Using Formative Assessment to Drive Learning
The Silicon Valley Mathematics Initiative: A Twelve-year Research and Development Project

By David Foster and Audrey Poppers
November 2009

Introduction

Ironically, American public education seems unable to learn and improve. Classroom instruction has remained virtually unchanged for decades, despite endless cycles of reform and a growing body of educational research (Weiss, Pasley, Smith, Banilower & Heck, 2003). This lack of progress is attributable to the current structure and culture of American education, which does not support rigorous practice. As each innovation gains widespread attention, a wave of superficial implementation efforts sweep across the educational community. Without the support required to do rigorous new work, the attempted innovation is stripped down to its simplest and most familiar elements; in the process, the most challenging elements, which are also necessary to the efficacy of the practice, are simply ignored (Fullan, 2007; Schmoker, 2006). Formative assessment is currently moving toward center stage on the national scene; and not surprisingly, it appears that most formative assessment efforts lack attention to the rigorous elements that are critical to potential effectiveness.

A growing number of researchers and educational leaders make a compelling case for the promise of formative assessment. Over the past twenty years educational research has pointed to the value of linking instruction to assessment (Marzano & Haystead, 2008; Reeves, 2004; Wiggins & McTighe, 1998), examining student work to inform instruction (Dufor, Dufor & Eaker, 2008; Schmoker, 2006), and using formative assessment practices to drive learning (Black, Harrison, Lee, Marshall & William, 2004; Black & Wiliam, 1998; Reeves, 2007; Thompson &Wiliam 2007). Black and Wiliam's review of the research cites compelling data to indicate that formative assessment focused on student thinking can inform future instruction and learning; however, there is little evidence that the analysis of student thinking is used to drive instruction in the typical mathematics classroom in the United States (Weiss et. al., 2003). Textbooks, pacing guides, state tests, and courses of study govern the topics that are taught, the time spent on a topic and the depth of what is taught. These practices continue despite the results of international, national, state and college entrance tests which indicate that American high school students have not learned adequate mathematics to support their own futures and career opportunities, or to support the future success of America in global competitive markets (Trends in International Mathematics and Science Study, 2007; Program for International Student Assessment, 2006; Friedman, 2005).

Definition of formative assessment

The information that is most valuable for teaching must focus on student thinking. Dylan Wiliam states that, "The central idea of formative assessment, or assessment for learning, is that evidence of student learning is used to adjust instruction to better meet student learning needs." He describes formative assessment practice as students and teachers using evidence of learning to adapt teaching and learning to meet immediate learning needs, minute-to-minute and day-by-day (Thompson & Wiliam, 2007). Most teachers don't actively use these practices. In fairness, very few teachers are trained to use formative assessment and have no apprenticeship implementing its use in classrooms.

Current state of assessment in the United States

Often educators and school districts claim to use formative assessments. Testing in schools is prevalent, yet most of the testing is conducted to produce grades, place or sort students, or to gauge or predict how students may score on a high-stakes exam – all summative approaches. Seldom do the findings from these summative assessments actually drive learning by informing teachers' instruction.

During this past decade of high-stakes accountability, states have developed standardized tests and schools have been held accountable for how their students have performed. The school's stakeholders have become very interested in the results and the improvement of student scores and achievement on the exams. As an example, benchmark tests are often used to predict how students might perform on a summative high-stakes exam. A significant cottage industry has grown up around these instruments. Some educators regard benchmark tests as formative assessments, but they don't match the true definition and purpose of formative assessment. Instead benchmarks are mini-summative tests that may accurately predict future success on another test, but do not provide specific information to guide instruction. The reports produced by benchmark tests usually provide a scale value on a continuum of scores and may list topics where students were successful and unsuccessful. This information is insufficient to truly help teachers understand what students know, what errors or misconceptions students are having, what approaches in reasoning were successful and where students met challenges and struggled – information essential to the successful design of future lessons.

Educational practice has a long history of substituting superficial practices for more rigorous approaches, thus rendering them ineffective. The subsequent lack of success is then regarded as proof that the recommended practice is ineffective. Continued use of the these mini-summative assessments and the mislabeling of them as formative assessments threatens to undermine the effective use of formative assessments and their potential for promoting improved instruction and student achievement.

There are multiple reasons for the superficial attempts to implement formative assessment in the United States, including beliefs about the nature of teaching and learning, the current structure of schooling, the lack of in-depth professional preparation for teaching, and the lack of sufficient numbers of teachers with a deep understanding of mathematics. In addition, there are few if any good models of formative assessment from which teachers can learn in widespread use in the U.S.

Silicon Valley Mathematics Initiative

In 1996, The Noyce Foundation created the Silicon Valley Mathematics Initiative (SVMI), a comprehensive effort to improve mathematics instruction and student learning.

Formed in partnership with the Santa Clara Valley Mathematics Project (California Subject Matter Project) at San Jose State University and 35+ member school districts, SVMI's primary components are formative and summative assessment systems, pedagogical content coaching, ongoing professional development and leadership training. The math performance assessment system goes by the name of Mathematics Assessment Collaborative (MAC). MAC contracts with the Mathematics Assessment Resource Service (MARS), a NSF funded international project that involves UC Berkeley and the Shell Centre in Nottingham, England to develop the performance exams and scoring materials.

It is a fundamental belief of SVMI that quality math performance assessments, coupled with effective professional development for classroom teachers and leaders, can support improved instruction and student achievement. To that end, SVMI has been engaged in a twelve-year research and development effort to create formative assessment practices and tools that support significantly improved teaching and learning.

In the fall of 2006, the Noyce Foundation created the First in Mathematics Collaborative (FiMC) in partnership with nine (9) SVMI member districts in order to test those SVMI tools and procedures in a more controlled approach. The FiMC collaborative pursued similar goals and strategies as SVMI, but incorporated a more intensive focus on supporting a structured implementation of formative assessment practices. The remainder of this article is devoted to a description of the SVMI development process and tools, followed by a description of outcomes achieved in FiMC.

FiMC theory of action

Over time, we have found that a majority of teachers lack an in-depth understanding of mathematical concepts and effective strategies for instruction. Without these in-depth understandings, it is impossible to design instructional experiences that drive significant student achievement. The FiMC theory of action was based on the premise that teachers can improve their instructional effectiveness by using a cycle of formative assessment practice. As they examine the student thinking revealed in the assessments and consider each student's current knowledge and misconceptions, teachers also clarify and strengthen their own understanding of mathematical concepts.

During the SVMI development work, we learned that effective formative assessment is more complicated than implementation of a system to assess student learning; we found that teachers needed considerable support in order to utilize the information revealed in the assessment process. As a result, the FiMC initiative was constructed to support two parallel lines of thinking with a complementary focus on both teacher learning and student learning. The FiMC process aimed to engage teachers in an integrated effort to deepen their understanding of the mathematics involved in a given concept and their students' thinking relative to this concept - including both accurate knowledge and misconceptions, and to use that information to design new instructional approaches that target students' current needs.

We agree with Dylan Wiliam that the most effective use of formative assessment practices occurs on a daily basis, supporting the teacher's minute-by-minute decision-making; however, very few teachers are adequately prepared to successfully implement that level of practice. In FiMC we intended to use tools and procedures to provide a structured approach in order to scaffold the learning for teachers.

This is not a quick intervention. In order to be successful and motivated to sustain their efforts, teachers need informed and skillful guidance until they attain some level of proficiency; however, as teachers begin to experience results in the form of deeper understanding for themselves and that of their students, they find the process to be highly engaging and generative. FiMC has gathered evidence that the process produces significant results. Educational research indicates that the most significant variable in student learning is the teacher (Boaler, 1998; Sanders, Horn, 1994; Schmidt, McKnight, Valverde, Houang & Wiley, 1997; Wright, Horn & Sanders, 1997). We believe that investments in rigorous teacher learning - structured to support application in the context of one's own classroom – can promote significant student learning.

The SVMI formative assessment cycle

The SVMI cycle mirrors other conceptions of formative assessment to be found in the literature; the SVMI contribution lies in the development of tools and procedures to support teacher learning and implementation. The general cycle involves four stages:

1. Selecting and administering a worthwhile assessment task
2. Examining and analyzing student work
3. Using the findings to inform and enhance teacher knowledge
4. Designing and teaching lesson(s) to address the learning needs of students.

