as required by law. All results will be presented anonymously. We will destroy all information that identifies individuals when our data analyses are complete.

There are two parts to the study: this written survey, and a follow-up telephone interview. The survey collects background information and asks you to evaluate a sample of potential portfolio tasks. The telephone interview, which we will schedule at your convenience, will focus on the way you select tasks and teach mathematics.

It is important that you complete the written survey and have it available at the time of the interview. It also would be helpful if you had access to tasks and student portfolios during the interview. When the interview is finished, mail the completed survey to RAND in the enclosed envelope.

If you have any questions, please call either of us, collect, at the numbers indicated below.

Thank you.

Brian Stecher (310) 393-0411, extension 6579
Karen Mitchell (202) 296-5000, extension 5855
RAND

Please write your name, address and social security number in the space below. We need your name to match your responses on this written survey with the information you provide during the interview. The other information is necessary to process the checks for the honorarium. All identifying information will be removed from the survey before the information is analyzed, and all links between this information and your survey responses will be destroyed when the analyses are completed.

- -

Name ______________________________ Social Security Number ____________________

Mailing Address ____

____

BACKGROUND (Fill in your response or circle the number that corresponds to your answer.)

1. How many years of teaching experience have you had? ______ years

2. How many years have you taught fourth grade? ______ years

3. Including this year, how many years have you participated in the Vermont mathematics portfolio assessment program? (Circle the number to the right of your answer.)

   One year..........1
   Two years..........2
   Three years..........3
   Four years..........4

4. Is your school "tracked" by ability level (i.e., are students of similar ability assigned to the same class)?

   No..........0
   Yes..........1

   If yes, in which track are the students in your class? ________________________

5. Do you specialize in mathematics or teach many subjects?

   Specialize in mathematics..........0
   Teach many subjects..........1

6. Were you an official scorer in the statewide mathematics portfolio scoring session in summer 1993 or the regional portfolio scoring sessions in summer 1992?

   During the summer of 1993?

   No..........0
   Yes..........1

   During the summer of 1992?

   No..........0
   Yes..........1

7. Did you attend the fall network training session in November? (The theme was identifying worthwhile tasks.)

   No..........0
   Yes..........1

8. Did you attend the summer 1993 Portfolio Institute? (The theme was integrating problem solving into the mathematics curriculum.)

No	0
Yes	1

9. Over the past two years what proportion of the network training sessions have you attended? (Note: There have been four sessions each year, in November, January, March and May. Last year the focus was on scoring and standards; this year the emphasis has been on mathematics instruction.)

None	1
Some	2
Most	3
All	4

10. How many students are in your class this year? ____________

b. How many are in the fourth grade? ____________

c. How many are compiling mathematics portfolios? ____________

11. Compared with the other fourth grade teachers you know, how would you rate your knowledge of mathematics?

Above average	1
Average	2
Below average	3

On the following pages we have reproduced mathematics tasks used by teachers in Vermont. Please review each task or set of tasks and answer the questions below it.

TASK NUMBER 1

Stickers and Brushes

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?

a. Is this the type of task that potentially would generate student best pieces? Why or why not?

b. What are the strengths of this activity as a source of best pieces?

c. What are the weaknesses of this activity as a source of best pieces?

d. What changes would you make to improve the task so it would lead to best pieces?

[If you need more room for any answer, write on the back or attach additional pieces of paper.]

TASK NUMBER 2

Raisins

No one knows why it happened, but on Tuesday almost all the students 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.

a. Is this the type of task that potentially would generate student best pieces? Why or why not?

____

____

____

____

____

____

b. What are the strengths of this activity as a source of best pieces?

____

____

____

____

____

____

____

c. What are the weaknesses of this activity as a source of best pieces?

d. What changes would you make to improve the task so it would lead to best pieces?

[If you need more room for any answer, write on the back or attach additional pieces of paper.]

TASK SET NUMBER 3

3a. Fractions Close to 1/2

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.

3b. Building Rectangles

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 blue
1/5 red, 4/5 yellow

a. Which task is a better instructional activity? Please explain.

____

____

____

____

____

____

[If you need more room, write on the back or attach additional pieces of paper.]

b. Which task would give students a better opportunity to produce a best piece that would receive high scores? Please explain why.

