Assessing Mathematical Proficiency
MSRI Publications
Volume **53**, 2007

Chapter 10
When Assessment Guides Instruction
Silicon Valley's
Mathematics Assessment Collaborative

DAVID FOSTER, PENDRED NOYCE, AND SARA SPIEGEL

Standardized testing for the purpose of accountability continues to dominate our nation's schools. Since we first reported on the Mathematics Assessment Collaborative [Foster and Noyce 2004], states have responded to the stringent testing requirements of the No Child Left Behind legislation by expanding annual testing in reading and mathematics to all students in grades 3 through 8. Currently states are adding science tests and more tests in high school. Working under cost constraints, most states elect to use multiple-choice tests, while some commentators such as Peterson [2006] detect a "race to the bottom" — a tendency to lower standards and simplify tests as a way of ensuring that more and more students can be deemed proficient. (See also [Fuller et al. 2006].)

How has assessment been used to inform instruction? A number of districts, challenged urban districts in particular, have responded to the need to boost student scores by increasing the frequency of benchmark assessments. Some districts developed assessments aligned with local curricula to help ensure that coverage and learning across schools. Other districts invested in technology-based programs that offer quarterly updates on student progress along a linear scale, based on easily scored (but often skills-oriented) computer multiple-choice assessments. These programs, while they may reassure a school's staff about student progress or alert them to trouble ahead, do little to inform teachers about how students are thinking, what they understand, where they are falling down, and how, specifically, teachers might change their own instructional practices to address students' difficulties.[1]

[1] For an account of a different use of technology to inform teachers about student performance, see Artigue's description of diagnostic tests in France and the LINGOT project in this volume.

For the past nine years a group of school districts in California's Silicon Valley have taken a different approach to mathematics assessment. These districts have supplemented the state testing system with a coordinated program of support and learning for teachers based on a common set of assessments given to students. In this chapter, we briefly review the history of the Mathematics Assessment Collaborative (MAC). We describe how the results of the annual performance assessment are used to guide professional development. We offer additional examples of what the MAC is learning about student understanding across the grades. We review trends in student performance and discuss the relationship between student performance on the MAC assessment and on the state tests.

A Brief History of the Mathematics Assessment Collaborative

In 1996, the Noyce Foundation formed a partnership with the Santa Clara Valley Mathematics Project at San José State University to support local districts with mathematics professional development. The new partnership was dubbed the Silicon Valley Mathematics Initiative. Its early work focused on providing professional development, establishing content-focused coaching in schools, and collaboratively examining student work to inform teachers of pupils' understandings.

At that time, the state of California was beginning a long and turbulent battle over the establishment of new state curriculum standards [Jacob and Akers 2000; 2001; Jackson 1997; Schoenfeld 2002; Wilson 2003]. Following the state board's adoption of standards in mathematics, the governor pressed to establish a high-stakes accountability system. For the first time, California would require a test that produced an individual score for every student. Because developing a test to assess the state standards was expected to take several years, the state decided in the interim to administer an off-the-shelf, norm-referenced, multiple-choice test — Harcourt's Stanford Achievement Test, Ninth Edition (known as the SAT-9) — as the foundation for the California Standardized Testing and Reporting (STAR) program. In the spring of 1998, students in grades 2 through 11 statewide took the STAR test for the first time.

In an effort to provide a richer assessment measure for school districts, the Silicon Valley Mathematics Initiative formed the Mathematics Assessment Collaborative (MAC). Twenty-four school districts joined the collaborative, paying an annual membership fee.

Selecting an Assessment

MAC's first task was to create a framework characterizing what was to be assessed. Keeping in mind William Schmidt's repeated refrain that the U.S.

curriculum is "a mile wide and an inch deep,"[2] MAC decided to create a document that outlined a small number of core topics at each grade level. The goal was to choose topics that were worthy of teachers' efforts, that were of sufficient scope to allow for deep student thinking, and that could be assessed on an exam that lasted just a single class period. Using as references, standards developed by the National Council of Teachers of Mathematics, by the state of California, and by the local districts, teacher representatives from MAC districts met in grade-level groups to choose five core ideas at each grade level.

Once the core ideas document was created, the next task was to develop a set of exams that would test students' knowledge of these ideas. MAC contracted with the Mathematics Assessment Resource Service (MARS), creators of Balanced Assessment,[3] to design the exams. (For a description of Balanced Assessment's design principles and the work of MARS, see Chapter 6 in this volume.) Each grade-level exam is made up of five tasks. The tasks assess mathematical concepts and skills that involve the five core ideas taught at that grade. The exam also assesses the mathematical processes of problem solving, reasoning, and communication. The tasks require students to evaluate, optimize, design, plan, model, transform, generalize, justify, interpret, represent, estimate, and calculate their solutions.

The MARS exams are scored using a point-scoring rubric. Each task is assigned a point total that corresponds to the complexity of the task and the proportional amount of time that the average student would spend on the task in relation to the entire exam. The points allocated to the task are then allocated among its parts. Some points are assigned to how the students approach the problem, the majority to the core of the performance, and a few points to evidence that, beyond finding a correct solution, students demonstrate the ability to justify or generalize their solutions. In practice, this approach usually means that points are assigned to different sections of a multi-part question. (For an example of such a rubric, see Burkhardt, this volume.)

