Designing and Using Alternative Assessments to Measure the Problem Solving Skills of Elementary Students

Sylvia R. Taube
Barbara E. Polnick
William A. Jasper
Sam Houston State University

Background
Teacher educators and school leaders are often challenged with providing relevant and meaningful professional development opportunities to improve teachers’ skills in teaching creative and critical thinking. The task becomes more daunting in the wake of the No Child Left Behind (NCLB) Act that holds school leaders accountable for improving the quality of their teachers and raising student achievement scores. Problem solving is emphasized on statewide assessments and has been endorsed by the National Council of Teachers of Mathematics (NCTM).

With the current focus on high stakes standardized testing, it is common knowledge that teachers feel pressured to prepare students to succeed on multiple-choice and select response test items. In such cases, much class time may be spent in teaching test-taking strategies, leaving less time for concept development and extended problem-solving experiences. These practices can conflict with standards-based teaching where students need even more class time to work on challenging and complex problems that require exploring multiple solutions and have students explain their reasoning (NCTM, 2000).

Alternative Assessments in the Classroom
Educators and researchers (e.g., Bryant & Driscoll, 1998) proposed that, in addition to the current standardized and select response tests, school districts need to support teachers in using open-ended assessment tasks and rubrics that both measure students' understanding of concepts and their ability to apply skills in problem solving situations. There are several reasons why these types of assessments are being recommended. One, state assessments alone cannot provide teachers with useful feedback to carefully diagnose and address the individual needs of all learners in their classroom. Secondly, if teachers use evaluation systems that highlight the importance of problem solving, reasoning, and communicating thinking, then students will value and attempt to demonstrate these processes. Thirdly, an over emphasis on standardized test preparation is likely to reinforce student beliefs that mathematics problems have only one solution and/or one correct answer. Szetela and Nicol (1992) further argued “students who are prone to make calculations without explanations often fail to reveal sufficiently the nature of the problem solver’s work and thinking” (p. 42).

Perspectives on the professional growth of mathematics teachers
Recent changes in areas that affect teaching practices such as an increased emphasis in implementing national standards, alternative assessments, as well as district and state accountability tests, call for a new vision for the professional development of mathematics teachers. Researchers (e.g., Fennema, et al., 1996; Hufferd-Ackles, et al., 2004) of standards-based teaching further underscore the need for teachers to increase their content and pedagogical knowledge, as well as opportunities to learn about students' thinking. Furthermore, the literature on professional development for teachers identifies research-based features of effective models for helping teachers. For example, the NCTM Yearbook (Aichele & Coxford, 1994) provided helpful guides in developing professional development models for elementary teachers with limited mathematics background. In the same source, Lester, et al., (pp.152-166) described researched-based strategies to help teachers teach via problem solving. The key features of their model included: a) encouraging reflections, b) modeling metacognition, c) acting as facilitator, d) supporting group problem solving, and e) developing a new vision for assessment. Furthermore, Borassi and Fonzi (2003) concluded that quality professional development programs are a) sustained and intensive, b) based on how people learn best, and c) focused on teaching and learning.

Design of the Study
A Partnership Model for Staff Development
The critical need for mathematics teachers to understand, in-depth, the nature of problem solving and the challenges of assessing student performance on problem-solving tasks presented an opportunity for collaboration between a regional university and school districts. Two university mathematics teacher
educators wrote a grant for Title II funding in collaboration with five suburban school districts located north of Houston, Texas, that had both diverse populations and a dire need to improve mathematical problem solving, as evidenced by campus results on the statewide and local assessments. In this paper, we will describe an 18-month program that supported the professional growth of fifteen elementary mathematics teachers from campuses with highly diverse student populations. This project consisted of three phases but only the first two phases will be addressed in this article.

Phase I

The first phase was a three-week summer institute in July 2002, and included 50 hours of instruction. This instruction specifically addressed the need for teachers to gain more in-depth knowledge in the mathematics content they teach as well as strategies that should be used in teaching standards-based curricula. This phase was critical, since teachers with strong knowledge of mathematics content and pedagogy are needed to make sound curricular and instructional decisions as well as to critique standards and challenge the assumptions and actions that affect their students (Cochran-Smith, 2001). During this period, teacher participants worked, in small cooperative groups, on a variety of problems such as making tables, looking for patterns, and working backwards. Assessment of the teachers’ knowledge and skills gained from the summer sessions included: pre and posttests, assessment projects, and teacher reflections on individual investigations of mathematics topics, including how to teach using a problem-based approach.