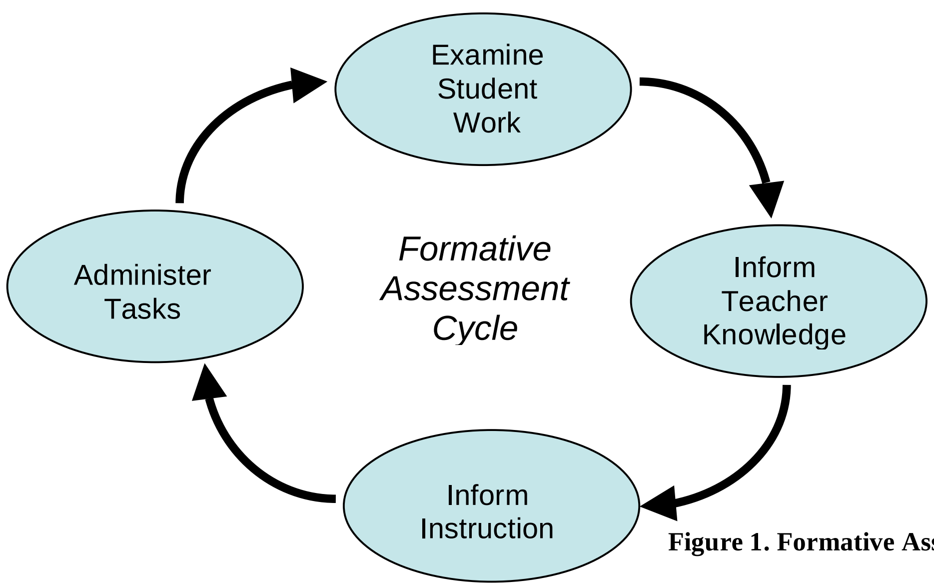


Figure 1. Formative Assessment Cycle

The SVMI development process

Initially we were mostly interested in finding worthwhile assessment tasks. We found the MARS tasks (Mathematics Assessment Resource Service). to be strong in assessing math concepts, relatively easy to use, and powerful in providing valuable feedback about student thinking and understanding. Our initial experiences with the tasks were not intentionally directed toward formative learning, but rather the use of performance assessments instead of traditional U.S. tests in order to support student achievement. We assumed that after teachers scored the student work and considered overall growth from one task to the next, they would be able to interpret the results and use them to design better instruction; however; using the assessment information to inform instruction proved to be very difficult for many teachers. In a later section, we will describe the tools and procedures developed to support instructional design.

As we began to analyze and report our growth findings from these early assessments in SVMI, we saw trends in student performance and realized that we could use these data to develop tools that would scaffold teachers' ability to understand and interpret their students' responses.

During the first year we looked carefully at a few tasks and documented the trends in student performance in the following ways: Common errors in students' answers and their problem solving processes; common methods that produced successful solutions; unique, interesting and/or unusual solutions. We were so pleased by the more thorough analysis of the findings, that we began to document the trends of student performance in all the tasks that were administered. This analysis process led to development of tools to support two major aspects of the cycle: 1) processes and tools for closely examining and charting student work in order to catalog and characterize performance; 2) general findings from the student work related to each particular task.

As we worked with teachers on each phase of the cycle, we learned that they needed significantly more support than we initially had assumed. Their responses and struggles provided formative information for our SVMI team, leading us to develop tools and procedures to scaffold teachers' understanding as they worked to develop both their formative assessment skills and their deeper understanding of mathematical ideas. Specifically, teachers needed support in all the following areas:

- selecting worthwhile tasks that were related to the concepts being addressed;
- analyzing and understanding students' responses to the assessments;
- reflecting on the analysis and revising one's mathematical and pedagogical thinking;
- designing instruction to address the learning needs identified in the analysis.

We knew this analytical information and the processes were invaluable to teachers, so we created an instrument called *Tools for Teachers*, often referred to as the toolkit. The toolkit contains data describing generalized performance of the students, as well as examples of student work with both well conceived and well communicated solutions and examples of common errors or invalid reasoning that resulted in unsuccessful attempts at solving the problems. The toolkit also features unique or unusual approaches that show interesting mathematical thinking. We developed commentary for each piece of student work – pointing out important features of the work and why it was selected. In addition to student work, the *Tools for Teachers* contains graphs illustrating the overall performance of students and explanations of the distribution of the data across the continuum of performances. Over the years, the *Tools for Teachers* document became more sophisticated and rich with information. The section on implications for instruction grew, providing numerous instructional suggestions and curricular resources.

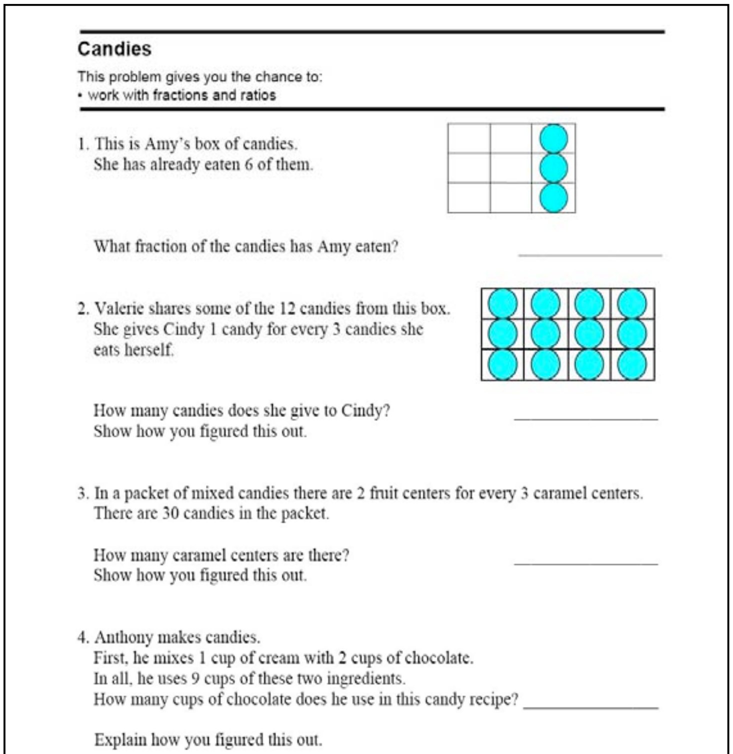
SVMI Tools and Procedures

A worthwhile task

The databank of MARS assessment tasks is a central feature in this system of support; it includes an extensive range of concepts and levels, covering the most important mathematical ideas from second grade through tenth grade**.** This rich resource of tasks was crucial to scaffolding the process for teachers. Good formative assessment tasks are difficult to find, and very difficult for the average teacher to create.

A worthwhile assessment task provides instructionally useful information for the teacher. The task must address the core mathematics and also capture the potential range of student performances. A task provides no formative value if students are totally unsuccessful with the mathematics - equivalent to receiving a zero score. A well-designed task provides access so nearly all students can demonstrate some level of success.

The MARS tasks are designed to assess across this range (Figure 2). This incremental design is represented by the ramp in Figure 3. The task gradually increases in



Candies

This problem gives you the chance to:
• work with fractions and ratios

1. This is Amy's box of candies.
   She has already eaten 6 of them.

   What fraction of the candies has Amy eaten? ______________

2. Valerie shares some of the 12 candies from this box.
   She gives Cindy 1 candy for every 3 candies she eats herself.

   How many candies does she give to Cindy? ______________
   Show how you figured this out.

3. In a packet of mixed candies there are 2 fruit centers for every 3 caramel centers.
   There are 30 candies in the packet.

   How many caramel centers are there? ______________
   Show how you figured this out.

4. Anthony makes candies.
   First, he mixes 1 cup of cream with 2 cups of chocolate.
   In all, he uses 9 cups of these two ingredients.
   How many cups of chocolate does he use in this candy recipe? ______________

   Explain how you figured this out.

Figure 2. MARS Task

"Candies" is a fifth grade task. It assesses students' knowledge of fractions, ratios and proportional reasoning. The first part of the task provides **access** for students. The prompt includes a visual model and asks the students to name the basic fraction of the candies eaten. One-third or an equivalent is acceptable. In parts two and three the **core** mathematics is assessed. Students must demonstrate knowledge of ratios and proportional reasoning. The fourth part is the **elaborated** level for fifth grade students. The students must solve a multi-step problem involving a proportional relationship and a sum of ingredients.

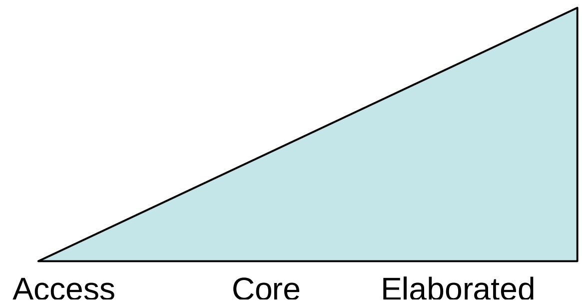


Figure 3. Design of a MARS Task

complexity and cognitive challenge as the core mathematics is assessed. A well-designed task goes conceptually deeper than the core math, probes for deeper understanding and requires explanations in order to provide evidence into students' thinking and knowledge. The final stage of the task requires an elaborated level of understanding of the core math; at this level the task may challenge students to demonstrate strong conceptual understanding, generalize a situation, justify a finding or make a connection. Success at the elaborated level is the best indicator of future success on related mathematical ideas.

Selecting a worthwhile task

Initially, teachers may need support in order to select an appropriate task. The MARS Task Anticipation Sheet was developed to assist teachers with this process (Figure 4). It is essential for the teacher to anticipate how students could be successful and to consider multiple ways students might approach the problem. The teacher should focus on what mathematical ideas and thinking students will encounter and what is required to mathematize the situation and engage in solving the problems. The teacher should also anticipate at what points students might struggle.

Teachers may work individually or in small groups to select a worthwhile task by carefully considering which task targets the mathematics and student thinking processes that the teacher wants to address. A promising method for considering a potential task begins with the teacher actually doing the task as a learner would. In this manner the teacher is able not only to examine the core mathematical ideas embedded in the task, but also to develop an appreciation and understanding of the mathematical thinking the students must engage in and the complexity of the steps and strategies students will encounter.