The combination of constructed-response tasks and weighted rubrics provides a detailed picture of student performance. Where the state's norm-referenced, multiple-choice exam asks a student merely to select from answers provided, the MARS exam requires the student to initiate a problem-solving approach to each task. Students may use a variety of strategies to find solutions, and most of the prompts require students to explain their thinking or justify their findings.

[2] William Schmidt, U.S. research director for the Third International Mathematics and Science Study, has made this statement in numerous places. See, for example, the press releases available on the Internet at http://ustimss/msu.edu.

[3] Balanced Assessment Packages [BAMC 1999–2000] of assessment items and sample student work were published by Dale Seymour Publications. Balanced Assessment tasks can be found at http://www.educ.msu.edu/mars and http://balancedassessment.gse.harvard.edu.

This aspect of the assessment seems impossible to duplicate by an exam that is entirely multiple choice. Details of the administration of the exams also differ from the state's approach, in that teachers are encouraged to provide sufficient time for students to complete the exam *without rushing*. In addition, students are allowed to select and use whatever tools they might need, such as rulers, protractors, calculators, link cubes, or compasses.

The Assessment in Practice

In the spring of 1999, MAC administered the exam for the first time in four grades — third, fifth, seventh, and in algebra courses — in 24 school districts. Currently the collaborative gives the exam in grades two through grade 8, followed by high school courses one and two. Districts administer the exam during March, and teachers receive the scored papers by the end of April, usually a couple of weeks prior to the state high-stakes exam.

Scoring the MARS exams is an important professional development experience for teachers. On a scoring day, the scoring trainers spend the first 90 minutes training and calibrating the scorers on one task and rubric each. After that initial training, the scorers begin their work on the student exams. After each problem is scored, the student paper is carried to the next room, where another task is scored. At the end of the day, teachers spend time reflecting on students' successes and challenges and any implications for instruction. Scoring trainers check random papers and rescore them as needed. Finally, as a scoring audit, 5% of the student papers are randomly selected and rescored at San José State University. Reliability measures prove to be high: a final analysis across all grades shows that the mean difference between the original score and the audit score is 0.01 point.

Along with checking for reliability, the 5% sample is used to develop performance standards for overall score reporting. The collaborative has established four performance levels in mathematics: Level 1, minimal success; Level 2, below standards; Level 3, meeting standards; and Level 4, consistently meeting standards at a high level. A national committee of education experts, MARS staff members and MAC leaders conducts a process of setting standards by analyzing each task to determine the core of the mathematical performance it requires. The committee examines actual student papers to determine the degree to which students meet the mathematical expectations of the task, and it reviews the distribution of scores for each task and for the exam as a whole. Finally, the committee establishes a cut score for each performance level for each test. These performance levels are reported to the member districts, teachers, and students.

Once the papers are scored, they are returned to the schools, along with a copy of the master scoring sheets, for teachers to review and use as a guide for further instruction. Each school district creates a database with students' scored results on the MARS exam, demographic information, and scores on the state-required exam. Using these, an independent data analysis company produces a set of reports that provide valuable information for professional development, district policy, and instruction.

How Assessment Informs Instruction

Over time, it has become clear that the tests, the scoring sessions, and the performance reports all contribute to MAC's desired outcome: informing and improving instruction. The scoring sessions are powerful professional development activities for teachers. To be able to score a MARS exam task accurately, teachers must fully explore the mathematics of the task. Analyzing different approaches that students might take to the content within each task helps the scorers assess and improve their own conceptual knowledge. The scoring process sheds light on students' thinking, as well as on common student errors and misconceptions. As one teacher said, "I have learned how to look at student work in a whole different way, to really say, 'What do these marks on this page tell me about [the student's] understanding?"' Recognizing misconceptions is crucial if a teacher is to target instruction so that students can clarify their thinking and gain understanding. The emphasis on understanding core ideas helps teachers build a sound sequence of lessons, no matter what curriculum they are using. All of these effects on instruction grow out of the scoring process.

The scored tests themselves become valuable curriculum materials for teachers to use in their classes. MAC teachers are encouraged to review the tasks with their students. They share the scoring information with their students, and build on the errors and approaches that students have demonstrated on the exams.

Tools for Teachers

Being data-driven is a common goal of school districts. In this day of high-stakes accountability, districts are awash with data, yet not much of it is in a form readily useful to teachers. To meet that need, MAC publishes *Tools for Teachers*, an annual set of reports derived from the results of each year's exam.

Along with broad performance comparisons across the collaborative's membership and analysis of the performance of different student groups, the reports provide a wealth of other information. A detailed portrait is compiled of how students approached the different tasks, with a description of common misconceptions and evidence of what students understand. The reports include student work samples at each grade level showing the range of students' approaches, successes, and challenges. (For examples, see Foster, this volume.) In addition, the reports educe implications for instruction, giving specific suggestions and ideas for teachers as a result of examining students' strengths and the areas where more learning experiences are required.