A pretest-posttest design measured gains in teachers’ knowledge and skills about problem solving, statistics, and probability. This instrument was criterion-based and consisted of Section I: Problem Solving and Section II: Applications of Statistics and Probability. The items measured teachers’ conceptual understanding of mathematics by posing problems requiring application, evaluation, and synthesis. A scoring rubric (see Appendix A) was developed to score the pre and posttest consisting of 15 open-ended problems.

Both sections of the pretest were administered at the start of summer component. At the completion of the summer course, the teachers took Section I of the posttest. At the last meeting in May 2003 the teachers took Section II of the posttest. Results on the posttest (Sections I) provided useful feedback for improving the sessions during the academic year.

Phase II

The second phase of the program involved another 50 hours of instruction and included: (a) two follow-up meetings each month during the fall and spring semester, (b) discussion and support for participants in the implementation of program objectives, and (c) classroom observations by an external evaluator. This phase of the program was designed to support and mentor teachers in classroom implementation of the program, including the development and use of alternative assessments to measure problem solving. During this phase the teachers learned to design mathematical problem-solving tasks for their students that included the development and utilization of a rubric to evaluate student work.

In the spring semester, the participants were observed by the evaluator to monitor their classroom implementation and how well their students responded to the lesson. The evaluator used a classroom observation form to record instructional processes and events occurring during the 40 to 50-minute lesson. She observed the extent to which teachers (and students) used (a) problem solving strategies, (b) technology, (c) data-based activities, (d) cooperative grouping, and (e) manipulatives. At the culmination of the lesson the teacher and evaluator held a short conference to reflect on the strengths and weaknesses of the lesson.

Evidence of teachers’ changing instructional practices was also collected through their reflections of their own implementation efforts and through their observations of their own students’ learning in the mathematics classroom. The impact on student learning was measured by the results on the assessment tasks designed by teachers.

Learning how to develop assessment tasks and rubrics

We stressed the development of assessments and rubrics in the implementation of standards-based teaching, especially in the area of problem solving. The NCTM strongly supports the use of authentic assessments to gauge students’ understanding of mathematics (1995; 2000). In order for teachers to understand both the purpose and benefits of using open-ended assessment tasks and rubrics, steps were carefully followed and successfully implemented during the follow up sessions held in the fall semester. Samples of trainer-teacher dialogue that occurred during each step are included to enhance the description of the steps that were undertaken as teachers were developing their own rubrics.

Step One: Develop areas to be scored
The instructors engaged the teachers in clarifying and identifying observable behaviors that students would be exhibiting if they were demonstrating understanding of a concept. The following is an excerpt of the dialogue during this inquiry:

**Trainer:** Think of a concept that you are currently teaching in which students often score low on state assessments.

**Teacher:** Fractions

**Trainer:** What’s important for students to know and do?

**Teacher:** Adding fractions

**Trainer:** What about “adding fractions”?

**Teacher:**

a) Apply the algorithm correctly; b) Recognize the relationship of place value/renaming.

**Trainer:** What behaviors, if students did them consistently, would greatly increase their level of understanding of concepts and their ability to apply the concepts in a problem-solving situation?

**Teacher:**

a) Have a plan; b) Select an appropriate strategy; c) Recognize important information; d) Eliminate erroneous information

**Trainer:** Specifically, what would you see and hear if students understood addition of fractions?

**Teacher:**

a) Steps would be recorded in some systematic way; b) The correct units are included in the answer; and c) The answer is written in a complete sentence

**Step Two: Determine relative value of each area**

Teachers were asked to think about how much each area should be valued relative to the other areas.

