AN ABSTRACT OF THE DISSERTATION OF

Jessica Strowbridge for the degree of Doctor of Philosophy in Mathematics presented on May 29, 2008.

Title: Middle School Teachers' Use of a Formative Feedback Guide in Mathematical Problem Solving Instruction.

Abstract Approved: _____

Thomas P. Dick

The purpose of this study was to address the implementation fidelity of one part of a professional development model developed by the Northwest Regional Educational Laboratory (NWREL). Specifically, this research investigates middle school teachers' use of a formative feedback guide developed by NWREL staff, examining the reliability with which teachers use the guide and how its use influences the follow-up instructional decisions they make. The subjects of the study were ten middle school mathematics teachers who participated in the NWREL professional development and the subsequent classroom problem solving activities. The primary data collected for the study were student solutions to problem solving tasks, teacher evaluations of student performance using the NWREL formative feedback guide, and teacher plans for sharing student work in a follow-up class.

Wiliam and Black et al provide frameworks for considering formative assessment that influenced this study, particularly around the nature of formative assessment practices that positively influence student learning. Stein et al's pedagogical model for

sharing student work in classroom discussions formed the basis on which follow-up instruction was considered in this study.

The results of this study suggest that the NWREL formative feedback guide functions more as a summative than a formative assessment guide. Teachers did not use the guide with high levels of reliability. The rationales teachers provided for selecting pieces of student work to share in follow-up classroom discussions were almost always related to aspects of mathematical problem solving, especially to show students a variety of strategies. While those rationales were usually reflected to some degree in the student solutions chosen, the comments that teachers wrote for students less frequently reflected the rationales.

Based on these results, three new guides are proposed for future professional development efforts for this component of the NWREL program: 1) a formative assessment written comment guide, 2) a scaffolding guide, and 3) a student work selection and sequencing guide.

Middle School Teachers' Use of a Formative Feedback Guide in Mathematics Problem Solving Instruction

by
Jessica Strowbridge

A DISSERTATION

submitted to

Oregon State University

in partial fulfillment of
the requirements for the
degree of

Doctor of Philosophy

Presented May 29, 2008
Commencement June 2009

Doctor of Philosophy dissertation of Jessica Strowbridge presented on May 29, 2008.

APPROVED:

Major Professor, representing Mathematics

Chair of the Department of Mathematics

Dean of the Graduate School

I understand that my dissertation will become part of the permanent collection of Oregon State University libraries. My signature below authorizes release of my dissertation to any reader upon request.

Jessica Strowbridge, Author

ACKNOWLEDGEMENTS

My deep gratitude is due to a number of people who provided both professional and personal support throughout this project. I certainly could not have completed it without their assistance

My major professor, Thomas P. Dick was a source of tremendous support and guidance over the course of this project. Tom involved me in the grant from which this research developed, and continued to provide me with opportunities to grow professionally. His insightful comments, expectations of quality, patience, and willingness to accommodate my hectic timeline are deeply appreciated.

I would like to thank the other researchers involved in this project for their guidance in the development of this research: Dr. Edith Gummer, Dr. Karen Marongelle, Claire Gates, and Sarah Enoch. Edith Gummer's comments regarding the methodology and results sections of this dissertation were particularly helpful. I would also like to thank my examination committee, Dr. Barbara Edwards, Dr. Karen Higgins, Dr. Bruce McGough, and Dr. Mina Ossiander.

I would like to thank the ten middle school teachers and their students who were involved in the project. The teachers deserve ample thanks for submitting the long list of data we requested from them.

I would like to thank Karen Barrese for her invaluable assistance analyzing the data for the study. Karen's thoughtful comments and commitment to faithfully representing the results were critical to the quality of this project. I am particularly

indebted to Karen for her observation linking the rationales teachers provided to the problem solving traits on the feedback guide.

I would like to thank two graduate students at Oregon State University, Shari Ultman and Gülden Karakok, for both their personal support and professional insight into this project. I would also like to thank Kayla Coe for her friendship, support and interest in my work. Thanks to Steve Cohen for his probing questions about my writing.

I would like to thank my parents, Ed and Judy Strowbridge, and my siblings, Stephanie, Michelle and Jonathan for supporting my academic endeavors. Thank you to my grandfather, Dr. Ed Strowbridge, for nurturing my early interest in mathematics education. Finally, my deepest gratitude to Joel Cohen. Without his support, faith, patience, understanding, assistance, encouragement and love this would never have been possible.

TABLE OF CONTENTS

TABLE OF CONTENTS (Continued)

TABLE OF CONTENTS (Continued)

LIST OF FIGURES

LIST OF TABLES

LIST OF TABLES (Continued)

Middle School Teachers' Use of a Formative Feedback Guide in Mathematics Problem Solving Instruction

Chapter 1 – Introduction

This work is based on the importance of two separate but interrelated concepts in mathematics education. One is the incorporation of formative assessment techniques into the mathematics classroom, and the other is the creation of mathematically productive, student-centered environments.

Research on formative assessment indicates that its use can improve student performance and also help teachers structure productive classroom sessions (Bangert-Drowns et al, 1991; Black et al, 2003; Elawar & Cordo, 1985; Wiliam, 2007) . Useable, constructive feedback to students can be beneficial to student learning, but the development of such comments is often challenging and time consuming for teachers (Black et al, 2003; Elawar & Cordo, 1985; Lee, 2006). The use of formative assessment information to make instructional decisions is also often challenging and time consuming for teachers, but making instructional adjustments to address students' learning needs has significant benefits (Black et al, 2003; Even, 2005; Fuchs et al, 1991). In both areas, professional development has been shown to have a positive effect on teacher practices (Black et al, 2003; Driscoll, 1999; Elawar & Cordo, 1985).

Mathematics education reform efforts also call for emphasis on creating student-centered classrooms using complex, non-algorithmic problem solving tasks. There is strong evidence that such classrooms are more conducive to productive student learning (Stein et al, 2000; Wertsch & Toma, 1995; ; Wood, 1995; Wood et al, 1995). Yet

research shows that the use of complex tasks and the development of student-centered classrooms can be both intimidating and challenging for teachers (Borko & Livingston, 1989; Chazan & Ball, 1999; Heaton, 2000). The work of Stein et al (in press) provides a framework for considering the elements of a classroom discussion based around sharing student solutions. The authors suggest that this framework may be useful to help teachers maintain control of mathematical learning goals while allowing students to play central roles in classroom discussions.

The Northwest Regional Educational Laboratory (NWREL) developed a Mathematics Problem Solving Model, designed to facilitate incorporation of problem solving activities into the classroom. The model also includes tools and professional development components to facilitate use of formative assessment practices and implementation of student-centered classrooms. This research is part of a larger project to address the effectiveness of the NWREL Mathematics Problem Solving Model. In particular, this dissertation seeks to determine the reliability of teachers' use of a formative feedback guide developed by NWREL to make formative assessment decisions. One component describes the ways in which teachers plan lessons around sharing student work, using the ideas of the pedagogical model suggested by Stein et al (in press).

The research on formative assessment, particularly relating to teachers' assessment of student performance, formative feedback to students, and the use of assessment information in instructional decisions, is described in Chapter 2 of this dissertation. The research about student-centered classrooms, especially relating to maintaining learning goals while allowing students to share in the mathematical authority

of the class, and using student work as the basis of a lesson, is described in Chapter 3. Chapter 4 provides background on the NWREL Mathematics Problem Solving Model and the goals of the larger research project of which this dissertation is part.

Chapter 5 describes the research methodology used. This includes a discussion of the treatment teachers underwent as part of the NWREL Mathematics Problem Solving Model, the tools provided to teachers as part of the model, and the data collection and analysis. The primary sources of data for this research were student solutions to problem solving tasks, teacher evaluations of student performance on those tasks using the NWREL formative feedback guide, and teacher reports of instructional plans using student solutions.

Three primary types of data analysis were used in the study. Teacher evaluations of student performance were compared to expert evaluation. The rationales teachers provide for selecting pieces of student work to share were categorized, as were the rationales for sequencing the pieces of work in a given order. The relationships between the pieces of student work selected for sharing, the rationales provided for selecting the student solutions, and the comments provided to the students by the teacher were described.

Chapter 6 reports the results of the study, and Chapter 7 discusses the study, summarizing and interpreting the results, identifying limitations of the study, and suggesting implications for teachers, professional developers, and researchers. I developed three new guides which are included in Chapter 7: 1) a new formative feedback guide, 2) a scaffolding guide, and 3) a debrief structuring guide. The new formative feedback guide is composed of stems for teachers to complete to provide

constructive, formative feedback to students on written solutions to their work. The scaffolding guide provides similar stems for teachers to use to give spoken comments while students are working on a task. The debrief structuring guide is a guide for planning a task debrief using shared student work, which maintains a mathematical goal. The last guide directly reflects Stein et al's (in press) framework for sharing student work in the classroom. Chapter 8 concludes the dissertation, summarizing the work.

Chapter 2 – Formative Assessment Literature Review

With the incorporation of the National Council of Teachers of Mathematics' *Curriculum and Evaluation Standards* (NCTM, 1989) into public school systems across the nation came a need for adjustments to assessment practices. The changed practices influenced by the standards are not always supported by traditional assessment methods. In response to this disconnect, the NCTM put forth its *Assessment Standards* (NCTM, 1995), prescribing the sort of assessment practices that would support reformed mathematics education. In this document, the NCTM identified four phases in the assessment process: planning assessment, gathering evidence, interpreting evidence, and using the results. These four phases, the NCTM emphasizes, are non-sequential and interrelated.

The phases in the assessment process serve as a loose framework for this chapter. As they are non-sequential and interrelated, it is not possible to divide the research literature into research on each of those four phases. It does somewhat parallel those phases, however, to consider the literature in four categories: selecting tasks for assessing student understanding, evaluating student responses, providing feedback, and using assessment information for instructional decisions.

The research literature indicates that the choice of mathematical tasks is essential for assessing student understanding. This chapter describes some of the characteristics of tasks identified to be useful in assessing student understanding and the benefits of such

tasks. It also acknowledges that such tasks are not yet fully integrated into classrooms and that teachers may find selecting or developing such tasks to be difficult.

Even given a quality classroom task, evaluating that task and providing information to students about their performance is another challenging area for teachers. The research literature suggests that teachers and students can both benefit from the use of rubrics to evaluate performance, but that being able to reliably use rubrics can require assistance. Useable, constructive feedback is shown to be beneficial to student performance, but those benefits are very dependent on the quality of the feedback. Professional development that helps teachers to compose such comments has been shown to improve teacher comments, change some classroom practices, and has been associated with improved student performance.

Perhaps the most challenging phase for teachers is making instructional decisions based on the data they collect from assessments. The research literature identifies the challenges, and investigates the results of some professional development designed to help teachers use assessment information to make instructional moves.

All four of these areas, selecting tasks for assessing student understanding, evaluating student responses, providing feedback, and using assessment information for instructional decisions, are documented to be challenging for teachers. The literature also shows that professional development can improve teachers' abilities in all areas, and thus affect student achievement and attitude improvements.

Definitions of Terms

In order to discuss assessment, it is first necessary to define the terms to be used in the discussion. The definition of *assessment* to be used originates in the NCTM's assessment standards (NCTM, 1995). Assessment is defined as *the process of gathering evidence about a student's knowledge of, ability to use, and disposition toward, mathematics and of making inferences from that evidence for a variety of purposes.* Assessment is further divided into summative and formative assessment. *Summative assessment*, sometimes called *assessment of learning* is a measurement of the level of accomplishment attained by a student (Woolfolk, 2004). *Formative assessment*, or *assessment for learning*, will be considered herein as any assessment that provides information to the teacher or the student about the student's performance, and that can be used to adjust practices, either on the part of the teacher or the student (Black & Wiliam, 1998; Wiliam, 2007). Further delineation of types of feedback will be included in this chapter.

Choosing Tasks for Assessment

The types of tasks that teachers choose for assessment purposes have a profound effect on the information gleaned from the assessment. The reform of mathematics education as articulated in the NCTM's (1989) *Curriculum and Evaluation Standards for School Mathematics* brought with it an impetus to reform the corresponding assessment. With a shift in curricular emphasis from procedural and algorithmic proficiency to reasoning and problem solving, assessment needed to follow suit.

In *Measuring What Counts,* the Mathematical Sciences Educational Board (MESB, 1993) provides a guide to mathematical assessment in the reform environment. The curricular shift toward problem solving and reasoning requires new methods of assessment. The MESB advises that assessment tasks should assess the mathematics that is most important for students to learn. This important mathematics includes nonroutine problems, mathematics that includes a communication component, and problems that require "habits of thinking" students' learn in their studies rather than techniques. The types of tasks that the MESB deems appropriate for assessing important mathematics are those that require students to articulate their thought processes, to draw connections between mathematical ideas, to consider mathematics in relevant contexts, and to experience solving nonroutine problems.

Performance Assessment

Performance assessment tasks are among the activities that can be used to evaluate students' problem solving abilities. Such tasks are complex, nonroutine tasks requiring reasoning, problem solving, and communication of thinking. Fuchs et al. (1999) investigated the use of performance assessment tasks on teachers' classroom practices. In the study, sixteen teachers were assigned randomly to two groups: those who were to use performance assessment-driven instruction in the classroom and those who were not. The teachers using performance assessment-driven instruction underwent professional development training on how to do so in the classroom.

The teachers involved in the study reported their perceptions of the amount of classroom time they spent on computation, math facts, word problems, and problem

solving after the study as compared to before. Positive effect sizes showed a greater decrease in time spent on computation and math facts for the performance assessment teachers than for those not using performance assessment. Moreover, positive effect sizes showed the performance assessment teachers reported a greater increase in teaching time spent on word problems and problem solving than the non-performance assessment teachers did. This is, however, based on teachers' perceptions of their own teaching time, not on some external measurement of time spent on each activity. Nonetheless, the teachers indicate that using such tasks focus more of their classroom time on those activities emphasizing reasoning and problem solving, as advised by the NCTM.

Assessing Students' Mathematical Understanding

While problem solving tasks help encourage more problem solving in the classroom, they may also give teachers better understanding of students' mathematical knowledge. Cai, Lane and Jakabcsin (1996) illuminated the importance of the task in assessing students' mathematical understanding. The authors compared a task in multiple choice format to a task assessing the same idea in an open-ended format (see Figure 2.1). The open-ended task requiring students' explanations resulted in a much deeper understanding of the students' knowledge and their misconceptions. In one student response shared in the chapter, the student chose .8 as the decimal with greatest value, and explained the solution in terms of place value, indicating quite clearly that he or she understood why .8 is larger than the others. Another student, however, chose the correct response, but justified it by stating "the more zero's, smaller number is." A third

student chose an incorrect response, and explained that "it doesn't matter how many zero's in front of the 8, it matters how many behind the 8 to figure out the place value."

While a multiple choice question would only have indicated whether students chose the correct or incorrect answer, the open-ended response provided important information to the teacher about student understanding. A correct response to a multiple choice question may not indicate that the student understands the concept. The open-ended responses provide indications of the students' understanding or misconceptions so that teachers can make instructional decisions suited to the students' needs. The opportunities for teachers to make instructional decisions are much greater with non-traditional tasks.

Figure 2.1. Multiple-Choice and Open-Ended Formats of a Task (Cai et al, 1996)

Multiple-Choice Format	Open-Ended Format
Which number has the greatest value? a. .08 b. .8 c. .080 d. .008000	Circle the number that has the greatest value. .08 .8 .080 .008000 Explain your answer.

Ball (1997) provides further evidence of the impact of the choice of task on teachers' understanding of students' knowledge. Ball describes a standard workbook task that asks students to shade one-third of a rectangle already divided into thirds, or three-fourths of a rectangle already divided into fourths. She describes a student who

completed that task successfully, but when given the same task with an undivided square or rectangle was unsuccessful. While Ball admits that the second task did not help her to understand *why* the student was incorrect, it did expose a lack of understanding that the more traditional first task did not. Ball describes an extensive classroom discussion that ensues from the incorrect response, and also explains that her own awareness of the misunderstanding guided her investigation of its roots in an in class discussion. Based only on the response to the first task, a subsequent in class discussion would have been far less likely to take place. The open-ended aspect of the task created far more instructional opportunities than a traditional task would.

For teachers whose curricula are focused on traditional tasks, finding opportunities to more deeply understand their students' mathematical knowledge is not always easy. In fact, Lester, Lamdin and Preston (1997) report that teachers find developing performance assessment tasks, or tasks that assess problem solving and reasoning skills, to be difficult. Quellmalz and Hoskyn (1997) outline the characteristics of tasks that may be suitable for assessing reasoning strategies. Such tasks should represent important issues and activities that recur in the curriculum. The reasoning required to complete the task should be sustained and integrate multiple skills and strategies. A task to assess reasoning should be able to be solved with multiple different approaches and strategies. The format of the task should require explanation of thinking and of the process the student used to solve the problem. The task should involve generalization or transfer of ideas and should require a coherent problem solving strategy.

Although it may be difficult to develop these tasks, Kulm (1994) showed that doing so has teaching benefits. Kulm studied teachers who developed and implemented

assessment tasks in their classrooms as part of a graduate seminar. Teachers involved in the seminar incorporated more higher-order thinking in their classrooms and reported that their own skills in recognizing student misconceptions, developing rubrics, evaluating problem solving skills, and understanding student thinking improved. The literature indicates, then, that implementing the right types of tasks in the classroom can help to promote assessment of students' mathematical knowledge and the reasoning and problem solving skills encouraged by the NCTM standards.

Evaluating Student Performance

Even given quality tasks designed to assess reasoning and problem solving, evaluating student performance on such tasks is not always easy for teachers. As part of their recommendations for assessment tasks to evaluate problem solving, the Mathematical Sciences Educational Board (1993) suggests that teachers may need scoring rubrics to help score nontraditional assessment activities.

Rubrics as Assessment Tools

In general, a rubric is defined to be a guideline for scoring a student response that lays out the expectations for the assignment given (Quinlan, 2006; Stevens & Levi, 2005). Rubrics may be holistic or analytic, generic or specific. A holistic rubric is used to evaluate the piece of work as a whole and assess an overall score or rating while an analytic rubric identifies components of the project to be evaluated and assesses a score or rating to each of those components. A generic rubric is one that may be used to

evaluate many tasks of a certain type, while a specific rubric is designed to be used for only one task.

Types of Rubrics

Stevens and Levi (2005) distinguish between *scoring guide* analytic rubrics and *three-to-five level* analytic rubrics. A scoring guide rubric uses categories, or components, on which the student work will be scored. For each of these categories, a scoring guide rubric identifies only the criteria for the highest level of performance. There is often room by each of the criteria for the rater to make notes and leave comments. A three-to-five level rubric uses categories as well, but for each of the categories the criteria for three to five levels of performance are identified. Three-to-five level rubrics will always list the criteria for the highest and lowest levels of performance. The criteria for one to three mid-range levels of performance will be listed as well. The number of mid-range levels depends on the range of performance the rubric developer expects on the task or tasks on which the rubric will be used.

Kulm (1994) also distinguishes further between types of rubrics. He suggests a rubric lying between holistic and analytic- the anaholistic rubric. Such a rubric assesses the work in different categories, but the scores for each of the categories are summed or otherwise combined to obtain a holistic score. Kulm also distinguishes a rubric by its purpose. A process rubric is designed to provide feedback on important mathematical processes such as reasoning, understanding problems, communicating, planning, and decision making. Process rubrics are a specific type of analytic rubric, as they require assigning a rating to each of the important mathematical processes. Kulm notes that to be

effective, a process rubric should describe each level of performance in each of the categories in sufficient detail so that qualitative judgments can be made. A process rubric could be used as an anaholistic rubric if the user chooses to sum or average scores or performance levels. Finally, it should be noted that Kulm uses *analytic* to refer to a task-specific rubric that scores steps or concepts in a specific mathematical task. This type of rubric breaks a task into steps in the problem solving process and scores the student on achieving certain of these steps.

Analytic rubrics, under the widely accepted definition (not Kulm's) provide more detailed analyses of performance and identify specific areas for improvement (Arter, 1993; Kulm, 1994; Quinlan, 2006). Perlman (2003) suggests that the more detailed information derived from rating using analytic rubrics is more useful to teachers than the information supplied by holistic rubrics for planning and improving classroom instruction. Kulm says that process rubrics, a specific variety of analytic rubrics, are "of the greatest importance for classrooms" (p. 92) because they enable teachers to evaluate progress on those important problem solving processes. Process rubrics also help students to understand mathematical thinking (Kulm, 1994).

Because Kulm's analytic rubric provides extremely focused information for rating mathematical processes, it tends to have high reliability. The task-specific criteria make this type of analytic rubric useful for large-scale assessment because they require less rater training than a more generic rubric might. Kulm's analytic rubric can provide feedback to students and teachers about specific mathematical concepts or procedural steps, but it doesn't give as useful feedback about the students' general problem solving skills because it is so focused on the specific steps in solving that task.

Stiggins (2001) and Arter and McTighe (2001) make suggestions for when task-specific rubrics or generic rubrics are more appropriate. Stiggins asserts that the best practice for using a rubric in the mathematics classroom is to have a generic rubric that is to be used to assess problem solving, and then teachers and students together should translate that rubric into a task-specific one to be used for a given task. Stiggins recognizes that task-specific rubrics can be very useful for evaluating a certain problem, and that the precise language used in them is helpful in communicating expectations to teachers and students. However, a task-specific rubric may not be something that can be shared with students before they solve a problem because it may give students too much information on how to approach the task. Furthermore, a task-specific rubric explains what is necessary for a good solution to a specific task, not what good problem solving looks like in general. For this reason, Stiggins indicates that it is best to build from a generic rubric.

Arter and McTighe (2001) assert that generic rubrics should be used to help students understand quality performance in generic, to assess complex skills that generalize across tasks, and when teachers are trying to consistently assess performance from students in different grades or at significantly different ability levels. The disadvantages of generic rubrics are that they are more difficult for teachers and students to learn to use and that, to be used on a specific task, they may still require some task-specific articulation.

Task-specific rubrics are more appropriate, according to Arter and McTighe, when it is important to teachers to score student work quickly, when teachers are assessing student knowledge of specific facts, equations, or methods, or when consistent

scoring is most important. Arter and McTighe echo Stiggins' concern that task-specific rubrics can not be shared with students beforehand without giving away solutions. In addition, the need to develop a new task-specific rubric for each task, the focus on a task-specific rather than general picture of quality, the possibility for correct solutions to be omitted from a task-specific rubric, and the lack of thinking about student understanding on the rater's part are among the disadvantages of a task-specific rubric.

Despite the disadvantages of using task-specific rubrics, there is general agreement that a rubric should be designed with a task or type of tasks in mind. There are several common threads in the literature on rubrics to be used in assessing tasks in all subject areas, including mathematics. Rubric design should be driven by objectives, whether they be learning objectives for a lesson, unit or course, or even state and national standards (Kulm, 1994; Leitze & Mau, 1999; Perlman, 2003; Quinlan, 2006; Stevens & Levi, 2005). Rubric design should include some reflection on what student work at the highest and lowest levels would look like (Kulm, 1994; Leitze & Mau, 1999; Perlman, 2003; Quinlan, 2006; Stevens & Levi, 2005). Rubric design should give attention to establishing well defined criteria statements for student performance using precise language (Perlman. 2003; Quinlan, 2006; Stevens & Levi, 2005).

Quinlan (2006) and Perlman (2003) both advise that rubric developers use samples of student work for the task, or type of tasks, they intend to evaluate to ensure that they have identified all important dimensions of the tasks and accurately assessed performance levels. Both Quinlan (2006) and Perlman (2003) suggest that rubrics should be evaluated, shared with colleagues and students, and revised during the development process. Stevens and Levi (2003) also recommend including colleagues and students in

the development process. The rubric development process may also often include modification of a pre-existing generic rubric to suit the purposes of a given task (Arter, 1993; Kulm, 1994; Quinlan, 2006; Stevens & Levi, 2005).

Development of rubrics for evaluating mathematical problem solving is based on the same foundation, but adds some more subject specific steps. Leitze and Mau (1999) address the steps in developing a problem solving rubric for a task. The first step in this process is solving the problem to be evaluated. This enables the evaluator to identify the critical elements of the problem and of the problem solving process. Next the rubric developer must "identify the phases of the problem solving process that are of interest." (Lietze & Mau, 1999 p. 305). This step determines the categories that will appear in the rubric. The third step is to determine the levels of work in each phase of the process so that student work can be evaluated at each phase. It should be noted that the authors' sample rubric does not use phases that depend upon the problem. Rather, they evaluate using phases that would likely appear in a wide range of tasks, such as 'understanding the problem,' or 'selecting the data to solve the problem.' The final step in Leitze and Mau's process is to assign point values to each phase. While this development process is clearly outlined, Leitze and Mau leave room for revising the resulting rubric. They present a situation in which a piece of student work being evaluated with the rubric identifies a weakness in the rubric, and suggest that the rubric could be rearticulated to better suit this newly realized aspect of the student work.

Teachers may need guidance in developing and implementing rubrics in their classrooms. Quinlan (2006) advises using model situations with teachers in which the teachers begin by evaluating a very simple task using an existing generic rubric. They

then use every day situations to create simple rubrics. Quinlan suggests teachers develop a rubric for evaluating the quality of a potato chip or a child's performance in cleaning his room. These simpler tasks provide teachers with insight into the process of comprehensively identifying the components of a task and exemplifying levels of performance.

Reliable Use of Rubrics

There is evidence in the research literature that teachers may need help using externally generated rubrics, as well. Koretz (1992, 1993) investigated rater reliability in a pilot study of the Vermont Portfolio Assessment Project. The Vermont Portfolio Assessment Project had 4th grade and 8th grade students prepare portfolios of five to seven of their best pieces of mathematical work on problem solving tasks. In the first year statewide implementation of the project, these portfolios were evaluated by trained teachers using a four-point scale in seven different areas of competency: language of mathematics, mathematical representations, presentations, understanding of the task, how: procedures, why: decisions, and what: outcomes.

The teachers had a generic rubric to use to determine how to assign scores from the four-point scale. For each student rated, each piece in the portfolio was rated on the seven criteria and then the scores from the five best pieces were combined using an algorithm developed by the mathematics portfolio committee, giving the student a single composite score for each of the seven criteria. The scores of independent raters were compared using reliability coefficients where a coefficient of 1.0 represents perfect rater

correlation and a score of 0.0 means no correlation. Table 2.1 summarizes the reliability coefficients in 1992, showing poor interrater reliability.

Table 2.1. Interrater Reliability Coefficients (Koretz, 1992)

Criteria	4th	8th
Language of math	.23	.28
Math representations	.33	.31
Presentations	.45	.42
Understanding of task	.26	.35
How: Procedures	.44	.30
Why: Decisions	.40	.31
What: Outcomes	.23	.35
Average	.33	.33

Koretz (1992) also explains the interrater reliability by looking at the percentage of student papers for which raters assigned the same score, as summarized in Table 2.2.

Table 2.2. Interrater Reliability Percent Agreement (Koretz, 1992)

Criteria	4th	8th
Language of math	52	49
Math representations	55	56
Presentations	54	51
Understanding of task	75	76
How: Procedures	66	62
Why: Decisions	47	49
What: Outcomes	81	89
Average	61	62

Here the average for both 4th and 8th grade papers is slightly over 60%. Koretz explains that this discrepancy is due to scores being heavily concentrated at one or two points on the four point scale. For example, the 8th grade <u>what: outcomes</u> criterion has a reliability coefficient of .35 and a percentage agreement of 89. But here, the first rater scored almost 92% of students as a 1, while rater 2 scored 93% of those 92% a 1. This could represent interrater reliability, but it could also be simple chance. Given the

concentration of scores here, pairing scores at random would result in an agreement rate of 85%, not far from the actual 89% of agreement.

While this study of the Vermont Portfolio Assessment Program did not have teachers scoring their own students' work in the classroom, the trained scorers were classroom teachers. Koretz suggested that the lack of reliability among the raters indicates that teachers may have some difficulty understanding the criteria or what evidence of a strong performance would look like. The 28 scale points on the rubric (four points on each of seven criteria) may be too complicated, or teachers may not be able to distinguish between the seven criteria. Koretz hypothesized that the teachers using the rubric may also have needed more training in its use.

In response to the lack of rater reliability in scoring portfolios for the 1991-92 school year, some changes were implemented for scoring portfolios from the 1992-93 school year (Koretz, 1994). Among these changes were that all scoring of portfolios was conducted over a five day period in a single location, so all raters were together, and the raters engaged in twice daily calibration sessions. For the 1992-93 school year, reliability did improve. The reliability for a single criterion ranged from a low of 0.35 in <u>what: outcomes of activities</u>, to a high of 0.60 in <u>mathematical representations</u>. The average reliability improved to 0.46 in 4th grade and to 0.50 in 5th grade. When scores were totaled, the reliability improved from 0.6 in 1991-92 to 0.72 in 1992-93 for 4th grade, and from 0.53 to 0.79 in 8th grade. An interrater correlation of 0.80 or higher is usually considered to be sufficiently strong for performance assessment.

Petit and Zawojewski (1997) discuss instructing teachers in use of the reformed Vermont state math assessment rubric. Problem solving in Vermont state testing was

then evaluated on five dimensions: 1) finding an effective problem solving approach, 2) demonstrating accuracy in executing the associated procedures, 3) making connections among the mathematical ideas in the problem solved and other mathematics experiences, 4) communicating with appropriate representations, and 5) presenting clear and complete solutions. The authors suggest that teachers begin to learn the rubric by focusing on one of the five categories. Teachers should be guided in a discussion of the highest level of performance for that category and should score sample student works together. This requires that teachers have access to problems that are appropriate for evaluation with the rubric and access to benchmark student papers for such problems. Furthermore, because the problems to be evaluated with the rubric are complex, teachers need access to good problems to use with their students and networks of communication with colleagues to share problem solving experiences.

The research shows, then, that choosing and developing good assessment tasks and evaluating them can be challenging for teachers and can require external assistance.

Facilitating Teacher Assessment Practices

In 1990, the American Federation of Teachers (AFT), the National Education Association (NEA), and the National Council on Measurement in Education (NCME) published "Standards for Teacher Competence in the Educational Assessment of Students." (AFT, NEA, NCME, 1990). The standards established seven areas of assessment competency for teachers:

1. Choosing assessment methods appropriate for instructional decisions
2. Developing assessment methods appropriate for instructional decisions

3. Administering, scoring, and interpreting the results of both externally produced and teacher produced assessment methods
4. Using assessment results when making decisions about individual students, planning instruction, developing curriculum, and improving schools.
5. Developing valid pupil grading policies
6. Communicating assessment results to students, parents, and other lay audiences, and educators.
7. Recognizing unethical, illegal, and other inappropriate methods and uses of assessment information

In 1992, the NCME set out to determine the assessment literacy of teachers in the U.S. (Plake & Impara, 1997). A multiple choice test consisting of a total of 35 questions, five in each of the seven competency areas, was created. The researchers contacted each state's education agency and requested a contact person in each of four randomly selected school districts. They then contacted the school districts and asked that 12 teachers across elementary, middle, and high school levels be randomly identified to participate in the survey. The multiple choice test was sent to each of the teachers identified.

Usable surveys were returned by a total of 555 teachers in 45 different states. The average score on the test was 66%. Teachers scored highest on the five questions about administering assessment, and lowest on the five questions about communicating assessment results. In addition to taking the test, teachers were asked to respond to survey questions, including one about course work in educational measurement. The researchers found that teachers who had taken a course or undergone professional

development in measurement scored significantly higher on the test than did teachers who had never had such a course. According to Plake and Impara (1997) the surveys indicated "woefully low levels of assessment competency for teachers." (p. 67). The superior performance of teachers who had coursework in measurement indicates that teachers may need, and may benefit from, training in assessment.