MARS Task Anticipation Sheet

Task Name:________________ Grade:_____ Year:_____ Tot Pts. _____ Core Pts._____

In anticipating the student work where will students show success?

What parts of the task will students be successful?	In terms of knowing and doing mathematics what does this indicate?

In anticipating the student work where will students struggle?

What parts of the task will students be unsuccessful?	In terms of knowing and doing mathematics what does this indicate? What understandings or skills do the students need to learn?

Considering strengths and weaknesses from students, what are plans for future teaching?

What are the implications for future instruction?	What specific instruction or lesson experiences will you design for students?

The MARS Task Anticipation Sheet is divided into three sections: areas where students are likely to be successful; areas where they are likely to struggle; and the proposed plan for teaching. Considering these issues, the teacher selects a worthwhile task that matches the learning trajectory of her students. These answers frame the mathematical ideas and process to be learned and assessed, and focus attention upon evidence of student thinking that will appear in their work.

Figure 4. MARS Tasks Anticipation Sheet

Analyzing student work

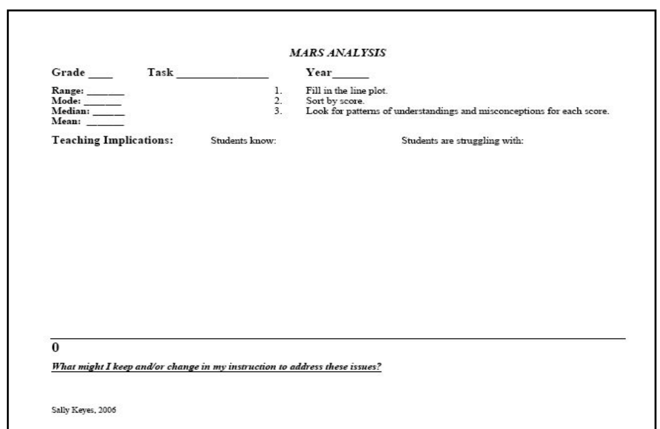
After the tasks are administered, the teacher possesses a gold mine of information that can drive significant learning for teachers as well as students; however, mining the student work is not simple. Examining student work to determine trends, identify misconceptions, categorizing successful strategies and extracting important findings is a learned art. Unfortunately teachers are provided little instruction or guided experience in analyzing student work. Those who have become accomplished at this important investigative work are self-taught - usually driven by passion for understanding how their students are thinking. In SVMI we believe this work is essential to becoming an accomplished and effective teacher.

In SVMI we created tools and protocols to help teachers analyze the student work. The analysis process usually begins with scoring of the student work. Collective scoring can assist teachers to learn the mathematical content, to recognize successful strategies and to develop a critical eye for examining student work. This process can initiate a transformational shift from traditional teacher scoring practice – that of merely correcting mistakes or assigning a grade to a student's work.

The MARS tasks come with a rubric and standardizing papers. Teachers learn to use an analytical rubric to score student work (Figure 5). The standardizing process guides them to match evidence in their students' papers to points on the rubric, calibrating their judgments in order to score reliably. In the scoring process, teachers start to recognize general trends in student responses, including both successful and unsuccessful approaches. It is helpful for teachers to see the variety of successful approaches, as well as common misconceptions and errors in unsuccessful efforts. These patterns are useful in characterizing performance. Uncommon but valid approaches can make scoring difficult; however, considering these solutions deepens the teacher's understanding of the mathematics and the range of student thinking. Actual scoring may not be necessary if teachers are sophisticated in examining and learning from student work. Instead the

Task 1: Candies	Rubric	
The core elements of performance required by this task are: • work with fractions and ratios Based on these, credit for specific aspects of performance should be assigned as follows	points	section points
1. Gives correct answer: **2/3** or **6/9**	1	1
2. Gives correct answer: **3** Shows work such as: 1 + 3 = 4 12 ÷ 4 = Accept diagrams.	1 1	2
3 Gives correct answer: **18** Shows work such as: 2 + 3 = 5 30 ÷ 5 = 6 6 x 3 = Accept diagrams.	2 1	3
4. Gives correct answer: **6** Gives a correct explanation such as: Anthony mixes a ratio of one cup of cream to two cups of chocolate. The ratio stays the same for different amounts. So I wrote the numbers in a chart like this 1 to 2 = a total of 3 2 to 4 = a total of 6 3 to 6 = a total of 9 Accept diagrams.	1 1	2
Total Points		**8**

Figure 5. MARS Rubric


MARS ANALYSIS

Grade ____ Task ______________ Year______

Range: ______
Mode: ______
Median: ______
Mean: ______

1. Fill in the line plot.
2. Sort by score.
3. Look for patterns of understandings and misconceptions for each score.

Teaching Implications: Students know: Students are struggling with:

0

What might I keep and/or change in my instruction to address these issues?

Sally Keyes, 2006

The SVMI Score Analyze is used to examine students' scores statistically, finding the measures of central tendency and examining the visual distribution of scores using a line plot. From these analyses, the teacher can arrange the papers by performance and look for patterns of understandings and misconceptions related to each score.

Figure 6. SVMI Score Analyzer

student work can be examined holistically after teachers have done the task themselves and surfaced the key mathematics.

After the papers are scored, tools such as the SVMI Score Analyzer assist teachers to identify trends in the students' performances (Figure 6).

The analysis process is often lead by a math coach or mentor until teachers have gained sufficient skills at analyzing and learning from student work. The MARS Task Analysis Sheet (Figure 7), similar in format to the Anticipation Sheet, provides a frame for tracking the student evidence to the mathematics. Listing important evidence and findings from student work is the first step in enhancing the teacher's ability to address the learning needs of the students.

MARS Task Analysis Sheet

Task Name:________________ Grade:__ Year:____ Tot Pts. ___ Core Pts.___ Core____%

In analyzing the student work where did students show success?

What parts of the task did students demonstrate success?	In terms of knowing and doing mathematics what does this indicate?

In analyzing the student work where did students struggle?

What parts of the task were students being unsuccessful?	In terms of knowing and doing mathematics what does this indicate? What understandings or skills do the students need to learn?

Considering strengths and weaknesses from students, what are plans for future teaching?

What are the implications for future instruction?	What specific instruction or lesson experiences will you design for students?

Figure 7. MARS Task Analysis

Informing teacher knowledge

In addition to the analytical tools presented above, we created a resource document called *Tools for Teachers*, a rich catalogue of information related to student performance on each task, gleaned from multiple years of data gathering. Developed as a resource to assist teachers in gaining mathematical and pedagogical knowledge, the toolkit is intended for hand-in-hand use with the teacher's own classroom papers to identify next learning steps and to address students' learning needs.

Toolkit excerpts are shown in Figures 8 – 12. The Graph & Analysis of MARS Task Data (Figure 8) was originally developed to capture the richness of the responses generated from a given task.

MARS Test Task 1 Frequency Distribution and Bar Graph, Grade 5

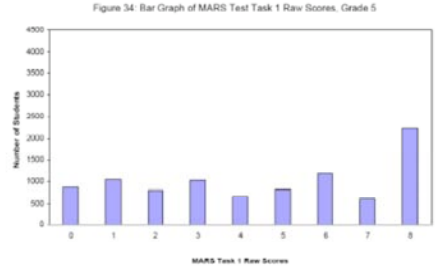
Task 1 - Candies

Mean: 4.50 StdDev: 2.79

Table 25: Frequency Distribution of MARS Test Task 1, Grade 5

Task 1 Scores	Student Count	% at or below	% at or above
0	883	9.5%	100.0%
1	1058	20.9%	90.5%
2	790	29.5%	79.1%
3	1039	40.7%	70.5%
4	661	47.8%	59.3%
5	819	56.7%	52.2%
6	1172	69.3%	43.3%
7	615	75.9%	30.7%
8	2230	100.0%	24.1%

Figure 34: Bar Graph of MARS Test Task 1 Raw Scores, Grade 5



The maximum score available for this task is 8 points.
The minimum score needed for a level 3 response, meeting standards, is 4 points.

Most students, 90%, could identify the fraction from a model. Many students, 70%, could identify the fraction of candy eaten in part 1 and find the amount of chocolate needed to extend a recipe from 3 cups to 9 cups. More than half the students,52%, could also find the amount of candies Cindy would receive in part 2 and explain how to figure it out. 24% of the students could meet all the demands of the task, including finding a scale factor and changing a ratio from 2:3 to 12:18, when the total candy would be 30 pieces. Almost 10% of the students scored no points on this task. 93% of the students with this score attempted the task.

Figure 8. Graph & Analysis of MARS Task Data

In the initial stage of development the toolkit was merely a report of findings from a significant sample set of students. As we became more sophisticated in using the tool, we adapted it to be used interactively with actual student work, allowing teachers to compare their current student work against the performance of the large normative set. By charting their own student work to see trends and patterns of performance, teachers can identify their students' strengths and varied approaches (Figure 9). They may also compare their students' errors and struggles to commonly identified misconceptions and challenges that appear in the larger sample. A teacher's own student work may be compared to other student responses that illustrate typical work as well as unique or productive approaches and solutions.