TASK SET NUMBER 4

4a. Weather

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

4b. What shows with 100 throws

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	///

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

a. Which task is a better instructional activity? Please explain why.

____

____

____

____

____

____

____

____

____

____

[If you need more room, write on the back or attach additional pieces of paper.]

b. Which task would give students a better opportunity to produce a best piece that would receive high scores? Please explain why.

Thank you very much for completing this questionnaire. Please keep the results until your telephone interview. Then mail the completed survey to RAND in the self-addressed envelope provided.

MATHEMATICS PORTFOLIO INTERVIEW

1993–94

(PURPOSE / CONFIDENTIALITY)

Thank you for agreeing to participate in this study about the effects of the Vermont portfolios on fourth grade math instruction. This study, like all of RAND's previous work in Vermont, is strictly confidential. No information about individual teachers or students will be shared, and no participants will be identified by name or location, except as required by law. All results will be presented anonymously. We will destroy all information that identifies individuals when our data analyses are complete.

Later in the interview I will refer to the sample tasks in the questionnaire we mailed you. Did you complete the questionnaire and answer the questions about the four groups of tasks?

> If no, **it is important for you to complete the written survey before we conduct the interview. When would be a good time for me to call back after you have completed the questionnaire?** Day ________ Date ________ Time ________
>
> If yes, **do you have those tasks handy so I can refer to them in the interview?**
>
> > If no, **can you please retrieve them now while I wait?**
> >
> > If yes, proceed

With your permission, I would like to tape record this interview so I can have an accurate record of your comments. The tapes will be kept strictly confidential; their sole purpose is to improve the accuracy of my notes and the subsequent analysis. When we are finished with the analysis they will be erased. Do I have your permission to record this conversation?

> If no, **then I will not tape record the interview.**
>
> If yes, **I am starting the recording now**. Begin recording.

The interview should last approximately 45 minutes. Do you have any questions before we begin?

(TASK CHARACTERISTICS AND SELECTION)

I. The first part of the interview is about mathematics portfolio tasks and the way you select them. I will use the phrase "portfolio task" or "task" to mean a math activity you assign with the intent that students produce scorable best pieces. Is that clear ?

> If no, **we are interested in talking about math tasks that might generate "best pieces" not about every single assignment. Those are the tasks we want to ask about.**

1. How do you select good math portfolio tasks (that is, tasks that you hope will produce best pieces)?

    a. **What characteristics do you look for in a math portfolio task? Tell me as many features of good tasks as you can.**

        [Optional: **For example, what about mathematical content?**]

        [Optional: **Student interest?**]

    b. **What other features do you look for when choosing tasks?**

    c. [Add, if difficulty was not mentioned above: **How difficult should portfolio tasks be compared to skill-oriented class work?]**

        [Optional: **Should tasks challenge students or give them the opportunity to demonstrate what they already know?]**

    d. [Add, if mathematical content was not mentioned above: **What mathematical content should tasks include?}**

        [Optional: **Should tasks be based on content already covered in class or should they contain new topics?**]

    e. [Add, if problem solving was not mentioned above: **What kind of problem solving skills should tasks elicit?**]

2. Let's consider a specific example.

    a. **What task or tasks did you assign most recently? Please describe it (them) briefly.**

        [Optional: **During the last week or two?**]

    b. **Why did you pick this task(s)?**

        [Optional: **What made it (them) appealing to you rather than some other task?**]

3. I'm sure not all tasks are good ones.

    a. **What types of tasks would you reject?**

b. **What features distinguish POOR mathematics portfolio tasks?**

[Optional: **For example, what about the amount of structure?**]

c. **Can you think of any other features of poor tasks?**

4. **Do the scoring criteria ever influence your choice of portfolio tasks?**

a. If yes, **how is your choice affected?**

[Optional: **Have you every selected or rejected a portfolio task because it fit or did not fit the scoring criteria?**]

b. **Can you give me a specific example of a task you selected or rejected because of the scoring criteria? [Please make a copy of the task, note what it represents, and mail it to us in the envelope with the survey.]**