The research evidence shows that professional development and other guidance for teachers around assessment can be beneficial. The Multicultural Reading and Thinking Program (McRAT) through the Arkansas department of education was a project to improve the reasoning skills of students, especially in the context of writing (Quellmalz & Hoskyn, 1997). This program had a significant professional development component wherein teachers were trained in all aspects of the project, including those related to assessment. The professional development was based on research studies, including Guskey, Osterman & Kottkamp, and Sparks & Simmons (as cited in Quellmalz & Hoskyn, 1997), indicating the need for extensive training, including modeling of effective practices, to achieve sustained change in teaching practices.

A portion of this professional development focused on teachers' assessment of reasoning strategies. Teachers spent a significant part of the training watching the instructor model scoring and also practicing scoring student work themselves until they could do so reliably. Teachers continued to receive practice scoring throughout the course of the program to ensure that they assessed their own students' work accurately. Through the professional development, teachers become proficient at assessing student reasoning by using the rating guides provided in the program.

Quellmalz and Hoskyn (1997) and Plake and Impara (1997) both indicate that some training in assessment, either through coursework or through professional development can improve teachers' abilities to assess student performance, including problem solving performance.

Feedback

One area of problem solving assessment that may merit extensive teacher training is feedback. The use of the term feedback in the educational setting is based on the notion of feedback as it was used in systems engineering (Wiliam, 2007). In this context, feedback refers to information that allows evaluation of a process and evaluation of adaptations. Based on the view of Wiliam (2007), this chapter will distinguish between *feedback,* any information provided to the student about performance, and *formative feedback*, information that identifies a difference between a student's performance and the ideal performance, but also provides some instruction, either implicit or explicit, on how to improve performance. Thus a grade would be feedback to a student, but not formative feedback. A written comment indicating a specific error could be considered formative feedback. It is important to note that what constitutes formative feedback in a particular classroom can be dependent on classroom norms. For example, in a classroom where standards of explanation have been clearly set, a comment like "give more detail" would be considered formative, but not in a classroom where the qualities of a complete explanation were not clearly established (Wiliam, 2007).

Among the aims of effective assessment, particularly that around problem solving, is to provide formative feedback to students. There is ample research evidence

that formative feedback is beneficial to student learning. Several meta-analyses of the research on feedback exist. Bangert-Drowns, Kulik, Kulik, and Morgan (1991) examined 40 studies of feedback in "test-like" events. The meta-analysis of these studies revealed that students benefit from feedback in the form of answers to questions only if they can not see the answer prior to attempting the question. Moreover, feedback that explained a correct answer was more effective than feedback that simply indicated correctness of a response. Combining the two aforementioned types of feedback into explanatory feedback presented after students had attempted the problem resulted in a mean effect size of 0.58 standard deviations. It should be noted that not all of the feedback provided in the individual studies is formative. The feedback provided to students in the studies considered in Bangert-Drowns et. al.'s (1991) meta-analysis indicates correctness but does not always provide students with instructions for improvement.

In contrast to the positive effects of feedback indicated in Bangert-Drowns et. al.'s (1991) analysis, Kluger and DeNisi's (1996) meta-analysis of feedback studies showed some *negative* effects of feedback. Kluger and DeNisi (1996) analyzed 131 reports regarding the effects of feedback on student performance. These reports included 607 effect sizes and examined a total of 12,652 participants. While the average effect size was 0.4, there was a standard deviation of almost 1 among the effect sizes, and it was found that nearly 40 percent of the effects were negative, meaning that feedback had a negative effect on student performance. However, the definition of feedback used to identify studies for inclusion was any collection and reporting of data. This would include grades or scores, information that certainly does not indicate strategies for improvement. The negative effects shown in Kluger and DeNisis's meta-analysis may be

attributable to the effects of grades, shown by Butler (1988), as discussed later in the chapter, to be detrimental to student performance. In fact, Kluger and DeNisi's analysis does not account in any way for the effects of the different types of information included under the umbrella of feedback.

Nyquist (as cited in Wiliam, 2007) conducted a meta-analysis of 185 studies of feedback in higher education that helps to address the effects of different types of feedback. He classified feedback into five categories:

- *Weaker feedback*: Feedback which only gives the student knowledge of a score or a grade
- *Feedback*: Feedback which gives the student knowledge of a score or grade and also provides clear goals to work for or knowledge of correct results, often in the form of correct answers to questions the student attempted.
- *Weak formative assessment*: The student is given information about the correct results along with some explanation
- *Moderate formative assessment*: The student is given information about the correct results, some explanation, and some suggestions for improvement
- *Strong formative assessment*: The student is given information about correct results, some explanation, and suggestions for specific activities to undertake in order to improve.

Nyquist categorized the studies in his analysis into those five categories based on the type of feedback students received. He then examined the average effect size for the studies in each category. The results are summarized in Table 2.3. All types of feedback show positive effect sizes, but strong formative assessment shows the greatest effect size.

Table 2.3. Effect Sizes of Types of Feedback (Nyquist, in Wiliam, 2007)

Type of feedback	**N**	**Effect**
Weaker feedback	31	0.16
Feedback	48	0.23
Weaker formative assessment	49	0.30
Moderate formative assessment	41	0.33
Strong formative assessment	16	0.51
Total	185	

It should be noted that strong formative assessment is the only category of feedback among Nyquist's five that would certainly be classified as formative feedback under the definition of this chapter, independent of classroom norms. Feedback, weaker formative assessment, and moderate formative assessment as defined by Nyquist may also qualify as formative feedback, depending on the classroom norms and the ability of students to make use of the information. Nyquist's meta-analysis does indicate that the quality of feedback students receive has a significant effect on achievement gains.

Elawar and Corno (1985) studied the effects of constructive comments on the performance of middle school students. Eighteen sixth-grade teachers in three Venezuelan schools received two days of professional development training in providing constructive written comments on homework to students. These written comments were to comment specifically on a student error, provide a suggestion on how to improve, and include at least one positive comment on the student's work. The teachers were encouraged to use four questions to prompt them in preparing feedback:

1. What is the key error?
2. What is the probable reason the student made this error?

3. How can I guide the student to avoid the same error in the future?
4. What did the student do well that could be noted?

The authors include an example of a constructive comment that addresses these key questions: "Juan, you know how to get percent, but the computation is wrong in this instance… Can you see where?"

The teachers were then randomly assigned to three groups: a treatment group that would provide constructive comments to their students, a control group that would provide the type of feedback that was typical in Venezuelan schools (i.e. just feedback about the number of problems answered correctly), and a "half-class" treatment group that would provide constructive comments to half the class and traditional feedback to the other half. The students were assigned to treatment in the half-class treatment group at random. The teachers were to collect and provide feedback on homework three times per week for ten weeks.

The researchers had teachers in the group and external experts evaluate the constructive comments given by the treatment and half-class treatment teachers. The feedback was consistently judged to be appropriate and responsive to the key questions. All students took pre- and post-tests and responded to attitude surveys. Students in the treatment and in the half-class treatment who received constructive comments outperformed control students with similar pre-test scores on achievement post-tests. They also reported more positive attitudes about mathematics after the ten week period than control students. The percentage variance in achievement that the researchers could attribute to being in the treatment group was 24%, and the percentage variance in attitudes attributable to treatment was 57%. The results of Nyquist's meta-analysis and

Elawar and Cordo's study indicate that formative feedback has been shown to improve both student achievement and student attitudes about mathematics.

Butler's Study of Grades as Feedback

Not only is formative feedback shown to be beneficial to students, but grades have been shown to have the potential to be detrimental to student performance. Butler (1988) investigated the relationship between the type of feedback given to students and their performance. Students from twelve 5th and 6th grade classes were given a booklet of divergent thinking tasks requiring creative thinking to complete. Students from four of the classes received feedback on their work in the form of a score between 40 and 99, four of the classes received only written comments, such as "you thought of quite a few correct words; maybe it is possible to think of some longer words," and four of the classes received both written comments and scores. When the work was returned, the students were asked to complete a second set of tasks and were told they would receive the same type of feedback on the second set as they received on the first. The work was again evaluated.

Butler examined the performance of the top quarter of students and the bottom quarter of students from each class, as determined by the students' class grades in mathematics and language at the beginning of the project. The study showed that students receiving only grades did not improve from the first task to the second. The students receiving only comments improved, on average, 30% on the second task. Students receiving grades and comments showed no improvement from the first task to the second.

Butler also looked at students' interest level in the tasks for which they were asked to perform. For students receiving only scores, the lower achieving students did not show interest in the second task, while the higher achieving students were still interested in the task. Both high-achieving and low-achieving students receiving only comments maintained interest in the second task, and some of them even showed increased interest. High-achieving students receiving both scores and comments reported interest in the second task, but low achievers receiving both scores and comments lost interest in the second task. Butler reported that students who received scores and comments only recalled the score they received. Assigning grades to students effectively overshadowed any comments teachers wrote.

Beneficial Feedback

Nyquist's meta-analysis and Elawar and Cordo's and Butler's studies indicate that the quality of feedback is critical in effecting student achievement gains. Formative feedback that provides students with instructions for improvement appears to be the most effective. To qualify as formative, however, feedback must be used by students to adapt their performance and by teachers to adapt their teaching (Wiliam, 2007). In her book, *Language for Learning Mathematics,* Lee (2006) states that feedback is only beneficial when the learner receiving the feedback uses it to improve. This confirms that the teacher must provide formative feedback that suggests actions that the student can take to improve performance. But to achieve the full effect of the feedback, students must be provided with opportunities to act on those suggestions. Students who understand the

learning goals of a task and understand how to recognize quality performance are more able to make use of formative feedback.

Black, Harrison, Lee, Marshall and Wiliam (2003) echo the notion of a feedback loop with the student as an important participant. They identify four key components in the operation of feedback: 1) feedback must provide data on the actual level of a measurable attribute, 2) feedback must provide data on the desirable level of that attribute, 3) there must be some mechanism for comparing the actual and desirable levels and measuring the gap between them, and 4) there must be a mechanism that can be used to bring the actual level closer to the desirable level. That is to say, in order to be effective, formative feedback needs to let students know how they measure up to a standard of performance that they understand. It also needs to provide them with some instruction on how to improve their performance to be closer to the standard.

Both Lee (2006) and Black et. al.(2003) recognize that providing the types of feedback they describe is not always easy for teachers. To help teachers improve formative feedback, Lee provides characteristics of useful comments and suggestions for formulating such comments. Black et. al. developed a professional development program to assist in the process. Lee suggests useful formative feedback comments should be focused on the learning goal, not the student. They should be about the mathematical learning, not the presentation of the solution. Comments should be clear about the student's level of performance and about what the student needs to do to improve. Comments should demand a response from the student and indicate what the student must do to respond.

Lee outlines types of comments that may be used to help students improve. Teachers may use comments to:

1. remind students of the learning goals or the criteria for successful performance, such as "remember what happens when you multiply two negatives," or "you have forgotten to multiply by the coefficients of *x*."
2. scaffold student responses by suggesting a direction to follow, like "do you think the length of this side is right? Should it be longer or shorter than the others?"
3. provide an example and have the students determine how to use that example, as in "use a table to work out the values for *x*. What numbers will you choose for *x*?"

The professional development program developed by Black et al. (2003) assists in the implementation of formative assessment in the regular practices of teachers, and investigates the advantages of doing so. The program involved 24 British teachers, two science and two mathematics teachers each from six schools with students in year 7, 8, or 10 (corresponding to American grades 6, 7, or 9). The six schools involved were chosen by local education authorities based on the researchers' criteria that the schools be "average," and the teachers included in the study were chosen by the department heads in their schools.

The professional development program was comprised of eleven full day and one half day inservice, as well as quarterly 'expert' visits to the classroom. The first three inservice sessions took place before the 1999-2000 school year began and was focused on developing action plans where teachers determined which aspects of formative

assessment they wanted to incorporate into their teaching practices. During this time, researchers introduced the teachers to the research on formative assessment and discussed means of introducing formative assessment into practice. The following four professional development sessions took place over the 1999-2000 school year. Over the course of the year, researchers visited the teachers' classrooms, and the teachers experimented with formative assessment techniques. Those four sessions, then, revolved around refining action plans based on the teachers' classroom experiences. They also provided opportunities for teachers to share ideas and discuss their experiences with the project.

As part of the action plans, teachers identified their own areas of focus in formative assessment, identifying areas such as effective questioning, providing comments instead of grades for feedback, sharing learning goals and success criteria with students, and student self-assessment. The remaining five professional development sessions involved incorporating new teachers into the program and determining how to implement the program throughout the schools.

Black et al. (2003) provide qualitative evidence of the teachers' changing practices, especially in relation to feedback. Early in the project, teacher feedback was typified by comments that were generally evaluative or were related to presentation. As teachers progressed, they reported that using comments to inform students what they have achieved and what they needed to do next became an important part of their teaching. They acknowledged the time consuming nature of feedback, but they also determined comments useful to students to be important enough to set aside time.

Quantitative results showed that students taught by teachers developing formative assessment practices outscored students in control classes within the same schools on externally produced and internally produced tests by 0.3 standard deviations. Wiliam (2007) notes that one year's growth in mathematics measured in the Trends in Mathematics and Science Study (TIMSS) is 0.36 standard deviations. Therefore, students taught by the teachers practicing formative assessment almost doubled the learning of control students, in the sense that they were achieving nearly an extra year's worth of growth in addition to the regular yearly growth in mathematics.

Black et. al.(2003) and Lee (2006) are not the only researchers addressing the difficulties teachers face in providing formative feedback to students. Even (2005) also recognized the need and used portfolio building to guide teachers in improving formative assessment practices. Sixteen 5th and 6th grade teachers chosen by their principals to lead mathematics education in their schools participated in a 5 month course on non-negative rational numbers. The course combined the content component of non-negative rational numbers with exposure to research in student understanding of the concept, and the misconceptions that students have in the area, as well as possible explanations for those misconceptions. The teachers were instructed, as a final project for the course, to create portfolios consisting of three pieces of student work, the teacher's assessments of the students' mathematical knowledge for each piece of work, and the teacher's reasons for including each piece of work.

Even claims that the portfolios showed that teachers shifted from assessing the final answer given by a student to assessing the solution process. The teachers reported that they were more likely to accept "non-conventional" answers as correct due to their

shift in focus toward assessing understanding. Teachers also used more open-ended tasks in their classrooms in order to assess student thinking. Thus professional development, like that implemented by Black et. al. (2003) and by Even (2005) appears to have the potential to improve teachers' beliefs about the usefulness of feedback, improve student achievement, and shift teacher attitudes from valuing solutions to valuing students' mathematical processes.

Effective Assessment Practices

While teacher educators can suggest and test many ways for teachers to improve assessment practices, it is also important that these practices be effective in the classroom and usable by teachers. Leahy, Lyon, Thompson, and Wiliam (2005) worked in collaboration with a group of K-12 teachers to attempt to determine what types of assessment practices were effective. The teachers participated in three days of professional development, studying assessment for learning and practices that could be implemented in the classroom. The teachers then tested out practices in their classrooms, and researchers met monthly with the teachers to discuss their experiences, to suggest ways to improve practice, and to observe the teachers' classrooms. Different strategies worked better for different teachers, but the researchers identified five broad strategies that they found to be equally effective at all grades and in all subjects. These strategies are:

1. Clarify and share intentions and criteria.

   These techniques include sharing and explaining scoring guides to students and having students analyze sample student works.

2. Engineer effective classroom discussion.

   This involves asking questions that prompt students to think deeply or that provide teachers with information about the students' understanding that can be used to shape instruction.

3. Provide feedback that moves learners forward.

   Comments written to students should address what the student must do to improve and also that cause students to think about their work. Linking these comments to rubrics was found by many teachers to be effective.

4. Activate students as owners of their own learning.

   Self-assessment and self reporting of confidence in a student's own knowledge can be helpful tools.

5. Activate students as instructional resources for one another

   Using rubrics to assess another student's work can be powerful in helping students to understand the rubric, what high quality work looks like, and how to evaluate their own performance.

While Leahy et al. determined that a set strategy for incorporating formative assessment into the classroom did not really exist, they have, with the help of teacher insights, composed a set of guidelines that can prove useful. These guidelines confirm and summarize much of the other research: that student understanding of performance standards is important, that feedback must be usable by the students, and that rubrics can be an important tool for both teachers and students in this process.

Using Assessment Information to Make Instructional Decisions

Knowing what sorts of tasks to provide opportunities for formative assessment and understanding how to assess and provide formative feedback to students is not sufficient. Teachers need also to be able to provide themselves with feedback that they then make use of for instructional decisions. Using the information gleaned from assessments to adjust instructional plans has not always been well incorporated into regular teaching practices. Peterson and Clark (1978) studied teacher practices when working with unfamiliar, volunteer student subjects. Some of the teachers involved in the study would adjust their instructional actions in reaction to student responses, while others did not. By investigating the teachers' attitudes and beliefs about teaching mathematics, Peterson and Clark were able to associate teacher's responses with their teaching beliefs. A lack of adjustment to instruction after an unsatisfactory student response - even when alternative methods of teaching the subject were available - tended to be associated with content-oriented planning, while adjustment tended to be associated process-oriented planning. Content-oriented planners are more concerned with correct solutions, while process-oriented planners are more concerned with reasoning and strategies. Moreover, the students whose teachers did not adjust instruction performed worse on achievement tests than those whose teachers did adjust, and they expressed poor attitudes toward the topic. While this study does not investigate students in a normal classroom setting, it does imply that failure of teachers to adjust instruction based on feedback from the student can be detrimental to students.

Putnam (1987) investigated the practices of teachers in a tutoring setting to better understand how teachers determine student knowledge and how they use that information

to adjust instruction. Putnam determined that detailed diagnoses of what students understand and what misconceptions they have were not a consistent part of the teachers' pedagogy. Moreover, the teachers in the study appeared to be gathering only enough information about student understanding to determine which steps in the curriculum were difficult for the student. The teachers would then reteach that part of the "curricular script" rather than using a different teaching technique or attempting to determine the roots of the students' misunderstandings.

Recognizing the difficulties teachers have colleting and integrating assessment information into classroom decisions, researchers have attempted to help make this process easier for them. Fuchs, Fuchs, Hamlett and Stecker (1991) used computer software in an attempt to improve teachers' adjustment of instructional practices. They used Curriculum-Based Measurement (CBM) programs in conjunction with Expert Systems (ExS) to help teachers make decisions. On its own, CBM has students take computerized tests composed of computational exercises. The program indicates a student's overall proficiency level based on the number of digits the student had correct in the skills test. The program will chart student performance over time, and allows a teacher to compare student growth in proficiency before and after instructional changes. CBM also provides a skills analysis that identifies a student's level of mastery in certain "mastery categories." When performance does not meet target improvement rates, CBM will notify the teacher that an instructional change should be made.

The ExS is a computer program that attempts to reproduce the instructional advice that a live expert might provide to a teacher, given a student's proficiency level and skills analysis. The program's responses are based on a network of rules for problem solving

developed by experts. A teacher can enter the CBM information on mastery and proficiency levels, percent of correct digits, the graphed pattern from the CBM, a judgment of the quality of the student's daily work, and information about the teacher's curricular priorities. The ExS then makes recommendations for actual instructional adjustments that the teacher could make, and then teachers can examine the resulting trends in student performance from those adjustments using CBM.

In Fuchs et al.'s study, 33 teachers in 15 schools in a southeastern metropolitan area were randomly assigned to three groups: teachers using both CBM and ExS, teachers using CBM without ExS, and control teachers using neither tool. The teachers chose two of their students in grades 2-8 who were chronic low achievers and had been officially identified as learning disabled or emotionally disturbed. Control teachers followed normal practices with their students. CBM-only teachers evaluated their students' performance using CBM and then made their own instructional changes when CBM notified the teacher that a student had not met the target improvement rates. CBM and ExS teachers would follow the instructional changes suggested by ExS when CBM notified the teacher that an instructional change was necessary.

The study showed that teachers using CBM, with or without ExS, made more instructional adjustments than did control teachers. Only a combination of CBM and some kind of expert consultation to support the teachers' instructional adjustment resulted in differential student performance. CBM alone did not. Teachers using both CBM and ExS showed greater student achievement (as measured on CBM tests) over a comparable number of instructional changes than did CBM teachers not using ExS, and control teachers.

Fuchs et al.'s study shows that help identifying student needs for instructional adjustments and help deciding how to adjust can be useful for teachers' instructional practices and for improving student achievement. However, this study investigates this assistance only on students identified as low achievers, either learning disabled or emotionally disturbed. There is no sense of how such a software system would affect the performance of more average students. Moreover, the problems involved in CAM are short, computational items. Making instructional decisions based on problem solving activities is probably a much more challenging activity, and may not be something software would even be able to work with.

The difficulty of making instructional decisions based on student responses to tasks is highlighted by Farrell's (1992) discussion of students' incorrect factoring of sums of squares. Farrell suggests that students may have created their own algorithms or rules that are correct for factoring other binomials, and that typical teaching tools for exposing student misconceptions about factoring may not be effective here. In order to understand student errors, Farrell asserts that teachers should carefully analyze those errors. Focusing on understanding the errors and the reasons for making the errors, as well as looking for what students *do* know, as opposed to what they don't can provide more insight into student errors. Farrell notes that asking students to write about their understanding can help illuminate misconceptions, but, it is important that teachers emphasize that such writing should be in the student's own words. Requiring students to reflect on their learning and to explain their reasoning may help teachers to better understand what their students know. It seems that open-ended tasks provide more

opportunities for students to explain their reasoning and may also help teachers to better discern the instructional changes necessary to suit students' needs.

The Classroom Assessment in Mathematics Project

In response to the difficulties faced by teachers in gathering and using assessment information to make instructional decisions, the Classroom Assessment in Mathematics (CAM) program was developed. CAM sought to help teachers to better understand how to use assessments to evaluate what students know and what they need to learn, select and invent curriculum and instructional strategies to meet students' needs, and to discern what constitutes worthwhile mathematics (Driscoll, 1999). The project was initially field tested between 1991 and 1993 in six cities. The 24 middle school teachers involved, four from two different schools in each city, participated with their supervisors in three national assessment workshops. One five-day session was held at the start of the first year, one three-day session was held at the end of the first year, and one two-day session was held at the end of the second year. The workshops focused on developing assessment tasks, scoring student work, and interactions with assessment consultants. Teachers would also meet locally in small groups to discuss the effectiveness of and make decisions for revising their assessments.

Teachers in the project were asked to choose tasks that would help them to make inferences about their students' understanding in several areas. In particular, an assessment task should help them make inferences about 1) the evidence in the student work of mathematical learning, 2) the student's opportunity to learn, 3) the suitability of a task's prompt to the task's purpose, and 4) the evidence in the student work of

misconceptions or misunderstandings. Teachers should consider what a particular piece of student work reveals about student understanding and what it reveals about the student's achievement of the desired outcome.

The researchers documented reflective discussions among teachers, reflecting on the quality of assessment tasks and the information about student understanding to be gleaned from a student response (Driscoll, 1999). The CAM field study was evaluated by external evaluators, and the evaluators determined that the teachers involved in the program benefited from the freedom they had to experiment with tasks, and the greatest benefit was in the form of increased understanding and appreciation of the usefulness of assessment information in making classroom decisions.

The evaluation indicated that there was variation in the amount of instructional change exhibited by teachers in the program, and suggested that teachers may need more teacher-to-teacher mentoring or classroom coaching to move from appreciation of assessment to integration of assessment in classroom practice. The teachers that participated in the project reported that it compelled them to add more problem solving tasks to the curriculum and to transition to a more student-centered classroom.

Forming reflective groups of teachers is a common theme in the research on promoting integration of assessment into the classroom. The Using Assessment in the Service of Instruction project also investigated how mathematics teachers can better use assessment to shape instructional practices (Wilcox & Zielinski, 1997). The project attempted to help teachers make sense of what and how their students understand, and what evidence exists of that understanding. Teachers in the project used analysis of materials that document student understanding to determine the next instructional move.

Teachers would meet to discuss assessment tasks they had used in their classrooms that they thought had been particularly effective or particularly problematic. Teachers would bring evidence of the assessment in the forms of sample student work, or audio or video recordings of whole class or small group discussions. The large group discussions would help to illuminate unseen weaknesses of tasks used for assessment or teachers' own misinterpretations of student understanding.

The results documented by Fuchs et. al., Driscoll, and Wilcox and Zielinski confirm that professional development can be beneficial in helping teachers to accurately draw conclusions from student assessments, and also make suitable instructional changes.

Summary

The research literature indicates that focusing on formative assessment can positively change classroom climates and improve student achievement. Yet the transition to formative assessment practices is not an easy one. Some of the key formative assessment practices that are particularly challenging for teachers are selecting tasks for assessing student understanding, evaluating student responses, providing feedback, and using assessment information for instructional decisions.

Research in all four aspects of assessment indicates that teachers can benefit from professional development that focuses on improving formative assessment practices. Helping teachers to select and create quality assessment tasks, to evaluate student performance on those tasks, to provide feedback to their students, and to make instructional decisions based on assessment results can lead to practices that are

associated with improved student achievement and attitudes and classrooms that are more in line with reform standards.

Chapter 3 – Student-Centered Classrooms Literature Review

The preceding chapter examined some of the effects of mathematics education reform on changing assessment practices. The reform efforts also have prescribed profound changes in the structure of the typical mathematics classroom. Where traditional classrooms are often teacher-centered, reform classrooms place much more emphasis on the student and classroom discourse. Implementing these changes without losing sight of mathematical goals poses a tremendous challenge to many teachers.

This chapter will examine some of the characteristics of a student-centered, reform classroom, and some of the challenges teachers face in creating such a classroom. A discussion of some of the ways in which expert teachers successfully create student-centered classrooms without neglecting important mathematics will follow. Finally, this chapter will investigate some suggested methods for helping teachers improve the quality of student-centered classroom episodes, particularly those in which shared student work is the central point of instruction.

Student-Centered Classrooms

The Professional Standards set forth by the National Council of Teachers of Mathematics (NCTM, 1991) help to paint a picture of the reform mathematics classroom. These standards emphasize the importance of student discourse in the classroom, and view the teacher as a facilitator of this discourse, with an aim of orchestrating discourse that promotes student learning. While the standards indicate the importance of the

teacher as skillful director of classroom discourse, the NCTM also places significant emphasis on the student-centered nature of classroom episodes. Rather than the teacher being the primary classroom speaker, introducing and explaining topics, in the reform mathematics classroom is characterized by the students taking on these roles.

Much of this emphasis on student-centered classrooms stemmed from a growing acceptance of constructivist theories of learning and knowledge. Constructivism suggests that students build up their own subject knowledge by actively reflecting on topics, rather than by passively receiving information from an authority (Wood, 1995). Accepting constructivist ideas reconceptualizes the mathematics classroom and the expectations of mathematics teachers. In the reform environment, then, the teacher becomes responsible for providing students with opportunities for learning and reflection, rather than delivering information.

Students' construction of mathematical knowledge is not a strictly internal process. For most students, the primary theater for learning mathematics is the mathematics classroom, featuring other students and the teacher. In this setting, the learning process is inherently social (Wertsch & Toma, 1995). The social nature of learning mathematics, in conjunction with students' development of mathematical understanding through reflection, leads to the importance of classroom discourse in the reform classroom.

Wood, Cobb and Yackel (1995) assert that discussions of mathematical ideas are not just helpful to student learning, but are essential. The quality of discourse, however, is critical. Students must do more than just sharing ideas and solutions, but must be expected to provide explanations and justifications of their work. They must also be

expected to listen to other students' ideas with the goal of understanding. These opportunities allow students to build understanding by assessing and incorporating the ideas of others to create ideas that are more mathematically advanced than those they created on their own.

Discourse is a sufficiently important topic in mathematics education to merit many different characterizations of the types of discourse that transpire in classrooms. This chapter discusses two such characterizations that may be helpful to teachers in cultivating productive classroom discourse.

Univocal and Dialogic Functions of Discourse

Lotman (as cited in Wertsch & Toma, 1995) distinguishes discourse (and text in general, including both verbal and nonverbal expressions) as having two distinct functions. The univocal function of discourse is to convey meaning, and depends upon the speaker and listener using language and communication in essentially the same way. The dialogic function of discourse is to generate meaning and depends upon viewing discourse as a tool for stimulating thinking.

The vast majority of classroom discourse has historically functioned more univocally than dialogically, as would be expected in a teacher-centered classroom. Wertsch and Toma (1995) suggest that in classrooms where the dialogic function of discourse dominates, opportunities for student learning are far greater. In such settings, the authors assert, discourse is treated by students and teachers as springboards for deeper thinking. Students consider contributions to classroom conversations, and rather than simply accepting them as factual, they are encouraged to extend, interpret them, and use

them to modify their own understanding. Teachers striving to create student-centered classrooms, then, should maximize dialogic discourse and minimize univocal discourse.

Types of Mathematical Communication

The distinction between the univocal and dialogic functions of discourse helps to provide a framework for the goal of utterances that facilitate student-centered classrooms. It is also valuable to understand the types of interactions that occur between discourse participants in order to identify the communication patterns that best encourage student-centered environments. Brendefur and Frykholm (2000) characterize these interactions through four types of mathematical communication: uni-directional, contributive, reflective, and instructive communication.

Uni-directional communication is communication dominated by the teacher, often characterized by lecturing. Such communication provides very few opportunities for students to communicate original thoughts and ideas. Contributive communication is characterized by greater interactions among students and between the teacher and students. This communication, however, is limited to sharing ideas or obtaining assistance, with little or no deeper thinking. In reflective communication, students share ideas and strategies, and these ideas and strategies are used as starting points for deeper investigations. Instructive communication extends reflective communication by placing more emphasis on the role of the teacher. Teachers take a more active role in shaping conversations toward mathematical goals and in using the student ideas revealed in such discussions to shape instruction.

The latter three types of communication – communicative, reflective, and instructive – all emphasize the role of the student in contributing ideas to the classroom. However, only reflective and instructive communication explicitly focus on building mathematical understanding without taking mathematical authority away from the students. Thus these forms of communication are most conducive to fostering a truly student-centered classroom environment.

Challenges Teachers Face

The emphasis on student-centered, discourse rich environments carries with it a significant shift in the expectations of the teacher's roles in the classroom. It is well documented that this shift creates difficulties both for teachers experienced in teaching using traditional methods and for novice teachers.

Chazan and Ball (1999), both practicing K-12 teachers, describe the reform environment as having created a perceived "no-telling" policy wherein teachers feel prohibited from directly providing students with information. In order to create student-centered classrooms without much guidance in how to develop them, many teachers reacted by stepping back from discourse and allowing classroom conversations to be strictly in the hands of students. As Chazan and Ball proclaim, a direction not to tell does not provide direction in what teachers *should* do in the classroom.

Ruth Heaton, an elementary teacher who investigated her own teaching beliefs while pursuing a graduate degree, echoed the challenges that Chazan and Ball report are common among teachers (Heaton, 2000). Heaton found that creating a classroom atmosphere in which students are major contributors and she was not the primary

authority brought with it some unique challenges. Particularly, balancing the desire to give students authority in the classroom, being open to student ideas, both right and wrong, and furthering the lesson's mathematical goals proved difficult.

It is not only experienced teachers accustomed to delivering material through a lecture who find adjusting to student-centered classrooms to be difficult. Novice teachers who have yet to establish a teaching style also find that student-centered classrooms can be difficult to facilitate. Borko and Livingston (1989) investigated the experiences of two preservice teachers during their student teaching experiences. Both preservice teachers were well acquainted with reform practices and indicated belief in student-centered classrooms as those most conducive to learning. Yet in the classroom, they had difficulty teaching in such a fashion, and often resorted to traditional, lecture-based teaching methods.

One of the preservice teachers, with support from researchers, successfully managed to incorporate interactive teaching into her lessons, although her journaling did indicate that this was not a simple process. The other preservice teacher was relatively unsuccessful in incorporating interactive teaching and reported significant discomfort in allowing students control of the classroom discourse. Indeed, the tension between covering curricular expectations and providing opportunities for student exploration was extremely challenging for the preservice teacher. Borko and Livingston (2000) affirm the need for teacher preparation programs that incorporate guidance in interactive, student-centered teaching.