Candies

Work the task and look at the rubric. What strategies do you think students might use to help them solve the task?

____

____

____

What are the big mathematical ideas being assessed?

____

____

In this task students need to think about fractions: part/whole relationships and ratios: part/part relationships. Look at student work in part 1. How many of your students put:

2/3 or 6/9	1/3 or 2/6	3/9	3/6 or 1/2	Whole number	Other

What does this show you about student understanding? Are students frequently asked to think about all the ideas in a model? (What fraction is shaded? What fraction is not shaded? What fraction is represented by all the parts?)

Next students are asked to think about a situation of distributing items in a ratio. How many of your students put:

3	4	9	1/4	6	4/12	1	12	Other

What is each student probably thinking? What are they confused about? How can these misconceptions be confronted? What task or problem might you pose for class discussion to help students clarify their thinking?

In part 3, students needed to think about how many candies made up a group, find the number of groups in the whole, and then use that scale factor to find the number of caramel candies. Now look at student work on part 3. Make a note of the types of models or strategies that students used and the answers they came up with. How many of your students put:

18	5	10	5	3	4	90	16/17	Other

What confused them?
What models did they use?

The *Tools for Teachers* contains a section where teachers can make sense of the results from their own class papers. Teachers use their class sets of student work to chart and record their students' approaches and solutions to the range of performances that were identified in the normative sample. The normative sample was derived from review of student papers from an original set of 8,000 – 10,000 students from over 30 school districts. Teachers categorize the student responses to see trends in the overall performance of their students in terms of common approaches, errors and misconceptions.

Figure 9. Questions for Teacher Reflection

The *Tools for Teachers* toolkit contains a large set of student work. The work displayed shows examples of well conceived and communicated solutions (Figure 10), examples of common errors or invalid reasoning that resulted in unsuccessful attempts at solving the problems (Figure 11), and unique or unusual approaches that show interesting mathematical thinking. These examples help teachers learn different methods that successful students use in problem solving. Often these solution paths are based on strong conceptual foundations and show sound logic and reasoning. Encouraging multiple solution strategies is important in class, and the toolkit can be a rich resource for ideas and approaches.

Looking at Student Work on Candies

Student A uses three different models to think about the ratios. In part 2 the student uses a dealing out strategy; 1,2,3 for Valerie then 1st for Cindy, etc. In part 3 the student uses a table, which probably uses groups of caramels and groups of fruit centers but also shows equivalent ratios. In the part 4 the student uses a scale factor of three to solve the problem.

Student A

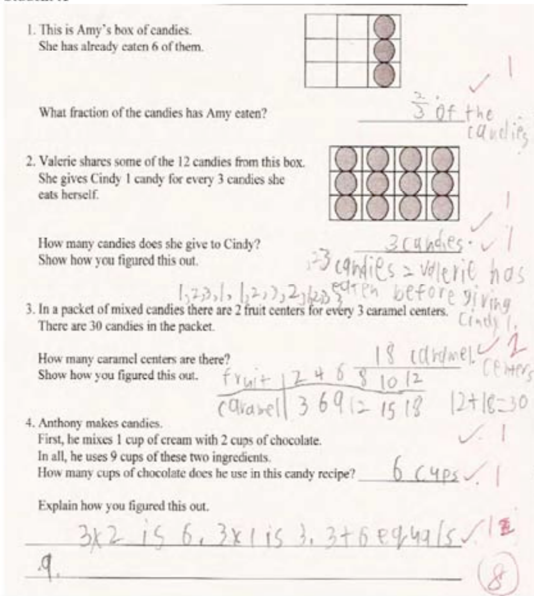


Figure 10. Student Work and Commentary

Students had difficulty understanding what to do with the scale factor. Student F is able to think about the part/part/whole relationship in 2 and 3 (3+1 makes a group of 4, 3+2 makes a group of 5). The student solves part 2 by counting the number of groups of 4, which is the same as the number of candies. The student then counts how many groups of 6 there are in 30. *Why does this strategy no longer work? What is different mathematically in part 3?*

Student F

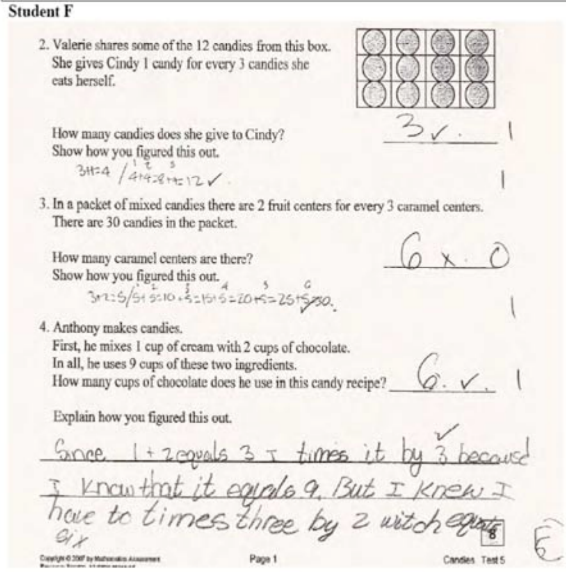


Figure 11. Student Work and Commentary

An important feature in the *Tools for Teachers* is a chart for each task,equating particular point scores with the common understandings and misunderstandings of students who attained the score (Figure 12). The matrix makes explicit the areas of student need.

The components of the *Tools for Teachers* provide a significant research resource for teachers. Few assessment reports are actually written with the teacher as the primary audience. This toolkit provides teachers with important information to enhance their mathematical and pedagogical content knowledge.

We assumed that this information would support and motivate teachers to target their classroom instruction to meet the needs of their current students.

Candies

Points	Understandings	Misunderstandings
0	93% of the students attempted the task.	Students confused the model in part one for what was in the box instead of what was missing. 14.5% put either 1/3 or 3/9 instead of 2/3. 8% made the model into a ratio, giving 3/6 or 1/2.
1	Students could name the fraction represented by the missing part in a model.	Students had trouble finding the amount of chocolate needed for a recipe. 8.5% of the students multiplied 2 x 9 to get 18. 5% confused scale factor, 3, with the amount of chocolate needed. Other common answers were 4,9, 2 and 5.
3	Students could name a fraction and use a ratio to find the amount of chocolate needed to make 9 total cups of a recipe.	Students had trouble thinking about a ratio for dealing out candies in a box of 12. 18% simply divided one of the numbers of the ratio into 12 to get 4. Other common answers were 2,9, 15 and 4/12.
4	Students could name a fraction, extend a ratio out a small distance, and find the number of candies for Cindy in part 2.	Students had difficulty explaining their thinking in part 2.
5		Students missed all of part 3, extending the ratio 2:3 out to 30 total candies. 8.5% could find the scale factor of 6, but didn't know what to do next. 7% thought about a ratio of 2 fruits and 1 caramel to make a total of groups of 3 instead of 2:3 giving a total of groups of 5. Other common answers were 5 and 3.
6		Students with this score missed either all of 2 or all of 4.
8	Students could identify a fraction from a model, apply a ratio to a model to find the number of candies Cindy would receive, find a scale factor to find the increased amount of a given ingredient.	

Figure 12. Summary of Students' Understandings and Misunderstandings

Over time, we realized that not enough teachers were using the information to address next steps in their classrooms. This realization led us to the next phase of tool development.

Informing instruction

The final phase of the cycle – supporting teachers to create follow-up learning experiences - took time to develop. Originally we were content to help teachers access important information, believing they would know how to use this information to improve learning. In reality most teachers did not know how or did not take the time to use the information to create a positive learning experience for their current students. In some instances teachers did learn about how students performed and tried to adjust instruction in subsequent years to prevent common errors, confusion or misconceptions, but there was little evidence of concerted efforts to use the information to improve the learning of the current students. When teachers were uncomfortable with what their class had learned, some would re-teach a lesson; but the re-teaching was usually a review of the original lesson or one very similar to the original. Instead of using the findings to define what specific experiences would promote further student learning, teachers regarded the findings as prompts to go back and do the lesson again with some minor changes in emphasis.

Re-engagement lessons

We realized that even when teachers were reteaching concepts, they often were not appreciating the need to engage their students in thinking about the concepts in a new way – the need to *re-engage* them differently in the mathematical ideas. In order to distinguish this type of lesson design from more traditional reteaching or review, we began to talk with teachers about this idea of re-engagement and to develop tools to support the practice of designing lessons that were directly tied to the results of formative assessments.

Re-engagement lessons are constructed to re-engage students in the core mathematical ideas of the assessment task in order to deepen their understanding of the core math and build a better conceptual foundation to learn further mathematics. The follow-up or re-engagement lessons featured in *Tools for Teachers* suggest ideas for designing lessons using the formative findings. It is common for test results in a class to range from students indicating little success to those students who successfully complete the task. A well-designed re-engagement lesson addresses students' learning needs across this continuum.

Structure of Re-engagement Lessons

Although there are no formal templates for constructing re-engagement lessons, certain principles might be considered. The lesson should address the learning needs of the entire class by:

- providing access into the task;
- solidifying foundational math concepts;
- addressing the core mathematics;
- surfacing errors and misconceptions in the students' work;
- using student work examples and thinking for others to critique;
- promoting higher thinking at the elaborated level of the task.