5. **Do you ever have problems knowing how difficult a portfolio task will be for your students?**

a. If yes, **when does this happen?**

b. **Are there types of tasks whose difficulty is hard to judge?**

6. **Can you give me an example of a task you thought would elicit scorable pieces that did not work well with your students?** If no, proceed with next question.

a. **Please describe the task.**

b. **What did you expect students to do in response to the task?**

c. **What did they do?**

d. **Why did this work so poorly?**

[Optional: **Due to features of the task or features of the lessons?**]

7. **I want to know about the sources you use for finding math portfolio tasks. A moment ago you described a task you assigned recently** [describe the task from question 2].

a. **Where did that task(s) come from?**

[Optional: **your own imagination, another teacher, training materials, supplemental books, etc.**]

b. **Over the course of the school year, approximately what percent of the tasks you assign come from each of the following four sources?**

____ Tasks you make up yourself
____ Tasks you obtain from other sources (teachers, network training, supplemental books, etc.)
____ Tasks you adapt from other sources
____ Other sources (Please describe ____________________________)

c. **Over the course of the school year, what percent of the tasks you assign are tasks that you've used in previous years?** ____________

8. **Do you assign different mathematics tasks to students based on their math or writing proficiency?**

a. If yes, **what is different about the tasks you assign to different students?**

b. If yes, **how do you match tasks to students of different ability?**

c. **Do you assign different tasks to different students for other reasons?**

d. **Which reasons?**

e. **Are there other ways you adapt to student differences when they are working on tasks for the mathematics portfolios?**

(UNDERSTANDING PROBLEM SOLVING)

II. The second topic of the interview is problem solving.

9. **Can you explain to me what "problem solving" is?**

[Optional: **What kind of problems are appropriate?**

What kinds of solving should students be able to do?]

a. **Is your view of problem solving different from the view of the Vermont portfolio assessment?**

If yes, **how do they differ?**

10. **What specific problem solving skills are you trying to teach?**

[Optional: **looking for a pattern?**]

a. **How are you trying to teach these skills?**

[Optional: **What kinds of instruction do you give? What activities do you provide?**]

11. **Are there any types of problem solving that are neglected by the tasks you assign?**

a. **Are there any types of problem solving that are widely neglected in Vermont?**

12. **What do you KNOW now about problem solving that you did not know prior to the portfolio assessment?**

13. **What do you DO now to foster problem solving that you did not do prior to the portfolio assessment program?**

14. **Please refer to Task #1 in the packet we mailed you.**

a. **How well would your students respond to this task?**

b. **What problem solving behaviors would they use?**

c. **If you were going to assign this problem next week, would you do anything in class now to prepare your students for it?**

15. **Please refer to Task #2 in the packet we mailed you.**

a. **How well would your students respond to this task?**

b. **What problem solving behaviors would they use?**

c. **If you were going to assign this problem next week, would you do anything in class now to prepare your students for it?**

16. Please look at Tasks 3a and 3b in the packet.

a. Which task demands a greater variety of problem solving skills? Which skills?

b. Which task would produce better scores on the portfolio criteria? Why would that happen?

17. Finally, lets review Tasks 4a, 4b and 4c in the packet.

a. Which task demands a greater variety of problem solving skills? Which skills?

b. Which task would produce better scores on the portfolio criteria? Why would that happen?

(END)

That is the end of our formal interview. Are there things you wanted to say that you did not have an opportunity to say? Are questions we should have asked that we failed to ask?

Thank you very much for your time. We will send you a copy of the report when it is completed, which should be in late summer or early fall. Your honorarium will be sent within four weeks. If you do not receive it or have any questions, please call me. Do you have my name and number?

To everyone, **Please do not forget to mail your answers to the written survey.**

[As appropriate: **Please do not forget to send copies of the specific tasks or student work we discussed and indicate clearly what they represent**]

APPENDIX C: CHARACTERISTICS OF GOOD PROBLEMS

State-sponsored training sessions offered teachers two relevant criteria for judging the quality of problems, and trainers spent considerable workshop time with teachers analyzing good and bad tasks. One definition of good problems is taken from Marilyn Burns (quoted from the "Fourth Grade Network Leader's Guide," October/November, 1993):

Criteria for Mathematical Problems

- There is a perplexing situation that the student understands.
- The student is interested in finding a solution.
- The student is unable to proceed directly toward a solution.
- The solution requires the use of mathematical ideas.