In order to create such programs, it is critical that the specific challenges that both preservice teachers and experienced teachers attempting to create student-centered

classrooms face are understood. Several researchers have contributed to the identification of these challenges. Chazan and Ball (1999) note that the management of classroom discussions can be a significant challenge for teachers.

Classroom discussions create opportunities for diverse, and sometimes incorrect, ideas to be exposed. An important challenge for teachers is to manage these discussions in a way that respects student ideas but also leads toward consensus that corresponds to that of the mathematical community. Teachers are still responsible for ensuring that students develop understanding of the definitions and concepts of mathematics that are already accepted. But this must be done while still acknowledging and incorporating the different ideas put forth by students.

Successful Practices

While the challenges in fostering student-centered classrooms where important mathematical learning takes place are significant, there are many accounts of teachers doing so successfully. Researchers have attempted to identify some of the practices teachers use in facilitating productive classroom discussions.

Magdalene Lampert is a teacher often cited as an example of a deft manager of a reform mathematics classroom. Leinhardt and Steele (2005) examined Lampert's work in depth with a focus on identifying her successful practices. A primary way in which Lampert handles the uncertainty of classroom discussions is by allowing flexibility in time to explore students' ideas. When unexpected ideas arose, Lampert addressed them by first attempting to understand the confusion through questioning and encouraging students to elaborate. She allowed the spontaneous, but potentially mathematically

significant, explorations to blossom until she fully understood the source of students' confusion or connections with such topics. At that point, she was then able to guide her students back to the topic she had intended by helping them to form the connections between their ideas and the mathematical concepts she wanted to get on the table. Certainly this is not a simple task, but Lampert provides an insightful example of a teacher that manages the uncertainty of a student-centered classroom.

Sherin (2002) suggests that the discomfort teachers feel facing uncertainty in the classroom can be abated by learning while they are teaching in the reform environment. Building new content knowledge through reflecting on classroom experiences can help teachers to more successfully navigate classroom discussions. Content knowledge includes both subject matter content knowledge, the understanding of mathematical facts and concepts, as well as pedagogical content knowledge, the understanding of teaching those facts and concepts.

Both types of knowledge must change in order to successfully be applied in a reform classroom. The pedagogical content knowledge that teachers possessed in how to deliver a lecture on a mathematical topic does not translate to that required to facilitate a discussion on that topic. Furthermore, subject matter knowledge must be modified, as open discussions where students introduce ideas will likely create connections between mathematical concepts that teachers have not formerly considered. Sherin suggests that being open to learning and consciously building on existing content knowledge can help teachers to eventually manage the uncertainty that arises in classroom discourse.

Chazan and Ball (1999) address this uncertainty, especially the potential for student disagreements that may take place in the classroom. Based on their own

successful experiences, the authors suggest techniques that can be used to "guide intellectual fermentation" during classroom discussions. They view the teacher's role as that of monitor and manager of classroom disagreement and suggest three considerations which teachers can make to determine the action that should be taken during a disagreement.

First, the mathematical value of the disagreement must be considered. If students are disagreeing on a point that is important for either immediate or long-term mathematical understanding, the teacher must step in to manage the disagreement. Secondly, the effect of the disagreement on the momentum of the discussion must be considered. The teacher should step in to manage disagreement when it is damaging the pace of the discussion by bogging students down on a mathematically insignificant, or immediately irresolvable, point. At the same time, the teacher may have to introduce disagreement when a discussion has essentially dead-ended because students have come to faulty agreement on a point. The third consideration is the tone of the disagreement. The teacher must intervene in disagreements that are overly emotionally charged or detrimental to the social climate of the classroom.

While uncertainty and dialogic conflict in the classroom is unavoidable, Chazan and Ball provide some tools for teachers to assess the necessity of intervention in disagreements. As teachers often cite the unpredictable nature of classroom discourse as an impediment to making it a central point of their classrooms, these considerations may provide some security when faced with the inevitable conflicts that will occur when students are encouraged to share ideas.

Sharing student ideas and solutions in the classroom is an essential aspect of the student-centered classroom, but possibly the most daunting for teachers. Using student work to guide classroom discussions in a purposeful fashion is an area that has recently received noticeable research attention. Researchers are emphasizing a shift away from, as Ball calls it, the "show-and-tell" policy that dominated sharing episodes toward more purposeful sharing of student work.

Sharing Student Work

The "show and tell" episodes that have typified sharing student work in the classroom are characterized by a series of students stating the work they have done with little or no explanation. In these episodes, the teacher's role is to simply call on students to share and to listen politely as they share. In the reform mathematics classroom, the teacher should play a more active role in managing these discussions of student solutions.

Ball (2001) asserts that teachers can build a lesson that is more than simply "show and tell" by composing "a mathematical discussion that takes up and uses the individual contributions." This requires that the teacher engages students in comparing and discussing the ideas presented while moving mathematical ideas along.

Lessons in which student work is shared productively are characterized by Yackel et al (1990) as having two distinct phases: one half of the class time is centered around small-group problem solving and the second half is a class discussion during which students share solutions. The teacher has important roles during both halves of the class time. During the problem-solving phase, the teacher circulates throughout the class, observing the work students are doing and intervening if necessary.

During the sharing phase, students share their solutions and the teacher facilitates the sharing by helping them to explain if necessary, clarifying their work, or suggesting that they share an alternate solution. The teacher's role during this phase is not to indicate the correctness of the students' solutions, but rather to encourage students to reflect on the solutions shared, and to come to agreement or discuss disagreement about mathematical ideas.

Kalathil (2006a) studied two second grade classrooms and characterized the discourse structures that these teachers and their students used during the group discussion phase of a classroom sharing episode. She identified six discourse structures and three discourse patterns that she believes can help teachers foster productive classroom discourse.

These six discourse structures are:

1) Answer and partial explanation- students display, but do not discuss their work.
2) Explicit explanation- students provide verbal explanations of their work.
3) Extension- students provide, often under prompting from the teacher, explanations that extend beyond discussion of their representations into the mathematical ideas and concepts inherent in the solution.
4) Comparison- students and the teacher create comparisons between different solution methods shared
5) Conjecture- students or the teacher hypothesize about the problem in a way that extends toward a more general idea

6) Justification- students or the teacher provide justifications of the conjecture made.

The discourse patterns Kalathil identified that unite these discourse structures are 1) getting started, 2) investigating by comparing, and 3) going deeper. Getting started involves a student display of his solution, a verbal explanation, and, if necessary, an extension of the explanation. Investigating by comparing occurs after the getting started pattern has been repeated for several solutions and involves making comparisons among the solutions shared. Going deeper is an extended discussion of the mathematically significant ideas exposed during the getting started and investigating by comparing phases. Conjecturing and justification are likely, but not necessarily, discourse structures included in this pattern.

The investigating by comparing pattern is sufficiently important to merit an extended discussion. Kalathil (2006b) suggests that the investigating by comparing pattern is the critical element in moving a discussion based on student solutions from a simple show and tell into a mathematically productive discussion. Simply inviting students to share, and even asking questions about their solutions, stops short of achieving the full potential of these episodes. Comparing and connecting student solutions are useful tools in drawing out the important mathematics.

Kalathil (2006b) identified two themes used by teachers to compare student responses during in class discussions. First, the discussions would build to more mathematically significant comparisons over time. The early comparisons made tended to be more around the surface features of the solutions, such as the choice of representation or the method of calculating. As the conversations progressed, the

comparisons among solutions became more mathematically significant, focusing more on the important mathematics than the surface characteristics of the solutions.

The second theme in the comparisons was the point at which comparisons were incorporated into discussions over the course of several discussions. During the early discussions, comparisons were usually initiated by the teacher as a way of wrapping up the discussion. In later discussions, however, teachers began encouraging comparisons much earlier in the discussions. These comparisons were used as ways to prompt further discussion around mathematically significant concepts.

Kalathil (2006b) suggests that the usefulness of discussions based on shared student solutions depend greatly on the teacher's thoughtful orchestration of these discussions. In addition to identifying discourse structures and patterns that make it clear that there can be much more to sharing student work than simply show and tell, Kalathil notes that when teachers take proactive roles in building these conversations, mathematical productivity is often enhanced. She suggests that this can be done by teachers carefully selecting the pieces of student work they would like to have shared. The teacher is then responsible for helping students to make the important connections between solutions, and also pushing students to consider alternate solutions. Teachers also must play a part in using student responses, both their solutions and their statements during the discussion, to explore the mathematically significant content of the problem.

The identification of discourse structures and patterns that are used successfully to facilitate classroom discussions surrounding shared student work can be useful in helping teachers determine how they may like to structure such episodes. Surely it can not be said that one should follow a set pattern of getting started, followed by investigating by

comparing, followed by going deeper. However, this does provide teachers with some structures and patterns they may use as appropriate for the class discussion to build mathematically productive episodes. Kalathil's work also helps to dispel the myth that student-centered classrooms are characterized by a free-for-all of student sharing. Instead, student work can be used skillfully and purposefully to guide discussions in a way that both maintain students' intellectual authority and forward mathematical goals.

This purposeful use of student work can be seen in Groves and Doig's (2006) description of an Australian teacher who uses student work to structure mathematical discussions. Where Kalathil's discussion emphasized the actions that teachers take during discussions centered on student work, Groves and Doig also delve into the actions of the teacher before the discussion.

While students worked on the task, Groves and Doig's teacher engaged in what they term "purposeful scanning." The teacher wandered throughout the class, observing the work students were doing and asking questions about their work, both to facilitate student work and to develop her own understanding of their solutions. The teacher used the information gathered during her purposeful scanning to select specific student papers she wanted shared and the order in which they would be shared. She also suggested to the students specific aspects of their solutions that she wanted shared.

Pre-selecting the students who will share work returns some control to the teacher, allowing her to determine in advance the mathematics included in the solutions that will be shared. Groves and Doig also indicated that the teacher had students share their work in a specific order, beginning with the least mathematically sophisticated solution and building to strategies of greater sophistication. They assert that sharing in

such an order allows all students an entry point into the conversation, through the less complicated solution, and creates a climate where the community can progress in understanding.

All of Ball, Yackel et al, Kalathil and Groves and Doig present useful examples of teachers successfully structuring classroom episodes around student work. Yet, they are really collections of practices that have been effective for others, not a prescription for incorporating effective practices into one's own classroom. Recognizing the difficulties teachers relate in their experiences sharing student work, Stein et al (in press) have attempted to address this need for help in structuring sharing episodes.

Before a productive discussion around student work can begin, Stein et al assert, students must be addressing a task that is sufficiently mathematically interesting to merit such a discussion. The authors have developed a framework that helps to understand the characteristics of discussion worthy tasks (Stein et al, 2000).

Cognitive Demand of Tasks

Stein et al (2000) have distinguished mathematical tasks based upon their cognitive demand into those of low cognitive demand and those of high cognitive demand. Tasks of low cognitive demand are subdivided into memorization tasks and procedures without connections tasks. Memorization tasks require students to reproduce previously learned facts or rules without connecting them to their corresponding mathematical concepts. Procedures without connections tasks require students to perform algorithmic procedures with a focus on obtaining the correct answer, and without

considering the connection between these procedures and the corresponding mathematical concepts.

Tasks of high cognitive demand are subdivided into procedures with connections tasks and doing mathematics tasks. Procedures with connections tasks require students to make use of procedures with a focus on building deeper mathematical understanding. These are not procedures that students can follow mindlessly, but instead demand students to think more deeply about the associated mathematical concepts. Doing mathematics tasks are tasks requiring complex, non-algorithmic thinking and can be approached in many different ways. They require students to consider mathematical concepts and relationships in a reflective fashion, and usually are challenging for students.

While tasks of low cognitive demand also have a place in the mathematics classroom (for example, learning terminology and notation allows for efficient and productive communication), they are not well suited to rich classroom discourse. Tasks of high cognitive demand are much more likely to promote engaging, mathematically significant conversations. Moreover, these tasks can be approached by students in many different ways, providing opportunities for discussions about different strategies, and also creating opportunities for mathematical misconceptions or new mathematics to be exposed.

However, simply having a task of high cognitive demand does not guarantee a mathematically productive classroom discussion. As previously discussed, teachers are often still unclear how to incorporate student responses to these tasks into discussions in a way that builds toward a mathematical point. Stein et al (in press) provide a set of five

practices that can help teachers to share student work in a more productive fashion. The departure from other accounts of successful practices is that the authors place significant emphasis on planning before the discussion begins. This helps to alleviate the anxiety that teachers often report in managing these discussions due to the uncertainty surrounding the use of student solutions.

The Five Practices for Sharing Student Work

Stein et al (in press) describe five practices that take the teacher through the process of using shared student work in a classroom discussion. The five practices they have identified are: 1) anticipating students' mathematical responses, 2) monitoring student responses, 3) purposefully selecting student responses for public display, 4) purposefully sequencing student responses, and 5) connecting student responses.

1) Anticipating Student Responses

The first of the five practices takes place before the task is even launched in the classroom. Anticipating student responses first requires the teacher to actually work the problem that will be assigned to students. This may help the teacher to identify areas that will likely be difficult for students, or to envision the type of student solutions that would be expected. The teacher should try to find as many different possible strategies for solving the problem as a way to prepare for the variety of responses students will generate. Anticipating student responses also entails considering the ways in which strategies used to solve a problem are related to the important mathematical concepts that the teacher wants students to understand.

Thinking in advance about the task, the expected responses to the task, the variety of possible strategies, the areas of difficulty, and the important mathematics in the task can help to eliminate some of the anxiety that a teacher may feel when faced with a discussion on the task. Thoughtful reflection about the task can identify many, if not most, of the strategies used and confusion encountered by the students. This preparation may help the teacher to feel more comfortable managing the variety of responses students generate.

2) Monitoring Student Responses

The second practice, monitoring student responses, takes place while, and after, the students are working on a task. This practice requires the teacher to pay careful attention to the solutions students generate, with special interest in the students' mathematical understanding. Monitoring student responses begins while students are working on a task, and is characterized by the "purposeful scanning" discussed by Groves and Doig (2006). In this practice, the teacher ask questions to better understand the solutions students generate and identifies the potential of various solutions for furthering the mathematical learning of the class.

Taking the time to anticipate student responses makes the interpreting that takes place while monitoring somewhat less daunting for teachers. Monitoring also allows the teacher time to consider responses to different pieces of student work that can be used to facilitate the discussion. This practice can help to eliminate much of the anxiety teachers experience due to the unpredictable nature of discussions. Considering in advance the important mathematics in student solutions and the teacher comments or questions that

will best draw the mathematics out can help the teacher to guide students' learning during sharing episodes.

3) Purposefully Selecting Student Responses for Public Display

The third practice, purposefully selecting student work, provides the teacher with an opportunity to manage classroom discussion in a productive, and, to the teacher, predictable way. This practice requires the teacher to choose specific students' solutions to be shared during the discussion. These pieces of work should be selected because they expose important mathematical ideas or misconceptions or present strategies that are dependent on those ideas. The key to purposefully selecting student work to share is that it must be *mathematically* purposeful; that is, that the pieces of work chosen are chosen because they further the *mathematical* goals of the discussion.

Selecting the pieces of student work that will be shared returns the sharing episode to the teacher's hands, enabling the teacher to build toward mathematical meaning. The unpredictability of traditional sharing episodes in which the class and the teacher are "at the mercy of whatever strategies the student volunteers present" is eliminated. The teacher is returned to the role of director of mathematical learning by choosing important pieces of mathematics to be shared. Moreover, knowing in advance which strategies will be shared helps the teacher to determine the questions that will be asked or the important ideas that need to be drawn out instead of having to make instant instructional decisions in response to shared work. Much of the discomfort of uncertainty can be eliminated through the practice of purposefully choosing student work.

4) Purposefully Sequencing Student Responses

Not only must the teacher purposefully select pieces of student work to be shared, he or she must also arrange for students to share in a purposeful order. Placing the pieces of work in a mathematically purposeful order can make it more likely that the intended mathematical goals of the discussion are met. Stein et al (in press) suggest several ways of sequencing student responses for maximum mathematical learning.

The teacher may wish to have the first solution presented be the most popular strategy so that students have an accessible point to the conversation. Starting with a more easily understandable strategy and moving to more difficult ones may also provide an entry point into the discussion for a wider variety of students. The teacher may wish to have a common misconception shared early in the discussion so that students' misunderstanding can be rectified, allowing them to learn from the other strategies. The authors also suggest that the teacher may wish to place related or contrasting solutions adjacent to one another in order to facilitate comparisons.

While many different sequencing rationales are possible, the authors note that little is know about the effectiveness of the different sequences. Research is still needed to understand what, and how, each type of sequence contributes to student learning or advancing of mathematical goals.

5) Connecting Student Responses

The fifth and final practice that Stein et al (in press) suggest is connecting student responses. This requires the teacher to help students draw connections between the different mathematical ideas presented during the discussion. This may involve drawing students' attention to the same mathematical ideas occurring in two solutions that appear

very different. The teacher may also wish to help students determine the effectiveness or usefulness of strategies shared. Connecting student responses helps to ensure that the mathematical goals of the discussion are met.

This practice returns some important pedagogical control to the teacher in that it allows the teacher to shape the discussion to achieve mathematical aims without being entirely dependent on student contributions. The planning of the discussion that occurs in the preceding four practices allows the teacher to shape the mathematical arc of the sharing episode, and connecting responses is a way to be sure that the mathematical arc is maintained.

Stein et al's (in press) practices provide a much needed guide for teachers to manage classroom discussions in a way that is mathematically productive. The authors provide some evidence, as well, that understanding the five practices has benefits for teachers in their own work. One of the authors modeled the practices in a course for preservice teachers and provided them with opportunities to reflect on the practices. Given pieces of student work to choose from, the preservice teachers were much more likely at the end of the course to select pieces of student work to share based on mathematical rationales, and to sequence those with a mathematical goal in mind, than they were at the start of the class. They also reported reflection about incorporating the five practices into their own teaching.

As of yet, no research has been conducted examining how well teachers actually incorporate these practices into their work. Much research is needed to determine if studying the five practices helps teachers to craft more mathematically productive sharing

episodes in their own classrooms. It is still unknown if teachers actually do feel more in control of classroom discussions when they utilize the five practices.

Conclusion

Mathematics education reform efforts call for classrooms to become more student centered. These classrooms are characterized by more open discussions over mathematically provocative tasks, and the expectations of the teacher are very different. Traditionally, teaches have been viewed as the classroom keepers and distributors of knowledge, while the reform environment expects teachers to take on the role of "guide on the side" (Leinhardt & Steele, 2005). For many teachers, this role is intimidating and confusing, and could be misinterpreted as meaning teachers hand control of the classroom over to students. While teachers understand that they are no longer supposed to deliver information through "telling" (Ball, 2002), there has been insufficient guidance on how they *are* supposed to manage mathematical goals and maintain student centered classrooms.

Accounts of expert teachers provide insight into some of the moves and practices that facilitate student-centered classrooms where important mathematical learning occurs. While teachers can certainly learn from these accounts, and perhaps even incorporate some of the practices into their own repertoire, expert accounts are far from pedagogical models. Stein et al (in press) provide a potentially powerful pedagogical model to help teachers plan and enact the use of student work in the classroom. The model could be an important tool incorporating student work in classroom episodes while still working toward a mathematical aim and minimizing uncertainty and anxiety.

Certainly, however, the effectiveness of the model in the hands of practicing teachers is still unclear. Sharing student work can be a powerful tool for creating a classroom where students have authority but the teacher still maintains mathematical control. It should be noted that all of the five practices Stein et al (in press) include in their model require the teacher to have a deep understanding of the mathematics involved in the task.

Anticipating the different responses and making sense of students' responses requires that teachers be not only comfortable with the mathematics but also aware of the place of the mathematical concept in the larger subject. Choosing the best pieces of work to achieve a mathematical goal requires teachers to be able to identify the most effective representations of a mathematical concept, or identify those that are less effective to draw attention to them. Ordering solutions in a way that best achieves a mathematical goal requires not only understanding of the mathematics involved, but also significant pedagogical content knowledge about the most effective way to build understanding of a concept. Making connections between different pieces of student work also requires significant content knowledge to understand how those pieces of work are or are not related.

It is worth investigating, then, the relationship between teachers' content knowledge and their effectiveness in implementing the model. Surely there is much more to look in to regarding the five practices, but they also certainly provide teachers with some concrete guidance in managing the uncertainty of classroom discussions.

Chapter 4 – Project Background

The Northwest Regional Educational Laboratory (NWREL) is a nonprofit organization based in Portland, Oregon that provides educational services and resources to the Northwest states. NWREL's Mathematics and Science Education Center offers professional development and educational materials to K-12 teachers and conducts educational research. In 1999, in response to the emphasis on increased problem solving in the classroom (NCTM, 1989) and the difficulties teachers encounter in implementing problem solving (Burkhardt, 1998; Schoenfeld, 1992), the Mathematics and Science Education Center developed the NWREL Mathematics Problem Solving Model (MPSM).

The original model had three broad goals: 1) to increase teachers' use of open-ended mathematics problems in their classrooms, 2) to improve their abilities to assess students' problem solving skills, and 3) to provide them with skills for teaching problem solving. The model attempted to address these goals through the use of classroom tasks, a scoring guide for assessing student performance, and sample student work. These tools comprising the model were accompanied by professional development around improving instructional strategies.

The materials were tested by NWREL during the 1999-2000 school year at two elementary schools in the state of Washington. This implementation of the project led to evidence-based modifications that resulted in the model inherited by the current researchers in 2004.

In 2006, these same researchers began a five-year long, National Science Foundation (NSF) funded study to both determine the effectiveness of, and make modifications to, the model as it was conceived at the time. The model implemented in 2006 had four major components: 1) a body of exemplary problem solving tasks, 2) a formative feedback guide for evaluating student work, 3) professional development instruction, and 4) internet-based communication tools.

Exemplary Problem Solving Tasks

Exemplary problem solving tasks, as defined in the model, are tasks grounded in the NCTM standards, open-ended, non-routine, and complex. Open-ended problems are those that allow for multiple approaches and, often, multiple different solutions. Non-routine problems refer to those that can not be solved by an algorithm known to the student. Complex tasks involve multiple steps, require students to explain and/or justify their thinking, and require cognitive work on the part of the student.

The NCTM standards (1989, 2000) identify developing problem solving skills as the central goal of mathematics education and emphasize the importance of solving open ended problems in fostering these skills. Problems that are open ended, non-routine, and complex are often identified as those best suited for developing mathematical understanding (Becker & Shimada, 1997; Hiebert et al, 1997; Stein et al 2000).

The NWREL model includes a body of tasks at levels appropriate for grades K-12. In addition to the tasks, detailed teachers' guides are provided to indicate the purpose and context of the task, as well as suggestions for introducing, discussing, and extending the task. The guides also identify the important mathematics highlighted in the task.

Formative Feedback Guide

The formative feedback guide is an amendment of the scoring guide that was part of the original Mathematics Problem Solving Model in 2000. The feedback guide helps teachers to evaluate student performance in each of five problem solving traits: 1) conceptual understanding, 2) strategies and reasoning, 3) computation and execution, 4) communication, and 5) insights. Criteria statements are used to identify the proficiency level of a student's performance in each of those five traits. The levels of performance are, from lowest level to highest: emerging, developing, proficient, and exemplary.

The five traits are grounded in the research literature. In particular, they share many similarities with Kilpatrick's five intertwined strands of mathematical proficiency (Kilpatrick et al, 2001): conceptual understanding, procedural fluency, strategic competence, adaptive reasoning, and productive disposition. The strands are not discrete problem solving skills or attitudes, but rather overlapping features of problem solving that interact with one another in the problem solving process. It was necessary, however, in order to provide feedback and evaluation to students through the feedback guide, to consider the five traits included in the guide separately.

Conceptual Understanding

The conceptual understanding trait examines the extent to which a student's response reflects the important mathematics in the problem in the representations and procedures used. Conceptual understanding, as conceived in the NWREL formative feedback guide, is analogous to Kilpatrick's (2001) conceptual understanding strand of proficiency. The NCTM *Principles and Standards* (NCTM, 2000) include understanding relationships, procedures, and representations in all of the mathematical content

standards. The elements of conceptual understanding, the use of appropriate procedures and representations, are tools for building and evidence of mathematical conceptual understanding (Hiebert et al, 1997; Schoenfeld, 1992).

Strategies and Reasoning

The second trait, strategies and reasoning, addresses the appropriateness of the strategy selected by the student to solve the problem, and the extent to which the strategy was well planned and logically implemented. This trait is similar to Kilpatrick's (2001) strategic competence strand. The NCTM's 2000 standard for Problem Solving expects students to "apply and adapt a variety of appropriate strategies to solve problems." Moreover, the choice and implementation of an appropriate strategy corresponds closely to Polya's second and third steps of the problem solving process: devising and carrying out a plan. Thus, emphasizing strategies and reasoning in student problem solving addresses Schoenfeld's (1992) recommendation that problem solving tasks make use of Polya's problem solving process.

Computation and Execution

Computation and execution refers to the accuracy and completeness with which the student implemented the approach selected. This trait is similar to Kilpatrick's (2001) procedural fluency strand. Several NCTM standards (2000) emphasize correct and accurate computations, including the Number and Operations standard that students should compute fluently and make accurate estimates. Hiebert notes that problem solving tasks should provide students with opportunities to use their "tools," the mathematical skills and procedures with which problems may be solved (1992).

Communication

Communication is concerned with the ease with which a reader could understand not only the student's solution process, but also the thinking behind the solution. The NCTM (2000) standards for communication expect students to use mathematical language precisely to communicate ideas, and also to be able to understandably communicate mathematical thinking to both teachers and peers. The research literature indicates that effective and clear communication of thinking is critical in building understanding, as it helps students to build ideas and think more deeply on mathematical concepts (Hiebert et al, 1997).

Insights

The final trait, insights, examines the extent to which the student demonstrated understanding of the deeper structure of the problem, and the connections of the problem to real world situations or previously learned mathematics. The NCTM (2000) standard for connection expects students to recognize and use connections among mathematical ideas, understand how mathematical ideas interconnect and build, and recognize and apply mathematics in other contexts. Polya (1945) includes reflection on one's own processes and relations to other, known ideas as part of devising a plan. Identifying similarities with known procedures or concepts can facilitate strategizing on an unknown problem. Moreover, recognizing these relationships can help students to build mathematical understanding (Hiebert et al, 1997).

Figure 4.1 NWREL Formative Feedback Guide

	Emerging	**Developing**	**Proficient**	**Exemplary**
Conceptual Understanding Key Question: *Does the student's interpretation of the problem using mathematical representations and procedures accurately reflect the important mathematics in the problem?*	1. Your mathematical representations of the problem were incorrect. 2. You used the wrong information in trying to solve the problem. 3. The mathematical procedures you used would not lead to a correct solution. 4. You used mathematical terminology incorrectly.	1. Your choice of forms to represent the problem was inefficient or inaccurate. 2. You used some but not all of the relevant information from the problem. 3. The mathematical procedures you used would lead to a partially correct solution. 4. You used mathematical terminology imprecisely.	1. Your choices of mathematical representations of the problem were appropriate. 2. You used all relevant information from the problem in your solution. 3. The mathematical procedures you chose would lead to a correct solution. 4. You used mathematical terminology correctly.	1. Your choice of mathematical representations helped clarify the problem's meaning. 2. You uncovered hidden or implied information not readily apparent. 3. You chose mathematical procedures that would lead to an elegant solution. 4. You used mathematical terminology precisely.
Strategies and Reasoning **Key Question***: Is there evidence that the student proceeded from a plan, applied appropriate strategies, and followed a logical and verifiable process toward a solution?*	1. Your strategies were not appropriate for the problem. 2. You didn't seem to know where to begin. 3. Your reasoning did not support your work. 4. There was no apparent relationship between your representations and the task 5. There was no apparent logic to your solution. 6. Your approach to the problem would not lead to a correct solution.	1. You used an oversimplified approach to the problem. 2. You offered little or no explanation of your strategies. 3. Some of your representations accurately depicted aspects of the problem. 4. You sometimes made leaps in your logic that were hard to follow. 5. Your process led to a partially complete solution.	1. You chose appropriate, efficient strategies for solving the problem. 2. You justified each step of your work. 3. Your representation(s) fit the task. 4. The logic of your solution was apparent. 5. Your process would lead to a complete, correct solution of the problem.	1. You chose innovative and insightful strategies for solving the problem. 2. You proved that your solution was correct and that your approach was valid. 3. You provided examples and/or counterexamples to support your solution. 4. You used a sophisticated approach to solve the problem.

Figure 4.1 NWREL Formative Feedback Guide (continued)

	Emerging	Developing	Proficient	Exemplary
Computation & Execution Key Question: *Given the approach taken by the student, is the solution performed in an accurate and complete manner?*	1. Errors in computation were serious enough to flaw your solution. 2. Your mathematical representations were inaccurate. 3. You labeled incorrectly. 4. Your solution was incorrect. 5. You gave no evidence of how you arrived at your answer.	1. You made minor computational errors. 2. Your representations were essentially correct but not accurately or completely labeled. 3. Your inefficient choice of procedures impeded your success. 4. The evidence for your solution was inconsistent or unclear.	1. Your computations were essentially accurate. 2. All visual representations were complete and accurate. 3. Your solution was essentially correct. 4. Your work clearly supported your solution.	1. All aspects of your solution were completely accurate. 2. You used multiple representations for verifying your solution. 3. You showed multiple ways to compute your answer.
Communication Key Question: *Was I able to easily understand the student's thinking or did I have to make inferences and guesses about what they were trying to do?*	1. I couldn't follow your thinking. 2. Your explanation seemed to ramble. 3. You gave no explanation for your work. 4. You did not seem to have a sense of what your audience needed to know. 5. Your mathematical representations did not help clarify your thinking.	1. Your solution was hard to follow in places. 2. I had to make inferences about what you meant in places. 3. You weren't able to sustain your good beginning. 4. Your explanation was redundant in places. 5. Your mathematical representations were somewhat helpful in clarifying your thinking.	1. I understood what you did and why you did it. 2. Your solution was well organized and easy to follow. 3. Your solution flowed logically from one step to the next. 4. You used an effective format for communicating. 5. Your mathematical representations helped clarify your solution.	1. Your explanation was clear and concise. 2. You communicated concepts with precision. 3. Your mathematical representations expanded on your solution. 4. You gave an in-depth explanation of your reasoning.

Figure 4.1 NWREL Formative Feedback Guide (continued)

	Emerging	Developing	Proficient	Exemplary
Insights Key Question: Does the student grasp the deeper structure of the problem and see how the process used to solve this problem connects it to other problems or "real-world" applications?	1. You were unable to recognize patterns and relationships. 2. You found a solution and then stopped. 3. You found no connections to other disciplines or mathematical concepts.	1. You recognized some patterns and relationships. 2. You found multiple solutions but not all were correct. 3. Your solution hinted at a connection to an application or another area of mathematics.	1. You recognized important patterns and relationships in the problem. 2. You found multiple solutions using different interpretations of the problem. 3. You connected your solution process to other problems, areas of mathematics or applications.	1. You created a general rule or formula for solving related problems. 2. You related the underlying structure of the problem to other similar problems. 3. You noted possible sources of error or ambiguity in the problem. 4. Your connection to a real-life application was accurate and realistic.

Professional Development

The third component of the mathematics problem solving model is the professional development surrounding the use of the tools included in the model. The basic elements of the professional development include 1) exposure to and reflection on the research literature regarding problem solving, 2) discussion of the origins and implementation of the problem solving tasks, 3) discussion of the origins and use of the formative feedback guide, and 4) ongoing coaching throughout classroom implementation of problem solving tasks.

1) Research literature regarding problem solving

Exposure to and reflection on the research literature on problem solving and assessment of problem solving is intended to help teachers understand the pedagogical reasons behind the model and its tools. This phase emphasizes the learning benefits of quality, open ended problem solving tasks and formative assessment practices. It also helps teachers to consider some successful practices for implementing problem solving in the classroom and assessing students' problem solving.