The lesson provides access and develops a conceptual foundation for understanding the core mathematics so that all students are better prepared to approach and attack future problems. This is often the goal in all re-teaching lessons, but it is usually unsuccessful. If students cannot connect the mathematics and develop an understanding of the core concept that underlies the problem, they will be doomed to forget and fail to solve future related problems. Surfacing the concept in a manner that helps students make connections and see relationships is a necessity for learning. Requiring critical thinking and diving deeper into the mathematics is important for all students. This is accomplished as students re-engage in the increasingly difficult levels of the task.

An interesting method for exploring these challenges is to use student work from the class or the toolkit. The work is transcribed to assure anonymity. Students are asked to examine the work, determine if it is mathematically sound and either to justify the findings or show where the work lacked mathematical accuracy or logic. The challenge of critiquing and explaining other students' thinking and misconceptions requires and develops high cognitive skills.

Asking the class to explore the student thinking in unique or mathematically interesting approaches, or in intuitive and logical approaches which contain mathematical flaws, are productive ways to deepen student understanding. Comparing alternative approaches is also valuable. The teacher may select a few different approaches and have the class examine and compare the methods to make connections between ideas and representations. The elaborated level of the task might also be explored through the examination of other students' thinking.

A sample re-engagement lesson

This example of a re-engagement lesson was developed as a result of fifth-grade responses to the MARS Candies Task (Figure 3). The core mathematics of the task involves understanding and using proportional reasoning, including the use of both part-to-part and part-to-whole ratios represented as fractions. At the 'elaborated' level of the task, students must work through multi-step proportional reasoning problems.

The first question from the assessment task shows a candy box with 3 candies and 6 missing pieces. The question states, "This is Amy's box of candies. She has already eaten 6 of them. What fraction of the candies has Amy eaten?" The responses of this fifth-grade class revealed that a few students struggled with the 'access' level of the assessment task, developing the ratio between the candies eaten and the total candies in the box. Some students counted the missing candies and compared that with the total numbers of candies to determine 6/9. Some other students used traditional procedures to simplify the fraction to 2/3. About 10% of the students were unsuccessful in arriving at a correct ratio.

At the beginning of the re-engagement lesson, it was important to revisit the 'access level' concept for the 10% who were confused, and to refresh the memories of those students who were successful. Understanding this level is important for providing 'access' to the more complicated understandings at the 'core' and 'elaborated' levels. The class was presented with the same illustration and the first question from the assessment, but this time the answers 6/9 and 2/3 were also shown (Figure 13). The teacher asked the class to use the drawing to illustrate how one could know the answer was 6/9. Students paired-up to discuss and share their answers, then the teacher gathered the class together to discuss the question as a whole group. Individuals shared that they counted the missing pieces to determine the 6 and then counted all the squares in the box to determine the total number of candies, which was 9; therefore, the fraction of eaten candies was 6/9.

Next the teacher asked the class where 2/3 was represented in the picture. After a think-pair-share process, a student described how she saw three separate columns. Two of the three columns had missing pieces; therefore, the fraction of eaten pieces to total pieces was 2/3. In this manner the simplification process was reinforced as a conceptually accurate solution to the original task, a connection that is often missing in math classrooms.

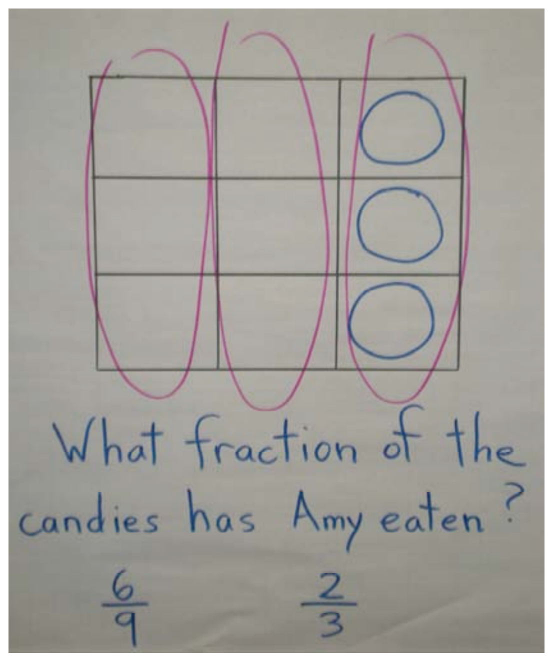


Figure 13. Re-Engagement Lesson, Problem 1.

The second question in the Candies task addresses the core mathematics – understanding ratios and solving a proportional reasoning problem. The task provides a new illustration of a rectangular candy box, arranged in three rows by four columns. The prompt states that, "Valerie shares some of the 12 candies from this box. She gives Cindy 1 candy for every 3 candies she eats herself. How many candies does she give to Cindy? Show how you figured this out."

In order for the class to consider different approaches to the second question, the teacher used two different student work samples. The teacher asked the class to look at the method the first student used to solve the problem (Figure 14). She asked the students whether it was correct, and if so, how the student was illustrating a correct solution. After pair-share discussions, one student suggested the following reasoning: The person must have worked down the column giving three to Valerie and one to Cindy; Next continuing down the second column the person gave three more to Valerie and one more to Cindy; The person continued this process until all the candies were distributed. By counting one could see that a total of 9 candies went to Valerie and 3 to Cindy.

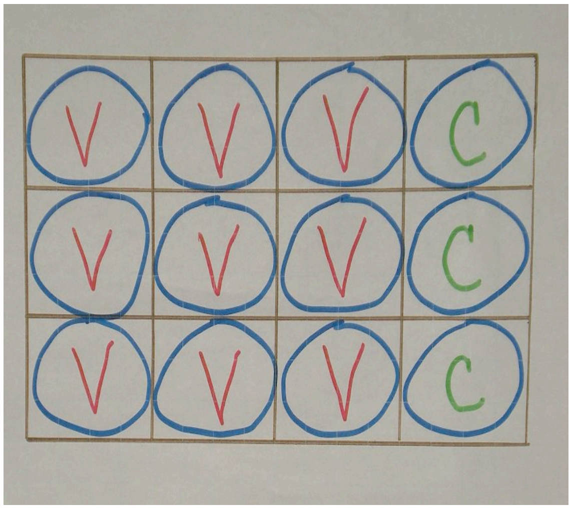


Figure 14. First Student Work for Problem 2, Re-engagement Lesson

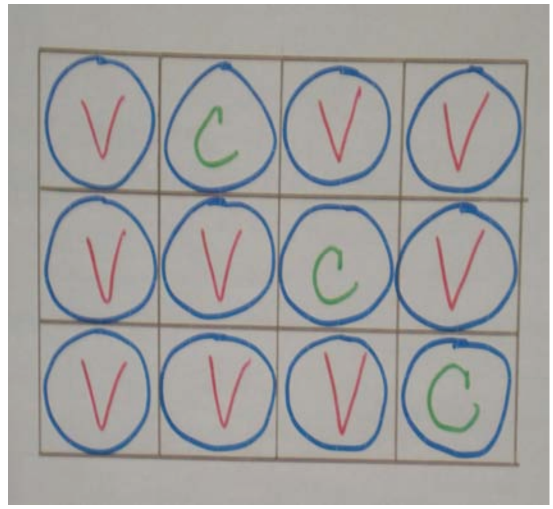


Figure 15: Second Student Work for Problem 2, Re-engagement Lesson

The teacher then showed the second poster (Figure 15) and students were asked again to consider how the students approached the problem. After the students had a chance to discuss the work in small groups, the students shared ideas. A common idea shared by the class was that the student used rows to first distribute 3 to Valerie and 1 to Cindy.

Operations on fractions are traditionally taught in fifth grade, but understanding a fraction as a ratio is often widely misunderstood (Fisher, 2007). Developing understanding of part-to-whole, part-to-part, and part-to-part-to-whole are important concepts for students. The third question of the task assesses the core mathematics of proportional reasoning at a deeper level. Proportional reasoning is a major topic in any sixth grade math curriculum and students in fifth grade should learn to understand this basic building block of mathematics.

The final problem in the Candies Task takes the core concept to an 'elaborated' level, and once again assesses students' conceptual understanding of proportional reasoning. The problem states, "Anthony makes candies. First, he mixes 1 cup of cream with 2 cups of chocolate. In all, he uses 9 cups of these two ingredients. How many cups of chocolate does he use in this candy recipe?" Instead of having students just re-work the problem, the teacher uses actual student work in the re-engagement lesson to promote deeper understanding. She shows two different solutions (Figure 16 and Figure17).

In each case the class was asked to consider how the student approached and solved the problem. The class needed to determine whether the solution was correct, and to explain either the correct reasoning or the faulty logic.

The class attempted to make sense of the student's reasoning shown in the first poster (Figure16). Originally, before the whole class discussion, the line did not appear on the poster. That line was added as members of the class explained where the student went wrong. Some classmates argued that the student was confused over the 9 cups in the problem. Instead of thinking it meant a total of nine cups, the student believed that it was 9 chocolate cups. The class reasoned that since the pattern of chocolate cups grew in an even number pattern of 2 cups, 4 cups, 6 cups, 8 cups, 10 cups, etc., that 9 cups would never appear in the table. Therefore they argued that the student must have stopped short of going over 9 at 8 cups. They pointed to the table and the highlighted 8 cups as their evidence.