This set of reports is distributed to teachers throughout the initiative. Each October, MAC presents large-scale professional development workshops to introduce the new *Tools for Teachers*. Many teachers use these documents to plan lessons, determine areas of focus for the year, and fuel future formative assessment experiences for their classes.

Using Student Responses to Inform Professional Development

The Mathematics Assessment Collaborative provides a broad range of professional development experiences for teachers and leaders. A significant design difference between the professional development provided through MAC and other professional development is that the MAC experiences are significantly informed by the results from the annual exams. This translates into workshops that target important ideas where students show areas of weakness. Here are three examples.

Proportional reasoning is a central idea in middle school. In 2001, seventh-grade students were given a task called The Poster (see next page). The task assesses students' ability to apply their understanding of proportion to a visual scaling situation.

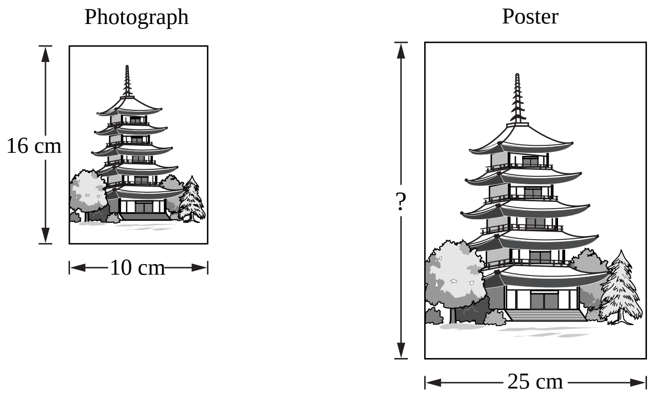
Only 37% of seventh graders were able to meet standard on the task, and only 20% could completely solve both questions in the task. Many (63%) of the students did not think of the problem as a proportional relationship; most used addition to find the missing measurement in the proportional situation. A typical misuse of addition in this problem is reproduced on the next page.

This student misunderstanding of proportional reasoning became a major focus of professional development planning for MAC middle school teachers. MAC institutes and workshops provided both content knowledge and pedagogical strategies for teaching proportional reasoning in middle school. These sessions for teachers made explicit the underlying concepts of ratios, rates, and proportions, including understanding proportions from a functional approach. At the professional development sessions, teachers practiced solving non-routine proportional reasoning problems. They made connections between representations and representatives[4] of these functions that used bar models, tables,

[4]Representatives of a function are descriptions that do not completely determine the function, for example, a finite table of values does not determine all values of a function that has infinitely many possible values.

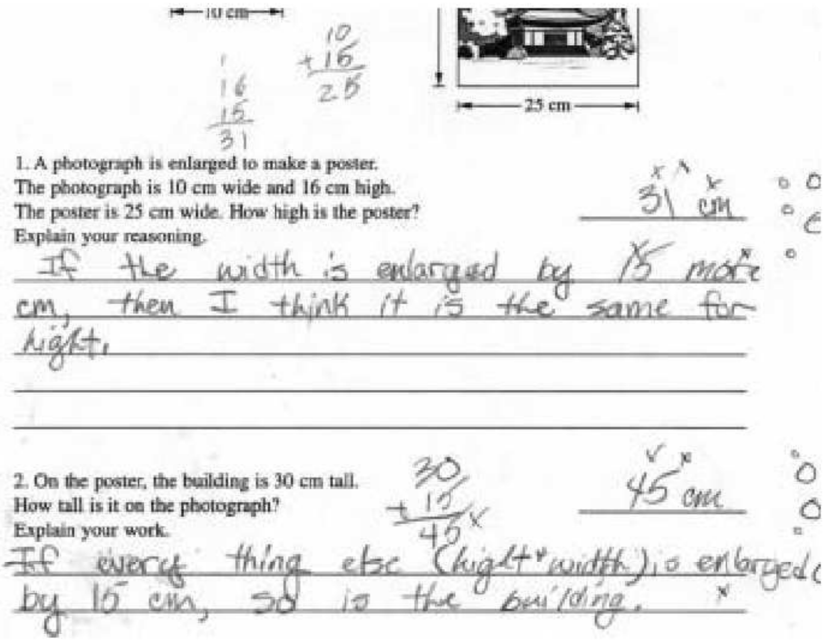
The Poster

This problem gives you the chance to:
- calculate sizes in an enlargement



1. A photograph is enlarged to make a poster.
The photograph is 10 cm wide and 16 cm high.
The poster is 25 cm wide. How high is the poster? ______________
Explain your reasoning.

2. On the poster, the building is 30 cm tall.
How tall is it on the photograph? ______________
Explain your work.


1. A photograph is enlarged to make a poster.
The photograph is 10 cm wide and 16 cm high.
The poster is 25 cm wide. How high is the poster? 31 cm
Explain your reasoning.
If the width is enlarged by 15 more cm, then I think it is the same for hight.