Another definition is drawn from the curriculum standards of the National Council of Teachers of Mathematics. It consists of a list of the features of worthwhile mathematical tasks (quoted from the "Fourth Grade Network Leader's Guide," October/November, 1993):

The teacher of mathematics should pose tasks that are based on:

- sound and significant mathematics;
- knowledge of students' understanding, interests, and experiences;
- knowledge of the range of ways that diverse students learn mathematics; and that
- engage students' intellects;
- develop students' mathematical understandings and skills;
- stimulate students to make connections and develop a coherent framework for mathematical ideas;
- call for problem formulation, problem solving, and mathematical reasoning;
- promote communication about mathematics;
- represent mathematics as an ongoing human activity;
- display sensitivity to, and draw on, students' diverse background experiences and dispositions; and
- promote the development of all students' dispositions to do mathematics.

REFERENCES

Burns, M. (1992). *About teaching mathematics: A K–8 resource*. White Plains, NY: Math Solutions Publications.

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

Flexer, R., Cumbo, K., Borko, H., Mayfield, V., & 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.

Gearhart M., Herman, J., Baker, E., & Whittaker, A. (1993). *Whose work is it?: A question for the validity of large-scale portfolio assessment* (CSE Tech. Rep. No. 363). Los Angeles: University of California, Center for Research on Evaluation, Standards, and Student Testing.

Goodnow, J., & Hoogeboom, S. (1987). *The problem solver 4: Activities for learning problem-solving strategies*. Mountain View, CA: Creative Publications.

Haladyna, T., Nolen, S., & Haas, N. (1991). Raising standardized test scores and the origins of test score pollution. *Educational Researcher*, *20*(5), 2–7.

Koretz, D., Klein, S., McCaffrey, D., & Stecher, B. (1993). *Interim report: The reliability of Vermont portfolio scores in the 1992–93 school year* (CSE Tech. Rep. No. 370). Los Angeles: University of California, Center for Research on Evaluation, Standards, and Student Testing. (Available from RAND as Reprint No. RP-260.)

Koretz, D., Stecher, B., Klein, S., & McCaffrey, D. (1994a). *The evolution of a portfolio program: The impact and quality of the Vermont program in its second year (1992–93)* (CSE Tech. Rep. No. 385). Los Angeles: University of California, Center for Research on Evaluation, Standards, and Student Testing.

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

Koretz, D., Stecher, B., Klein, S., McCaffrey, D., & Deibert, E. (1993). *Can portfolios assess student performance and influence instruction? The 1991–92 Vermont experience* (CSE Tech. Rep. No. 371). Los Angeles: University of California, Center for Research on Evaluation, Standards, and Student Testing. (Available from RAND as Reprint No. RP-259.)

Lane, S., Lio, M., Stone, C., & 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.

McKnight, C., & Cooney, T. (1993). Content representation in mathematics instruction: Characteristics, determinants and effectiveness. In L. Burstein (Ed.), *The IEA Study of Mathematics III: Student growth and classroom process*. New York: Pergamon.

Messick, S. (1975). The standard problem: Meaning and values in measurement and evaluation. *American Psychologist, 30*, 955–966.

Messick, S. (1984). The psychology of educational measurement. *Journal of Educational Measurement*, *21*, 215–237.

National Council of Teachers of Mathematics. (1989). *Curriculum and evaluation standards for school mathematics.* Reston, VA: NCTM.

Polya, G. (1980). On solving mathematical problems in high school. In S. Krulik (Ed.), *Problem solving in school mathematics*, *1980 Yearbook of the National Council of Teachers of Mathematics.* Reston, VA: NCTM.

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

Robitaille, D. (1993). Contrasts in the teaching of selected concepts and procedures. In L. Burstein (Ed.), *The IEA Study of Mathematics III: Student growth and classroom process.* New York: Pergamon.

Shepard, L., & Dougherty, K. (1991). *Effects of high stakes testing on instruction.* Paper presented at the annual meetings 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., & 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.