2) Implementation

Beyond understanding the research-based motives for including open ended problems in the classroom, teachers are also given experience with these problems. Not only do teachers have the opportunity in the professional development to solve and discuss some of NWREL's tasks, but they also learn to identify the quality of a mathematical task. Using Stein et al's (2000) mathematical task framework, teachers participating in the professional development learn to distinguish between tasks that

encourage students to build deeper mathematical understanding and those that are of lesser value in building understanding.

3) Formative feedback guide

Through the exploration of research literature about problem solving and formative assessment, professional development participants learn the research-based origins of the formative feedback guide and also have opportunities to reflect on its usefulness. Along with the examination of the NWREL problem solving tasks, the teachers also have the opportunity to put the feedback guide to use on sample student responses to those tasks. Extensive practice using the feedback guide accurately is included in the model.

4) Coaching

Finally, as the teachers implement the problem solving tasks in their classroom, professional developers make on site visits to coach teachers. Some of this coaching may be around launching, facilitating, or debriefing problem solving tasks, or it may relate to an area of improvement specified by the teacher.

Internet-Based Communication Tools

The fourth component of the mathematics problem solving model is access to and structured use of internet-based communication. The model uses Moodle, a course management system for online courses, to create an online learning community, the members of which are the teachers involved in the professional development. This creates the opportunity for teachers in different locations to communicate with one another outside of the professional development meetings. Teachers use Moodle to

respond to discussion questions posed by the professional developers, and to respond to the postings of others.

The teachers participate in reflective inquiry groups via Moodle. These groups are formed during the professional development based upon areas of improvement identified by the teachers. Through Moodle, the teachers in a reflective inquiry group are able to share and discuss challenges and successes they faced while working to improve in their focus area.

Finally, Moodle is a convenient place to post the tasks, feedback guide, data collection documents, and other documents that the teachers might need to use the tasks and participate in the research. Teachers may also use Moodle to ask questions of one another and the professional developers or researchers outside of professional development meetings.

Research on the Model

In 2006, a group of researchers, commissioned by the Northwest Regional Educational Laboratory and funded by the National Science Foundation, launched a five year study to determine the effectiveness of, and make evidence based modifications to, the NWREL Mathematics Problem Solving Model. The researchers identified five primary research questions to be addressed in the study, the results of which would inform ongoing changes to the model to be evaluated over the course of the study.

The Five Research Questions

Question 1:

Does the implementation of the NWREL Mathematics Problem Solving Model have a positive effect on student achievement in and attitudes toward mathematics and mathematical problem solving?

Question 2:

Do teachers use the NWREL problem solving formative feedback guide in valid and reliable ways for formative assessment decisions?

Question 2a:

Do teachers attain a high degree of interrater reliability in the use of the formative feedback guide?

Question 2b:

Do the teachers' abilities to assess students' problem solving skills improve with the implementation of the program?

Question 2c:

What is the evidence of the teachers' use of assessment information from students to inform instructional plans?

Question 3:

Does the professional development result in changed teacher practices?

Question 3a:

Does the follow-up instruction use the information that teachers identified in student performances in the way that the teacher had planned?

Question 3b:

Do the teachers' uses of the formative feedback guide change the classroom discourse?

Question 4:

How do teachers use the web-based technological support of rubric-based discourse with middle school mathematics students? Does this contribute to increased implementation and student achievement?

Question 5:

Do teachers incorporate the elements of the NWREL Mathematics Problem Solving Model into ongoing instructional practices that are evident after interactions with the trainers have ended?

The second question, investigating the teachers' use of the NWREL formative feedback guide for assessment decisions, is elaborated on in the following chapter, and is the principal research question of this dissertation.

Chapter 5 – Methodology

Principal Research Question

As part of the larger research project, this work seeks to address the effectiveness of the formative feedback guide and the professional development components of the NWREL Mathematic Problem Solving Model in facilitating formative assessment in the classroom.

Principal Research Question: Do teachers use the NWREL formative feedback guide in valid and reliable ways for formative assessment decisions?

To fully address the question, three particular issues are examined:

- *Inter-rater reliability:* Do teachers attain a high degree of inter-rater reliability in the use of the NWREL formative feedback guide?
- *Improvement over time:* Do the teachers' abilities to assess students' problem solving skills improve with the implementation of the model?
- *Follow-up Instruction:* What is the evidence of the teachers' use of assessment information from students to inform instructional plans?

This chapter begins by discussing the treatment undergone by the teachers involved in the first year of the study as it relates specifically to the principal research question. The data collection instruments and processes are then described. The final part of the chapter discusses the methods of data analysis.

Participants

Ten teachers from different middle schools in a western state participated in the professional development and the associated research study. These teachers became aware of the project through advertisements in a regional mathematics teaching journal or through supervisors contacted by NWREL. The teachers were self-selected and informed that a stipend would be paid for their participation in the project.

The participating teachers were from schools in midsize towns or suburbs, located within a 100-mile radius of a western city. One teacher was from an alternative middle school for students who had been expelled from the regular regional middle schools. The remaining teachers were from traditional middle schools. The schools varied in ethnic and socioeconomic diversity, with some schools being fairly diverse, and others more homogenous.

The teachers were asked to select one of their mathematics classes with which to implement the mathematics problem solving model as prescribed in the professional development, and on whom data would be collected. The researchers requested that the class selected not be a remedial class, but otherwise placed no restrictions on the class selected. Therefore, the participating classes were at mixed levels and incorporated all middle grades. The sizes of the classes selected by the participating teachers varied from 8-10 students at the alternative middle school, to 42 students at a suburban school. It should be noted that the class size and student composition at the alternative middle school varied throughout the school year as students returned to the regular middle schools or new students expelled from the regular middle schools began attending the

alternative school. Moreover, the teachers were allowed and encouraged to implement the model in several of their classes, yet data were collected only on the class identified as the research class by the teacher.

A group of ten control teachers were recruited from middle schools within the 100-mile radius, from schools deemed to be similar to those of the treatment teachers. However, those teachers have no bearing on the portion of the study discussed herein, and so the details of their participation will not be outlined.

Treatment

The treatment teachers participated in several workshops in preparation for implementing the mathematics problem solving model in their classrooms. The first workshops took place in late March and early April of 2006. Teachers were asked to attend one of two Saturday sessions during which the researchers provided an overview of the project. These workshops were used to discuss the research component of the project, including the primary research questions and the nature of data to be collected. The characteristics of the NWREL exemplary tasks were included in the project introduction.

A second set of workshops took place in early May. Again, these were two Saturday sessions, of which the teachers were asked to select one to attend. During these workshops, the primary focus was building understanding of summative and formative assessment practices.

Spring Workshop

During the second set of workshops in May, the teachers were first asked to define assessment. This led to an activity where the teachers created a Venn diagram relating assessment, evaluation, and testing. This launched a discussion about the vague nature of the terminology used to describe facets of assessment. Teachers discussed their own classroom practices as they related to assessment issues. As a group, the teachers and NWREL staff generated a list of purposes of assessment. The main purposes, including determining grades, as tools for student learning, and evaluating one's own teaching, were prioritized by the teachers to identify the most important uses of assessment. In both sessions, the teachers revealed that they attached significant importance to grading as a purpose of assessment. The teachers sorted the purposes of assessment into formative and summative uses of assessment, and the workshop concluded with a discussion of beliefs about and importance of formative and summative assessment.

The Summer Workshop

The bulk of the teacher inservice education took place during a five day workshop in July, 2006. The workshop was attended by all ten teachers, as well as two other teachers who did not complete the project and were not included in the research. In the description below, the workshop activities pertinent to the dissertation research question are detailed, while other workshop activities are briefly outlined for completeness.

Day One:

The first day began with a discussion to establish norms for group work during the course of the workshop. One of the researchers led a group discussion about the five strands of mathematical proficiency as described in *Adding it Up* (Kilpatrick et al, 2001). The teachers spent 90 minutes working on one of the NWREL problem solving tasks, solving the task in groups, discussing the solutions generated by the teachers, discussing the solutions their students might generate, and identifying the important mathematical ideas that were part of the task.

The teachers were then guided in a brief discussion of two papers about formative assessment, and encouraged to reflect on their own assessment practices. This was followed by an introduction to the NWREL formative feedback guide. The teachers discussed the relationship between the traits in the feedback guide and Kilpatrick et al's (2001) strands of proficiency, as well as the state problem solving assessment evaluation rubric. Using sample student responses to a NWREL problem solving task, the teachers then practiced using the NWREL feedback guide to evaluate those solutions. This process included lengthy discussion about the observations made of the student work. Attention was also drawn to using the student papers and the feedback guide to determine subsequent instructional steps.

Day Two:

The second day of the summer workshop began with a 90 minute investigation of another of the NWREL problem solving tasks. As on the first day, the teachers prepared solutions to the task, discussed the different solution methods generated, and were led by the researchers in an investigation of the important mathematics in the task. The

remainder of the day focused on components of the project not directly relevant to the principal research question of this dissertation: a discussion of classroom discourse, an introduction to reflective inquiry, and an introduction to Moodle.

Day Three:

Day three began with an exploration of Stein et al's (2000) framework of mathematical tasks. The teachers were given problems to sort into the four categories of mathematical tasks as set forth in the framework, and also asked to consider the cognitive demand of tasks that they had used in their own classrooms prior to the professional development. The teachers then engaged in another 90 minute session during which they solved another NWREL problem solving task, discussed solutions, and investigated more deeply the mathematical content of the problem. Another 90 minute session had the teachers using the feedback guide to evaluate sample student papers on the same task they solved.

The researchers asked the teachers to provide written explanations of their understanding of both the plan for the project and the reasons behind the project so that any misunderstandings could be rectified before the end of the summer professional development session.

Day Four:

The teachers began the fourth day by solving the remaining middle school level NWREL problem solving tasks in groups. Four groups each solved three of the remaining six tasks, so that every task had been solved by some of the teachers. The whole group discussed the solutions generated, and also identified the cognitive demand (Stein et al, 2000) of the NWREL tasks.

The teachers, as a group, selected five of the nine NWREL middle school level tasks to implement during the course of the school year. The research study is based on the implementation of these selected NWREL tasks: 1) Cross Country Scoring, 2) Design a Dartboard, 3) Kyle's Tiles, 4) Mathemagicians' Cookies, and 5) Spinner Elimination (see the Appendix for actual copies of the tasks as they appear in the NWREL materials).

The Cross Country Scoring task gives students the finishing order of runners on four different cross country teams and asks the students to develop four different scoring methods, one so that each team would win. The Design a Dartboard task asks students to create a dartboard in a shape that is neither square nor circular such that 15% of the total area is in an inner section, 25% is in a middle section, and 60% is in an outer section. Kyle's Tiles asks students to design the "best" decorative tiled portion of a wall measuring 28" x 30", using any combination of 4"x4" tiles costing $2, 4"x6" tiles costing $3, and 6"x6" tiles costing $4. The Mathemagicians' Cookies task asks students to package 360 cookies costing $.10 each in three different sized packages, and price the packages so they would make a minimum $50 profit on the sale.

The Spinner Elimination Task has students create a spinner with eight equally sized spaces, filling the spaces with their choice of integers between 0 and 9 inclusive. The students use their spinners to play a game where they may spin the spinner as many times as they like, marking the product of their spins off on a chart of numbers between 1 and 50. They are then asked to make recommendations about the best way to design a spinner to attain the most values between 1 and 50.

In addition to choosing the tasks, the teachers were also guided in the identification of the important mathematics in each of the tasks, as well as the state

mathematics standards which each task addressed. The remainder of the day was spent on inservice activities not related to the research question addressed here: an activity around peer assessment and further development of the teachers' reflective inquiry questions.

Day Five:

On the fifth day, the teachers individually solved all five of the tasks they had selected to implement in their classrooms throughout the course of the project. Special attention was paid to identifying the important mathematics in the tasks, and also anticipating the types of solutions students may generate. Furthermore, the teachers discussed the curricular units into which the tasks would fit, so that they could integrate the tasks into the mathematics already being taught, rather than conceptualizing the tasks as separate activities, unrelated to regular classroom functioning. The remainder of the activities on the fifth day included an activity about classroom questioning, further refinement of reflective inquiry questions, and wrap up of the week's activities.

Classroom Implementation

Over the course of the 2006-2007 school year, the treatment teachers were expected to: 1) implement the five NWREL problem solving tasks in the identified class, 2) evaluate student performance on each task using the formative feedback guide, and 3) debrief the task in a follow-up class.

Task Implementation

The teachers implemented the five tasks during the year at times they deemed appropriate for continuity of mathematical ideas within their classrooms. Since the teachers were from different school districts with different curricula, the tasks were implemented at different times. For example, the task one teacher implemented first in the school year may have been implemented last by another teacher.

In addition to the five tasks, the teachers were provided with teacher guides to the tasks (see Appendix) which identified the important mathematics of the task, provided suggestions for launching the task, and suggested extensions to the task. The researchers left most of the implementation decision making to the teachers, allowing them to launch the task as they saw fit and to determine if students would work individually or in small groups.

Performance Evaluation

The teachers were asked to evaluate every piece of student work from the class they identified on each task using the NWREL formative feedback guide. The researchers did not require the teachers to address every problem solving trait on every task, but rather allowed teachers to identify a subset of traits on which they would focus for a given task. Teachers were allowed to provide feedback to students in any way they chose, including providing students with the feedback guide; using a student-worded version of the feedback guide; and providing written comments from, or not from, the feedback guide.

Task Debrief

The researchers requested that the teachers debrief each task in a timely fashion after the students solved the task. The debrief was expected to be conducted in a whole class discussion format. It was suggested that tasks be launched on a Friday and debriefed on a Monday so that teachers would have time to use the feedback guide on student papers, but the teachers were not required to follow this format.

The teachers were asked, prior to the debrief, to identify pieces of student work to be shared, to sequence those pieces of work in some order, and to identify rationales both for the selection of the individual pieces of work and for the sequencing. The teachers were given freedom to decide if the student would present his or her own work, or to present it on the student's behalf, although it was expected that the whole class be able to see the student's solution.

Ongoing Professional Development

The professional development component of the project continued throughout the course of the 2006-2007 school year, with in-school coaching sessions, whole group workshops, online discussions, and classroom visits from researchers.

In-School Coaching Sessions

A NWREL staff member visited each teacher's classroom several times during the school year. During these visits, the staff member would act as a coach to the teacher, working with the teacher to improve the practices he or she identified in the reflective inquiry question. These coaching sessions were largely determined by the teacher's

needs, and varied between teachers. The coach engaged in such activities as teaching the teacher's class to model effective practices, and observing the teacher teaching and providing feedback and constructive suggestions. Some coaching sessions took place during the launch or debrief of the NWREL tasks, but some sessions were during regular classroom episodes.

Whole-Group Workshop Sessions During the School Year

Three whole-group Saturday workshops were conducted throughout the course of the school year. During the first workshop in October, 2006, the teachers shared their experiences implementing the NWREL mathematics problem solving model in their own classrooms. They discussed the modifications they were making to their classroom practices, especially those around formative assessment and providing feedback to students. A discussion of the data collection, both to ensure teacher understanding of the expectations, and to allow teachers input into the data collection methods, followed.

The second workshop took place in February of 2007. Researchers observing the teacher's debriefs had noticed, both through in-class observations and discussions with teachers, that the debriefs were often not centered around important mathematical ideas, but rather around display of diverse strategies. To address this, the second workshop focused on identifying the important mathematics in the tasks, and using the important mathematics to guide a task debrief.

The researchers purposefully selected several pieces of student work from a task that most of the teachers had implemented to illuminate the mathematical ideas upon which the problem was based. The ensuing discussion directed teacher focus on those

mathematical ideas, and their usefulness in structuring a debrief. This was contrasted with a debrief focused on displaying different problem solving strategies. The researchers were careful to reiterate the goals of sharing student work in the task debrief discussion, particularly as tools to illuminate important mathematics or debunk common misconceptions.

The final workshop took place in June, 2007. Final data were collected from the teachers at this time. The teachers presented the results of their reflective inquiries, and the group discussed the usefulness of the NWREL formative feedback guide in providing constructive comments to students, the effectiveness of using student work to structure debrief classes, and the adequacy of the problem solving tasks. This provided the researchers with feedback useful in making modifications to the model before its implementation with the second cohort of teachers in the following school year.

Online Discussions

As mentioned previously, Moodle served as a tool for ongoing communication among teachers and between teachers and researchers. The researchers used Moodle to further provide educational experiences for the teachers, particularly in the form of discussions around assigned readings or tasks. As the teachers were able to create their own discussions, either within their reflective inquiry groups or for the entire cohort, they had the opportunity to direct their own learning throughout the course of the project.

Classroom Visits from Researchers

A researcher attended every debrief of nine of the ten participating treatment teachers. The tenth teacher was not observed due to the unexpected move of one of the research staff members. During the debrief, the researchers acted as non-participating observers of the class. However, the contact with researchers allowed teachers to share concerns or questions about both the implementation of the model and data collection on a regular schedule. The researchers addressed some data collection issues at the time, and brought any remaining issues raised to the attention of the NWREL staff so that common or more complicated concerns could be addressed in whole group workshops. The researchers did not critique or provide suggestions or assistance to the teachers in these visits.

Data Collection

The data collected to address this question was motivated by the particular issues examined. This led to collection of three primary types of data: 1) written student work, 2) teacher reporting of student performance, and 3) teacher plans for sharing student work. Data of these types was collected from nine of the ten treatment teachers, due to the unexpected departure of one NWREL staff member.

Written Student Work

For each of the five tasks, teachers were asked to submit photocopies of every piece of unmarked student work. This included each student's entire solution without

any markings made by the teacher. This enabled the researchers to examine the students' solutions without influence from teacher comments.

Not every teacher submitted data on every task. Written student work was collected from eight teachers on the Cross-Country Scoring problem, from six on the Design a Dartboard problem, from all nine teachers on the Kyle's Tiles problem, from seven on the Mathemagicians' Cookies problem, and from eight teachers on the Spinner Elimination problem. The data collection is summarized in Table 5.1 on page 96.

Teacher Reports of Student Performance

If the teachers provided written comments to their students, they were asked to submit photocopies of the marked student work as well as the unmarked. In addition, the teachers were asked to provide their evaluation of student performance on the five traits of the feedback guide for every student on every task. This information was collected in several forms, either through a researcher-created proficiency reporting form, as recorded directly on the NWREL formative feedback guide, or in a spreadsheet.

Student Proficiency Reporting Data

The researchers created a form for the teachers to indicate student performance in each of the five traits (Figure 5.1). This form asked the teacher to circle a student proficiency level for the trait, and to circle the criteria statements from the feedback guide that supported this decision. Space was provided for the teacher to add comments, either to the student, or to justify the proficiency level or criteria statements chosen. While the researchers intended this to serve only as a data collection tool, some teachers did use the form to provide feedback to their students.

Some of the teachers chose to report student proficiency levels by circling levels and criteria statements directly on the NWREL formative feedback guide. Again, some of the teachers shared this information with their students; others provided it only to the researchers. One teacher created his own spreadsheet to report student proficiency levels on the problem solving traits. This spreadsheet consisted of one row for each student and one column for each of the five traits. In the intersection of each row and column, the teacher indicated the student's proficiency level on that trait. The spreadsheet was shared only with the researchers. (This teacher also created student-friendly restatements of several of the criteria statements, and used those to report performance information to students.)

Some reporting of student proficiency was submitted by every teacher from whom written student work was collected on every task except for the Spinner Elimination problem. For the Spinner Elimination problem, one of the eight teachers who submitted written student work did not report his students' proficiency levels on any of the five traits.

Moreover, as previously mentioned, the teachers were not required to evaluate student performance on each of the five traits for each of the tasks. Table 5.1 summarizes the reporting of proficiency levels by teacher by task for the traits. (The names of the teachers have been changed to protect anonymity.)

Figure 5.1: Proficiency Reporting Form

Mathematics Problem Solving Feedback Guide Student Levels

Instructions: For each trait, please circle **one** proficiency level. Then circle all criteria statements for that level that you observed in the student's work.

Student:_____

Conceptual Understanding	**Emerging**	**Developing**	**Proficient**	**Exemplary**
Criteria Statements	**1**	**2**	**3**	**4**

Notes/ Comments:_____

_____

_____

Strategies and Reasoning	**Emerging**	**Developing**	**Proficient**	**Exemplary**		
Criteria Statements	**1**	**2**	**3**	**4**	**5**	**6**

Notes/ Comments: _____

_____

_____

Computation and Execution	**Emerging**	**Developing**	**Proficient**	**Exemplary**	
Criteria Statements	**1**	**2**	**3**	**4**	**5**

Notes/ Comments:_____

_____

_____

Communication	**Emerging**	**Developing**	**Proficient**	**Exemplary**	
Criteria Statements	**1**	**2**	**3**	**4**	**5**

Notes/ Comments:_____

_____

_____

Insights	**Emerging**	**Developing**	**Proficient**	**Exemplary**
Criteria Statements	**1**	**2**	**3**	**4**

Notes/ Comments:_____

_____

_____

Table 5.1: Teacher Reporting of Proficiency Levels by Task

	Cross-Country Scoring	**Design A Dartboard**	**Kyle's Tiles**	**Mathe-magician's Cookies**	**Spinner Elimination**
Alex	CU, SR, CE, Comm, Ins		CU, SR, CE, Comm, Ins	CU, SR, CE, Comm, Ins	
Brian	CU, SR, CE, Comm, Ins	CU, SR, CE, Comm, Ins	CU, SR	CU, SR, CE, Comm, Ins	CU, SR, CE, Comm
Carmen	CU, SR, CE, Comm, Ins	CU, SR, CE, Comm, Ins	CU, SR, CE, Comm, Ins	CU, SR, CE, Comm, Ins	Comm, Ins
Daphne			CU, SR, CE, Comm, Ins		
Ellie	CU, SR, CE, Comm, Ins	CU, SR, CE, Comm, Ins	CU, SR, CE, Comm, Ins	CU, SR, CE, Comm, Ins	CU, SR, CE, Comm, Ins
Fatima	CU, SR, CE, Comm, Ins	CU, SR, CE, Comm, Ins	CU, SR, CE, Comm	CU, SR, CE, Comm, Ins	CU, SR, CE, Comm, Ins
Grace	CU, SR, CE, Comm, Ins		CU, SR, CE, Comm, Ins		CU, SR, CE, Comm, Ins
Helen	CU, SR, CE, Comm, Ins	CU, SR, CE, Comm, Ins	CU	CU, SR, CE, Comm, Ins	CU, SR, CE, Comm, Ins
Ian	CU, SR, CE, Comm, Ins	CU, SR, CE, Comm, Ins	CU, SR, CE, Comm	CU, SR, CE, Comm	

The problem solving traits are abbreviated as follows:

- Conceptual Understanding (CU)
- Strategies and Reasoning (SR)
- Computation and Execution (CE)
- Communication (Comm)
- Insights (Ins)

Teacher Plans for Sharing Student Work: Data Collection

The researchers developed an Instructional Sequence Analysis form to collect teacher information on their plans for sharing student work during the debrief session of the class (see Appendix B for the actual form). Using this form, teachers identified the student papers they planned to have shared, and the order in which they planned to have them shared. The teachers used the form to provide rationales for choosing each piece of student work, and for sequencing the solutions in the order chosen. The teachers were asked to complete this form prior to the debrief class for each of the five tasks. The submission of instructional sequencing plans is summarized in Table 5.2.

Table 5.2: Summary of Teacher Reporting of Instructional Sequences

	Cross-Country Scoring	Design a Dartboard	Kyle's Tiles	Mathe-magicians' Cookies	Spinner Elimination
Alex	X		X	X	X
Brian	X	X	X	X	X
Carmen	X	X	X	X	X
Daphne			X		
Ellie	X	X	X	X	X
Fatima	X	X		X	X
Grace	X		X		X
Helen	X	X	X	X	X
Ian	X	X	X	X	X

Informed Consent

Both the teachers participating in the study and the parents of their students were informed prior to participation in the study of the intended use of data collected and the potential for its use in publication. NWREL ensured that the privacy of the teachers participating in the study and the privacy of their students were protected through the use of pseudonyms for the teachers and no publication of student names or pseudonyms. Parents or legal guardians consented in writing to have their students' work used in the research. All student papers examined, included in data analysis, or referred to in this work are from students whose guardians consented to their participation in the study and were informed of the full scope of the research and use of data.

Data Analysis

The principal research question was addressed through examination of three issues: inter-rater reliability, improvement over time, and follow-up instruction. The data analysis methods are explained as they relate to each of these three issues.

Inter-rater Reliability

Do teachers attain a high degree of inter-rater reliability in the use of the NWREL formative feedback guide?

The unmarked pieces of student work and the teachers' reports of student proficiency levels on the five traits included in the feedback guide are the data used to address this question. The researcher and an undergraduate student assistant researcher began by developing a task-specific feedback guide for each of the five tasks. This step was strongly grounded in research indicating that while a general rubric is best suited for teaching problem solving skills, a task-specific rubric is more useful in evaluating student performance on a task (Arter & McTighe, 2001; Stiggins, 2001).

The development of each task-specific feedback guide began with careful consideration of the important mathematics in the problem, and led to a restatement of the original criteria statements in the NWREL feedback guide in such a way that they appropriately reflected levels of performance on the specific task. The researchers independently used the task-specific feedback guide to examine twenty to thirty pieces of student work on the task. Based on observations of student solutions, both researchers developed recommendations for modifications to the first draft of the task specific feedback guide. They then met to discuss these recommendations and to determine modifications to the first draft. The agreed upon modifications were incorporated and a working draft of the task-specific feedback guide was established. The finalizing of a usable task-specific feedback guide involved extensive discussion of the restatements of

the criteria statements, and the researchers were reasonably confident that they both had the same understanding of the task-specific feedback guides.

Once a task-specific feedback guide was established, the two researchers independently analyzed every piece of student work submitted for that task. Each researcher assigned a proficiency level for each trait for which the teacher had evaluated proficiency on every piece of student work. For example, since Brian had reported his students' proficiency levels on Kyle's Tiles only on the traits of conceptual understanding and strategies and reasoning, the researchers each assigned proficiency levels on only those same traits as well.

After assigning proficiency levels for the student work on that task, the researchers then met to compare evaluations. In the case of agreement between the two researchers, that proficiency level was recorded as the "expert" analysis. In the case of disagreement, the researchers discussed their differences in interpretation and came to consensus on a proficiency level. This consensus level became the expert analysis of the student performance.

The expert analyses of performance were compared to the teacher analyses. That is, for every piece of student work, and for every trait analyzed by the teacher, the teacher's determination of proficiency level on that trait was compared to the expert's. The comparisons were analyzed in three ways: (1) simple percentage agreement between the expert and teacher assignments of proficiency levels, (2) calculating the Cohen's kappas for measurements of reliability on each task, and (3) the distribution of the differences between teacher assignments of proficiency levels and expert assignments.

Percent Agreement

For each task, the instances of agreement between the expert analysis of student proficiency and the teacher analysis of proficiency on each trait were counted. Percent agreement was calculated by dividing instances of agreement by total reports of student proficiency on that trait for that task. Total percent agreement was calculated for each task, each trait, and over all traits and tasks.

Cohen's Kappas

Cohen's kappa is a statistical measurement of inter-rater reliability which accounts for the possibility of scoring agreement due to chance. The Cohen's kappas were calculated for each task in each trait, totals for each task over all traits, totals for each trait over all tasks, and a total over all tasks and traits.

Distribution of Differences

Percent agreement indicates the extent to which teachers' ratings of student work compared to the expert rating, but because there are four proficiency levels, the differences in levels between teacher and expert ratings provide useful information as well. The differences were determined by comparing the teacher's rating of proficiency to the expert's, and assigning a numerical value to that comparison. These values range from -3 to 3. A value of -3 would indicate that the teacher scored the student work 3 proficiency levels lower than the expert (a teacher rating of emerging and an expert rating of exemplary), and a value of 3 would indicate that the teacher scored the student work 3 proficiency levels higher than the expert (a teacher rating of exemplary and an expert rating of emerging). The relative frequencies of these differences are examined by task and by trait.

Furthermore, mean level differences were calculated by using the absolute value of the differences in order to reflect the magnitude of the mean difference, rather than possibly being minimized by offsetting differences in the positive and negative directions. The mean level differences were calculated for each trait and for each task.

Examining all three measures of agreement among expert and teacher raters provides a more robust picture of the inter-rater reliability in use of the NWREL formative feedback guide.

Improvement over Time

Do the teachers' abilities to assess students' problem solving skills improve with the implementation of the model?

The second research issue involved investigating the changes in the treatment teachers' use of the NWREL formative feedback guide over the course of the school year during which they participated in the research project. To address this, the same data used to determine inter-rater reliability is analyzed. This analysis, however, focuses on the teachers' individual reliabilities, and the trends over time exhibited in their use of the guide.

Two of the same measurements used to determine inter-rater reliability were considered in this analysis. The percent agreement of each teacher's ratings with the expert rating was considered in the order in which the tasks were implemented to determine if any trends were observed over time. The mean level of difference was calculated for each teacher on each task, and then the individual teacher's mean levels of

difference were considered in the order in which the tasks were implemented to determine if any trends were observed over time.

Follow-Up Instruction

What is the evidence of the teachers' use of assessment information from students to inform instructional plans?

The third research issue involved describing the evidence of the teachers' use of assessment information to make follow-up instructional decisions. To describe this evidence, the researcher used both the teachers' reported instructional sequencing from the Instructional Sequence Analysis form and the written student work.

First, the responses on the Instructional Sequence Analysis forms were analyzed to determine categories of rationale for selecting particular pieces of student work and for sequencing in a particular order. To determine categories of rationale for selecting pieces of student work, the researcher and the undergraduate assistant researcher began by identifying key words or phrases in the rationales as written by the teachers. Based on these key words and phrases, the two researchers identified themes and overarching categories of rationales. They sorted the rationales provided by the teachers into these overarching categories, and made modifications to the overarching categories to include all types of rationales provided.

Within several of the overarching categories, the primary researcher determined that subcategories were necessary to more accurately categorize the data. The data were then sorted, when appropriate, into the subcategories, and the researcher made modifications to the subcategories to include all types of rationales provided.

Essentially the same process was followed for the categorization of the teachers' rationales for sequencing the shared student work in the order chosen. In this case, the primary researcher examined the sequencing rationales to identify themes and overarching categories. The sequencing rationales were sorted into those categories, and then modifications were made until the categories included all types of sequencing rationales indicated by the treatment teachers.

Additionally, the Instructional Sequence Analysis forms were used to determine the extent to which the teachers appropriately chose pieces of student work based on the rationale they indicated. For each piece of student work identified by the teacher to be shared in the debrief, the researcher examined the corresponding piece of student work to determine to what extent the rationale indicated by the teacher was evident in the piece of student work. To describe this evidence, the researcher categorized the student work as exemplifying the rationale strongly, some, or not at all. In the case where the teacher provided comments to the student, the researcher also investigated the written comments to determine the extent to which the comments corresponded to the rationale cited. This correspondence was also categorized as strong, some, or none.

Finally, for the cases for which it was determined that no evidence of the rationale indicated by the teacher was present in the student work identified, the researcher examined the set of class papers to determine if a better example of a student work displaying that rationale was present.

The results of these analyses are summarized in the following results chapter.

Chapter 6 – Results

The principal research question for this dissertation is "Do teachers use the NWREL formative feedback guide in valid and reliable ways for formative assessment decisions?" In Chapter 5, three particular issues to be addressed were identified: 1) the inter-rater reliability attained by the teachers in the use of the NWREL formative feedback guide, 2) the improvement over time in the teachers' assessment abilities, and 3) the evidence of the teachers' use of assessment information in making instructional decisions. The results of the investigations of each of the three parts are discussed here.

Inter-Rater Reliability:

Do teachers attain a high degree of inter-rater reliability in the use of the NWREL formative feedback guide?