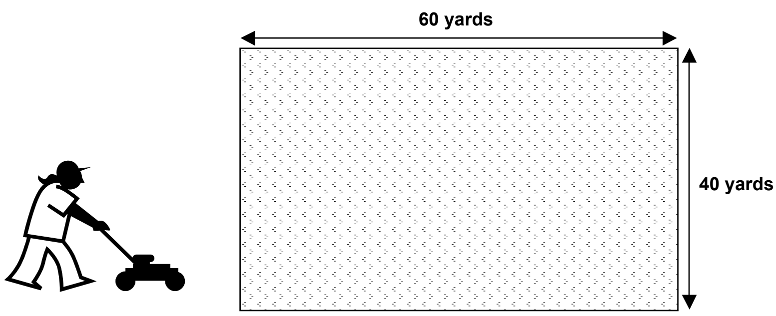
2. On the poster, the building is 30 cm tall.
How tall is it on the photograph? 45 cm
Explain your work.
If every thing else (hight + width) is enlarged by 15 cm, so is the building.

graphs, and equations. Teachers were encouraged to use non-routine problems in their classroom and to promote the use of different representations and multiple strategies to find solutions.

Four years later, seventh-grade students were given Lawn Mowing to assess proportional reasoning. The task involved making sense of rates, a special type of proportional reasoning that students traditionally struggle with.

Lawn Mowing

This problem gives you the chance to:
- solve a practical problem involving ratios
- use proportional reasoning



Dan and Alan take turns cutting the grass.
Their lawn is 60 yards long and 40 yards wide.

1. What is the area of the yard? ________________ square yards

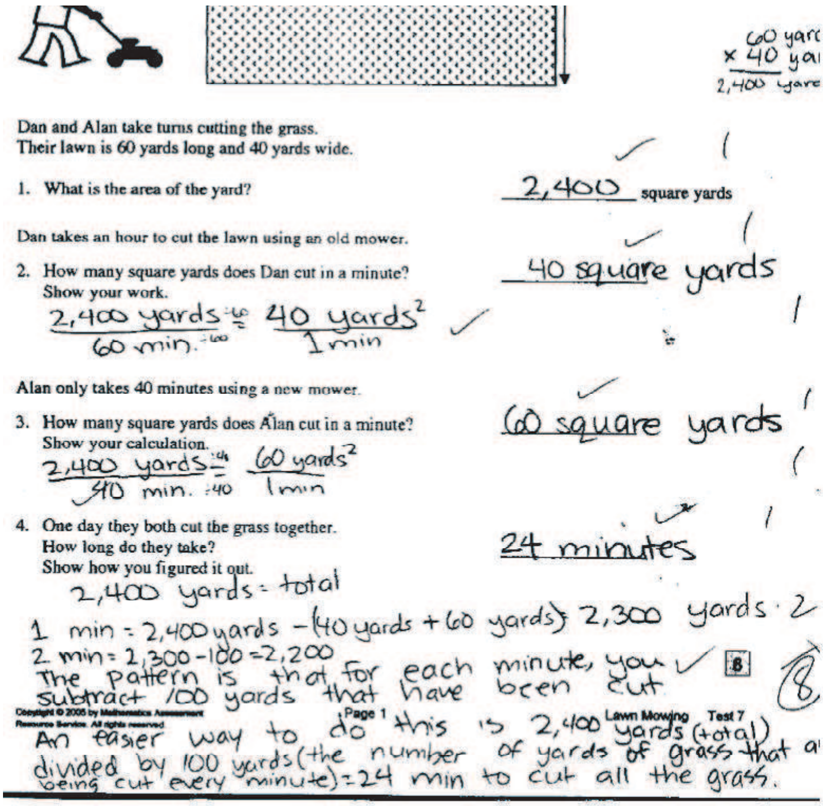
Dan takes an hour to cut the lawn using an old mower.

2. How many square yards does Dan cut in a minute? ________________
   Show your work.

Alan only takes 40 minutes using a new mower.

3. How many square yards does Alan cut in a minute? ________________
   Show your calculation.

4. One day they both cut the grass together.
   How long do they take? ________________
   Show how you figured it out.



60 yards
× 40 yards
2,400 yards

Dan and Alan take turns cutting the grass.
Their lawn is 60 yards long and 40 yards wide.

1. What is the area of the yard? 2,400 square yards

Dan takes an hour to cut the lawn using an old mower.

2. How many square yards does Dan cut in a minute? 40 square yards
Show your work.
2,400 yards ÷ 60 / 60 min. ÷ 60 = 40 yards² / 1 min

Alan only takes 40 minutes using a new mower.

3. How many square yards does Alan cut in a minute? 60 square yards
Show your calculation.
2,400 yards ÷ 40 / 40 min. ÷ 40 = 60 yards² / 1 min

4. One day they both cut the grass together. 24 minutes
How long do they take?
Show how you figured it out.
2,400 yards = total
1 min = 2,400 yards − (40 yards + 60 yards) = 2,300 yards · 2
2 min = 2,300 − 100 = 2,200
The pattern is that for each minute, you subtract 100 yards that have been cut
An easier way to do this is 2,400 yards (total) of grass that a divided by 100 yards (the number of yards of grass that a being cut every minute) = 24 min to cut all the grass.