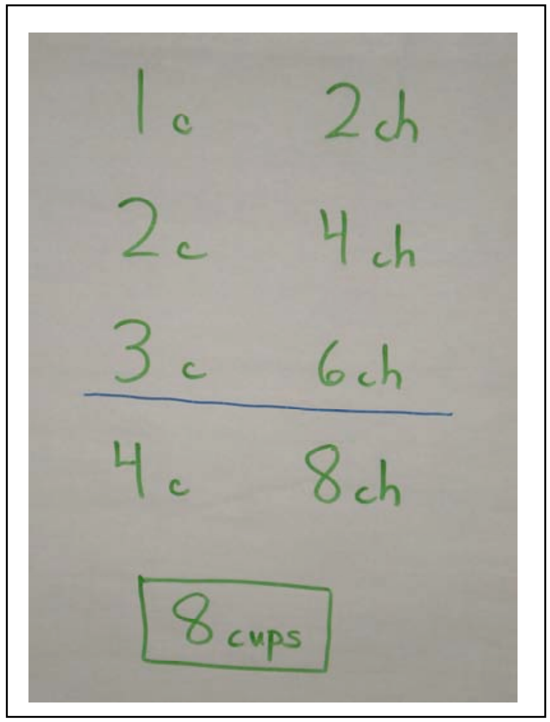


Figure 16: Student Work for Problem 4, Re-engagement

Next the teacher asked the class to consider a second solution to Anthony's problem (Figure 17. After puzzling over the thinking behind the student's approach, the class suggested that the student must have been confused by what was meant by the total number of cups being 9. Instead of understanding that there were just 9 cups in the batch, the student must have thought that each ingredient was a factor times 9. As the students in the class made sense of the error, the teacher recorded where the actual nine occurs by circling the 3 sets of 3 cups that total 9.

This level of reasoning about another student's work is a high cognitive skill. As students learn to reason and justify, it is equally important for them to know why some arguments are mathematically invalid. A re-engagement lesson deepens the mathematical understanding rather than merely revisiting the original content.

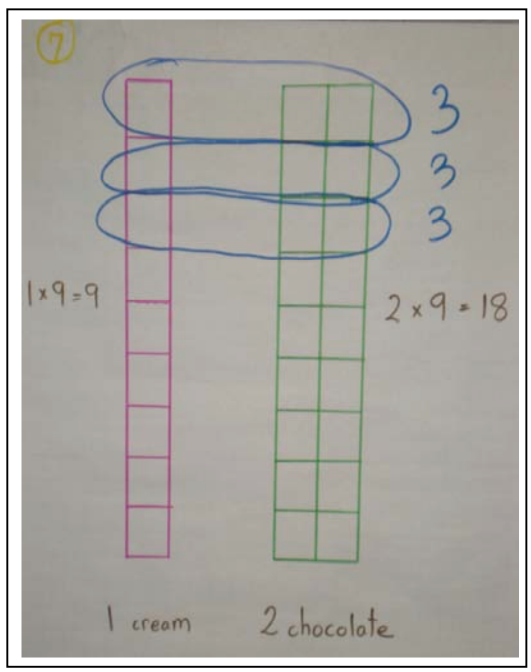


Figure 17: Student Work for Problem 4, Re-engagement

It is important that those students who were unsuccessful develop an understanding of the core mathematics in the task, and that all students go deeper and develop conceptually sound foundations. This work prepares all students for more advanced thinking in the future.

A case study of formative assessment

Achievement results from the Silicon Valley Mathematics Initiative (SVMI) support Paul Back and Dylan Wiliam's claim that formative assessment practice improves student learning significantly. During the 2006–07 school year, SVMI selected nine member school districts (later reduced to eight) to engage in an intensive formative assessment program. These nine districts were identified as the First in Mathematics Consortium (FiMC). The goal of the program was to improve mathematics instruction at the middle grades in order to increase student learning as measured by multiple assessments. The theory of action called for intensive professional development around formative assessment practices and support for collaboration within math departments as they engaged in the complete cycle. Over two years, the middle school teachers attended professional development and engaged in a minimum of fourteen cycles of formative assessment. Working together within their math departments, teachers selected and administered a series of common MARS performance assessment tasks; met to score and analyze the student papers; and identified successful practices, common errors and misconceptions. They used *Tools for Teachers* documents to research student performance and compare their students' work to the general findings. This process helped teachers gain greater knowledge about both math content and pedagogy and to craft and teach new lessons that re-engaged their students in the key math concepts.

In this case study, we will compare the achievement results of the eight FiMC districts with those of 19 non-FiMC districts in SVMI. Baseline data were collected in 2006 in order to determine if the groups were similar in terms of demographics achievement levels and enrollment in higher level math courses at that time. The demographic composition of the two groups was highly similar on those demographic indicators that correlate highly to student achievement: the percentage of students who qualified for free and reduced lunch; the percentage of English language learners; the ethnic composition; and the percentage of students of parents with no college (see Figure 18).

Student Demographics	Non-FiMC	FiMC
Percent of Students that Qualify for NSLP	30%	27%
English Language Learners	17%	21%
American Indian, African American, Hispanic/Latino, Pacific Islander and Filipino	62%	65%
Parent Education - No College	39%	38%

Figure 18. Student Demographics Table

The student achievement of the two groups was also quite similar. In 2006 twenty-eight SVMI school districts administered the spring MARS summative performance assessment exam; achievement by students in the FiMC districts was similar to achievement by students in the other nineteen SVMI districts (See Figure 19). The number of students enrolled in college prep mathematics in eighth grade (Algebra I and Geometry) and their level of success in those courses was also of interest. Baseline data from the spring, 2006, indicated that the two sets of districts were enrolling similar numbers of students and their students were achieving at a similar rate. Given the similarity of the two groups, it seemed reasonable that the nineteen non-FiMC districts could be used as a comparison group to the FiMC treatment group.

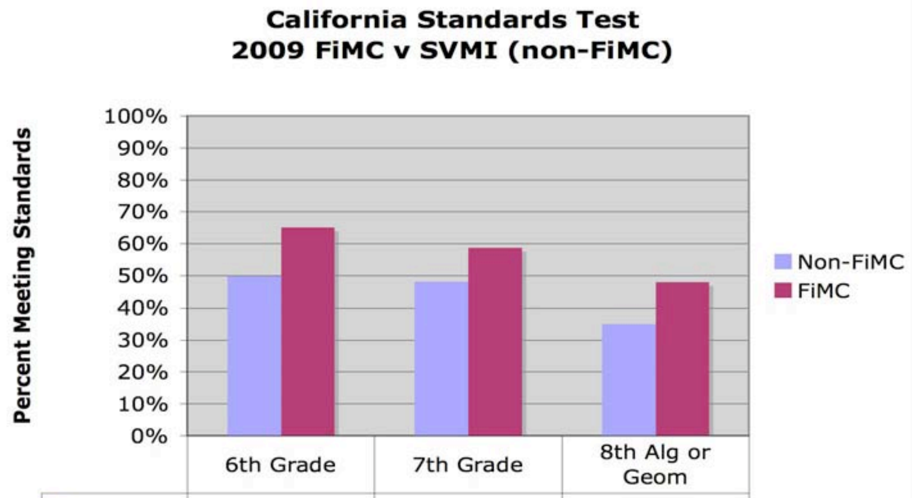
Baseline year 2006	**SVMI**	**FiMC Districts**	**Non-FiMC**
Enrollment in 8th Grade	6778	2489	4289
Percent of 8th Graders enrolled in Algebra 1 or Geometry	51%	52%	51%
Percent of 8th Graders Proficient or Advanced in Algebra or Geometry	34%	32%	35%

Figure 19. Enrollment and Achievement Table (baseline year)

The first year of the initiative was spent recruiting, selecting and inducting districts into FiMC, followed by two years of intensive professional development, utilizing the formative assessment cycle as the major strategy. At the end of the third year, student achievement in FiMC districts improved dramatically while achievement in the other SVMI districts grew at a much slower rate, as indicated on both test measures - the state multiple-choice California Standards Test (CST) and the summative MARS performance assessment exam (see Figures 20 & 21). A similar pattern of growth occurred at all grade levels - sixth, seventh and eighth grade.

In eighth grade, a higher bar was set for meeting standard. Only eighth graders who were proficient or advanced on the CST or who achieved Levels 3 or 4 on the MARS Algebra 1 or Geometry exam were counted as meeting standard. Eighth-grade students who were enrolled in eighth-grade math and who therefore took the general eighth-grade MARS exam were not counted as proficient, even if they achieved a proficient or advanced score on the test. All the increases in student achievement were statistically significant as measured by t-test ($p<0.5$).