To determine the teachers' inter-rater reliability in the use of the NWREL formative feedback guide, three measurements were investigated: percent agreement between teacher and expert ratings, Cohen's kappa coefficients of agreement, and level differences between teacher and expert ratings.

Percent Agreement

The agreement percentages in each task for each trait, over all traits for each task, and over all tasks for each trait, are summarized in Table 6.1 . Over all five NWREL tasks, and over all five of the problem solving traits evaluated with the

Table 6.1. Percent Agreement, Teacher Rating and Expert Rating

	Cross-Country Scoring	Design a Dartboard	Kyle's Tiles	Mathe-magicians' Cookies	Spinner Elimination	Over all tasks
Conceptual Understanding	.56	.74	.62	.67	.69	.65
Strategies & Reasoning	.53	.56	.69	.75	.67	.64
Computation & Execution	.48	.65	.69	.57	.57	.60
Communi-cation	.56	.57	.69	.61	.52	.60
Insights	.48	.36	.55	.42	.62	.58
Over all traits	.52	.58	.66	.61	.61	.60

feedback guide, the teachers achieved 60% agreement with the expert rating. The trait with the greatest agreement percentage was conceptual understanding, with 65% agreement, while the lowest agreement was observed in insights, with 58% agreement. Kyle's Tiles exhibited the greatest agreement over all the tasks, with 66% agreement. The lowest percentage of agreement by task was the Cross-Country Scoring task, with 52% agreement.

Generally, a minimum 70% agreement is considered acceptable; however, some standardized tests expect 85% to 90% agreement (Goldberg & Michaels, 1995). Using the 70% agreement standard, the agreement between teacher and expert rating was acceptable only when evaluating conceptual understanding on Design a Dartboard and strategies and reasoning on Mathemagicians' Cookies. Nearly 70% agreement was achieved on several tasks under several traits: strategies and reasoning, computation and

execution, and communication on Kyle's Tiles; and conceptual understanding on Spinner Elimination.

Cohen's Kappa Coefficients of Agreement

Measuring simple percentage agreement does not account for the agreement that would be expected even if ratings were assigned at random. Cohen's kappa coefficient is a statistical measurement that measures the agreement that can be attributed to factors other than chance, and is considered to be more robust in measuring reliability than simple percent agreement. The range of kappa coefficients indicating acceptable inter-rater reliability is a subject of debate by some researchers. However, Watkins and Pacheco (2001) established a scale that is considered by many to be an acceptable measure of quality of agreement (see Table 6.2).

Table 6.2. Cohen's Kappa Coefficients

Cohen's kappa coefficient (κ)	Level of agreement indicated
< .40	poor
.40 - .59	fair
.60 - .74	good
.75 - .99	excellent
1.0	perfect

Table 6.3 summarizes the Cohen's kappa coefficients for the tasks and problem solving traits evaluated by the teachers. The Cohen's kappa coefficient over all the traits

and all the tasks is 0.34. The task with the highest Cohen's kappa coefficient is Kyle's Tiles, with a kappa of 0.45, and the task with the lowest Cohen's kappa coefficient is Cross-Country Scoring, with a kappa of 0.25. The problem solving trait with the highest Cohen's kappa coefficient is conceptual understanding, with kappa 0.5. Insights has the lowest kappa of 0.16.

Table 6.3. Cohen's Kappa Coefficients (κ): Agreement of Teacher/ Expert Ratings

	Cross-Country Scoring	Design a Dartboard	Kyle's Tiles	Mathe-magicians' Cookies	Spinner Elimination	Over all tasks
Conceptual Understanding	0.26	0.50	0.54	0.47	0.36	0.50
Strategies & Reasoning	0.21	0.28	0.37	0.56	0.34	0.39
Computation & Execution	0.03	0.37	0.50	0.28	0.01	0.26
Communi-cation	0.30	-.02	0.43	0.31	0.14	0.28
Insights	0.11	0.01	0.24	0.18	0.36	0.16
Over all traits	0.25	0.29	0.45	0.38	0.30	0.34

Using Watkins and Pacheco's scale of agreement, overall agreement is poor, the agreement for all but one of the tasks is poor, and the agreement for all but one of the traits is poor. Agreement calculated over all traits for Kyle's Tiles is fair, and agreement for conceptual understanding over all five tasks is fair. Notably, no agreement in any of the categories is better than fair. In fact, only 20% of the trait/task evaluations exhibit even fair agreement.

Level Differences Between Teacher and Expert Ratings

Both percent agreement and Cohen's kappa coefficients measure the extent of exact agreement between raters. However, since the NWREL formative feedback guide provides a framework for evaluating student problem solving performance over a range of four proficiency levels, examining the distribution of differences between teacher ratings and expert ratings provides information on the degree of disagreement. Table 6.4 summarizes the distribution of these differences by trait. Table 6.5 summarizes the distribution of differences in ratings by task.

Table 6.4. Differences between Teacher & Expert Ratings by Problem Solving Trait

Problem Solving Trait	Percentage with level of difference							Mean level of difference
	-3	-2	-1	0	1	2	3	
Conceptual Understanding	0	0	.11	.70	.18	.01	0	.31
Strategies & Reasoning	0	.01	.12	.65	.22	.01	0	.38
Computation & Execution	0	.01	.23	.60	.16	0	0	.41
Communication	0	.01	.19	.60	.18	.01	0	.31
Insights	0	.03	.31	.48	.16	.02	0	.57

Table 6.5. Differences between Teacher & Expert Ratings by Task

	Percentage with level difference							Mean level of difference
Task	-3	-2	-1	0	1	2	3	
Cross-Country Scoring	0	.01	.21	.52	.25	.01	0	.5
Design a Dartboard	0	.01	.15	.58	.24	.01	0	.33
Kyle's Tiles	0	0	.16	.69	.14	.01	0	.32
Mathemagicians' Cookies	0	.03	.03	.77	.17	0	0	.26
Spinner Elimination	0	.01	.17	.61	.12	0	0	.31

The significant majority of rating differences on the problem solving traits are within one proficiency level. This also provides more information about the problem solving trait insights, the trait with the lowest percent agreement. A difference of (positive) 1 represents a teacher rating of student performance one proficiency level higher than the expert. This occurs with similar frequency on all problem solving traits. However, a difference of -1 (teacher rating one level below the expert) occurs with much greater frequency on the insights trait.

The mean level of difference is calculated by using the absolute value of the differences so that it reflects the magnitude of the mean difference, rather than possibly being minimized by offsetting differences in the positive and negative directions. While the percentages of agreement and the Cohen's kappa coefficients have shown poor to fair exact agreement, the mean levels of difference indicate that the level of disagreement is not severe. On all the tasks, and for each of the traits, the mean levels of difference are less than one, indicating that the teachers are generally rating student papers within one proficiency level of the expert rating. Moreover, for all of the problem solving traits

other than insights, the mean levels of difference are within 0.5. For all of the tasks, the mean levels of difference are at most 0.5.

Furthermore, examination of the mean levels of difference for each teacher indicates that the teachers individual ratings are similar to the means (see table 6.6). The teachers individually also display mean levels of difference of at most 0.5, so it is reasonable to conclude that the cumulative data reported for the problem solving traits and the tasks are not skewed by one teacher having a very low or very high mean level of difference in rating.

Table 6.6. Percentage Differences by Teacher

	Percentage with level of difference							
Teacher	-3	-2	-1	0	1	2	3	Mean level of difference
Alex	0	.02	.25	.60	.11	.01	0	.42
Brian	0	.02	.20	.54	.22	.02	0	.5
Carmen	0	.01	.22	.60	.15	.02	0	.43
Ellie	0	.01	.18	.63	.17	0	0	.37
Fatima	0	0	.08	.66	.25	0	0	.33
Grace	0	0	.14	.61	.24	.01	0	.4
Helen	0	.01	.16	.60	.21	.01	0	.41
Ian	0	.02	.25	.61	.12	0	0	.41

Improvement Over Time:

Do the teachers' abilities to assess students' problem solving skills improve with the implementation of the model?

To determine if the teachers improved in their use of the NWREL formative feedback guide over the course of the project, the teachers' percentages of agreement and mean levels of difference on the tasks over the course of the project were examined. Data were collected on more than one task for eight of the nine teachers. Of those eight teachers, data were collected for five on all tasks, and for the remaining three on three of the five tasks. The order in which the teachers implemented the tasks in their classrooms is summarized in Table 6.7.

Table 6.8 shows the percentage agreement of each teacher with the expert rating on each of the tasks implemented in the sequence in which they were implemented.

Examining the trends in percentage agreement reveals few observable patterns over time. The mean percentages of agreement indicate that agreement does not improve over time, but does stay within a 15% range. No individual teacher exhibits improvement in percentage agreement over time, nor is there evidence of the teachers' agreement with the expert decreasing over time.

Table 6.7. Task Implementation Order by Teacher

	First Task	Second Task	Third Task	Fourth Task	Fifth Task
Alex	Kyle's Tiles	Cross Country Scoring	Mathemagi-cians' Cookies		
Brian	Kyle's Tiles	Design A Dartboard	Cross Country Scoring	Spinner Elimination	Mathemagi-cians' Cookies
Carmen	Kyle's Tiles	Mathemagi-cians' Cookies	Cross Country Scoring	Spinner Elimination	Design A Dartboard
Ellie	Mathemagi-cians' Cookies	Cross Country Scoring	Spinner Elimination	Kyle's Tiles	Design A Dartboard
Fatima	Spinner Elimination	Kyle's Tiles	Cross Country Scoring	Mathemagi-cians' Cookies	Design A Dartboard
Grace	Kyle's Tiles	Cross Country Scoring	Spinner Elimination		
Helen	Cross Country Scoring	Kyle's Tiles	Spinner Elimination	Design A Dartboard	Mathemagi-cians' Cookies
Ian	Kyle's Tiles	Design A Dartboard	Mathemagi-cians' Cookies		

Table 6.8. Percentage Agreement with Expert Rating by Teacher

	Task 1	Task 2	Task 3	Task 4	Task 5	Mean
Alex	.65	.53	.68			.62
Brian	.78	.77	.43	.61	.55	.63
Carmen	.61	.81	.48	.64	.32	.58
Ellie	.61	.52	.60	.76	.62	.62
Fatima	.64	.79	.85	.78	.41	.69
Grace	.69	.56	.52			.57
Helen	.52	.67	.69	.55	.71	.63
Ian	.75	.72	.45			.64
Mean	.66	.67	.59	.67	.52	.62

The teachers' mean levels of difference from the expert ratings were also considered over time, as summarized in Table 6.9. As with percentage agreement, there is no trend of decreasing levels of difference over time, either with individual teachers or with the averages of their mean levels of difference. There is no evidence that the mean levels of difference increase over time, either.

Table 6.9. Mean Levels of Difference from Expert Rating

	Task 1	Task 2	Task 3	Task 4	Task 5	Mean
Alex	.37	.50	.36			.41
Brian	.28	.23	.57	.38	.47	.39
Carmen	.41	.18	.55	.36	.77	.45
Ellie	.41	.49	.40	.24	.41	.39
Fatima	.38	.21	.15	.22	.63	.31
Grace	.31	.44	.52			.42
Helen	.52	.33	.31	.47	.31	.39
Ian	.25	.28	.58			.37
Mean	.37	.33	.43	.33	.52	.40

The discussion chapter will outline some possible causes of the lack of evidence of improvement over time in the teachers' use of the NWREL formative feedback guide.

Follow-Up Instruction:

What is the evidence of the teachers' use of assessment information from students to inform instructional plans?

Each teacher was asked to indicate instructional plans for the task debriefs through the use of an Instructional Sequence Analysis form, identifying student solutions to be shared. These forms were submitted by eight teachers for the Cross Country Scoring task and for Kyle's Tiles, by seven teachers for Mathemagicians' Cookies, and

by six teachers for Design a Dartboard and Spinner Elimination, for a total of 35 instructional sequence analysis forms submitted. In total, 192 pieces of student work were identified to be shared. It should be noted that for some pieces of student work identified to be shared more than one rationale was provided, so a total of 230 rationales were provided by the teachers.

The rationales provided by the teachers for selecting each piece of student work to be shared were analyzed to determine overarching categories into which they could be sorted. This process resulted in the determination that the five problem solving traits from the NWREL formative feedback guide – conceptual understanding, strategies and reasoning, computation and execution, communication, and insights – when augmented with a sixth category to encompass non-mathematical, social rationales, were suitable overarching categories. The rationales were sorted into the six categories and the five mathematical categories were then subdivided to better reflect the rationales indicated by the teachers. In some cases, subcategories were further divided to indicate positive and negative versions of the same subcategory.

In the case where a teacher provided more than one rationale for selecting one piece of student work, the rationales were considered separately and each were categorized where appropriate. Furthermore, some rationales provided by the teachers were simply too vague to be categorized without making significant inferences, so a seventh overarching category was established to note incidences of rationales that could not be categorized by the researchers.

The resulting categories, with descriptions, are:

- Conceptual Understanding
  - Mathematical concepts
    - *Important mathematics:* Display or introduce an important mathematical idea or concept.
    - *Misconception:* Display or debunk a misconception or misunderstanding of a mathematical idea or concept.
  - Mathematical terminology
    - *Precise language:* Display an example of precise, correct use of mathematical language, especially around important mathematical ideas.
    - *Incorrect language:* Draw attention to misuse of mathematical language.
  - Mathematical representations
    - *Appropriate:* Display appropriate or particularly illuminating mathematical representations.
    - *Inappropriate:* Draw attention to inappropriate or unnecessary mathematical representations.
  - Mathematical procedures
    - *Appropriate:* Display mathematical procedures which are particularly appropriate or effective for addressing the mathematics in the problem.
    - *Inappropriate:* Draw attention to mathematical procedures which are inappropriate for addressing the mathematics in the problem.

- Relevant information: Draw attention to relevant information not used in a student's approach to the task.

- Strategies and Reasoning
  - Strategy type
    - *Variety:* Display a different strategy for solving the problem.
    - *Common:* Display a strategy that was generated by many students.
    - *Simple:* Display a relatively simple or unsophisticated strategy.
    - *Sophisticated:* Display a sophisticated or more complex strategy.
    - *Non-mathematical:* Display a strategy that did not require the use of mathematical procedures.
    - *Successful:* Display a successful strategy, often in a class where few students successfully solved the problem.
    - *Incomplete:* Display an incomplete strategy, sometimes with the intention to discuss steps that should be taken to complete.
  - Reasoning and justification
    - *Strong:* Display an example of reasoning or justification that is particularly strong.
    - *Weak:* Display an example of reasoning or justification that falls short, often to launch a discussion of the additional reasoning needed.
  - Compare and connect: Display a strategy that can be used to make connections or comparisons with other strategies shared.

- Expansion
    - *Clarify:* Display a piece of work for the purpose of clarifying the student's strategy through discussion. This rationale is often provided when the teacher has questions about the student's strategy.
    - *Invite input:* Display a piece of work for the purpose of inviting comments from other students, often to improve or clarify strategy.
    - *Application:* Display a piece of work for the purpose of determining under what circumstances a strategy can be used.

- Computation and Execution
    - Correctness
        - *Correct:* Display solution with correct, clear computations.
        - *Incorrect:* Draw attention to incorrect or unclear computations.
    - Completeness
        - *Complete:* Display particularly well documented or thoroughly explained computations.
        - *Incomplete:* Draw attention to incomplete or undocumented computations.
    - Clarify: Display a piece of student work for the purpose of clarifying the student's computations through discussion. This rationale is often provided when the teacher has questions about the student's computation.

- Communication
    - *Strong:* Display an example of a particularly well-organized or well-communicated solution.

  - *Weak:* Draw attention to a solution that needs improvement in organization or communication.
- Insights
  - Connect to other mathematics: Display a solution that draws connections to other, previously seen mathematics.
  - Connect to real world
    - *Reasonable:* Display solution which accurately reflects or draws connections to a real world situation.
    - *Unreasonable:* Draw attention to a solution which does not reflect a real world situation.
- Social: This category includes all social motivations for selecting a piece of work, including gender, confidence, frequency of sharing, ability to take criticism, and others.
- Undetermined: This category includes all rationales too vague to be sorted into the other categories without significant inference on the part of the person categorizing, given knowledge of the task and the type of solutions generated by the students.

The 230 rationales provided by the teachers were sorted into the categories established. Table 6.10 summarizes the sorting of the rationale into the overarching categories. Social reasons, or rationale that could not be determined account for less than 10% of the rationales provided, indicating that more than 90% of the rationales provided by the teachers correspond to the five problem solving traits identified in the NWREL formative feedback guide. Conceptual understanding and strategies and reasoning based

rationales are the most common of the problem solving trait-based rationales, accounting for 60% of all the rationales provided, and for 66% of the problem solving trait- based rationales provided.

Table 6.10 Rationales by Overarching Category

Overarching Category	Frequency	Relative Frequency
Conceptual Understanding	53	.23
Strategies and Reasoning	86	.37
Computation and Execution	18	.08
Communication	37	.16
Insights	15	.06
Social	15	.06
Undetermined	6	.03

Table 6.11 and Table 6.12 summarize the distributions of rationales for the categories falling under conceptual understanding and strategies and reasoning, respectively. The majority of the conceptual understanding rationales, 74%, are related to a mathematical concept – either displaying an important mathematical concept, or drawing attention to a mathematical misconception. Mathematical concepts, either correct or incorrect, account for 17% of all rationales provided in every category.

The large majority of the strategy and reasoning rationales, 80%, related to the type of strategy chosen, sometimes to simply display a variety of strategies, or sometimes based on the complexity, simplicity, or regular occurrence of a strategy. The type of strategy accounts for 30% of all rationales provided in every category. It was most common for teachers to choose a strategy to share in order to display a broad range of

solutions generated during the debrief session, rather than to show a particularly important, elegant, or effective strategy.

Table 6.11. Conceptual Understanding

Category	Frequency	Relative frequency over category	Relative frequency over all categories
Mathematical concepts	39	.74	.17
Important mathematics	17	.32	.07
Misconception	22	.42	.10
Mathematical Terminology	2	.04	.01
Precise language	1	.02	.004
Incorrect language	1	.02	.004
Mathematical Representations	9	.17	.04
Appropriate	6	.11	.03
Inappropriate	3	.06	.01
Mathematical Procedures	1	.02	.004
Appropriate	0	0	0
Inappropriate	1	.02	.005
Relevant Information	2	.04	.01

Table 6.12. Strategies and Reasoning

Category	Frequency	Relative frequency over category	Relative frequency over all categories
Strategy Type	69	.80	.30
Variety	37	.43	.16
Common	1	.01	.004
Simple	11	.13	.05
Sophisticated	12	.14	.05
Non-mathematical	5	.06	.02
Successful	1	.01	.004
Incomplete	2	.02	.01
Reasoning and Justification	6	.07	.03
Strong	4	.05	.03
Weak	2	.02	.01
Compare and Connect	2	.02	.01
Expansion	9	.10	.04
Clarify	4	.05	.02
Invite input	3	.03	.01
Application	2	.02	.01

Examples of Rationales and Categorizations

To better illuminate the coding of rationales, some examples of rationales provided by the teachers and the categories into which they were sorted are discussed.

Fatima- Cross Country Scoring: Fatima provided this rationale for sharing a particular group's work on the Cross Country Scoring problem:

> "The mean computation of the first four finishers is incorrect with this group. I wonder if any of the other students will catch it. It makes Applebee the winner when actually Clarke should be. This group also used the range in calculating DeWitt to be the winner, so it's a good time to discuss and review range."

Fatima's rationale is considered in two parts: the first part, addressing the incorrect computation of the mean, and the second part related to reviewing range. The incorrect

computation of mean is categorized as Computation- Incorrect, while the discussion of range is categorized as Conceptual Understanding- Mathematical concepts- Important mathematics, since range is an important mathematical idea that could be applied to this task.

Carmen- Cross Country Scoring: Carmen's rationales provided were much terser than Fatima's. For choosing one piece of student work, Carmen simply writes "Averaging is addressed in method 3." Because Carmen is noting one of the important mathematical concepts of the task, this rationale is classified as Conceptual Understanding- Mathematical concepts- important mathematics. However, the terseness of Carmen's rationales does not always make it easier to classify her rationales.

For the same set of student papers, Carmen indicates "Creative, but mathematically based," as a rationale for sharing another piece of work. On its own, this rationale would be difficult to classify. It is not identifying a specific mathematical concept, nor does it refer to computation or quality of communication. "Creative, but mathematically based" could describe a strategy, or refer to an insight that the student has displayed, but either choice would require inferring Carmen's intentions. However, on the Instructional Sequence Analysis form, the teachers identify the rationale for sequencing the pieces of student work in the order they determined. Carmen's sequencing rationale states "the sequence is only to show a variety of strategies encountered." From this sequencing rationale it becomes more evident that the rationale for choosing the piece of work can be categorized as Strategies and Reasoning- Strategy type- variety.

Ellie- Kyle's Tiles: Ellie's rationales for selecting pieces of student work tended to be quite terse as well. However, they require less inference or investigation of her intentions to categorize because they are clearer statements. On the Kyle's Tiles task, students are asked to design a tiling for a wall of fixed dimensions, with choices of three different sizes of tiles, all differently priced. They are required to create the "best" design, and justify why the design generated is "best."

The rationales Ellie provided for two different pieces of student work reflected both the particulars of the task and the problem solving norms for her classroom. The first student was chosen because he "computed accurately, talked about symmetry." This statement was categorized in two parts. Accurate computations clearly are categorized as Computation and Execution- Correctness- correct. The reference to symmetry is categorized as Insights- Connect to other mathematics, as symmetry is a mathematical concept the students have previously encountered.

Daphne- Kyle's Tiles: Daphne's students generated some different ideas from Ellie's, creating opportunities to select student work for other reasons. Daphne indicated that she selected one piece of student work to share because the student "hinted at the idea of unit rate, and used mostly 6x6 (tiles)." Unit rate is an important mathematical concept that may be drawn from the task, particularly in determining the least expensive way to tile the wall. Moreover, the 6x6 tiles are those with the lowest unit rate. Therefore the rationale is categorized as Conceptual Understanding- Mathematical concepts- important mathematics.

Ian- Mathemagicians' Cookies: The Mathemagicians' Cookies task requires students to package 360 cookies into packages of three different sizes, and price the packages to

attain a minimum profit. For one piece of student work Ian chose to share, he provided the rationale "Difference in cost. $2+$4 = $6, 4+8 = 12 cookies, $6 = 10 cookies." Familiarity with the task facilitates interpretation of Ian's statements. The student has priced a box of 4 cookies for $2, and 8 cookies for $4, so 12 cookies could be purchased for $6 in two separate packages. However, the third package contains 10 cookies and is priced at $6. This rationale was categorized as Insights- Connect to real life- unreasonable. While it is mathematically possible to price in such a way, and the student can still sell the required number of cookies and attain the minimum profit, the pricing scheme is unrealistic.

Grace: Spinner Elimination: The spinner elimination problem provided opportunities to expose some student misunderstandings, both about factors of numbers, but also about probability. Grace noticed such a misunderstanding, and used a piece of student work to expose it. She chose to share the work of two students because they were "based on 'luck' and 'good' numbers, no connection to multiples." Grace is exposing the misunderstanding some students had of the spinner elimination game depending on lucky spins or the choice of lucky numbers within the spinner. This rationale is categorized as Conceptual Understanding- Mathematical concepts- misconception.

Sequencing Student Work

In addition to investigating the types of rationales the teachers provided for selecting pieces of student work, the rationales for sequencing pieces of student work in the order given were analyzed. Of a total of 36 instructional sequence analysis forms submitted by the teachers, 11 did not indicate any sequencing rationale at all. Two of the

remaining 25 rationales were strictly for social reasons, leaving only 23 rationales that were related to mathematics problem solving.

Examination of the sequencing rationales provided led to the establishment of the following categories for sequencing rationales.

- From Less to More: This category encompasses all rationales for sequencing in increasing degrees of some quality. The specific qualities include:
    - Building from partial solutions to complete solutions
    - Building from simpler strategies to more complex strategies
    - Building from apparently random strategies to more structured strategies
    - Building from incorrect solutions to correct solutions
    - Building from less detailed solutions to more detailed solutions
    - Building from less reasonable solutions to more reasonable solutions
- Highlighting a Mathematical Idea: This category encompasses all rationales for sequencing which have a goal of discussing, drawing attention to, or exposing an important piece of mathematics. Highlighting a Mathematical Idea also includes drawing attention to a mathematical misconception, especially if the misconception is debunked before sharing important mathematical concepts.
- Variety of Strategies: This category includes all rationale related to simply displaying different methods of solving the problem that do not fall under "From Less to More" or "Highlighting a Mathematical Idea." Examples of sequencing rationales included in this category include:
    - To show a variety of strategies

- Random ordering of strategies
- To share common strategies

The majority of sequencing rationales, 14 of the 23 mathematical rationales provided, fall into From Less to More. Five of the 23 mathematical rationales pertain to Highlighting a Mathematical Idea. Four of the rationales are included in Variety of Strategies.

Connections Between Rationale, Comments, and Student Work

The pieces of student work identified to be shared by the teachers were examined to determine the extent to which they reflected the rationales provided by the teachers. In the case where written comments from the teachers were available, those comments were also examined to determine the extent to which they reflected the rationales provided by the teachers.

The degree to which each student's solution reflected the rationale identified by the teacher was ranked on a scale of three possibilities: *not at all*, *somewhat*, or *very well*. If the connection between the rationale provided and the student work could not be identified, the degree was noted as not at all. If every element of the rationale could be clearly identified in the student work, the degree was noted as very well. Correspondence between the student work and part of the rationale provided, or correspondence that was not particularly strong was noted as somewhat. In order for correspondence to be determined to be strong, the rationale had to be sufficiently specifically stated.

For example, rationales of "inaccurate" or "different strategy" would, at best, be somewhat reflected in the student work, as it is difficult to determine exactly what

portion of the student work the teacher is referring to. Rationales of ".25/cookie strategy" on Mathemagicians' Cookies or "0 and 1 addressed" on Spinner Elimination could be strongly reflected in the student's work because they are sufficiently specific to be identifiable in the student's work. A specific selection rationale is a necessary, but not sufficient, requirement for a strong reflection of the rationale in the corresponding student work. The rationale must be very clearly included in the students' work, without a need for interpretation on the part of the researchers, in order for the rationale to be strongly, rather than somewhat, reflected in the student work. For example, if a student places 0 and 1 on the spinner in the Spinner Elimination task, but says in the discussion of a good strategy for building a spinner, "0 doesn't help, and 1 can be helpful but only for small numbers," the student did not address the usefulness of 0 and 1 sufficiently to warrant strong reflection of the rationale.

Table 6.13 summarizes the correspondences between student work and rationales identified by task. Over all the tasks, 47% of the rationales provided by the teachers were reflected very well in the students' solutions. The rationales identified by the teacher were somewhat reflected in 40% of the student solutions shared. No correspondence between rationales and student solutions was identified for only 12% of the student solutions shared. It should be noted that in at least 9 cases where the student work did not reflect the rationales provided at all, the rationales were strictly social and not related to problem solving.

Table 6.13. Correspondence between Student Work and Rationales

Task	Not at all	Somewhat	Very Well	Totals
Cross Country Scoring	2	15	22	39
Design A Dartboard	1	14	16	31
Kyle's Tiles	6	22	13	41
Mathemagicians' Cookies	5	13	18	36
Spinner Elimination	8	8	16	32
Total	22	72	85	179

In the case where written comments from the teachers to the students were submitted, the comments were examined to determine the extent to which they reflected the rationales provided by the teachers for selecting the piece of student work. The degree to which the comments correspond to the rationale indicated by the teacher was ranked on a scale of three levels: *not at all*, *somewhat*, or *very well*. If no apparent connection existed between the rationale indicated and the written comments, the degree of correspondence was recorded as not at all. If the written comments very clearly reflected every aspect of the rationale indicated, the degree of correspondence was recorded as very well. If the written comments reflected some parts of the rationale provided, or corresponded to the rationale, but not very strongly, the level of correspondence was recorded as somewhat. As with the correspondence between

rationale and student work, it is a necessary, but not a sufficient, condition that a rationale be specifically stated in order for the comments to reflect the rationale very well.

For example, comments would, at best, reflect rationales of "inaccurate" or "different strategy" somewhat. For the rationale ".25/cookie strategy," a comment such as "Your use of a .25/cookie strategy is interesting here," would reflect the rationale very well, whereas a comment such as "The strategy you chose was effective," would only reflect the rationale somewhat. For the "0 and 1 addressed" rationale, the comment "It looks like you understand that 0 and 1 are not very useful," would reflect the rationale very well. The comment "I like your explanation of important numbers," would only somewhat reflect the rationale.

Tables 6.14 and 6.15 summarize the correspondences between the written comments indicated on the student work and the rationales identified by the teacher for sharing that piece of student work. For nearly 17% of the student papers identified to be shared, the teachers either did not provide written comments to the students or did not submit the written comments to the researchers. In the case when comments were provided, nearly 48% of the comments were determined to have no correspondence to the rationales indicated. The comments have some correspondence to the rationales 32% of the time, and 20% of the comments correspond strongly to the rationales.

Table 6.14. Correspondence between Comments and Rationales

Task	No comments	Not at all	Somewhat	Very Well	Totals
Cross Country Scoring	0	23	10	6	39
Design A Dartboard	8	15	7	1	31
Kyle's Tiles	7	16	11	7	41
Mathemagicians' Cookies	7	7	13	9	36
Spinner Elimination	8	10	7	7	32
Total	30	71	48	30	179

Table 6.15. Relative Frequency of Correspondence between Comments and Rationales

Task	No comments	Not at all	Somewhat	Very Well
Cross Country Scoring	0	.59	.26	.15
Design A Dartboard	.26	.48	.23	.03
Kyle's Tiles	.17	.39	.26	.17
Mathemagicians' Cookies	.19	.19	.36	.25
Spinner Elimination	.25	.31	.22	.22
Total	.17	.40	.27	.17

In the course of evaluating the extent to which the written comments and the student work reflected the rationales indicated by the teachers, impressions of the choices the teachers made were noted. Several noteworthy trends emerged in this observation.

In particular, it was observed that, while a piece of student work may somewhat or strongly reflect the rationale provided by the teacher, there may be elements of the student's solution that would distract from the learning benefits of the solution. For example, Brian chose two student papers to share on the Mathemagicians' Cookies task. One of them was chosen because the student used a unit pricing strategy; the other was chosen because the unit price decreased as the package size increased. While both pieces of student work reflected the respective rationales, both student solutions were disorganized and had unexplained or inaccurately explained statements and calculations. The potential to confuse other students with these examples is great, and so, while the correspondences between the rationales and the student work were good, that does not necessarily indicate that they are papers that are beneficial to share.

Brian is not the only teacher who selected student solutions for sharing that may have been more confusing than helpful. Ellie's selections for the Cross Country Scoring task raised similar concerns. She shared one student's work because the student selected the team with the most finishing places that were composite numbers. Her comments to the student indicated that she deemed the composite number strategy "Very Creative!" However, selecting a winning team based on the number of composite number finishing places would not be reasonable from the perspective of reflecting the order of finishers. This solution could be confusing to other students unless it was accompanied by a discussion of the fairness of scoring in such a way.

Ellie selected another piece of student work because the student made use of the median score as a measure for selecting a winner. However, the student selected the team with the highest median to win. Ellie's comments to the student did not question

the reasonableness or fairness of this method. Again, to share such a method without drawing attention to its reasonableness in the context of the problem (a goal not indicated by the teacher's plan) could introduce confusion or misconceptions in the whole class discussion. Ellie and Brian were not the only teachers to choose to share solutions of this type, but they provide interesting examples of the possible difficulties introduced by solutions chosen for sharing.