8

Copyright © 2005 by Mathematics Assessment Resource Service. All rights reserved. Page 1 Lawn Mowing Test 7

Fifty-nine percent of the students met standard on the task, a big improvement over students' performance on the proportional reasoning task in 2001. We see above work in which the student set up a ratio between the size of the lawn and the minutes it took to cut the lawn in order to find the unit rates. Then the student used two different reasoning strategies to determine and confirm the amount of time it takes for each person to cut the lawn. This paper is typical of the way students' approaches to proportional reasoning problems have improved over the years.

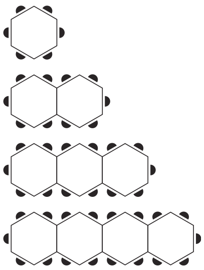
We believe, based on survey feedback and student achievement data, that MAC's explicit focus on proportional reasoning with middle school teachers contributed to this improvement in student achievement. Using student results from the MARS test to tailor and inform professional development for the following year has become a cornerstone of MAC strategy.

The MARS assessment provides a valuable perspective on students' understanding from grade level to grade level. This vertical view over the grades allows us to investigate how student performance related to a particular mathematical idea, such as patterns and functions, changes over time — or doesn't change. One worrisome trend is that students often are able to learn foundational skills in an area of mathematics in the early grades but then are unable to move beyond their basic understanding to apply, generalize, or justify solutions. This trend is illustrated in a set of similar tasks concerning linear functions in fourth grade, seventh grade, and Algebra 1 (Hexagon Desks, Hexagons, and Patchwork

Hexagon Desks

This problem gives you the chance to:
- find and extend a number pattern
- plot and use a graph

Sarah finds how many students can sit around a row of desks. The top surface of each desk is a hexagon, and the hexagons are arranged in rows of different shapes.



1 desk 6 students

2 desks 10 students

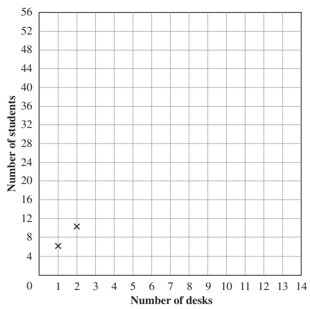
3 desks 14 students

4 desks

1. Complete Sarah's table.

Number of desks in a row	Number of students
1	6
2	10
3	
4	
5	
6	

2. On the grid, plot the results from the table you completed in question 1. The first two points have already been plotted for you.



3. Sarah says that 47 students can sit around a row of 11 desks. Without drawing the desks, explain how you know that Sarah is wrong.

____

____

____

____

How many students can sit around a row of 11 desks? __________

10

Hexagons

This problem gives you the chance to:
- recognize and extend a number pattern in a geometric situation
- find a rule for the pattern

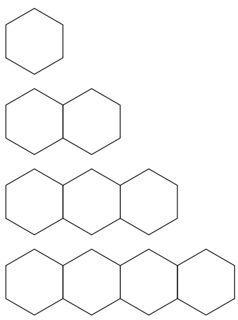
Maria has some hexagonal tiles.
Each side of a tile measures 1 inch.
She arranges the tiles in rows; then she finds the perimeter of each row of tiles.

1 tile
perimeter = 6 in.

2 tiles
perimeter = 10 in.

3 tiles

4 tiles



Maria begins to make a table to show her results.

Number of tiles in a row	Perimeter in inches
1	6
2	10
3	
4	

1. Fill in the empty spaces in Maria's table of results.
What will be the perimeter of 5 tiles? ______________ inches

2. Find the perimeter of a row of 10 tiles. ______________ inches
Explain how you figured it out.

____

____

____

3. Write a rule or formula for finding the perimeter of a row of hexagonal tiles when you know the number of tiles in the row.
Let n = the number of tiles, and p = the perimeter.

____

____

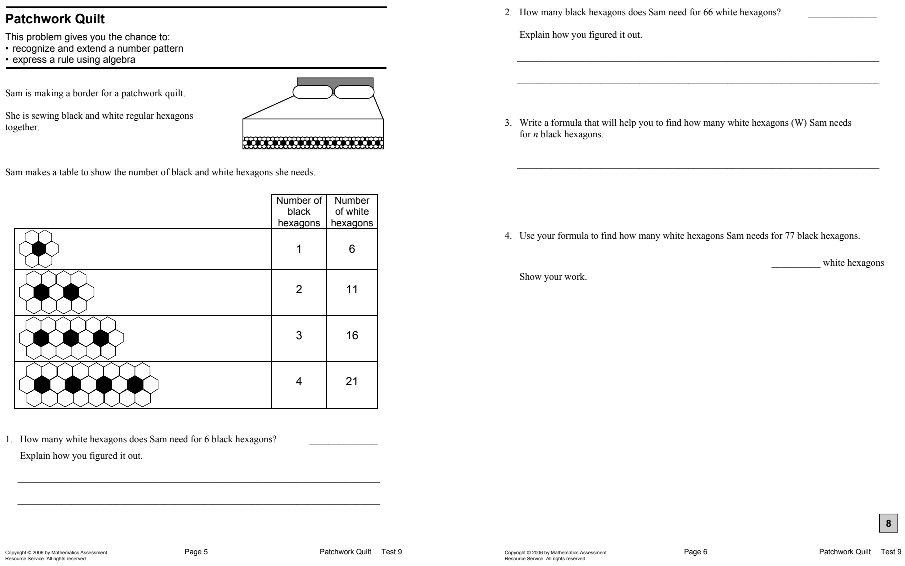
____

____

4. Find the perimeter of a row of 25 hexagonal tiles. ______________ inches
Show your work.

5. The perimeter of a row of hexagonal tiles is 66 inches.
How many tiles are in the row? ______________

10



Patchwork Quilt

This problem gives you the chance to:
• recognize and extend a number pattern
• express a rule using algebra

Sam is making a border for a patchwork quilt.