**Trainer:** If all areas are equally important, then each area would receive the same value.

**Teacher:** What do you mean by “value”?

**Trainer:** For example, is “getting the correct answer” worth as much as “using more than one strategy?”

**Teacher** a) Well, I think getting the correct answer is important. Students have to do that on the TAKS (state) test; b) I think that leaving off the units is not as important as systematically recording all the steps.

In some cases, it was easier for the teachers in their initial development to assign equal weight or value to each area assessed. No actual points were assigned at this stage.

**Step Three: Assign maximum points for each area to be scored**

Once teachers have determined the relative value of each problem-solving area being assessed, the next step is to assign a maximum number of points to each area being assessed. Teachers, for whatever reason, found this part of building the rubric to be particularly challenging. Because of this, some teachers found it useful to assign these maximum values based on a 100-point system. So if the teachers were evaluating four areas and they were all determined to be of equal value in Step Two above, then each area would be worth a maximum of 25 points. If on the other hand a teacher wanted an area such as “including a visual or diagram” to count more (for example, 40 points), the points for the other three areas would have to be adjusted so that the sum was 100 (e.g., 20 points, 20 points and 20 points).

**Step Four: Determine a range of possible scores for each area.**

Establishing a range of possible scores for each area helps to develop a standardized system for assigning points to students’ work. Since not all students will necessarily get the maximum number of points for each area, it makes sense to spread the points out over a range of possible responses. This requires the teachers to think in terms of dividing the maximum number of points into increments. For example, most teachers divided the point range into four or five increments. Using the 100-point system, if Area One was assigned 25 points with five incremental values, the point spread might be: 5, 10, 15, 20 and 25. However, some teachers wanted to start with zero, which meant that the increments would not necessarily be at equal intervals. An example of assigned values across four increments with a maximum value of 25 might be 0, 10, 20, and 25, as seen in Figure 1.
Step Five: Describe what the student work would look like for each score.

This step gets at the heart of what it takes to be able to diagnose student strengths and weaknesses by systematically examining students’ work. By building student behaviors into the rubric, teachers are better able to communicate their expectations to the students and to the parents. For students, understanding the rubric is a first step to understanding problem solving and what it really means to be a good problem solver. Completing this step in the rubric requires teachers to set high expectations and to establish each set of behaviors in reasonable increments, each set building on the other. Communicating what are reasonably expected behaviors at each incremental level is critical during this step. Crucial to the effective utilization of the rubric with all students, including those who are struggling, is establishing criteria that is both attainable and developmental, as exemplified in the partial rubric shown in Figure 1.

<table>
<thead>
<tr>
<th>Example</th>
<th>0</th>
<th>10</th>
<th>20</th>
<th>25</th>
</tr>
</thead>
<tbody>
<tr>
<td>a) Include a visual or diagram when solving a problem</td>
<td>None included</td>
<td>Diagram included, but does not relate to student's work</td>
<td>Diagram included, relates to student's work but inaccurate</td>
<td>Diagram included, supports student's ideas and work with no errors</td>
</tr>
</tbody>
</table>

*Figure 1. A section of a teacher-developed rubric*

*Classroom implementation*

After going through the steps in developing rubrics, the teachers were required to design and implement lessons involving problem solving with probability or statistics as the content. A major component of the problem-solving lessons involved developing and implementing assessment tasks with corresponding scoring rubric appropriate for open-ended problems. The full description of this particular assignment is in Appendix B.

*Results*

**Teachers’ conceptual understanding of mathematics**

The pre and posttest data for the problem-solving portion were analyzed by using a paired t-test technique. A detailed analysis of the pre and posttest results appears in another paper (in press) and will not be elaborated here since the focus of this paper is on teachers’ classroom implementation of the concepts learned during the intense summer session. Overall, there was an average gain of 21.7 points on the 60-point posttest as compared to the pretest for the 14 teachers who took both tests. This difference was significant at the 0.05 which indicated a strong improvement in problem-solving abilities of the teacher participants.

**Evidence of classroom implementation by teachers**

The results discussed in this section are based on the first cohort consisting of 15 teachers (2 males, 13 females) whose classroom assignments ranged from Grade 2 to Grade 7. These teachers successfully completed the project that was conducted from May 2002 to September 2003. A discussion of classroom observation results, assessment projects and teacher reflections is included in this section.