In addition, it was observed that the teachers would occasionally indicate a rationale for sharing a particular piece of student work that did appear in the work; however, there may have been a better reason to share that piece of work than the one identified by the teacher. For example, Ian chose to share a student solution to the Kyle's Tiles task to expose an error the student made in calculating area. However, the student had mistakenly calculated the cost of cutting tiles, leading to a total cost of $230,762.50, when most students obtained total costs between $95 and $135. Ian commented about the high cost on the student's paper, but did not refer to it in his rationale for choosing the piece of work. There are many learning opportunities that could stem from this mistake. Ian could have launched a discussion of considering the reasonableness of solutions to mathematical problems, or he could have used the student's solution to explore confusion students had about calculating the total cost. The student's error in calculating the area is only one part of the learning potential in the solution.

The final observation was that correspondence between the student's solution, the rationale provided, and the written comments from the teacher can be attributed to teacher scaffolding of the problem. Fatima often provided her students with a structured worksheet for responding to a task, which forced solutions to take on a very organized

format. She also selected pieces of student work to share because of their organization, a rationale reflected in the student papers. Yet Fatima's scaffolding of the task made it nearly impossible for a student response not to be well organized.

The ten cases where the rationale provided by the teacher had no correspondence to the piece of student work chosen, and the rationale was not social and was easily understood were considered. The classes from which those eleven pieces of student work were chosen were examined to determine if a piece of work better reflecting the rationale was available for selection by the teacher. The cases are examined by task.

Cross Country Scoring

Grace selected a piece of work to be shared because of a "totally random selection for winners." The student actually had strategies that were mathematically valid and possibly fair, including choosing the winning team to be the team with the lowest sum of the first two placers, or the team with the lowest average. The student did select the winning team as the team with the lowest place out of each team's fifth place finisher.

Figure 6.1. Student Solution to Cross Country Scoring Task

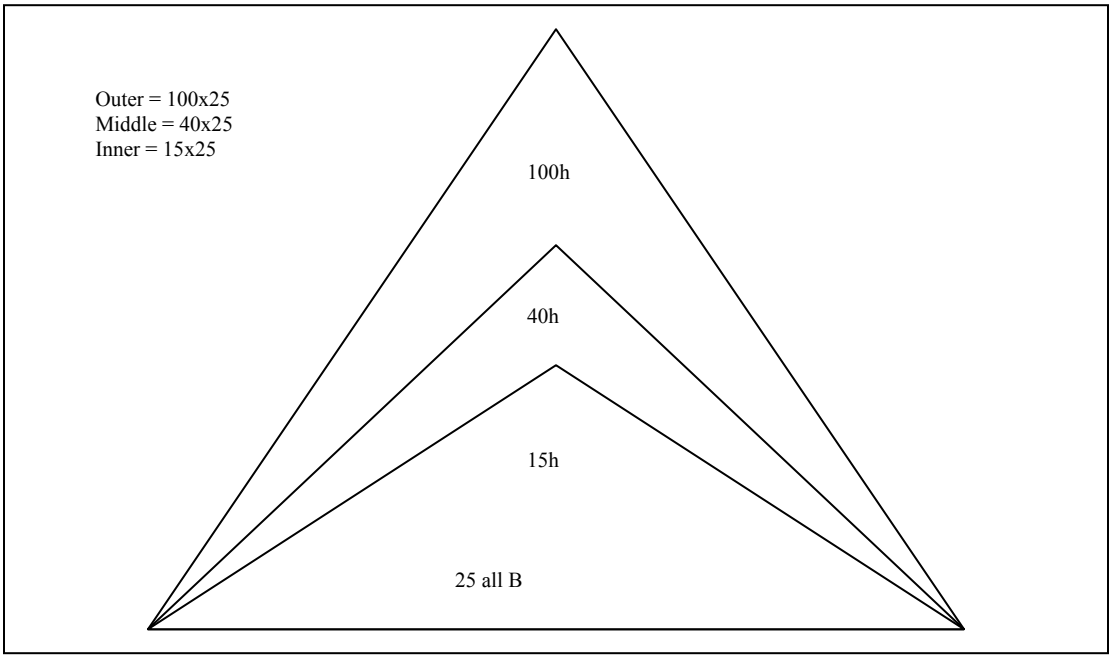
	Applebee	**Burnside**	**Clarke**	**Dewitt**	
Strategy 1	3 5 10 18 3rd	4 8 9 21 4th	1 2 3 1st	6 7 13 2nd	I took the first ten runners and who ever had the lowest won
Strategy 2	36 ÷ 3 12	34 ÷ 4 8.5 1st	35 ÷ 1 35	29 ÷ 6 4.8 2nd	I took the bottom numbers and ÷ by the top and whoever had the lowest won
Strategy 3	20	21	19	17 1st	I took the 5th runner in each group and whoever had the lowest won
Strategy 4	14.3 1st	18.3	16.6	17.4	I added all the numbers together and ÷ by 10.

Grace may have interpreted this as random, but this is quite close to the lowest median of each team, and would be worth examining further. An examination of the student papers for the remainder of the class does not reveal a student solution that appears to be more random than the one selected. However, Grace's class had only six students at the time of the task, so the breadth of student solutions was not very great.

Design a Dartboard

Ellie chose to share a piece of student work on the Design a Dartboard task because the solution used a triangle shape and because of the dimensions used to represent the different areas.

Figure 6.2. Student Solution to Design a Dartboard Task

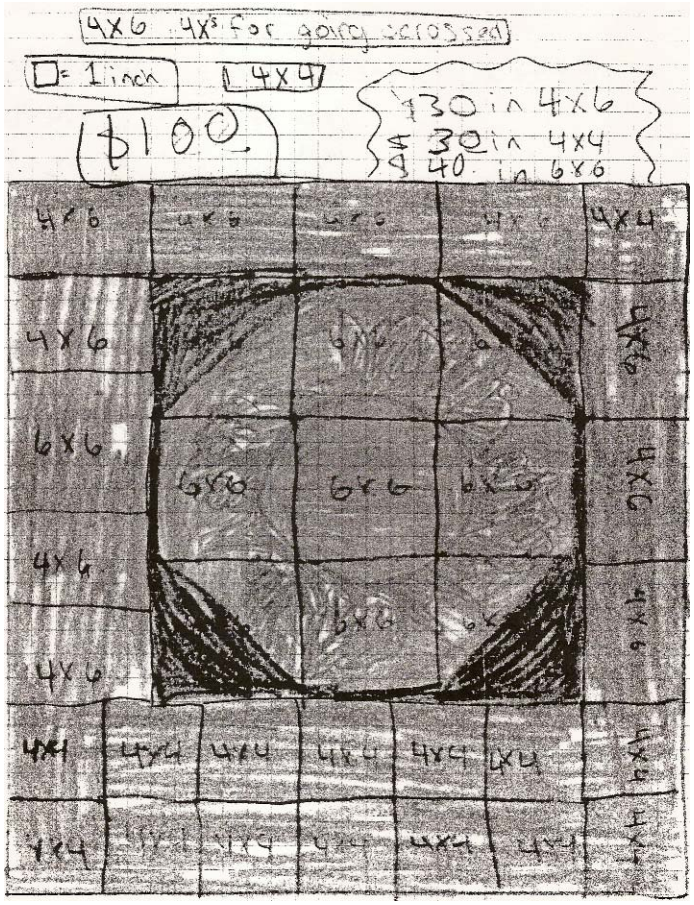


However, the dimensions that Ellie identifies are not correct for the student's work. In fact, the student clearly did not model the situation correctly, with an inner area that is significantly larger than the middle area, though it is supposed to reflect a smaller percentage of the total area. Examination of the solutions generated by the class indicated that at least three other students used a triangular shape and also correctly modeled the areas as established in the task requirements.

Kyle's Tiles

Carmen identified a piece of student work to be shared on the Kyle's Tiles class because the student's strategy appeared to be a random arrangement of tiles.

Figure 6.3. Student Solution to Kyle's Tiles Task



However, examination of the student's solution indicates that she actually created a pattern with the tiles, indicating that the tiles may not be randomly arranged. Examination of the class papers reveals three solutions that do appear to be randomly arranged. No pattern is observed in the layout of the tiles, and the students' explanations do not indicate that the arrangement was anything other than random.

Mathemagicians' Cookies

Brian chose a piece of student work to share on the Mathemagicians' Cookies task because it used a "different strategy." However, the student's strategy involved two pages of subtractions until she had determined the number of packages she would need to sell (see Figure 6.4 for a reproduction of part of the student's tables). As such, the student was not using an appropriate strategy for the task, and the work should not be considered a strategy to be shared. The class papers revealed several other solutions using strategies with some differences from the previously shared solutions, yet using mathematically appropriate strategies for solving the problem.

Figure 6.4. Partial Student Solution to Mathemagicians' Cookies Task

Cookies	**Package Size**	**Cost of $**	**Result**
10	Large	$ 3.00	350
5	Medium	$ 1.50	345
3	Small	$ .75	342
10	Large	$ 3.00	332
5	Medium	$ 1.50	327
3	Small	$ .75	324
10	Large	$ 3.00	314
5	Medium	$ 1.50	309
3	Small	$ .75	306
10	Large	$ 3.00	296
5	Medium	$ 1.50	291
3	Small	$ .75	288
10	Large	$ 3.00	278

Alex chose a piece of student work to be shared because it "needed a small reality check." However, nothing about the students' solution appeared to be unrealistic (see Figure 6.5). The students selected a unit price per cookie that would result in a profit just a few dollars more than required, and they packaged the cookies in reasonable quantities. Examination of the class papers indicated that one other group of students had made a mistake with the per package pricing, creating a situation in which buying a small and medium package yields more cookies at a lower price than buying a large package. This likely would have been a better choice of a solution needing a "small reality check."

Figure 6.5. Second Student Solution to Mathemagicians' Cookies Task

We need to make $86 to reach are goal. If we sell 360 cookies for $.25 a piece then we will get $90.

Size #	**Amount per package**	**Price per package**
1	1	$.25
2	2	$.50
3	3	$.75

If all the cookies are sold we will replace all the money we used to make the cookies and get $54 for the club. We picked the sizes we did so people can buy them individually or multiple cookies. If we sell 360 cookies for $.24 we will end up with $86.40. The prices we chose were reasonable.

Helen's rationale for choosing a piece of student work did not correspond with the student's solution in one instance. It appears that she simply misunderstood the student's work, leading to the disconnect between rationale and student solution. Helen chose the solution because "he sold 360 cookies in 3 different ways." The student does include

some calculations which indicate what would happen if he sold different quantities of each size of package. Initially, it appears that he may be attempting to sell all his cookies in one package. However, after two pages of tables (see Figure 6.6), the solution concludes with the quantities of each size package he will sell, totaling 360 cookies together. There were two student solutions in the class papers that did sell 360 cookies in three different ways, so solutions better reflecting the rationale were available.

Figure 6.6. Sample of Third Student Solution to Mathemagicians' Cookies Task

Packages of A Dozen Cookies	Costs to Make	$1.20	$2.40	$3.60	$4.80	$6	$7.20	$8.40
	Money Made	$6	$12	$18	$24	$30	$36	$42

Packages of a Dozen Cookies	Dozens of Cookies Sold	10	20	30	40	50	60	70
	Total Profit	48	96	144	102	240	288	330

I sold 20 dozen packs = 240
sold 15 half dozen = 90
330
And 29 singles + 29
359
and I eat one + 1
360

20 dozen = 96$
15 half dozen = 9$
29 single = 4.25
109.25

$109.25 total profit

Spinner Elimination

Carmen selected a piece of work to share because the student addressed the importance of large and small numbers in building a spinner. However, the piece of

student work shared presents a strategy for playing the game, not for designing a spinner (see Figure 6.7). The student discusses some specific numbers, as do nearly all solutions, but does not address the difference between large and small numbers in building the best spinner.

Figure 6.7. Student Solution to Spinner Elimination Task

1. We used the small numbers first then we started multiplying and using are second spin to get bigger numbers. I would recommend people use the small numbers first, like if you role a 9 don't spin again.

2. We lost because we used all our big numbers to get closer to 50. But we lost because we went over 50 4 times. I recommend that you don't role more than once until you used up your small numbers.

Carmen identified several other pieces of work to share for the same reason, and those do address the magnitude of the numbers chosen, so more suitable pieces of student work were available to her from the solutions generated by the class.

Grace identified a piece of work to be shared for its "clear explanation of what numbers would work better." However, the student solution has a one sentence explanation of the numbers that should be chosen, not a clear explanation of the best numbers to choose (see Figure 6.8). The explanations generated by Grace's class were generally very brief and none comprehensively addressed the quality of each digit for use on the spinner. Grace had only 8 student papers to choose from for the class, none of which were particularly well communicated, reasoned, or explained.

Figure 6.8. Second Student Solution to Spinner Elimination Task

1. Yes, because I had all high numbers and you would get 20 and then you would go again and multiply by 9.
2. Stay with the low numbers

Alex chose a piece of student work from the Spinner Elimination task to be shared because it had "some real insight." His sequencing rationale indicates that the insights he hopes to expose are related to the redundancy of larger composite numbers when their prime factors appear on the spinner. The piece of work he identified for its insight suggests that even numbers are more useful than odd numbers, yet only one even number is prime. The remaining justifications provided by the student surround choosing smaller numbers over larger numbers, but includes 4 in that list of numbers.

Figure 6.9. Third Student Solution to Spinner Elimination Task

Pick lower numbers like 2 and 4 more than 5 or 7. You should have low numbers. After awhile you will have to spin more and then you will go over. You need low numbers, especially evens. You won't go over so much with 2 or 4.

Alex's rationale is sufficiently vague that it becomes difficult to determine if student papers existed which better exhibited insight. Several other student papers referred to lower numbers being better than higher numbers, but no other papers addressed the issue of prime and composite numbers. Nonetheless, the solution chosen to be shared is in conflict with the idea of not choosing large composite numbers, as it specifically mentions 4 as a good choice for the spinner. The opinion of the researcher,

then, is that a better piece of student work could have been selected from the solutions generated by Alex's class.

It appears that in these cases superior examples of the rationales identified were available for the teachers to select to share. However, choosing pieces of work to share is often a challenging process and requires identifying and interpreting subtle characteristics of the solutions.

Summary

The results of this study reveal that the NWREL formative feedback guide is not being used by the teachers with high levels of reliability. The teachers' use of the NWREL formative feedback guide does not appear to improve over the course of the project, nor does it appear to deteriorate. An investigation of the rationales provided by the teachers for selecting pieces of student work to share revealed that they were very often related to aspects of mathematical problem solving, and could be categorized using the five NWREL problem solving traits as a basic framework. The sequencing rationales provided by teachers were less often related to aspects of mathematical problem solving.

The relationships between rationales, student work, and written teacher comments were examined. While rationales are either reflected strongly or somewhat in 87% of the corresponding student work selected, rationales are reflected strongly or somewhat in the written comments provided to students only 52% of the time. An analysis of the cases where rationales did not correspond at all to the pieces of student work selected to share revealed that teachers nearly always had pieces of student work available to select that better characterized the rationales.

Chapter 7 – Discussion

The preceding results chapter summarized the results of data analysis in response to the three particular issues identified to address the principal research question. The first section of this chapter interprets the results. The second section identifies some possible limitations of the study. The final section of the chapter discusses the implications of this study for practitioners, professional developers and researchers.

Interpretation of Results

Inter-Rater Reliability

Do teachers attain a high degree of inter-rater reliability in the use of the NWREL formative feedback guide?

To determine the inter-rater reliability of the teachers' use of the NWREL formative feedback guide, teacher ratings for each piece of student work were compared to the corresponding expert ratings. Over all of the teachers from whom data were collected, percent agreement between teacher and expert ratings was below 70% on every task and for every trait. The Cohen's kappa reliability coefficients were below 0.6 on every task and for every trait. Under both measurements this indicates agreement below standard acceptable levels (Watkins and Pacheco, 2001; Goldberg, 1995) .

However, the mean levels of difference between teacher and expert ratings were less than or equal to 0.5 on all tasks and on four of the five traits. The mean level difference for the fifth trait was 0.6. While the levels of agreement do not indicate

strong inter-rater reliability, the teachers tended to rate their students' performances within one half level of the expert rating.

Among the possible reasons for the low inter-rater reliability is the complexity of the NWREL formative feedback guide (see Chapter 4). The guide uses 84 criteria statements over the five problem solving traits and four proficiency levels. On any trait, a teacher using the feedback guide would have to consider between 13 and 20 criteria statements to determine a student's proficiency level. The sheer number of statements to consider may be overwhelming for teachers. Moreover, the distinction between proficiency levels is dependent on these criteria statements, which are not always well defined. For example, the following appear as criteria statements for conceptual understanding:

- **Emerging:** You used mathematical terminology incorrectly.
- **Developing:** You used mathematical terminology imprecisely.
- **Proficient:** You used mathematical terminology correctly.
- **Exemplary:** You used mathematical terminology precisely.

The distinction between incorrect and imprecise use of terminology, or between correct and precise use is not obvious, and different interpretations of the distinction would be reflected in different judgments of the proficiency level of a student's performance.

A student's proficiency level in conceptual understanding is not based simply on use of mathematical terminology. Use of representations, consideration of relevant information, and choice of mathematical procedures also are considered in judging proficiency level. These additional facets of the single trait conceptual understanding may be judged by the teacher to reflect different levels of proficiency. For example, a

student solution might use mathematical language correctly, but also apply a mathematical procedure that would lead to a partially, but not completely, correct solution. Such student work thus exhibits characteristics of both a developing and a proficient level of conceptual understanding, according to the guide. Yet the teacher must choose a single proficiency level for this trait.

Examination of the tasks and traits for which agreement was lowest provides further insight into some of the causes of low inter-rater reliability.

Possible Causes of Low Reliability for a Specific Task: Cross Country Scoring

The Cross Country Scoring task had the lowest percent agreement, lowest Cohen's kappa coefficient, and highest mean level of difference of all five NWREL tasks implemented by the teachers in this study. Information gathered during the February, 2007 workshop provided insight into the possible causes of the low inter-rater reliability on that task. In the classroom observations prior to the February workshop, NWREL staff members had observed that a significant number of the teachers were not structuring task debrief sessions with easily articulated mathematical goals in mind. The NWREL staff used the Cross Country Scoring task as the basis of an activity during the February workshop designed to address the potential misunderstanding of the intended use of the debrief sessions. The goals for the activity were to help the teachers identify mathematical goals for a discussion and select and sequence pieces of student work to better achieve those goals.

The task had been selected by the NWREL staff members in part because solution strategies for the task could use several measures of central tendency. The teacher guide

for the task (see Appendix) suggests that the content strand for the task is Probability and Statistics, and that the task is intended to develop understanding of central tendency. The teacher guide also states that the task allows students to explore mean, median, weighted average, variance, and range.

The teachers were provided with sample student solutions to the task and asked to use the solutions to identify a mathematical goal for a debrief session, and to select and sequence pieces of student work which would help achieve the mathematical goal. When the teachers were asked to identify the important mathematics of the task, the NWREL staff were surprised to learn that many of the teachers did not identify measures of central tendency. In fact, the whole group discussion of the important mathematics in the Cross Country Scoring task became a rather heated exchange, during which one teacher insisted that the task was not suited to exploring central tendency.

Not all of the teachers agreed that the task was inappropriate for teaching measures of central tendency. However, most expressed some difficulty identifying measures of central tendency other than mean and median that would be useful for the task. When one NWREL staff member suggested that some measure of variance in the finishing places could be useful for devising a scoring system, several of the teachers balked.

Evidently, the important mathematical concepts and ideas that the Cross Country Scoring task could be used to highlight were not clear to the teachers. The discussion during the workshop revealed that many solution strategies that the NWREL staff would consider appropriate or even sophisticated were viewed by some of the teachers as inappropriate. The difficulties teachers experienced identifying the important

mathematics of the task would likely be reflected in the use of the feedback guide. It would be expected, then, that the teachers would evaluate student performance differently than the expert, especially if there is disagreement about the mathematical concepts in the task.

The complexity or subtlety of the mathematics inherent in the task may contribute to lower agreement between teacher and expert ratings. The task with the highest agreement between teacher and expert ratings was the Kyle's Tiles task. The teacher's guide for Kyle's Tiles suggested that the task is appropriate for the content strands of Number Theory and Operations and Geometry. The mathematical ideas identified in the teacher's guide as part of the task include factors and multiples, spatial reasoning, and addition of decimals. These ideas may be better understood by teachers than measures of central tendency, and so may make it easier for teachers to accurately evaluate student performance than on the more mathematically sophisticated Cross Country Scoring task.

Possible Causes of Low Reliability for a Specific Trait: Insights

Across all tasks, the insights problem solving trait had the lowest percent agreement, lowest Cohen's kappa coefficient, and the highest mean level difference of all five problem solving traits. Throughout the course of the project, the teachers regularly identified insights as the most challenging problem solving trait on which to evaluate student performance. Insights was the most common trait on which student performance was not evaluated by teachers who submitted proficiency level ratings on other traits. During the summer workshop, the Saturday sessions during the school year, and the classroom observations, several teachers expressed difficulty understanding the definition

of insights as a problem solving trait, identifying insightful student solutions, and accurately judging the insightfulness of solutions.

The expert raters experienced similar challenges. When articulating the task specific feedback guides used by the expert raters to evaluate student solutions, it was found that articulating the characteristics of insightful solutions was particularly challenging on some of the tasks. If this is a challenge for the experts, it would be expected that it is a challenge for the teachers, and likely would result in variability in rating student performances.

General Versus Task Specific Guides

Another source of low inter-rater reliability may be due to the use of a general feedback guide to evaluate student performance. The research literature (Arter & McTighe, 2001; Stiggins, 2001) indicates that general guides are best suited for teaching general problem solving skills, such as identifying the characteristics of a well-communicated solution, or the characteristics of a solution that displays understanding of the important mathematical concepts. The literature indicates that a task specific guide is best suited for summative evaluation of student performance on a particular task (Arter & McTighe, 2001; Stiggins, 2001). The use of a general guide by the teachers for evaluating their students' performance, rather than a guide articulated for the particular task may be an additional source of low reliability.

Improvement Over Time

Do the teachers' abilities to assess students' problem solving skills improve with the implementation of the model?

To determine if the teachers improved in use of the NWREL formative feedback guide over the course of the project, their percentages of agreement with the expert rating and their mean levels of difference from the expert rating were examined over the year. No evidence of improvement over time was seen from any of the teachers, nor did the teachers' abilities to assess students' problem solving skills over time deteriorate.

One possible cause for a lack of trends over time is that the extent to which a teacher uses the NWREL formative feedback guide reliably depends more upon the task than on the number of times he or she has used the guide. It is possible that the tasks are sufficiently unique that the use of the guide on one task has little or no effect on the use of the guide on a subsequent task.

Follow-Up Instruction

What is the evidence of the teachers' use of assessment information from students to inform instructional plans?

To address the teachers' use of assessment information to inform instructional plans, the selection and sequencing rationales provided by the teachers were examined. It was determined that 90% of the rationales provided for selecting pieces of student work were not social, and instead were related to the problem solving traits identified in the NWREL formative feedback guide. More than 90% of the sequencing rationales provided were related to mathematical problem solving, although only about 70% of the

teachers' planning reports included a rationale for sequencing student work to be shared in the order selected.

While the rationales for both selecting and sequencing pieces of student work were mostly around mathematical problem solving, the frequency of rationales highlighting specific mathematical concepts was much lower. Mathematical concepts account for 17% of the rationales identified. Five of the 23 sequences related to mathematical problem solving identified highlighting a mathematical idea or moving past a misconception as reasons for the sequence. It was more common for teachers to use student work to display a variety of solution strategies or to share student papers that were examples of good problem solving elements, such as good communication or reasoning. While the rationales for selecting and sequencing student work are most often related to facets of mathematical problem solving, they are not frequently deeply mathematical. Many of the rationales do not clearly work toward mathematical goals, as defined in Stein et al's (in press) pedagogical model for sharing student work.

This could be because teachers are more easily able to discern different strategies than to identify the important mathematics in a task or the important mathematics in a student's solution. Identifying the important mathematics or mathematical misconceptions in a student's solution may require significant mathematical content knowledge on the part of the teacher. Determining how to use student papers to expose mathematical ideas or misconceptions may require significant pedagogical content knowledge on the part of the teacher. Simply sharing a variety of strategies may require less sophisticated content knowledge.

In the case where teachers provided written comments to the students, the written comment strongly or somewhat reflected the rationales provided for 52% of the student papers. While the level of correspondence between the rationales and the corresponding student work is reasonable, the evidence that the rationales and the comments provided to students correspond is not as high.

Several patterns were noted in examining the correspondences between rationales, comments, and student solutions selected for sharing which provide evidence of the teachers' use of assessment information to inform instructional decisions. Some pieces of student work selected by teachers to be shared were consistent with the rationales provided, but included elements that could distract from the learning benefits. Other pieces of student work were consistent with the rationales provided, but also exhibited other important mathematical ideas for which they could have been shared which were unrelated to the rationales. In some cases, a student work may be consistent with the rationale provided, but the correspondence may be due more to the way in which the teacher organized the problem than the student's solution process.

All three patterns indicate some lack of use of assessment information to inform follow-up instruction. In some cases, this is attributable to a focus on sharing a variety of strategies without attending to reasonableness, as seen in the student work Ellie shared on the Cross Country Scoring task. Ellie chose two pieces of work, one for what she deemed a creative strategy, and the other for its use of median. However, Ellie's feedback to both students did not attend to the use of unreasonable strategies in each case. This results in pieces of student work being used to plan a follow-up lesson that may not reflect the teacher's mathematical goals.

In other cases, the apparent lack of use of assessment information to inform follow up instruction may be due to the teacher's choice of important elements of a student's solution. For example, Ian chose to share a student solution to the Kyle's Tiles task due to an error in area calculation, on which he commented on the student's paper. However, Ian also commented on the student's paper about the student's extremely unreasonable total cost, yet did not refer to it in the rationale for sharing the solution. The cost was so unreasonable that it would possibly overshadow and distract from the incorrect area calculation. This indicates that a teacher may observe information in assessment, but choose to overlook that information for use in follow-up instruction.

In the case where teacher organization of the problem creates the rationale for which a solution is selected to be shared, the issue is more attributable to the teacher's structuring of the task than to the use of assessment information. For example, Fatima often selected student solutions to share because they were well organized. However, Fatima launched the tasks by providing her students with a worksheet with blanks, clearly identifying which portions of the solution students needed to address. In such a situation, the rationales are not necessarily reflecting the students' performance, but perhaps reflecting the way in which the teacher implemented the task.

Limitations of the Study

Inter-Rater Reliability and Improvement Over Time

One limitation of the inter-rater reliability and improvement over time portions of the study is the size of the sample used for making quantitative conclusions. With only 8 teachers from whom proficiency level data were collected, there may be too few teachers

to draw conclusions about the reliable use of the formative feedback guide. The representativeness of the group of teachers who participated in the study is a limitation as well. The NWREL Mathematics Problem Solving Model is designed for K-12 use, so to consider the reliable use of one of the tools (the feedback guide) with only middle school teachers may not provide a complete picture of the tool. As individual teachers, the representativeness of the 8 teachers may be a limitation as well. With such a small sample size, the inclusion of a teacher from an alternative middle school, with 8-10 students who change throughout the term may further make generalizations difficult. Even without the teacher from the alternative school, the teachers in the study tended to be from suburban or semi-rural schools, which may not be representative of the effectiveness of the model in an urban setting.

The analysis of trends over time in the teachers' use of the feedback guide was complicated by the freedom given to teachers to implement the tasks in different orders. However, the teachers' differing curricula make it impractical to demand they follow a predetermined order of implementation.

Follow-Up Instruction

One limitation of the study regarding instructional plans is the nature of the data collected. The rationales the teachers provided were often not as detailed as the researcher would have hoped, leaving some room for interpretation. However, understanding of the task, an awareness of the sort of solutions generated by students, and the incorporation of sequencing rationales minimized the extent to which inferences were required.

A further limitation is that not every teacher submitted all the requested data on every task. In some cases comments provided to students were not available for comparison to rationales. The small sample size further compounds the effects of missing data.

Critical Discussion of the NWREL Formative Feedback Guide

Chapter 2 defined summative assessment as the measurement of a level of accomplishment attained by a student, while formative assessment is assessment that provides information to the teacher or student about student performance and can be used to adjust practices. The NWREL formative feedback guide is a collection of general feedback statements that have the form of statements from teacher to student. These statements are organized into a matrix – one dimension representing the five problem solving traits, the other representing the four levels of proficiency (emerging, developing, proficient, and exemplary).

As such, the feedback statements serve as criteria statements to evaluate student proficiency levels in five problem solving traits. This essentially provides a score for the student work in each of the five areas. While that information could be used by teachers or by students to make decisions and adjust practices, by itself, the use of the guide suggests a summative assessment of performance rather than being a source of formative assessment comments to students.

The NWREL guide does, at some levels of performance, identify differences between student performance and ideal performance but does not provide guidance as to how performance can be improved. Moreover, since the criteria statements of the

feedback guide are not task specific, they may not give students information about the specifics of mistakes. The criteria statements, in the form found on the guide, certainly do not qualify as constructive feedback, as defined in Chapter 2, as they do not provide specific information for students on how to improve.

The organization of the guide by traits can serve a useful purpose in providing a framework for teachers to structure comments. However, the organization of these statements by proficiency level strongly suggests a guide for assigning scores to student problem solving performance. Indeed, we note that the discussion of inter-rater reliability addresses the extent to which the guide is used reliably as a *summative assessment* guide. The guide may help provide feedback to teachers and students, but that depends very much in its use by teachers and students. Without interpretation or augmentation by teachers and students, the NWREL formative feedback guide is more accurately a summative assessment guide.

Consideration of the inter-rater (or intra-rater) reliability of the teachers' use of the guide as a source of formative feedback comments is extremely difficult because of the nature of the guide. The comments provided to students by the teachers can be analyzed to determine the extent to which they are formative feedback, but the extent to which comments are or are not formative feedback can not be attributed to the feedback guide. The guide does not, on its own, help teachers provide formative feedback to students.

Implications

Reflections on the Study

Addressing teachers' use of assessment information to inform follow-up instructional plans is a multi-layered process. In this case, the tools incorporated in the NWREL Mathematics Problem Solving Model, particularly the NWREL formative feedback guide may not be the best tools for helping teachers translate assessment information to instructional plans. The research literature and the evidence from teacher practices in this study indicate that providing formative feedback to students and using assessment information to plan follow-up instruction is challenging for teachers.

To ask the question, "Do teachers use the NWREL formative feedback guide in valid and reliable ways for formative assessment decisions?" presumes to some extent that the guide is suitable for such decisions, and its use depends upon the teachers. However, since the guide in its prescribed use is much more summative than formative, one could suggest that its use in formative assessment decisions would be similar to the use of an exam or the results of state assessments – sources of general performance information from which a teacher could choose to focus on a specific instructional area, but do not provide any information for how to do so.

Given the tools included in the NWREL Mathematics Problem Solving Model, and the nature of the NWREL formative feedback guide, we can conclude that the teachers involved in the study did not use the guide in particularly valid or reliable ways for making formative assessment decisions. However, this may be more of a function of the NWREL formative feedback guide than of the teachers' abilities to use the guide. In

the remainder of this chapter, suggestions are made for improvements to the model to better facilitate the formative assessment decisions made.

Implications for Practitioners

Teachers in many states are required to administer problem solving tasks to students for state assessment purposes, and these tasks are often evaluated through the use of a scoring guide. In the climate of high stakes testing, student performance on these assessments can be particularly important to teachers, administrators, and school districts. This study suggests that the use of a general problem solving scoring guide does not, by itself, provide significant formative assessment information to teachers or students, under the definitions of formative assessment set forth in this dissertation. In order to provide formative feedback to students, the information they receive needs to indicate some possible ways to improve performance.