She is sewing black and white regular hexagons together.

Sam makes a table to show the number of black and white hexagons she needs.

	Number of black hexagons	Number of white hexagons
	1	6
	2	11
	3	16
	4	21

1. How many white hexagons does Sam need for 6 black hexagons? ________
Explain how you figured it out.

Copyright © 2006 by Mathematics Assessment Resource Service. All rights reserved. Page 5 Patchwork Quilt Test 9

2. How many black hexagons does Sam need for 66 white hexagons? ________
Explain how you figured it out.

3. Write a formula that will help you to find how many white hexagons (W) Sam needs for *n* black hexagons.

4. Use your formula to find how many white hexagons Sam needs for 77 black hexagons.
________ white hexagons
Show your work.

8

Copyright © 2006 by Mathematics Assessment Resource Service. All rights reserved. Page 6 Patchwork Quilt Test 9

Quilt). Each grade's task concerns a finite pattern to be examined, extended, and explored. Students are also asked questions about values in the domain or range of the linear function associated with the pattern. At the secondary level, students are asked to write a formula for this relationship.

MAC analyzed student papers at these three grades to determine where students were successful with functions and where they struggled. Here is the percentage of students successful on each element of the task:

	% successful		
Task element	Grade 4	Grade 7	Algebra
Extend the pattern	84%	82%	87%
Given a value in the domain, find the value	57%	58%	53%
Given a value in the range, find a value in the domain	40%	35%	68%
Write a formula	n/a	27%	27%

The exam results show that student performance on most elements of these tasks did not improve as the grade level increased. Students in all three grades were similarly successful at extending the pattern. The percentage of students successful at using a functional relationship to find a value was similar at all grade levels, although the algebra task asked students to use a larger value. The algebra students were more able to determine the inverse relationship but showed no more success in writing an algebraic equation than the seventh graders. These results — flat performance on function tasks across grade levels — point to a need to go beyond asking students to extend patterns by teaching them to reason and generalize at more complex levels. Without instruction asking students to go more deeply than their first solution, their mathematical development may stall at the basic level that they attained years earlier. These statistics help us understand the challenges of teaching upper-grade students. Not only do students need to learn more mathematical ideas and language, but they also need explicit learning experiences in reasoning, generalizing, and justifying. These higher-level thinking skills should be routinely addressed in mathematics classes.

Using the MARS exam has also helped MAC make inroads on student misunderstandings related to mathematical conventions, language, or notation. When we focus professional development on one of these misunderstandings, we often see dramatic changes in student responses. One common early problem, observed in student solutions involving multiple operations was the use of mathematical run-on sentences. Consider the problem of how to calculate the number of feet in a picture of a girl walking three dogs. A typical (incorrect) student response reads: $4 \times 3 = 12 + 2 = 14$. This is a "mathematical run-on sentence": 4×3 does not equal $12 + 2$. The solution steps should have been written out:

$$4 \times 3 = 12 \qquad 12 + 2 = 14$$

At first glance, this correction may seem like nit-picking. But the problem with the notation is more than just sloppiness; a run-on sentence betrays a common misconception. Instead of understanding that the equal sign indicates that expressions on the two sides of the sign have the same value, students using such run-on sentences take the equal sign to signal that an operation must be performed: "The answer is . . . " [Siegler 2003]. This view contributes to further confusion as students learn to generalize and work with expressions containing variables in later grades.

We found that this error in notation occurred regularly throughout the tested population. On further investigation, we learned that teachers commonly allowed this notation to be used in classrooms, or even used it themselves when demonstrating solutions to multi-step problems. The assessment report for that year pointed out the problem and announced that solutions using run-on sentences would no longer receive full credit. Subsequent professional development showed teachers how such notation led to student misconceptions. Within a year, the collaborative noted a dramatic change in the way students in 27 districts communicated mathematical statements.

This matter of notation was just one example of how analyzing patterns of student error led to improvements in instructional practice. Other areas of improvement include differentiating between continuous and discrete graphs, noting and communicating the units in measurement problems, distinguishing between bar graphs and histograms, understanding correlation trends in scatter-plots, and developing understanding of mathematical justifications. Examining MARS results has also led teachers to confront significant chunks of unfamiliar mathematical content. Discussing the tasks and student responses often uncovers the fact that, for many topics and concepts in algebra, geometry, probability, measurement, and statistics, teachers' understanding is weak. Uncovering these gaps in teachers' content knowledge is central to improving instruction.