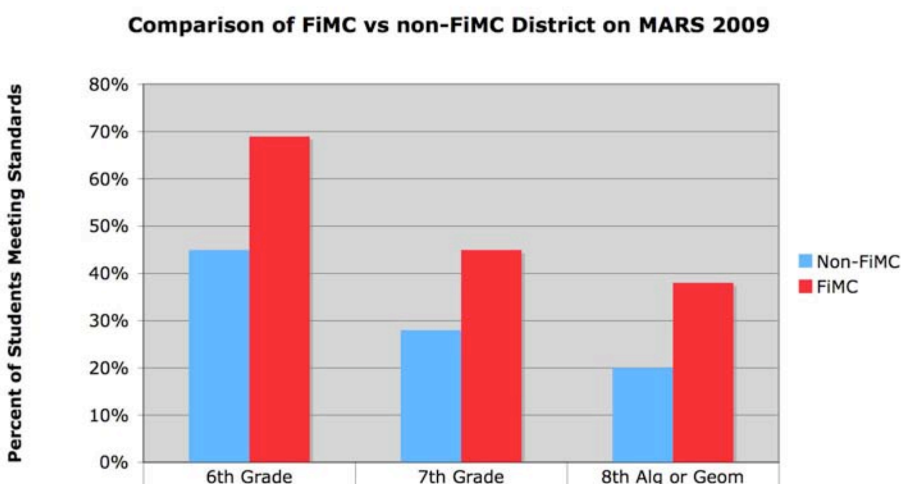
The net positive difference between FiMC students and non-FiMC students achieving grade-level standard (proficient or advanced) on the CST test was fifteen (15) percentage points in sixth grade, eleven (11) percentage points in seventh grade and thirteen (13) percentage points for eighth-graders meeting standards in college prep mathematics.



	6th Grade	7th Grade	8th Alg or Geom
Non-FiMC	50%	48%	35%
FiMC	65%	59%	48%

Figure 20. California Standard Test spring 2009

The net differences in student performances between the two sets of districts were even greater on the MARS performance assessment exams. Those net differences were twenty-four percent (24%) at sixth grade, seventeen percent (17%) at seventh grade and eighteen percent (18%) at eighth-grade for students meeting standard on college prep mathematics. The larger differences on the MARS versus the CST is impressive, yet might not be surprising since the professional development focused directly on similar types of performance assessment tasks. Students, as well as teachers, became stronger at working at high cognitive levels, using non-routine approaches, applying conceptual understanding, and explaining and justifying their conclusions.



	6th Grade	7th Grade	8th Alg or Geom
Non-FiMC	45%	28%	20%
FiMC	69%	45%	38%

Figure 21. Math Assessment Resource Service (MARS) Exam spring 2009

In eighth grade, the percentage of students enrolled in Algebra 1 and Geometry in FiMC districts also increased dramatically over the three years, growing from 52% to 76%, a 46% increase. The non-FiMC districts' Algebra 1 and Geometry enrollment also increased, but at a much slower rate, growing from 51% to 60% in the three years. For the FiMC districts, even though more students were enrolling in Algebra 1 and Geometry, the percent of students meeting standard on both the CST and MARS tests also increased.