**Open-ended assessment tasks and scoring rubrics**

Two examples of the teachers’ work using the assessment tasks are described in this section, along with samples of student work and analysis of class results.

*Teacher A.* A fourth-grade teacher developed an assessment task that required students to find all possible combinations for ordering “Daily Specials” from a diner if there were three choices of entrees,
vegetables, and desserts. She administered this assessment to 21 students after giving them instructions. Using the rubric (See Appendix C) that she designed, Teacher A reported that 34% of the students tried more than one strategy, 71% showed their work, 24% checked their work, 81% did not use any math vocabulary, and 19% explained what they did to solve the problem. From the results, she made insightful recommendations such as: improve the tests and homework to include at least one problem that requires students to use more than one strategy; use the same problem each day of the week and challenge students to use a different strategy each day; and go over the rubric with the students before giving the assessment.

Teacher B. This teacher developed a rubric (see Appendix D) for scoring a performance-based assessment titled, “What is probability?” for 7th grade students identified as “gifted and talented”. This assessment included finding real-life problems that had five possible outcomes, listing permutations of possible outcomes, and drawing a tree diagram to determine probability values. The teacher wanted to collect evidence to show that her students had gained a conceptual understanding of the term “probability”. She was interested in looking at students’ level of understanding, concept clarity, and interest. Summarizing the results, she reported that 11 out of 20 students met the “Excellent” mark in all three categories. She also found out that some students were still confused about drawing tree diagrams and could benefit from reteaching. Teacher B, after analyzing students’ work, decided to make changes such as giving students a choice to pick their topics and requiring them to write a reflection on the activities.

Teachers’ Reflections

Data collected during the spring semester showed evidence that teachers had gained both understanding and application of the targeted mathematical concepts as well as the ability to design and implement rubrics to assess students’ problem-solving skills. This was evident in their projects, which included designing the problem or task to be administered to students, developing the rubric, analyzing and scoring students’ work, and describing and reporting class data. The following section highlights teachers’ thoughts about their experiences in designing and implementing alternative assessments, and analyzing student responses to inform teaching.

Overall, I felt very confident going into this assessment and came out of it with some things to work on. The results had shown me what areas of the TEKS (Curriculum) my students are strong in and areas we needed to address. A change that I would make to my assessment would be to make it mandatory to use hands-on bills and coins and to have them work in a group to show how they could pay for the items being purchased. (3rd grade teacher)

I feel that all students needed to be encouraged to take risks so that they can become successful problem solvers. The problem-solving assessment and rubric helped me determine how well my students understood the task and it helped guide my decision making with regard to future problem-solving activities. (3rd/4th Grade Teacher/Skills Specialist)

Out of the 43 students assessed, I found that 2% of the students did not show any work. I will continue to model and instruct ways to organize work. Showing their answers is another organizational skill students need to master. They need to get in the habit of making their solution clearly available. I enjoyed doing this project because it allowed me to analyze each student…the few students who did not show mastery of this objective were pulled together during small group time for re-teaching. (3rd Grade Teacher)

This project I developed was very challenging for students. For the most part I was pleased with the enthusiasm of the group and the success of the project. It took a while to grade these projects because I really wanted to know if the students had a clear understanding of the differences between concepts, especially between permutations and combinations and between theoretical and experimental probability. I also needed to know if they understood how a tree diagram works in displaying data. To get a score of “excellent” my students had to go above and beyond the “normal” evaluation of their data and give me something special. (6th/7th grade teacher)

Emerging themes

From the teachers’ self-reports of their classroom implementation emerged at least three themes that revealed not only their understanding of alternative assessments and rubrics but also how the results were used to benefit them and their students. First, teachers’ values were
reflected in the rubrics as well as in their analysis of students’ work. For example, one teacher made explicit her high expectations by challenging her students to turn in exemplary work while another teacher valued organization of students’ written work. Second, teachers focused on students’ performance on the tasks and identified ways to improve student learning and how the teacher could facilitate this. Lastly, the teachers felt confident and not threatened by the new assessment method.