MAC Assessment and State Testing

The quality of information that the Mathematics Assessment Collaborative has provided to its member districts has helped the districts maintain their commitment to professional development that concentrates on improving teacher understanding. California offers significant incentives and sanctions for student achievement on the state STAR exam, and many districts across the state are thus tempted to embrace narrow quick-fix methods of test prep (drill on practice tests and focus on strategies for answering multiple-choice tests) and "teaching to the test."

To counter this temptation, MAC has been able to show that, even when a significant number of students are improving on the state test, their success may not translate into greater mathematical understanding as demonstrated by success on the more demanding performance assessments. The statistics also indicate that, as students move up the grades, the disparity increases: more and more students who appear to be doing well on the state exam fail to meet standards on the performance exam. Conversely, success on the MARS exam becomes an ever *better* predictor of success on the state's STAR exam. By

		STAR	
	MARS	Basic or below	Proficient or above
grade 3	Below standard	23%	7%
	Meets or exceeds standards	12%	58%
grade 7	Below standard	46%	11%
	Meets or exceeds standards	6%	37%

grade 7, students whose teachers have prepared them to perform well on the MARS exam are extremely likely to perform above the fiftieth percentile on the STAR exam. The table on the preceding page compares success rates on the 2004 MARS and STAR exams for grades 3 and 7.

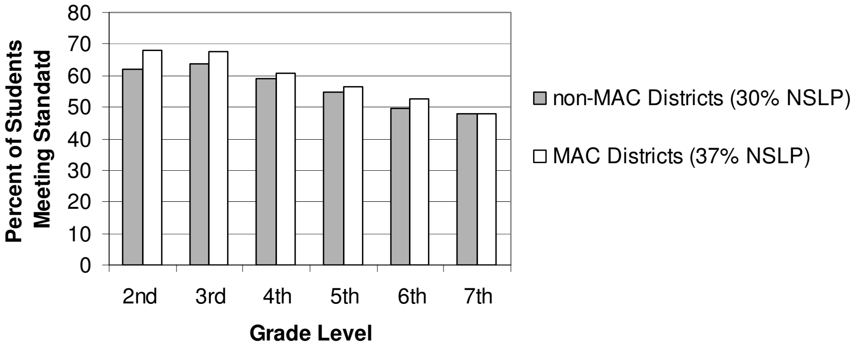
The Mathematics Assessment Collaborative has been able to demonstrate to the satisfaction of superintendents and school committees that high-quality professional development significantly enhances student achievement. District case studies show that students whose teachers participate in intensive MAC professional development achieve higher averages on both the state mathematics test and the MARS exam than students whose teachers who are less involved. As a result, districts have continued to invest in mathematics professional development and formative assessment. The number of students assessed and teachers and grade levels involved has grown every year, even as MAC has kept the number of member districts relatively constant. In 2006, more than seventy thousand students of 1300 teachers in thirty-five districts participated in the MARS exam.

The performance of MAC district students on the STAR exam has continued to rise. For example, while 53% of MAC district third graders performed above the fiftieth percentile on the state mathematics test in 1998, 68% met standard (Proficient or Advanced) on the more challenging California Standards Test for mathematics in 2005. Similar growth has occurred in the other grades, with a minimum of 52% of MAC district students meeting standard at each grade level.

There are a wide number of variables to consider when comparing student achievement statistics. In an effort to eliminate as many variables as possible and still compare performance on the STAR exam between students of teachers involved in MAC programs and students of other teachers, we analyzed statistics from nineteen districts in San Mateo County. Of the nineteen districts, ten are member districts of MAC.

The analysis compared student achievement on the 2005 STAR exam for students in second grade through seventh grade. The MAC students as a group are generally poorer than the comparison group, with 37% of the MAC students qualifying for the National School Lunch Program (NSLP) compared to 30% of the non-MAC students. Both groups have 26% English Language Learners. The data set consists of 21,188 students whose teachers are not members of MAC and 14,615 students whose teachers are involved in MAC programs. The figure at the top of the next page indicates that a larger percentage of students from MAC teachers met standards on the 2005 STAR exam than students from non-MAC teachers at every grade level except seventh grade, where the percentage was the same.

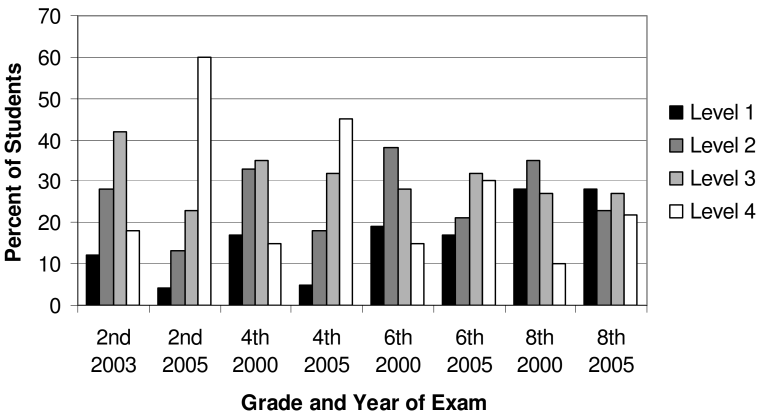
These statistics are encouraging because the population served by the MAC schools is slightly poorer, a demographic factor that has been shown to limit



2005 student performance on the STAR exam

achievement. These findings are promising in that they reinforce that students of MAC teachers, who typically engage their students in high level problem solving and more open, constructed-response tasks, outperform other students on the more procedurally oriented STAR exam despite challenging economic factors.