Problem solving skills are a critical part of students' mathematical understanding. The NCTM Standards (NCTM, 2000) stress the importance of students developing strong problem solving skills. State problem solving assessments reflect the same goals as the standards, but they shape instruction only in the sense of defining a target. It could be argued that summative assessment techniques such as the use of a scoring guide could play a feedback role if they are used periodically leading up to a higher stakes assessment (whether it be a state exam or a task used toward determining an evaluative guide or made in a course in school). The idea is that the preliminary assessments identify deficiencies or areas where improvement is possible. The difficulty lies in the lack of direction for *how* to eliminate the deficiencies or *how* to make improvement. That is,

simply labeling a performance as "developing" is feedback only in the weak sense, for it does not include direction for "how to develop further." Improvement in problem solving skills may be facilitated by the use of formative feedback, another reason why teachers would want to improve the quality of feedback they provide to students.

This study shows that Stein et al's model (in press) has promise for helping teachers select and sequence student work, to create task debrief discussions that use pieces of student work for some mathematical reasons. Teachers incorporating this model should strive to select and sequence student work based on mathematical learning goals, rather than simply sharing the variety of approaches taken by the class.

The levels of reliability with which the teachers used the NWREL formative feedback guide and the patterns observed in the selections of student work for sharing indicate the importance for teachers in understanding the important mathematics of a task. Identifying the important mathematics in the task being implemented is a critical first step for teachers, and teachers should be clear on the mathematical ideas prior to evaluating student performance or orchestrating a follow-up discussion.

The patterns observed in the selection of student work for sharing indicate the importance of attending to all facets of student solutions, particularly if they are to be shared in a whole class discussion. Teachers should carefully consider the entire solution before selection, and be prepared to address aspects of the solution other than those for which the piece of work was selected. Certainly, the process of evaluating student performance, and selecting and sequencing pieces of student work to be shared requires considerable work on the part of teachers, and depends greatly on the mathematical and pedagogical content knowledge required to interpret student solutions.

Implications for Professional Developers

The most significant implication for the NWREL professional developers is that the expectations for the use of the NWREL formative feedback guide in making formative assessment decisions are somewhat unrealistic based on the nature of the guide. Even with instruction in the use of the guide throughout the inservice activities, the guide appears to be sufficiently complicated and, at times, ill defined, for the teachers to use it with high levels of reliability. Professional developers could choose to focus on improving instruction in the use of the guide, possibly through more practice evaluating student performance, or through an attempt to reach consensus on the definitions and applications of individual criteria statements. Professional developers could move toward task-specific feedback guides to improve reliable use of the guide, but, based on the research literature, this would likely not have significant impact on improving student problem solving skills in general.

Proposal for a New Formative Feedback Guide

I have created a guide proposed to replace the NWREL formative feedback guide, which appears on the following pages (Figure 7.1). This guide consists of generally worded statement stems which may be filled in to customize each comment for use to address a specific issue with a piece of student work. The proposed guide is intended to help teachers construct comments for students that not only indicate the deficiency in the student's performance, but also suggest possible avenues for improving the deficiency.

The structure of the new guide is somewhat different from the original NWREL guide. One difficulty I believe was inherent in the original guide was the overlap in the

problem solving traits. With the current guide, determining whether a particular issue on a student solution related to, for example, conceptual understanding or strategies and reasoning, was often a challenge. The proposed replacement guide attempts to address these challenges by asking teachers to begin by identifying the NCTM Process Standard to which the issue to be addressed best relates: Problem Solving, Reasoning and Proof, Communication, Connections, or Representations. After determining the Process Standard that best fits the issue, the guide then uses key questions to identify the NWREL problem solving trait best incorporating the issue. Selection of the Process Standard and appropriate problem solving trait reveals several stems for comments which teachers can complete to provide personalized comments to students.

As an illustrative example of the use of this proposed guide, consider the Cross Country Scoring task. The Cross Country Scoring task presents finishing places for four different high school cross country teams. Students are asked to define four different scoring methods so that each team wins under one of the four methods.

Most Multiples of Five: One student solution devised a scoring strategy in which the team with the most finishers whose places are multiples of 5 wins. This issue relates to the way in which the student chose to solve the problem, so we look first under Problem Solving. The issue relates to the strategy the student chose, so we consider the comments under Strategies & Reasoning. An appropriate formative comment would be, "Is your use of (multiples of 5) the fairest way to (make Team A win)? Can you think of a way that would make more sense?"

Figure 7.1. Proposed Formative Feedback Guide

Problem Solving Does the issue relate to the way in which the student chose to solve the problem?	**Conceptual Understanding** Does the issue relate to the mathematical ideas & procedures used to solve the problem?	• Your use of __*(mathematical concept)*__ doesn't make sense to me here. Can you see why/ convince me? • I'm not sure your solution addresses ___*(relevant information from problem)*__. Can you change your solution to address that? • Are you sure __*(mathematical procedure)*__ makes sense here? Would __*(something else)*_ be better? • Did you consider __*(mathematical idea or procedure)*__ when you solved the problem?
	Strategies & Reasoning Does the issue relate to the appropriateness of the strategy the student used to solve the problem?	• I'm not sure your use of __*(strategy or strategy part)*__ makes sense here. Can you think of a way to fix it? • Is your use of __*(strategy or strategy part)*__ the best (easiest, most reasonable) way to solve the problem? Can you think of a better way? • I don't think __*(strategy or strategy part)*__ works to solve the problem. Try changing __*(part of strategy)*__ to see if that helps. • Can you think of another way to solve this problem? How does it relate to the strategy you chose? Try __*(step or technique)*__ if you need help getting started.
	Computation & Execution Does the issue relate to the correctness of the computations or execution of procedures the student chose to solve the problem?	• You made some mistakes calculating __*(calculation)*__. Can you see how to correct them? • You didn't use __*(procedure)*__ correctly. Can you fix it? • I don't think you finished calculating __*(calculation)*__. Try going back to see if you can make it work.
Reasoning and Proof Does the issue relate to the reasoning that the student provided?	**Strategies & Reasoning** Does the issue relate to the mathematical reasoning and logic the student provided to support the solution?	• I'm not entirely convinced by ___*(argument)*__. Can you find a better way to convince me that your strategy works? • You seem to skip from __*(step) to (step)*__. Can you fill in that gap to make your argument stronger? • Can you think of an example to show that __*(step or strategy)*__ works? • Can you think of a counterexample to show that __*(step or strategy)*__ is better than/ more appropriate than __*(other step/ strategy)*__. • Can you find a way to convince me that __*(strategy)*__ is the best way to solve the problem?
	Computation & Execution Does the issue relate to the degree to which the student's computations support the solution?	• Are you sure __*(calculation)*__ supports your solution? • Can you think of another __*(calculation)*__ to provide evidence for your solution?
	Insights Does the issue relate to the conjectures, predictions, or generalizations that the student made?	• Do you have any more evidence to support __*(prediction/ conjecture/ generalization)*__ ? • Can you think of how ___*(strategy/ procedure/ pattern)*__ might apply to other situations? • Can you come up with a general rule for __*(pattern or relationship)*__ identified? • Did you consider __*(mathematical concept or counterexample)*__ when you made your conjecture or prediction?

Figure 7.1. Proposed Formative Feedback Guide (continued)

Communication Does the issue relate to the way in which the student organized and explained the solution?	**Conceptual Understanding** Does the issue relate to the student's use of mathematical terminology?	• I'm not sure you used __*(mathematical term)*__ correctly. Check the definition and see if there is a better word to use. • You said you used __*(mathematical term)*__ but I think you might have used __*(term)*__. Compare the definitions to see which one works better here. • What does __*(term)*__ mean? Did you use it correctly here? • Is there a mathematical term to describe __*(something student uses/ states)*__?
	Communication Does the issue relate to the organization or clarity of the language used to explain the solution?	• Can you find a better, clearer way to explain __*(strategy or steps)*__? • I couldn't understand what you did in __*(step or result)*__. Can you explain it better? • Could you find a better way to organize your solution? Think about the steps you took in your strategy, and see if you can explain them in order.
Connections Does the issue relate to the connections the student made or didn't make to real life and/ or other mathematics?	• Does __*(answer student attained)*__ make sense in real life? If not, what's wrong with it? How can you fix it? • Did you think about how __*(problem aspect)*__ works in real life? Can you change __*(your strategy)*__ to reflect that? • Does __*(solution, strategy, procedure)*__ remind you of any math you've done before?	
Representations Does the issue relate to the way(s) in which the student represented the problem?	**Conceptual Understanding** Does the issue relate to the degree to which the representations reflect the mathematics of the problem?	• Does your __*(representation.)*__ work for the problem? Check the __*(mathematical issue)*__ and see if you can make your __*(representation.)*__ work better. • Can you think of a __*(representation)*__ that would help explain or support your solution? • I'm not sure your __*(student's representation)*__ is the best choice here. Have you thought about using __*(alternate representation)*__?
	Strategies & Reasoning Does the issue relate to the relationship between the strategy and the representation?	• How does your __*representation)*__ relate to the strategy you chose? Can you find a better representation? • Your __*(representation)*__ doesn't fit with __*(strategy or step)*__. Can you change it to fit better?
	Computation & Execution Does the issue relate to the accuracy or correctness of the representation?	• I'm not sure your __*(feature of representation)*__ is correct. Can you see how to fix it? • You need to __*(completion/ correction)*__ to make your __*(representation)*__ work. • Your __*(feature of representation)*__ is not very precise/ accurate. Can you see how to fix it?
	Communication Does the issue relate to the labeling or explanation of the representations?	• Can you label your __*(representation or feature of representation)*__ better/ more clearly? • I'm not sure what your __*(representation)*__ shows. Can you explain it more clearly? • Can you identify __*(important mathematics/ strategy detail)*__ in your __*(representation)*__? • Where does your __*(representation)*__ show __*(element of solution/ problem)*__?

Terminology Issues- Median and Mean: Several student solutions misidentified a calculation as the median when in fact, it was the mean. This is a communication issue, and because the focus is on mathematical terminology, the issue relates to conceptual understanding. The teacher may choose from several of the comments, depending on the learning goals. The comment, "You said you used (median) here, but I think you may have used (mean). Compare the definitions to see which one works better here," would be appropriate if the teacher wants to draw attention to the correct choice of terminology. If the teacher wants the student to determine the correct terminology herself, the comment, "What does (median) mean? Did you use it correctly here?" would be more appropriate.

Highest Mean: Some students calculated the means of each team's finishing places, and selected as the winner the team with the highest mean. A teacher may wish to draw attention to the unfairness of this scoring strategy. Such a comment relates to the connections the student is making to real life situations, so the comment stem would be found under Connections. An appropriate comment could be, "Did you think about how (winning a race) works in real life? Can you change your (use of highest mean) to better reflect that?"

Plots of Data: One student solution made use of box and whisker plots showing the upper and lower quartiles and medians of the teams' finishing places, but did not refer to the plots in the solution. The issue is with the student's representation, and most closely a representation issue related to the strategy the student used. The teacher may wish to comment on this by writing, "I'm not sure what your (box and whisker plots) show. Can you explain them or their connection to the strategy more completely?"

Not every comment stem will be useful for every task, but these can provide guidance for teachers to construct comments that are truly formative for students. These formative comments also give students more specific suggestions for improvement. The nature of the comments could lead to a revision process, where students submit revised solutions attending to the comments provided by the teacher, much as students write and revise drafts in an English class. Developing and incorporating such a process into mathematics classroom might be a useful practice for both teachers and professional developers.

Proposal for a Scaffolding Guide

Comments that are appropriate to make to students after solutions have been submitted may also be appropriate for commenting to students during the problem solving process. I have made slight revisions to the proposed formative feedback guide to make it useful as a scaffolding guide, providing teachers with comment stems they can customize to assist students in the problem solving process (see Figure 7.2). The proposed scaffolding guide works in essentially the same way as the proposed formative feedback guide: teachers identify the area of the student's solution process that needs to be addressed, identify the problem solving trait into which the issue falls, and select and customize the appropriate comment stem.

For example, as students are working on the Cross Country Scoring task, the teacher can comment on the misuse of median at the time, using a similar comment to the one that would be written on the student's work. The teacher may also be able to comment to help students who are having trouble starting the task. A teacher may choose

to say, “I can see you’re having some trouble getting started here. Have you thought about (what it means to score in a fair way)?”

Teachers may find that using a scaffolding guide, especially when it is quite similar to the guide they use to comment on student work, helps to devise questions that help students understand the problem, but do not necessarily reduce the cognitive demand of the task. Professional developers may wish to use the scaffolding guide to teach the type of scaffolding they would like teachers to implement in the classroom, particularly if the professional developers are concerned with instructing teachers in maintaining high cognitive demand of rich tasks.

Figure 7.2. Proposed Scaffolding Guide

Problem Solving Does the issue relate to the way in which the student is solving the problem?	**Conceptual Understanding** Does the issue relate to the mathematical ideas & procedures the student is using (or not using) on the problem?	• Your use of __*(mathematical concept)*__ doesn't make sense to me here. Can you see why/ convince me? • I'm not sure your solution addresses ___*(relevant information from problem)*__. Can you change your solution to address that? • Are you sure __*(mathematical procedure)*__ makes sense here? Would something else be better? • Have you considered __*(mathematical idea or procedure)*__ while you've been working on the problem? • Does this problem remind you of anything you've seen before? • Have you thought about how __*(mathematical idea or procedure)*__ we've been working on in class might fit in here? • Can you tell me what you mean by __*(mathematical concept being misused)*__?
	Strategies & Reasoning Does the issue relate to the appropriateness of the strategy the student is using to solve the problem?	• I'm not sure your use of __*(strategy or strategy part)*__ makes sense here. Can you think of a way to fix it? • Is your use of __*(strategy or strategy part)*__ the best (easiest, most reasonable) way to solve the problem? Can you think of a better way? • I don't think __*(strategy or strategy part)*__ works to solve the problem. Try changing __*(part of strategy)*__ to see if that helps. • Can you think of another way to solve this problem? How does it relate to the strategy you chose? Try __*(step or technique)*__ if you need help getting started. • It looks like you're having trouble getting started. Have you thought about trying __*(strategy/ step)*__? • It looks like you're having trouble getting started. What are you thinking about the problem?
	Computation & Execution Does the issue relate to the correctness of the computations or execution of procedures the student is using to solve the problem?	• You made some mistakes calculating __*(calculation)*__. Can you see how to correct them? • You didn't use __*(procedure)*__ correctly. Can you fix it? • I don't think you finished calculating __*(calculation)*__. Can you fix it? • You might want to check your __*(calculation/ procedure)*__ at this step. • I'm not sure what you have done here (*in a calculation or procedure).* Can you explain it to me?
Reasoning and Proof Does the issue relate to the reasoning that the student is using?	**Strategies & Reasoning** Does the issue relate to the mathematical reasoning and logic the student is providing to support the solution?	• I'm not entirely convinced by ___*(argument)*__. Can you find a better way to convince me that your strategy works? • You seem to skip from __*(step) to (step)*__. Can you fill in that gap to make your argument stronger? • Can you think of an example to show that __*(step or strategy)*__ works? • Can you think of a counterexample to show that __*(step or strategy)*__ is better than/ more appropriate than __*(other step/ strategy)*__. • Can you find a way to convince me that __*(strategy)*__ is the best way to solve the problem? • Does your strategy always work? How do you know? What constraints are there? What if you had __*(some change in situation)*__?
	Computation & Execution Does the issue relate to the degree to which the student's computations support the solution?	• Are you sure __*(calculation)*__ supports your solution? • Can you think of another __*(calculation)*__ to provide evidence for your solution? • You need more support for your solution. What calculations and procedures did you use to show it worked?
	Insights Does the issue relate to the conjectures, predictions, or generalizations that the student is making?	• Do you have any more evidence to support __*(prediction/ conjecture/ generalization)*__ ? • Can you think of how ___*(strategy/ procedure/ pattern)*__ might apply to other situations? • Can you come up with a general rule for __*(pattern or relationship)*__ identified? • Did you consider __*(mathematical concept or counterexample)*__ when you made your conjecture or prediction? • Did you notice any patterns here? • How would this work in the *nth* case (or some other generalization)?

Figure 7.2. Proposed Scaffolding Guide (continued)

Communication Does the issue relate to the way in which the student is organizing and explaining the solution?	**Conceptual Understanding** Does the issue relate to the student's use of mathematical terminology?	• I'm not sure you are using __*(mathematical term)*__ correctly. Check the definition and see if there is a better word to use. • You say you are using __*(mathematical term)*__ but I think you may be using __*(term)*__. Compare the definitions to see which one works better here. • What does __*(term)*__ mean? Are you using it correctly here? • Is there a mathematical term to describe __*(something student uses/ states)*__?
	Communication Does the issue relate to the organization or clarity of the language the student is using to explain the solution?	• Can you find a better, clearer way to explain __*(strategy or steps)*__? • I don't understand what you are doing in __*(step or result)*__. Can you explain it better? • Can you find a better way to organize your solution? Think about the steps you are taking in your strategy, and see if you can explain them in order.
Connections Does the issue relate to the connections the student is making (or not making) to real life and/ or other mathematics?	• Does __*(answer student attained)*__ make sense in real life? If not, what's wrong with it? How can you fix it? • Did you think about how__*(problem aspect)*__ works in real life? Can you change your strategy to reflect that? • Does ___*(solution, strategy, procedure)*__ remind you of any math you've done before?	
Representations Does the issue relate to the way(s) in which the student is representing the problem?	**Conceptual Understanding** Does the issue relate to the degree to which the representations reflect the mathematics of the problem?	• Does your __*(representation.)*__ work for the problem? Check the __*(mathematical issue)*__ and see if you can make your __*(representation.)*__ work better. • Can you think of a _*(representation)*__ that would help explain or support your solution? • I'm not sure your __*(student's representation)*__ is the best choice here. Have you thought about using __*(alternate representation)*__?
	Strategies & Reasoning Does the issue relate to the relationship between the strategy and the representation?	• How does your __*representation)*__ relate to the strategy you chose? Can you find a better representation? • Your __*(representation)*__ doesn't fit with __*(strategy or step)*__. Can you change it to fit better?
	Computation & Execution Does the issue relate to the accuracy or correctness of the representation?	• I'm not sure your __*(feature of representation)*__ is correct. Can you see how to fix it? • You need to __*(completion/ correction)*__ to make your __*(representation)*_ work. • Your __*(feature of representation)*__ is not very precise/ accurate. Can you see how to fix it?
	Communication Does the issue relate to the labeling or explanation of the representations?	• Can you label your __*(representation or feature of representation)*__ better/ more clearly? • I'm not sure what your __*(representation)*__ shows. Can you explain it more clearly? • Can you identify __*(important mathematics/ strategy detail)*__ in your __*(representation)*__? • Where does your__*(representation)*__ show __*(element of solution/ problem)*__?

Proposal for a Debrief Structuring Guide

Teachers may need further guidance in structuring debrief class sessions, particularly when they are trying to use student work as the foundation of the debrief. This study indicated that structuring a debrief class with a mathematical goal, around the solutions generated by students is a challenging and complex task. The proposed debrief structuring guide here is based on the steps identified in Stein et al's (in press) pedagogical model (see Figure 7.3). The guide is essentially a series of questions for teachers to respond to which, together, create a plan for sharing student work with a mathematical goal in mind.

To illustrate, again consider the use of the debrief structuring guide in the context of the Cross Country Scoring task to plan a hypothetical debrief session.

Anticipating: The first step in the guide is Anticipating student work on the task. The teacher solves the task and considers the types of strategies, the mathematical ideas, and the misconceptions that he or she expects to see. The teacher also identifies the mathematical goals of the task.

In the Cross Country Scoring task, suppose the teacher anticipates that students will consider mean and median, and will probably make some mistakes in choosing highest mean, or in looking at total scores although the teams had different numbers of runners. The teacher may decide to use the task to compare the situations in which mean, median, or mode might be most appropriate and to introduce weighted averages.

Figure 7.3. Proposed Debrief Structuring Guide

I. Anticipating

1. Solve the problem in as many ways as you can think of. What are the important mathematical ideas in your solutions?

2. What would the best solution your students generate look like? What would the worst one look like?

3. What common strategies do you expect your students to use?

4. What common difficulties do you expect your students to have in solving the problem? What mathematical misconceptions do you expect to see?

5. Why are you using this task at this time in your classroom? What mathematical ideas and concepts do you want your students to learn from the task?

II. Monitoring

1. What are the common strategies your students are using? What are the important mathematical ideas in their strategies?

2. What kinds of difficulties are your students having solving the problem? What mathematical misconceptions are they revealing?

3. What are the mathematical concepts or ideas that are appearing in your students' solutions?

III. Selecting

1. What are your mathematical learning goals for the debrief?

2. Which pieces of student work have misconceptions or mistakes that were common or important and need to be addressed to achieve your mathematical goals?
    a. Of those pieces, choose the ones that will best help you address the misconceptions.
    b. For each piece you have selected, what are some questions you can ask the student or the class during the discussion to expose and/ or debunk the misconception?

Figure 7.3. Proposed Debrief Structuring Guide, Continued

3. Which pieces of student work incorporate important mathematical concepts, ideas, or procedures that can help you meet your mathematical goals?
    a. Of those pieces, choose the ones that best help you address the important mathematics.
    b. For each piece you have selected, what are some questions that you can ask the student or the class during the discussion to expose or illuminate those important ideas?

4. Is there anything important you want your students to consider that you did not find in a student solution?
    a. Can you use a piece of student work from another class to share this idea?
    b. Can you generate a solution that you can use to share this idea?
    c. Can you develop an activity that will help expose this idea?

IV. Sequencing

1. What is the most effective way to sequence student work to meet your mathematical goals?
    a. Do you need to address a common or important misconception before you can move on to the mathematical ideas you want your students to learn?
    b. Do you need to start with a common strategy to create an entry point into the discussion so students can understand the more mathematically sophisticated strategies?
    c. Do you want to place related or contrasting strategies next to one another so that you can build or compare mathematical ideas?
    d. Is there another sequencing strategy that makes sense for achieving your mathematical goals?

V. Connecting

1. What questions will you ask between solutions to make connections and expose ideas?
    a. Do two (or more) of the solutions you selected look different but have the same basic ideas? What can you ask to help students see that?

Figure 7.3. Proposed Debrief Structuring Guide, Continued

b. Are there strategies that were shared that would be more suitable than others for other situations? What can you ask to help students see that?
c. Are there strategies that are very different but both work well? What can you ask to help students see that?
d. Are there any other questions you need to ask to ensure that your mathematical goals will be reached?

2. Do you think your students may need more than the discussion to meet your mathematical learning goals?
    a. Can you develop an activity, or use one you already have to help them understand the important mathematics?

Monitoring: The next step is Monitoring student work on the task. The teacher is observing students at work on the problem, or examining the written solutions after they have been completed. Here the teacher notes the strategies students are using and the important ideas in the strategies, and identifies the common or important misconceptions or mistakes students are making.

In the case of Cross Country Scoring, suppose, the teacher notices that students are referring to mean and median, as she anticipated, and students are also choosing highest means and medians, and comparing sums of finishing places. She is surprised to notice that students have selected some strategies that are unreasonable and not at all related to measures of central tendency, such as choosing the team with the most even number finishers as the winner. She also notes that none of her students have used any type of weighted averaging to solve the problem.

Selecting: In Selecting student work to share, the teacher first identifies the mathematical goals for the debrief discussion. She then selects pieces of student work that best expose

the misconceptions or mistakes she wants to address, and notes the questions she will ask the student or the class to expose and resolve those misconceptions. She selects the pieces of student work that best share the important mathematical ideas she wants to address, and plans the questions she will ask to facilitate this. Finally, if there are any important mathematical ideas that she wanted to share that did not appear in any of her student solutions, she could choose a piece of student work from another class, generate one, or develop an activity to address these ideas.

Consider the Cross Country Scoring task. Suppose the teacher's mathematical goals are to address the use of nonmathematical strategies, the appropriateness of mean, median, and mode, and introduce weighted averages. The common misconceptions she identifies to discuss are the nonmathematical strategies, the use of highest mean to select the winner, and the use of sums of places. She selects the pieces of student work she believes will best expose each misconception. She also prepares some questions to ask, such as "Is it fair to select the winner based on the types of numbers the finish places are (i.e. even, prime, etc)?" or, "Is it fair to select the team with the lowest sum as the winner? What about the different number of runners on each team?"

She has identified the use of mean and median as the strategies she wants to share to discuss mathematical ideas, and selects three sample solutions to be shared. She prepares some questions to ask, such as "Why did you decide to use median or mean? Are they fair ways to select the winner? Why?" or "Did anyone consider using mode here? Why or why not?" She might also decide to ask the students if they can think of situations when mean would be more appropriate than median, or when mode would be more appropriate than the other two.

She also wanted to address weighted average, but no student has used a weighted average strategy. She writes her own solution and plans to present it as one she saw in another class. She plans to ask the students what they think about the fairness of using weighted average, and if they can think of any other situations where it is used or would be a good choice.

Sequencing: In the Sequencing step, the teacher determines a sequence for the pieces of student work to share that is will help meet her mathematical goals.

In the case of the Cross Country Scoring task, she may decide to address the misconceptions first, then share the successful strategies (starting with the more commonly used mean, then addressing median), and, finally, she shares the solution using weighted averaging.

Connecting: In the Connecting step, the teacher plans questions to ask between the solutions or after several solutions to draw connections between them. The teacher also decides if a follow up activity may be necessary.

In the case of the Cross Country Scoring task, she may decide that the questions she already planned comparing mean, median, and mode will draw attention to some connections between the student papers. She also plans to ask a question that compares weighted average to mean and asks students to identify the similarities and the differences between the two ways of averaging.

The debrief structuring guide can help teachers make some of the instructional decisions required to facilitate the use of student work. The guide may help minimize some of the discomfort teachers feel (see Chapter 3) when using student solutions as the

foundation of a class discussion, particularly by placing more control of the discussion in the hands of the teacher. The debrief structuring guide could serve as a tool for professional developers to assist teachers in their efforts to structure debrief sessions around student work, and to increase the chances that those debrief sessions are focused on mathematical goals, rather than just displays of strategies generated.

Implications for Researchers

A natural next step in research would be to investigate teachers' use of the three new proposed guides: the formative feedback guide, the scaffolding guide, and the debrief structuring guide. I suggest that the same principal research question would be appropriate with the new set of tools: Do teachers use the *new formative feedback guide, the scaffolding guide, and the debrief structuring guide* in valid and reliable ways to make formative assessment decisions? The three issues identified in the original principal research question would be reasonable avenues of research as well.

- *Inter-Judge Reliability:* Do teachers use the new feedback guide and the scaffolding guide validly and reliably in making comments to students?
- *Improvement Over Time:* Do teachers' abilities to make appropriate comments using the new feedback guide and the scaffolding guide improve over time?
- *Follow-Up Instruction:* What is the evidence of teachers' use of *formative* assessment information to make follow up instructional plans.

New research questions emerge as well. For example, *to what extent does the debrief structuring guide affect the mathematical quality of debrief sessions*? Does use of the

debrief structuring guide influence teachers to select pieces of student work to share for mathematical reasons? It may be worthwhile to investigate the selection rationales teachers provided in this year of the study in comparison to the selection rationales provided by teachers with access to the debrief structuring guide to observe any differences or trends.

The new formative feedback guide provides opportunities to conduct research on the process of revising student work in a mathematics classroom. This would entail teachers providing formative, constructive comments to students on their problem solving and students revising their solutions to reflect the comments provided. One possible research question would be *does the practice of revising solutions based on constructive feedback from the guide improve student's problem solving abilities*?

An additional strand for research could be to investigate the connections between the principal research question of this study and the Question 3 of the NWREL research project, the question investigating teacher practices. For example, *how are the plans teachers identify for sharing implemented in the debrief session*? What evidence is there of the rationales teachers provide for selecting and sequencing student work in the follow-up instruction as implemented? What evidence is there of the instructional plans provided by the teachers in the teacher discourse during the debrief session?

Aspects of this study, particularly the difficulties teachers had using the NWREL formative feedback guide and structuring debrief sessions with mathematical purposes, could have been attributed to the mathematical complexity of the task and the teachers' mathematical understanding. This suggests that mathematical content knowledge and pedagogical content knowledge might play an important role in teachers' reliable use of

assessment tools or use of assessment information to plan follow-up instruction. One direction for future research might be to investigate the connections between teachers' content knowledge and use of the new feedback, scaffolding, and debrief structuring guides.

Chapter 8 – Conclusion

This study investigated middle school teachers' use of a formative feedback guide for mathematics problem solving instruction. The research is part of a larger project investigating the effectiveness of a professional development model for incorporating problem solving into the classroom. This part of the research project is motivated by the research literature indicating that formative assessment is an important, but challenging facet of classroom practices (Black et al, 2003; Lee, 2001; Wiliam, 2007) and uses Stein et al's (in press) pedagogical model as a framework for analyzing follow-up instruction.

Results

The principal research question investigated in this study is: **Do teachers use the NWREL problem solving formative feedback guide in valid and reliable ways for formative assessment decisions?**

The question was addressed by examining three issues: the inter-rater reliability of the teachers' use of the feedback guide, any trends in the teachers' use of the feedback guide over time, and the evidence of teachers' use assessment information to inform follow-up instruction.

Inter-Rater Reliability

Do teachers attain a high degree of inter-rater reliability in the use of the NWREL formative feedback guide?

The study suggests that teachers do not attain high levels of inter-rater reliability in the use of the NWREL formative feedback guide. While the levels of exact agreement between teacher and expert ratings are low, the teacher ratings tended to differ from the expert ratings by less than 0.5 levels.

Improvement Over Time

Do the teachers' abilities to assess students' problem solving skills improve with the implementation of the model?

The study provides no evidence to indicate that teachers' abilities to assess students' problem solving skills improve over the course of the implementation of the model, nor is there any evidence that teachers' abilities deteriorate over the course of the project.

Follow-Up Instriuction

What is the evidence of the teachers' use of assessment information from students to inform instructional plans?

The study suggests that teachers participating in the research project based instructional plans on aspects of mathematical problem solving most of the time. However, the aspects of mathematical problem solving on which the plans were based were not often deep mathematical ideas and concepts, as suggested in the pedagogical model of Stein et al (in press).

Implications

The study suggests that the NWREL formative feedback guide may not be a tool for providing constructive, formative comments. In response to this difficulty, a proposed formative feedback guide was developed to better provide these types of comments. To accompany the proposed feedback guide, a proposed scaffolding guide and a proposed debrief structuring guide were developed. The proposed guides suggest further research into the principal research question based on teacher use of the proposed guides. Other research areas of investigation include the use of formative comments to help students revise solutions, the relationship between teacher content knowledge and reliable use of the proposed guides, and the relationship between teacher plans for follow-up instruction and the implementation of those plans.

References

American Federation of Teachers, National Council on Measurement in Education, National Education Association (AFT, NCME, NEA). (1990). Standards for teacher competence in the educational assessment of students. *Educational Measurement: Issues and Practice, 9* (4), 30-32.

Arter, J. (1993 April). *Designing scoring rubrics for performance assessments: The heart of the matter.* Paper presented at the annual meeting of the American Educational Research Association, Atlanta, Georgia.

Arter, J., & McTighe, J. (2001). *Scoring rubrics in the classroom: Using performance criteria for assessing and improving student performance.* Thousand Oaks, CA: Corwin Press, Inc.

Ball, D. L. (1997). From the general to the particular: Knowing our own students as learners of mathematics. *Mathematics Teacher 90* (9), 732-737.

Ball, D. L. (2002). Teaching, with respect to mathematics and students. In T. Wood, B.S. Nelson, & J. Warfield (Eds.) *Beyond classical pedagogy: Teaching elementary school mathematics* (11-22). Mahwah, NJ: Erlbaum.

Bangert-Drowns, R. L., Kulik, C.-L. C., Kulik, J.A., & Morgan, M. T. (1991). The instructional effect of feedback in test-like events. *Review of Educational Research, 61* (2), 213-238.

Becker, J.P., & Shimada, S. (1997). *The open-ended approach: A new proposal for teaching mathematics.* Reston, VA: National Council of Teachers of Mathematics.

Black, P. & William, D. (1998). Assessment and classroom learning. *Assessment in Education: Principles, Policies, and Practice, 5* (1).