**Implications for research and practice**

If teaching practices are changing, it is but logical to change classroom assessments (Lambdin, 1993). Mathematics teachers, however, should not be expected to implement classroom change alone. They need varied types of support from their leaders, mentors, and peers. These include professional development programs that meet not only their needs but also the needs of their students. Critical elements of this support system for mathematics teachers must address the following:

- Providing in-depth mathematical knowledge and processes and flexible processes;
- Developing a mathematics community that fosters and values mathematics ideas;
- Helping students build mathematical ideas as well as understanding;
- Designing assessments that are meaningful and can provide feedback to students.

Meanwhile, university professors of teacher preparation programs must play a role in developing and sustaining a strong partnership with school districts for the continuous professional development of teachers. With their expertise and leadership they can help teachers gain ownership of the new ideas through effective mentoring. Empowering teachers to implement new standards for teaching and assessing mathematics learning is a process that requires a strong collaboration between university and school district personnel.
### Problem Solving Rubric

<table>
<thead>
<tr>
<th>POINTS</th>
<th>METHOD</th>
<th>ANSWER</th>
<th>EFFORT</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>None</td>
<td>None</td>
<td>None</td>
</tr>
<tr>
<td>1</td>
<td>Inappropriate</td>
<td>Incorrect</td>
<td>Good</td>
</tr>
<tr>
<td></td>
<td>Inappropriate</td>
<td>Correct</td>
<td>Weak</td>
</tr>
<tr>
<td>2</td>
<td>Inappropriate</td>
<td>Correct</td>
<td>Good</td>
</tr>
<tr>
<td></td>
<td>Acceptable</td>
<td>Incorrect</td>
<td>Weak</td>
</tr>
<tr>
<td>3</td>
<td>Acceptable</td>
<td>Incorrect</td>
<td>Good</td>
</tr>
<tr>
<td></td>
<td>Acceptable</td>
<td>Correct</td>
<td>Weak</td>
</tr>
<tr>
<td>4</td>
<td>Acceptable</td>
<td>Small errors</td>
<td>Good</td>
</tr>
<tr>
<td>5</td>
<td>Acceptable</td>
<td>No errors</td>
<td>Great</td>
</tr>
</tbody>
</table>

**METHOD** includes choosing a problem-solving strategy that is appropriate for that particular problem. Guess and check is not normally an appropriate strategy, unless combined with another strategy such as making a table. Focused guesses (when the student makes the next guess based on the results of the previous guess) are much better than random guesses.

**ANSWER** is the correct number obtained from solving the problem. Even though the answer is either correct or incorrect, a solution process that includes only a small calculation error is better than answers obtained by performing the wrong operation or strategy.

**EFFORT** is a measure of how hard the student worked while solving the problem and how much of the work was actually shown. A great rating means all work was accurate and was shown. A good rating means that work was accurate and most of it was shown.
Math Problem Solving Assessment Project

1. Design or select an assessment question or problem that will assess problem solving as it relates to understanding and applying a concept or skill (or group of skills) that are in the TEKS for your grade level.

2. The problem needs to be written in a meaningful, challenging context that allows for different strategies and approaches to be used in reaching a solution(s).

   Example: Objective for 4th Grade: Solve problems with decimals (or money).
   Problem: Looking at the coins in my pocket, I noticed that I could pay the exact price for any item from one cent up to and including one dollar without receiving any change. What coins do I have in my pocket?

   Note: This problem has more than one solution, but we could have asked for the least number of coins possible and there would have been only one solution.

3. Design the rubric that you will use to assess student understanding of the concept as well as whatever problem solving behaviors you want to assess. The rubric must contain at least four things you are assessing and must include four levels of concept or skill attainment.

4. Administer the assessment to your class. (Follow the process we used during class by giving the student clear instruction, clarify the steps with examples if needed and administer the test in a standardized manner.)

5. Score the student responses using the rubric.

6. Analyze the class data from the assessment and determine the strengths and weaknesses of students (based on what you are evaluating), including areas for improvement or reteaching.

7. Prepare two reports on the data you collected from the results of the assessment, one to your principal and one for the parents of the students. Use graphics to display your data. Include specific examples from student work on the assessment that substantiates your analysis.