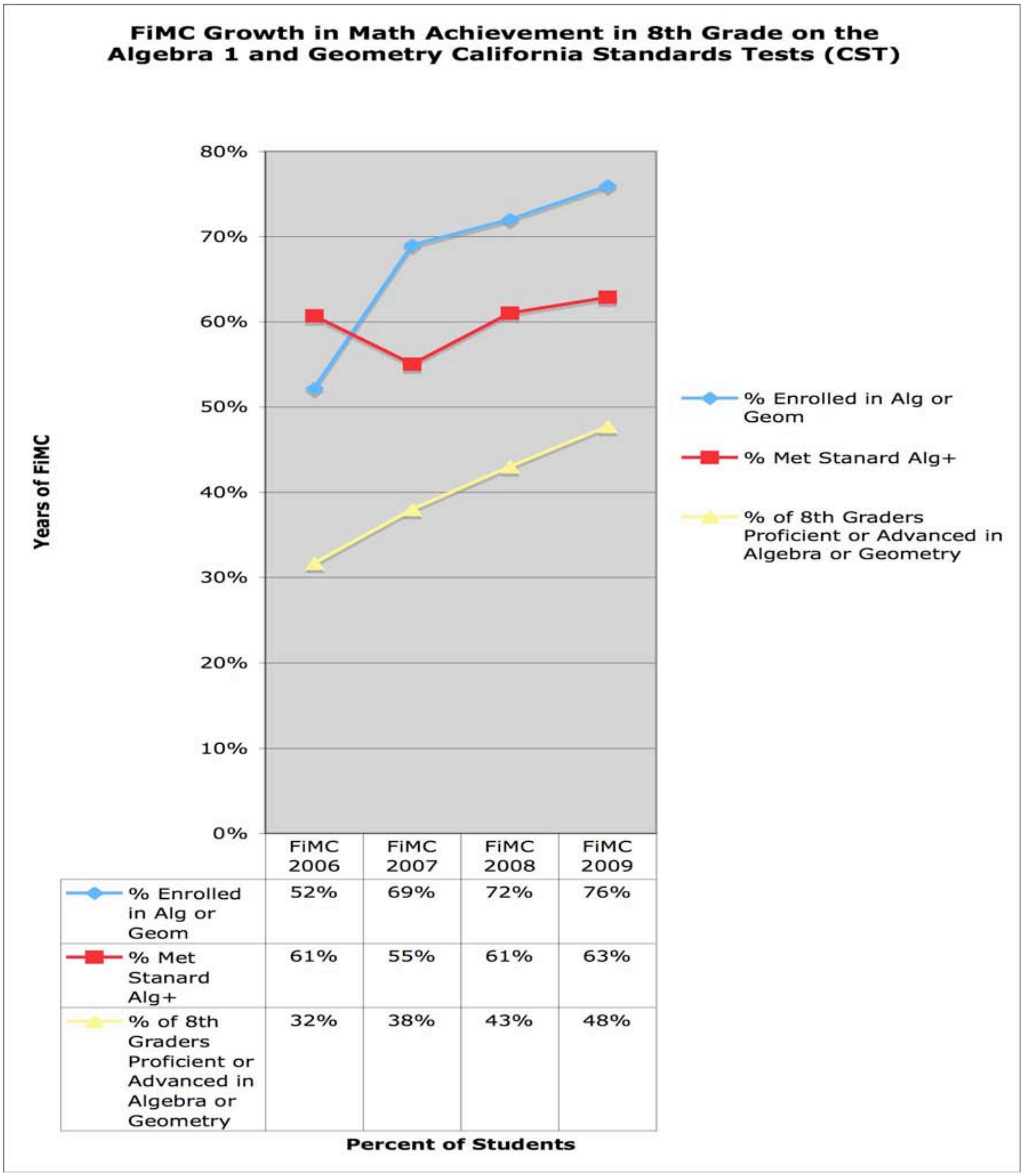


Figure 22. Enrollment and Math Achievement by FiMC on CST 2006 - 2009

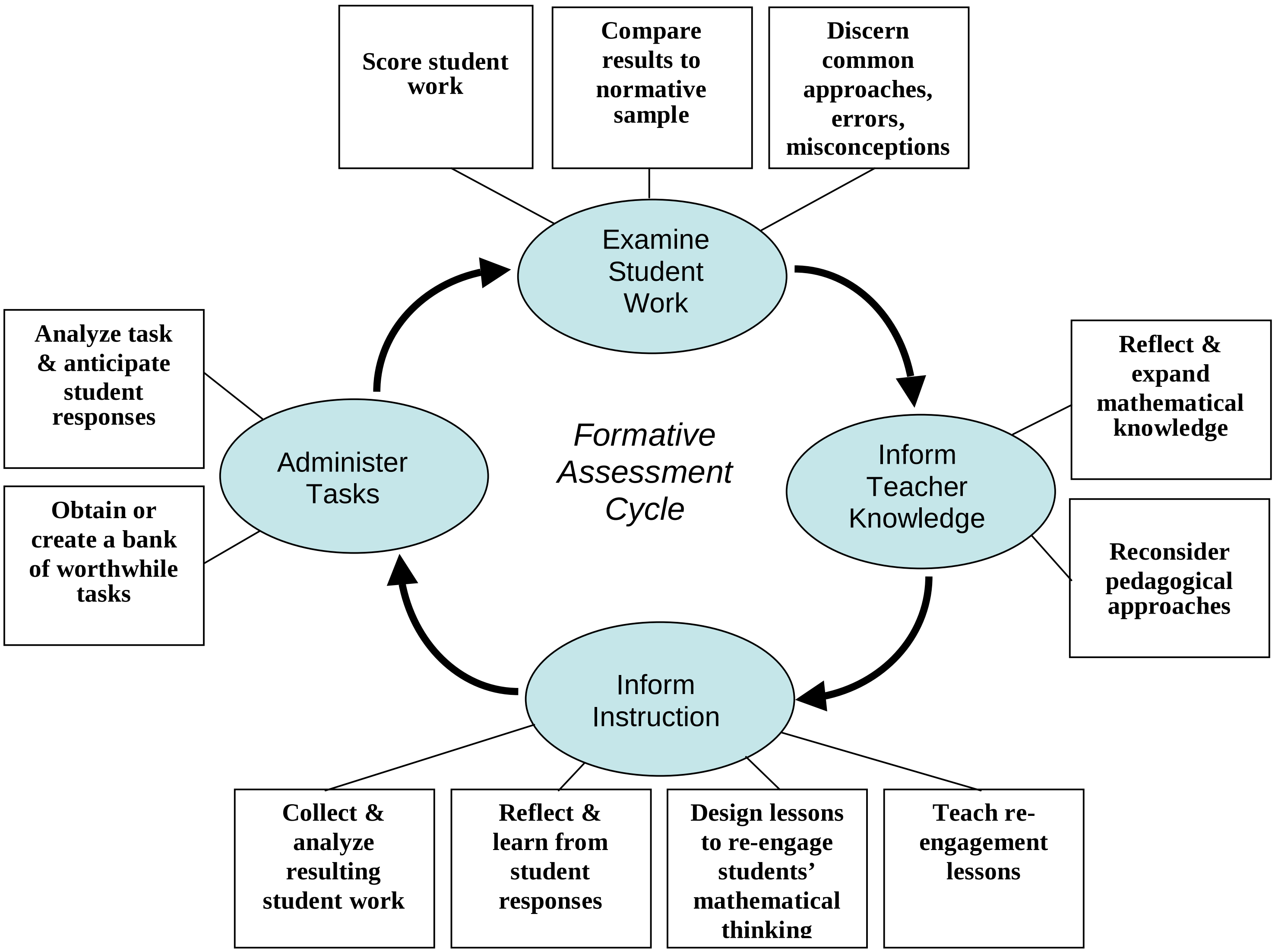
This growth in the percent of students meeting standard at the same time that enrollment percentages were rapidly increasing makes these findings quite significant. A further analysis of the interaction of eighth-grade enrollment and achievement across the years provides additional insight. Over the course of the three years (05-06 baseline, 06-07 recruit/selection/introduction year, 07-08 intensive PD year 1, 08-09 intensive PD year 2) there were increases in both enrollment and numbers of eighth-graders meeting standard on college prep courses, as described above. The graph in Figure 22 illustrates the changes from spring 2006 to spring 2009. One line represents student enrollment in Algebra 1 and Geometry. The second line depicts the percent of Algebra 1 and Geometry students that met standard on the CST**,** and it shows that initially, the percent of students meeting standard on the CST declined, as districts enrolled more students in Algebra 1 and Geometry in 2006-07. This was at the start of the project prior to intensive work in formative assessment. In the two subsequent years student achievement increased significantly within those courses. The third line shows the combined impact of enrollment increases and the percentage of students meeting the performance standard, illustrating the percent of all eighth graders who were proficient or advanced on the CST Algebra 1 test or CST Geometry test. This factor shows steady growth increasing at an approximate rate of 6% of all 8^{th} graders per year**.**

These findings support the claims made by other researchers that rigorous implementation of formative assessment techniques can improve student achievement. Implementation rates of formative performance assessment practices were high in the FiMC middle schools. Ninety-percent (90%) of the FiMC middle school math teachers reported engaging in formative assessment practices using the MARS bank of released tasks. Two-thirds (67%) of the FiMC middle school teachers reported completing the formative assessment cycle, including designing and teaching re-engagement lessons. In addition teachers who engaged in this intensive professional development reported that they plan to continue to use formative assessment. Each of the participating middle schools is continuing to administer and analyze student work using MARS performance assessment tasks as regular part of their mathematics curriculum during 2009-2010 school year. These commitments were made, despite the fact that formal funding and professional development sponsored by the Noyce Foundation ended with the close of the 2008-09 school year. This might be an indication that the powerful practices of formative assessment are becoming systemic within the middle schools of FiMC districts.

Conclusion

During the SVMI research and development work, we realized that our initial conception of a useful system of formative assessment, as pictured in Figure 1, was too simplistic to adequately guide and support a rigorous and effective process. Figure 23 provides an elaborated view of the cycle, including those activities which proved to be crucial to the efficacy of each of the steps, and for which teachers needed models and guided practice.

The tools developed in SVMI were employed in the FiMC work to support teacher learning and systematic implementation of formative assessment practices. The tools included in the *Tools for Teachers* toolkit and illustrated in previous sections of this paper provide a useful model for teachers as they move toward the independent and daily use of formative assessment practices described by Dylan Wiliam.



Key activities which are subsets of each of the major steps in the cycle are pictured in the boxes. SVMI tools provide a scaffold for learning as teachers move through each of the steps or phases of the cycle.

Figure 23. SVMI Formative Assessment Cycle

Although the Noyce Foundation sponsorship of the FiMC and SVMI initiatives was brought to a close in June 2009, the work lives on as the Silicon Valley Math Initiative and more information is available at www.svmimac.org. A rich archive of tasks and video of re-engagement lessons can be viewed at www.insidemath.org.

References

Black, P., Harrison, C., Lee, C., Marshall, B., and Wiliam, D. (2004). Working inside the black box. *Phi Delta Kappan*, 9-21.

Black, P., and Wiliam, D. (1998). Inside the black box: Raising standards through classroom assessment. *Phi Delta Kappan*, 80, 139-144.

Boaler, J. (1998). Open and closed mathematics: Student experiences and understandings. *Journal for Research in Mathematics Education*, 29, 41-46.

Dufour, R., Dufour, R., and Eaker, R. (2008). *Revisiting Professional Learning Communities at Work: New Insights for Improving Schools*. Bloomington, Indiana: Solution Tree.

Fisher, L. (2007). *Mathematics Assessment Collaborative Final Report*. Palo Alto, California: The Noyce Foundation, https//www.noycefdn.org/math/

Foster, D. (2007) When assessment guides instruction: Silicon Valley's mathematics assessment collaborative. *Assessing Mathematical Proficiency*. Berkeley, California: Mathematical Sciences Research Institute Publications, 53.

Foster, D., and Noyce, P. (2004). The Mathematics Assessment Collaborative: Performance testing to improve instruction. *Phi Delta Kappan*, 85, 367-374.

Friedman, T. (2005). *The World is Flat: A Brief History of the Twenty-first Century*. New York: Farrar, Straus and Giroux.

Fullan, M. (2007). Changing the terms for teacher learning. *Journal of Staff Development*, 28, 35-36.

Marzano, R., and Haystead, M. (2008). *Making Standards Useful in the Classroom*. Alexandria, Virginia: Association for Supervision and Curriculum Development.

Mathematics Assessment Resource Service. *Balanced Assessment Tasks*. Nottingham, England: The Shell Centre. MARS tasks can be found at http;//www.Nottingham.ac.uk/education/MARS/

Asked to identify the ways that MARS assessment services are unique, users of the MARS test have cited the intellectually rigorous nature of the tasks and rubrics, the willingness of MARS to customize assessments to client needs, and the value of scoring tasks as professional development. See St. John et al., 2000; Foster & Noyce, 2004; Foster et al., 2007).

The Program for International Student Assessment (PISA). (2006). International Association for the Evaluation of Educational Achievement. PISA is a system of international assessments that focus on 15-year olds' capabilities in reading literacy, mathematics literacy, and science literacy.

Reeves, D. (2004). *Accountability For Learning: How Teachers and School Leaders Can Take Charge*. Alexandria, Virginia: Association for Supervision and Curriculum Development.

Reeves, D., (Ed.) (2007). *Ahead of the Curve: The Power of Assessment to Transform Teaching and Learning*. Bloomington, Indiana: Solution Tree.

Sanders, W.L. and Horn, S.P. (1994). The Tennessee Value-Added Assessment System (TVAAS): Mixed model methodology in educational assessment. *Journal of Personnel Evaluation in Education*, 8, 299-311.

Schmidt, W.H., McKnight, C.c., Valverde, G.A., Houang, R.T., Wiley, D.E. (1997). *Many Visions, Many Aims, Volume 1: A Cross-National Investigation of Curricular Intentions in School Mathematics*. Norwell, Massachusetts: Kluwer Academic Publishers.

Schmoker, M. (2006). *Results Now: How We Can Achieve Unprecedented Improvements ion Teaching and Learning*. Alexandria, Virginia: Association for Supervision and Curriculum Development.

St. John, M., Houghton, N., Tambe, P. (2000). *A Study of the MARS Project: The Contributions to Clients*. Inverness, California: Inverness Research Associates, www.inverness-research.org

Thompson, M., and Wiliam, D. (2007). *Tight But Loose: Conceptual Framework for Scaling Up School Reforms*. Washington, D.C.: American Educational Research Association.

Trends in International Mathematics and Science Study (TIMSS). (2007). International Association for the Evaluation of /Educational Achievement. The 2007 TIMSS Report is the fourth comparison of mathematics and science achievement carried out since 1995.

Weiss, I. R., Pasley, J. D., Smith, P. S., Banilower, E. R., Heck, D. J. (2003). *Looking Inside the Classroom: A Study of K-12 Mathematics and Science Education in the United States*. Chapel Hill, North Carolina: Horizon Research Inc.

Wiggins, G. and McTighe, J. (1998). *Understanding By Design*. Alexandria, Virginia: Association for Supervision and Curriculum Development.

Wright, S.P., Horn, S.P., and Sanders, W.L. (1997). Teacher and classroom context effects on student achievement: Implications for teacher evaluation. *Journal of Personnel Evaluation in Education*, 11, 57-67.