Black, P., Harrison, C., Lee, C., Marshall, B. & Wiliam, D. (2003). *Assessment for learning: Putting it into practice.* Berkshire, England: Open University Press.

Borko, H. & Livingston, C. (1989). Cognition and improvisation: Differences in mathematics instruction by expert and novice teachers. *American Educational Research Journal, 26* (4), 473-498.

Brendefur, J. & Frykholm, J. (2000). Promoting mathematical communication in the classroom: Two preservice teachers' conceptions and practices. *Journal of Mathematics Teacher Education, 3,* 125-153.

Burkhardt, H. (1988). Teaching problem solving. In H. Burkhardt, S. Groves, A. Schoenfeld, & K. Stacey (Eds.), *Problem solving—A world view. Proceedings of the problem solving theme group, ICME 5* (pp. 17-42). Nottingham, England: University of Nottingham, Shell Centre for Mathematical Education.

Butler, R. (1988). Enhancing and undermining intrinsic motivation: The effects of task-involving and ego-involving evaluation on interest and performance. *British Journal of Educational Psychology, 58* (1), 1-14.

Cai, J., Lane, S., & Jakabcsin, M. S. (1996). The role of open-ended tasks and holistic scoring rubrics: Assessing students' mathematical reasoning and communication. In P. Elliot & M. Kenney (Eds.), *Communication in mathematics, K-12 and beyond* (137-145). Reston, VA: The National Council of Teachers of Mathematics, Inc.

Calfee, R. C., & Masuda, W. V. (1997). Classroom assessment as inquiry. In G. Phye (Ed.), *Handbook of classroom assessment* (69-102). San Diego: Academic Press, Inc.

Chazan, D. & Ball, D. (1999). Beyond being told not to tell. *For the Learning of Mathematics 19* (2), 2-10.

Driscoll, M. (1999). Crafting a sharper lens: Classroom assessment in mathematics. In M. Solomon & A. Costa (Eds.), *The diagnostic teacher: Constructing new approaches to professional development* (78-103). New York: Teachers College Press.

Elawar, M. C. & Corno, L. (1985). A factorial experiment in teachers' written feedback on student homework: Changing student behavior a little rather than a lot. *Journal of Educational Psychology, 77* (2), 162-173.

Even, R. (2005). Using assessment to inform instructional decisions: How hard can it be? *Mathematics Education Research Journal 17* (3), 45-61.

Farrel, M. A. (1992). Learn from your students. *Mathematics Teacher 85* (8), 656-659.

Fuchs, L. S., Deno, S. L., & Mirkin, P. K. (1984). Effects of frequent curriculum-based measurement and evaluation on pedagogy, student achievement, and student awareness of learning. *American Educational Research Journal, 21* (2), 449-460.

Fuchs, L. S., Fuchs, D., Hamlett, C. L., & Stecker, P. M. (1991). Effects of curriculum – based measurement and consultation on teacher planning and student achievement in mathematics operations. *American Educational Research Journal, 28* (3), 617-641.

Fuchs, L. S., Fuchs, D., Karns, K., Hamlett, C. L., & Katzaroff, M. (1999). Mathematics performance assessment in the classroom: Effects on teacher planning and student problem solving. *American Educational Research Journal, 36* (3), 609-646.

Goldberg, G. L, & Michaels, H (1995). Same-scorer judgments on multiple content area items in integrated performance assessment. Proceedings of AERA '95: *The Annual Meeting of the American Educational Research Association*.

Groves, S. & Doig, B. (2006). Progressive discourse in mathematics classes- the task of the teacher. Proceedings of PME '04: *The 28th Conference of the International Group for the Psychology of Mathematics Education* (2), 495-502.

Guskey, T. R. (1986). Staff development and the process of educational change. *Educational Researcher, 15* (5).

Heaton, R. M. (2000). *Teaching mathematics to the new standards: Relearning the dance.* New York: Teachers College Press.

Hiebert, J., Carpenter, T.P., Fennema, E., Fuson, K.C., Wearne, D., Murray, H., Olivier, A., & Human, P. (1997). *Making sense: Teaching and learning mathematics.* Portsmouth, NH: Heinemann

Kalathil, R.R. (2006a). Characterizing the nature of discourse in the mathematics classroom. Proceedings from ICLS '06: *The 7th International Conference on Learning Sciences*. Bloomington, IN: The International Society of the Learning Sciences.

Kalathil, R.R (2006b). Using comparisons of alternate strategies to promote discourse in mathematics classrooms. Proceedings from ICLS '06: *The 7th International Conference on Learning Sciences*. Bloomington, IN: The International Society of the Learning Sciences.

Kilpatrick, J., Swafford, J., Findell, B. (2001). *Adding it up: Helping children learn mathematics.* Washington D.C.: National Academy Press.

Kluger, A. N., & DeNisi, A. (1996). The effects of feedback interventions on performance: A historical review, a meta-analysis, and a preliminary feedback intervention theory. *Psychological Bulletin 119* (2), 254-284.

Koretz, D. A., Klein, S., McCaffrey, D., & Stecher, B. (1994). *Interim report: The reliability of Vermont portfolio scores in the 1992-1993 school year.* Santa Monica, CA: RAND.

Koretz, D., McCaffrey, D., Klein, S., Bell, B., & Stecher, B. (1992). *The reliability of scores from the 1992 Vermont portfolio assessment program.* Santa Monica, CA: RAND Institute on Education and Training.

Kulm, G. (1994). *Mathematics assessment: What works in the classroom.* San Francisco: Jossey-Bass Publishers.

Leahy, S., Lyon, C., Thompson, M., & William, D. (2005). Classroom assessment, minute by minute, day by day. *Educational Leadership, November 2005*, 18-24.

Lee, C. (2006). *Language for learning mathematics.* Berkshire, England: Open University Press.

Leinhardt, G. & Steele, M. D. (2005). Seeing the complexity of standing to the side: Instructional dialogues. *Cognition and Instruction, 23*(1), 87-163.

Lester, F. K., Jr., Lambdin, D. V., & Preston, R. V. (1997). A new vision of the nature and purposes of assessment in the mathematics classroom . In G. Phye (Ed.), *Handbook of classroom assessment* (287-319). San Diego: Academic Press, Inc.

Lietze, A. R. & Mau, S. T. (1999). Assessing problem-solving thought. *Mathematics Teaching in the Middle School, 4* (5). 305-311.

Mathematical Sciences Education Board (MSEB) (1993). *Measuring what counts: A conceptual guide for mathematics assessment.* Washington D.C.: National Academy Press.

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

National Council of Teachers of Mathematics (1995). *Assessment Standards for School Mathematics.* Reston, VA: NCTM.

Nyquist, J. B. (2003). *The benefits of reconstruing feedback as a larger system of formative assessment: a meta-analysis.* Unpublished master's thesis, Vanderbilt University, Nashville Tennessee.

Osterman, K. F., & Kottkamp, R. B. (1993). *Reflective practice for educators: Improving schooling through the professional development.* Newbury Park, CA: Corwin Press.

Peressini, D. & Bassett, J. (1996). Mathematical communication in students' responses to a performance-assessment task. In P. Elliot & M. Kenney (Eds.), *Communication in mathematics, K-12 and beyond* (146-158). Reston, VA: The National Council of Teachers of Mathematics, Inc.

Perlman, C. C. (2003). Performance assessment: Designing appropriate performance tasks and scoring rubrics. In J. E. Wall & G. R. Walz (Eds.), *Measuring up: Assessment issues for teachers, counselors, and administrators.* Greensboro, NC: CAPS Press.

Petit, M. & Zawojewski, J. S. (1997). Teachers and students learning together about assessing problem solving. *The Mathematics Teacher, 90* (6). 472-477.

Peterson, P. L. & Clark, C. M. (1978). Teachers' reports of their cognitive processes during teaching. *American Educational Research Journal, 15* (4), 555-565.

Plake, B. S., & Impara, J. C. (1997). Teacher assessment literacy: What do teachers know about assessment? In G. Phye (Ed.), *Handbook of classroom assessment* (53- 68). San Diego: Academic Press, Inc.

Putnam, T. (1987). Structuring and adjusting content for students: A study of live and simulated tutoring of addition. *American Educational Research Journal, 24* (1), 13-48.

Quellmalz, E., & Hoskyn, J. (1997). Classroom assessment of reasoning strategies. In G. Phye (Ed.), *Handbook of classroom assessment* (103-130). San Diego: Academic Press, Inc.

Quinlan, A. M. (2006). *A complete guide to rubrics: Assessment made easy for teachers, K-college.* Lanham, Maryland: Rowman & Littlefield Education.

Schoenfeld, A.H. (1992). Learning to think mathematically: Problem solving, metacognition, and sense making in mathematics. In D.A. Grouws (Ed.), *Handbook of research on mathematics teaching and learning* (pp. 334-368). New York, NY: Macmillan.

Sherin, M.G. (2002). When teaching becomes learning. *Cognition and Instruction, 20*(2), 119-150.

Sparks, G. M., & Simmons, J. M. (1989). Inquiry-oriented staff development: Using research as a source of tools, not rules. In S. Caldwell (Ed.), *Staff development: A handbook of effective practices.* Oxford, OH: National Staff Development Council.

Stevens, D. D. & Levi, A. (2005). *Introduction to rubrics: An assessment tool to save grading time, convey effective performance, and promote student learning.* Sterling, Virginia: Stylus Publishing.

Stein, M.K., Smith, M.S., Henningsen, M.A., & Silver, E.A. (2000). *Implementing standards-based mathematics instruction: A casebook for professional development.* New York: Teachers College Press.

Stein, M. K., Engle, R. A., Hughes, E. K., & Smith, M. S. (in press). Orchestrating productive mathematical discussions: Five practices for helping teachers move beyond show and tell. *Mathematical Thinking and Learning*.

Stiggins, R. J. (2001). *Student-involved classroom assessment*. New Jersey: Prentice-Hall, Inc.

Watkins, M. W. & Pacheco, M. (2001). Interobserver agreement in behavioral research: Importance and calculation. *Journal of Behavioral Education 10* (4), 205-212.

Wertsch, J. & Toma, C. (1995). Discourse and learning in the classroom: A sociocultural approach. In L. Steffe and J. Gale (Eds.), *Constructivism in education* (159-174). Hillsdale, NJ: Erlbaum.

Wilcox, S. K. & Zielinski, R. S. (1997). Using the assessment of students' learning to reshape teaching. *Mathematics Teacher 90* (3), 223-229.

Wiliam, D. (2007). Keeping learning on track: Classroom assessment and the regulation of learning. In F. Lester (Ed.), *Second handbook of research of mathematics teaching and learning.* Reston, VA: National Council of Teachers of Mathematics.

Wood, T. (1995). From alternative epistemologies to practice in education: Rethinking what it means to teach and learn. In L. Steffe and J. Gale (Eds.), *Constructivism in education* (331-340). Hillsdale, NJ: Erlbaum.

Wood, T., Cobb, P., & Yackel, W. (1995). Reflections on learning and teaching mathematics in elementary school. In L. Steffe and J. Gale (Eds.), *Constructivism in education* (401-422). Hillsdale, NJ: Erlbaum.

Woolfolk, A. (2004). *Educational Psychology, 9th Edition.* Boston: Allyn and Bacon.

Yackel, E., Cobb, P., Wood, T., Wheatley, G., & Merkel, G. (1990). The importance of social interaction in children's construction of mathematical knowledge. In T.J. Cooney (Ed.) *Teaching and learning mathematics in the 1990s* (12-21). Reston. VA: The National Council of Teachers of Mathematics.

Appendix

NWREL TASKS

Cross-Country Scoring

Here's something to think about…

Four high schools, Applebee, Burnside, Clark, and DeWitt competed in a cross-country race. Each school entered up to 10 runners in the race. The table below shows the finish position of each runner and the school they attended.

Applebee	Burnside	Clarke	DeWitt
3	4	1	6
5	8	2	7
10	9	12	11
13	15	14	16
20	21	19	17
26	28	24	18
30	31	27	22
36	33	32	23
*	34	35	25
*	*	*	29

Can you devise 4 different scoring systems where each team could be a winner?

Explain each ranking system and tell the order of finish that would result.

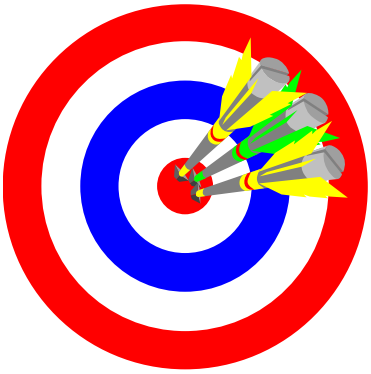
- Show **all** work using pictures, numbers and words.
- Which ranking system do you think is most fair? Why do you think so?

Cross-Country Scoring (continued)

Did you remember to:

- ☐ Show your solutions with a diagram or chart?
- ☐ Explain your strategy and thinking?
- ☐ Choose the ranking system that was best and explain why you chose it?

Design a Dartboard



As some of you may know, the traditional dartboard is made up of concentric circles or squares. As a new twist on the traditional dartboard your company wants to make a dart board of some other shape. You are in charge of designing the board. Be sure to use a shape other than a square or circle. The board should have three major sections. The area of the board should be divided so the area has:

15% for the inner section
25% for the middle section
60% for the outer section

- Draw a design for your dartboard
- Show **all** your work using numbers, drawings, and words
- Explain the strategy you used to get your answer

Design a Dart Board Continued…

Did you remember to:

Show your work using numbers, drawings, and words.

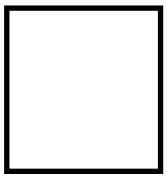
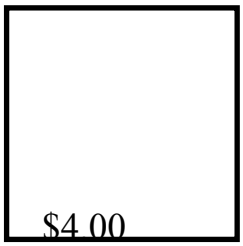
Explain the strategy you used to get your answer.

Kyle's Tiles

Here's something to think about…

Kyle is planning to put decorative tile on a wall. The space is 28 inches

wide and 30 inches high. The tiles come in these sizes and costs:

4"x4" squares	**6"x6"squares**	**4"x6" rectangles**
		
$2.00	$4.00	$3.00

If Kyle needs to cut a tile, it costs $.50 for each cut.

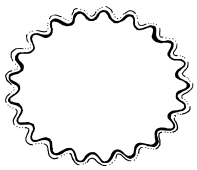
- Design a layout that you think would be best for the tiles.
- Explain why you decided it was "best".
- Find the total cost for your design and show all your work.

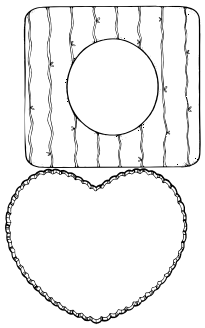
Kyle's Tiles (continued)

Did you remember to:
- Show your solution with a picture or table
- Describe why you chose your design
- Explain your strategy and thinking
- Check that your answer makes sense

The Mathemagicians' Cookies

Here's something to think about…



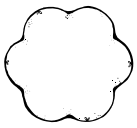


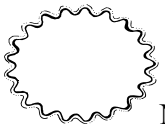
The Mathemagicians are having a bake sale to raise at least $50 for some new math manipulatives.

They all got together and made 360 cookies.

It cost them $.10 per cookie to make them. Now they are ready to package and price the cookies. They want to sell them in three different sized packages (each size package has a different number of cookies).

- **How could the cookies be packaged?**
- **How much should they charge for each size package?**
- **If they sell all the packaged cookies how much <u>profit</u> will the Mathemagicians make?**
- **Show your work using pictures, tables or numbers.**
- **Explain why you picked the size of packages you did.**
- **Is your pricing reasonable? Explain why you priced them the way you did.**





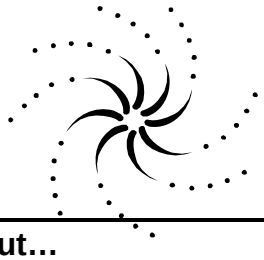
NSF

Cookies (continued)

Did you remember to:

- Show your solution with a picture or table?
- Explain your strategy and thinking?
- Describe your package pricing?
- Determine the profit?
- Check that your answer makes sense?

The Spinner Elimination Game



Here's something to think about…

Try to design a spinner that will let you cross out more squares than your opponent on a 50's chart. (See rules for the Spinner Elimination Game.) After you finish a game, decide if you want to change the numbers on your spinner.

After you have played several games, answer the following questions:

1. Would you change the numbers on your spinner? Why or why not?
2. What advice would you give to someone who wants to win the game for strategies on how to choose numbers for their spinner? Support your argument.

Rules for the Spinner Elimination Game

1. **Divide your spinner into eight equal sections.**
2. **You may choose up to 8 numbers (from zero to nine) to put on your spinner.**
3. **You may put them on any space you choose on the spinner and you may use the same number as many times as you like.**
4. **You eliminate squares on the 50's chart by spinning your spinner as many times as you choose and multiplying the product of the spins (E.g. If you spin three times and you spin a 4, then a 3 and then a 4 you would get $4\times3 = 12$, $12\times4 = 48$. You would eliminate 48 from the 50's chart. If choose to spin only one time and get a 4, then you would eliminate the 4.). Each time you eliminate a square counts as one turn.**
5. **If your spin creates a product greater than 50 you lose that turn and the next player spins.**
6. **You can only cross off one number per turn.**
7. **After 20 turns, the player with the most squares eliminated on their 50's chart wins the game.**

1	2	3	4	5	6	7	8	9	10
11	12	13	14	15	16	17	18	19	20
21	22	23	24	25	26	27	28	29	30
31	32	33	34	35	36	37	38	39	40
41	42	43	44	45	46	47	48	49	50

1	2	3	4	5	6	7	8	9	10
11	12	13	14	15	16	17	18	19	20
21	22	23	24	25	26	27	28	29	30
31	32	33	34	35	36	37	38	39	40
41	42	43	44	45	46	47	48	49	50

NWREL TEACHER GUIDES

Teachers Guide for

Cross-Country Scoring

Grade Level	Content Strand	Timing	Grouping	Materials
6 - 8	Probability and Statistics	Approximately 1 hour	Teams of 2-3 or individual	None

Here's something to think about…

Four high schools, Applebee, Burnside, Clark, and DeWitt competed in a cross-country race. Each school entered up to 10 runners in the race. The table below shows the finish position of each runner and the school they attended.

Applebee	Burnside	Clarke	DeWitt
3	4	1	6
5	8	2	7
10	9	12	11
13	15	14	16
20	21	19	17
26	28	24	18
30	31	27	22
36	33	32	23
*	34	35	25
*	*	*	29

Can you devise 4 different scoring systems where each team could be a winner?

Explain each ranking system and tell the order of finish that would result.

- Which ranking system do you think is most fair? Explain why you think so.
- Show all your work using charts, diagrams, numbers and words.

Purpose
Students explore sets of data using a problem solving application to develop their understanding of central tendency and fairness. Students will:

- Interpret data from a table
- Design systems that result in "fair" ranking systems
- Develop intuition for the significance of mean and median values of a data set

This problem allows students to explore concepts of weighted average, variance and the range of a set of data

Context
This problem could serve as an introduction to a unit on measures of central tendency (mean, median, mode, weighted mean, etc.) and variance. It provides students a context for seeing the importance of determining the central value and the variance of a data set. Students should be familiar with the concepts of average and have experience reading data from tables.

Presenting the Task
Explain to the class that in cross-country meets, several teams can compete at once and each team can have several runners. There are numerous ways that a winning team could be determined. There are some ways that are more fair than others.
This is a good activity to have students work in groups of two or three. For students that are having trouble getting started you might need to prompt them to:

- Think of a ranking system that would be completely unfair and why it is not fair
- Decide which team they think *should* win. Why did they choose this team?

After they have determined four systems in which a different team wins under each system, have each group share the system they felt was "best" and why they chose that system. Since cross-country is a sport that values participation and personal growth over individual excellence, ranking systems that take into account the placing of all the runners are generally considered more fair than those that heavily weight the top finishers. Probe their thinking to include a discussion of which systems motivate ALL team members to do their best.

Extensions
Students may want to find out how their cross-country team is actually scored. Do they use the same system for league meets and the state meet? How about track meets? A group of students could assist in the scoring of a meet.

Connections

Some groups may use the concept of average finish to determine ranking while others may use a weighted average system where proportionately more points are given for higher finishes. This provides an opportunity to discuss weighted average (for example in grading systems). Other methods of ranking may involve finding the median of each list or the most within a certain range. You could show how the rankings change when you create a frequency distribution from the data.

This problem has connections to many fields in the social sciences. For example, in economics a similar problem could be devised for comparing the "average" income of citizens from different countries. This could lead to a discussion on the effect of outliers on an average such as Bill Gates' salary.

Teacher's Guide for Design a Dartboard

Grade Level	Content Strand	Timing	Grouping	Materials
6-12	Geometry and Probability and Statistics	2 to 3 days	Teams of two or individual	Graph paper or Rulers

As some of you may know, the traditional dartboard is made up of concentric circles or squares. As a new twist on the traditional dartboard your company wants to make a dart board of some other shape. You are in charge of designing the board. Be sure to use a shape other than a square or circle. The board should have three major sections. The area of the board should be divided so the area has:

15% for the inner section
25% for the middle section
60% for the outer section

- Draw a design for your dartboard
- Show **all** your work using numbers, drawings, and words
- Explain the strategy you used to get your answer

Purpose

Students use their knowledge of the area of various polygons to explore the relations between linear dimensions and area.
Students will:

- Perform multiple calculations using geometric formulas
- Develop spatial sense and intuition about relative areas
- Develop probability concepts
- Practice drawing and measuring various geometric figures

This problem helps build a conceptual understanding of the relationship between linear dimensions and area as well as an increased understanding of probability. It provides a context for practicing computation and estimation strategies and in using different formulas for the areas of polygons.

Context

Prerequisites include the ability to calculate the area of various polygons and familiarity with the different polygons. This activity could provide a good introduction or extension to a study of the area of polygons or could be used in a unit on probability.

Presenting the Task

Introduce the problem with a discussion of how the point system of a dartboard is based on the probabilities of a dart landing in the different regions. It might be useful if you have a dartboard or two to use as visual aids in your discussion.

Let students know that you want them to be creative in their design of dartboards as well as accurate in their computations of the relative areas.

Discussion

After students have finished the task have them share their solutions and discuss their methods and findings. Look for diverse approaches to the problem. Since the vast majority of dartboards are circular, discussion could center on if this is justified or if there are there other shapes that would be just as good or better.

Extensions

You could have students discuss or research how they determine the probability of winning in other games (roulette, poker, the lottery, etc.). Another question to pursue would be, does the shape of the dartboard influence the probabilities or is it strictly a function of area.

Connections

This problem is related to other problems involving geometric construction given certain perimeter constraints where you need to find an optimal shape for the situation.

Teachers' Guide for

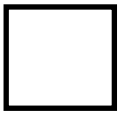
Kyle's Tiles

Grade Level	Content Strand	Timing	Grouping	Materials
4-8	Number Theory and Operations, Geometry, Money	Approximately 1 hour	Teams of two or individual	Optional graph paper

Here's something to think about…

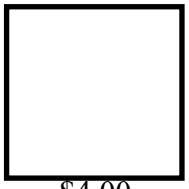
Kyle is planning to put decorative tile on a wall. The space is 28 inches wide and 30 inches high. The tiles come in these sizes and costs:

4"X4" squares



$2.00

6"X6" squares



$4.00

4" by 6" rectangles

$3.00

If Kyle needs to cut a tile, it costs $.50 for each cut.

* Design a layout for the tiles.
* Explain why you thought it was "best".
* Find the total cost for your design and show all your work.

Purpose

Students use division, factors, or multiples to solve an area problem and determine associated costs. Students will:

- Find common multiples or use repeated addition to fill a space in two dimensions
- Use spatial reasoning to divide a given area
- Explore how dimensional changes affect area
- Add decimals by calculating costs based on area designs
- Consider criteria for making a design choice

This problem gives students an opportunity to develop and build on the concepts of multiplication, division, area, and money.

Context

This task fits well into a unit on finding common multiples or factors, as well as work with area. Students will explore thinking about measurement in two dimensions. Prerequisites include a conceptual understanding of area, ability to use repeated addition or multiplication and division, and facility with adding decimals in money.

Presenting the Task

Pose the problem to students as a design challenge. You may want to mention that their solutions do not need to include space for grout between tiles (as they would in an actual tile problem). As students experiment with tile layouts, it might be helpful for some students to have graph paper available to visualize their designs.

As students begin writing up their solutions remind them to include a:

- diagram of their design including tile sizes
- explanation of why they think this is the "best" solution
- materials list of tiles needed and costs
- cost calculation of additional cuts
- total cost for their design

In your group discussion of the task, provide students with the opportunity to share what they thought would be the "best" layout. Students will have varying interpretations of which layout is "best". Encourage them to accept their peers' interpretation as long as they used their criteria for "best" in finding the solution. Discuss their strategies for arriving at a layout (how did they decide which tiles to use). Analyze cost as a factor. Some students will have considered cost in their "best" design, what impact did that have on the layout? Determine which tiles give you the most tiles per dollar. This may lead back into a discussion of aesthetically pleasing designs as compared with cost-efficient designs. Encourage students to think about what would have made the task easier numerically.

Extensions

- Challenge students to figure out the least expensive tile pattern (if they haven't already).
- Challenge students to find other size areas that the same pattern would fit into.
- What would happen if the dimensions of the tiles were twice as big?
- You may want to extend this project by having students show their solutions with color "concept drawings" of their designs.
- A bulletin board section or hallway could be used to show actual size designs.
- Use the sample of the class' design costs to determine a fair average cost per square inch for a tile company.

Northwest Regional Educational Laboratory
Mathematics and Science Education Center

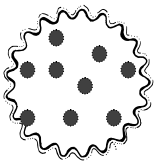
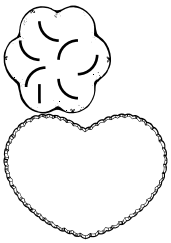
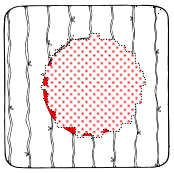
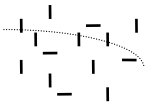
Connections

This task has numerous applications in the "real world". You may have students generate a list of situations where a comparable problem could emerge. Problems related to flooring, ceiling tiles, siding, windows, building blocks, and fabric design will come up. Of course there are innumerable examples of cases where "best" needs to be considered in the context of for whom it would be best.

Teachers' Guide for
The Mathemagicians' Cookies

Grade Level	Content Strand	Time Required	Grouping	Materials
5-7	Patterns and Algebra	Approximately 1 hour	Teams of two or individual	None

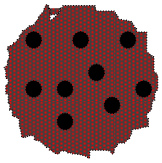
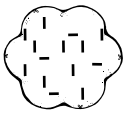
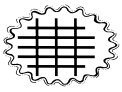
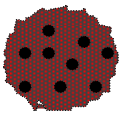
Here's something to think about…

The Mathemagicians are having a bake sale to raise $50 for some awesome math materials. They all got together and made 360 cookies. It cost them $.10 per cookie to make them.

Now they are ready to package and price the cookies. They want to sell them in three different sized packages (each size package has a different number of cookies).

- How could the cookies be packaged?
- How much should they charge for each size package?
- If they sell all the packaged cookies how much profit will the Mathemagicians make?
- Show your work using pictures, tables and numbers.
- Explain why you picked the size of packages you did?
- Explain why you priced them the way you did?

Purpose: Students use a concrete situation to investigate number patterns and find variables to solve the problem. It provides an opportunity for students to develop a mathematical expression to represent their solution. Students will:

- Gain experience in solving a multi-step problem and organizing information

- Use factoring, division, or repeated addition to find missing values
- Use calculators to find and check solutions
- Use tables and/or equations to communicate their solutions
- Determine production costs and set prices associated with variables
- Calculate profit based on their solutions
- Determine reasonableness of their pricing strategies

Context

Students should be able to perform basic operations with whole numbers and decimals (money). This task requires students to use patterns and functions to model the problem, or develop mathematical rules and algorithms to solve the problem algebraically. The problem lends itself to intensive calculator use.

Presenting the Task

As an introduction to this task you may want to discuss pricing considerations with your students. They will be familiar with the concept but may not have thought about strategies for packaging, setting prices, and profit margins. You could use the example of a lemonade stand to provide a familiar context. Encourage them to consider issues such as:

- Target customers -- which would affect the size of cups sold (or in the case of the cookies, the number of cookies in the package).
- Covering costs -- which means they need to make at least as much money as they spent to make the lemonade.
- Competitive pricing -- recognizing that if prices are set too high consumers won't buy the product, as well as developing a consistent pricing strategy. Should an 8 oz. cup be twice as expensive as a 4 oz. cup, or slightly less than twice – making it a better value for customers but less profit for the seller?

Remind students that there are many reasonable solutions to this problem, emphasizing the importance of consistent reasoning about the task, organization of their solution and communicating their solution clearly and convincingly.

In your follow-up discussion share the spectrum of solutions from the class. Students will benefit greatly from not only seeing the range of possible solutions but the variations in how students represented and communicated their answers. Students may have used graphic representations, tables, organized lists, or equations to show their solutions. Talk about strategies students used to solve the problem. Ask students:

- How they calculated the number of each package needed to make 360 cookies
- What strategy they used for pricing (the unit price of the cookies has to be greater than $.10/cookie to cover costs)
- How they calculated their profit

Extensions

- Model how a change in one quantity can result in a change in another such as an increase in costs
- Compute the percent, fraction, and/or decimal for each package size, out of all the cookies
- Have students challenge each other by giving clues and having the class find their solution. For example, a student might say, "I packaged the cookies in bags of 6, 12, and 24. Before they were packaged, there was an equal number of cookies in each group. How many bags of each size did I need?"

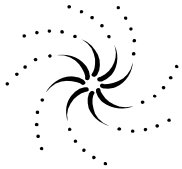
Connections

This task connects to competitive pricing and marketing. Students could find advertisements for identical items at various stores and compare prices or calculate the difference in mark-up. If you have a convenience store or parent employed in the retail business, you might want to invite them in to share how items are priced at their store, and how they decide which items to order and the quantity to order.

Teacher's Guide for Spinner Elimination

Grade Level	Content Strand	Time Required	Grouping	Materials
5-8	Probability Number Theory and Operations	Approximately 2-3 days	Teams of two	Spinners made from cardboard and paper clips

Try to design a spinner that will let you cross out more squares than your opponent on a 50's chart. (See rules for the Spinner Elimination Game.) After you finish a game, decide if you want to change the numbers on your spinner. After you have played several games, answer the following questions:



3. Would you change the numbers on your spinner? Why or why not?
4. What advice would you give to someone who wants to win the game for strategies on how to choose numbers for their spinner?

Purpose

Games are an excellent way for students to practice and reinforce concepts as well as to discover new relationships and insights. Spinner Elimination helps students explore prime and composite numbers as well as to experience the effects of probabilities.

Context

This activity can be done as part of a unit on probability and/or when studying prime and composite numbers and prime factorization. Students should be fairly competent in multiplying numbers up to 50.

Presenting the Activity

Using an overhead spinner or a large sample spinner, model playing the game with one of your students. Talk about the fact the each of the eight sections on the spinner is equally likely. Think aloud as you are choosing your numbers and playing the game or have students help you choose numbers to put on your spinner and ask them why they chose those numbers.

When students have played the game a number of times conduct a discussion with them about what they have learned. Questions to ask include the following: Which numbers did they use and why? Did anyone try one or zero? How did that work out? What changes did they make to their spinners as they learned more about what worked well? Is there a perfect choice for the numbers on a spinner? Why or why not?

Extensions

When students have had plenty of practice using a 50's chart have them try playing the game with a 100's chart and analyze how that changes their decision about what numbers to use on their spinner.

Connections

You might want to follow up this game with the Factor Game, another two-player game involving prime and composite numbers and using a 50's chart. The first player chooses a number and the second player circles all the factors of that number. The first player gets a point for each factor circled. Then the second player circles a number and the first player circles all of its factors in a different color. The play continues until only prime numbers remain.