8. Write 2-3 sentences that reflect your thoughts on the scores from the assessment. Examples: Were there any surprises in the students' responses?

9. Did the assessment question or problem assess what you wanted it to assess?

10. What changes, if any, would you make to the assessment question or problem?

11. Be prepared to give an oral version of your written report (with overhead transparencies, charts or PowerPoint) to your peers.
## Appendix C
### Teacher A Rubric

<table>
<thead>
<tr>
<th>Categories</th>
<th>Criteria</th>
<th>Points</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>4</td>
<td>3</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Strategies</th>
<th>More than two strategies tried</th>
<th>Two strategies tried</th>
<th>One strategy tried</th>
<th>No strategies tried</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mathematical Representation</td>
<td>All work shown and checked</td>
<td>Most work shown and all checked</td>
<td>Some work shown, not all checked</td>
<td>Little or no work shown and very little if any checked</td>
</tr>
<tr>
<td>Explanation</td>
<td>All work and strategies explained in clear manner</td>
<td>Most work and strategies explained in clear manner</td>
<td>Some work and strategies, incomplete explanation</td>
<td>Little or no work and strategies, incomplete explanation</td>
</tr>
<tr>
<td>Solution</td>
<td>All solutions correct and checked</td>
<td>Most solutions correct and checked</td>
<td>Some solutions correct, most checked</td>
<td>No solutions correct, none checked</td>
</tr>
<tr>
<td>Communication</td>
<td>All math vocabulary and symbols correct</td>
<td>Most math vocabulary and symbols correct</td>
<td>Some math vocabulary and symbols correct</td>
<td>Very little or no math vocabulary and symbols incorrect</td>
</tr>
</tbody>
</table>
## Appendix D

### Teacher’s B Rubric

<table>
<thead>
<tr>
<th>Criterion</th>
<th>Excellent 31-35 points maximum</th>
<th>Very Good 26-30 points maximum</th>
<th>Satisfactory 21-25 points maximum</th>
<th>Unsatisfactory 16-20 points maximum</th>
</tr>
</thead>
<tbody>
<tr>
<td>Presentation</td>
<td>Project was organized, and all parts included as stated in directions.</td>
<td>Project was organized, and all parts included as stated in directions.</td>
<td>Project was fairly organized, and all parts were included as stated.</td>
<td>Project was not organized, and some parts were not included</td>
</tr>
<tr>
<td></td>
<td>Took time to complete project</td>
<td>Took time to complete project</td>
<td>Took a reasonable amount of time to complete project</td>
<td>Time was not taken to complete project satisfactorily</td>
</tr>
<tr>
<td></td>
<td>Project was colorful, and creative, and accurate.</td>
<td>Project was colorful, and creative, and accurate.</td>
<td>Project was somewhat creative.</td>
<td>Project was not very creative.</td>
</tr>
<tr>
<td></td>
<td>Student went beyond expectations to create a unique project.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Written Mechanics</td>
<td>All written work followed directions</td>
<td>All written work followed given instructions</td>
<td>Most of written work followed directions</td>
<td>Written work did not follow directions</td>
</tr>
<tr>
<td></td>
<td>Spelling and grammar was completely error free.</td>
<td>Spelling and grammar mostly free of error</td>
<td>Spelling and grammar had errors</td>
<td>Spelling and grammar were full of errors</td>
</tr>
<tr>
<td>Charts and Diagrams</td>
<td>All diagrams and table were accurate</td>
<td>All diagrams and tables were accurate.</td>
<td>Diagrams and table were accurate.</td>
<td>Diagrams and tables were not accurate.</td>
</tr>
<tr>
<td></td>
<td>Data was correctly displayed</td>
<td>Data was correctly displayed in the charts and diagrams.</td>
<td>Data was correctly displayed in charts and diagrams with minimal errors</td>
<td>Data was not correctly displayed in charts and diagrams.</td>
</tr>
<tr>
<td></td>
<td>All directions were followed</td>
<td>All directions were followed.</td>
<td>Most of directions were followed.</td>
<td>Most directions were not followed.</td>
</tr>
<tr>
<td></td>
<td>Charts and diagrams showed exceptional creativity and detail.</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### References