Success on the state test is politically important, but knowing the percentage of students that exceeds a single score level tells us little about how scores are distributed. More telling as a measure of progress in student learning over the whole range of performance is the growth that students demonstrate on the MARS performance exam. Over time, grades 3 through 6 have shown considerable increases in the percentage of students meeting standard on the MARS exam. In addition to seeing more students in MAC achieving at the highest performance level, we also find fewer students who are performing at the lowest level. The figure below shows the percentage of students at each



Performance on MARS Exam in year of implementation and in 2005 performance level in the first year of implementation versus the current year. In second grade, 60% reached level 4 in 2005, up from 18% in 2003. During that same period, the percentage of second graders in the lowest performance level decreased, with only 4% at the lowest level in 2005 compared to 12% in 2003. In fourth grade, there is a net change of 30% more students achieving level 4 in 2005 compared to 2000. More fourth graders have moved up from the lowest performance level, with numbers decreasing from 17% in 2000 to 5% in 2005. What is happening in practice at these elementary grades is an upward shift in performance for students at every achievement level. Even the lowest-achieving students are addressing the complex MARS tasks and demonstrating some level of understanding. In grades 6 and 8, on the other hand, there are gains in achievement at the highest level with an increase of 15% and 12% respectively, but there is little change in the lowest performance level.

These findings have convinced district leaders to embrace the theory central to our work. This theory states that when teachers teach to the big ideas (e.g., the core topics in the MAC framework), participate in ongoing content-based professional development, and use specific assessment information to inform instruction, their students will learn and achieve more.

Teachers benefit from this approach as much as students do. Just as formative assessment sends a different message to students than do final exams and grades, a collaborative system of formative performance assessment sends a different message to teachers than does high-stakes summative assessment. Teachers laboring to improve student performance on a high-stakes exam can come to feel isolated, beaten down, and mystified about how to improve. Because they are rewarded for getting higher percentages of students over one score bar in one year, they may be tempted to focus primarily on the group of students nearest to that bar or to grasp at a set of narrow skills and procedures that will allow students to answer a few more questions correctly. The exigencies of test security mean that teachers often receive little specific information about where their students' performance excelled or fell short. When high-stakes test results come back, often months after the exam, teachers can do little with the results but regard them as a final grade that marks them as a success or failure.

The Mathematics Assessment Collaborative fights teachers' sense of isolation and helplessness by sharing everything it learns about students. It identifies common issues and potential solutions. It helps teachers understand how learning at their particular grade level is situated within a continuum of students' growing mathematical understanding. It promotes communication across classrooms, schools, and grade levels. It encourages teachers to take a longer, deeper view of what they are working to achieve with students.

Assessment that requires students to display their work is not a panacea that suddenly transforms student learning. Rather, it is a tool for building the capacity of the teaching community to improve its work over time. The discipline of exploring together both the core mathematics we want students to know and the evidence of what they have learned is simultaneously humbling and energizing. Knowing that they are always learning and improving creates among educators a healthy, rich environment for change. To improve instruction requires that teachers become wiser about the subject they teach and the ways that students learn it. Performance assessment of students, with detailed formative feedback to teachers accompanied by targeted professional development, helps to build the teacher wisdom we need.

References

[BAMC 1999–2000] *Balanced Assessment elementary/ middle grades/ high school/ advanced high school assessment packages 1 & 2*, Balanced Assessment Project, White Plains, NY: Dale Seymour Publications, 1999–2000.

[Foster and Noyce 2004] D. Foster and P. Noyce, "The Mathematics Assessment Collaborative: Performance testing to improve instruction", *Phi Delta Kappan* **85**:5 (2004), 367–374.

[Fuller et al. 2006] B. Fuller, K. Gesicki, E. Kang, and J. Wright, *Is the No Child Left Behind Act working? The reliability of how states track achievement*, Working paper **06-1**: Policy Analysis for California Education, 2006. Available at http://pace.berkeley.edu/NCLB/WP06-01_Web.pdf. Retrieved 2 Jul 2006.

[Jackson 1997] A. Jackson, "The math wars: California battles it out over mathematics education", *Notices of the American Mathematical Society* **44** (1997), 695–702, 817–823.

[Jacob 2001] B. Jacob, "Implementing standards: The California mathematics textbook debacle", *Phi Delta Kappan* **83**:3 (2001), 264–272. Retrieved July 23, 2006 from http://www.pdkintl.org/kappan/k0111jac.htm.

[Jacob and Akers 2000] B. Jacob and J. Akers, "Research based mathematics education policy: The case of California 1995–1998", *International Journal for Mathematics Teaching and Learning* (2000). Available at http://www.cimt.plymouth.ac.uk/journal/bjcalpol.pdf. Retrieved 23 Jul 2006.

[Peterson and Hess 2006] P. E. Peterson and F. H. Hess, "Keeping an eye on state standards: A race to the bottom?", *Education Next* **6**:3 (2006), 28–29. Available at http://www.educationnext.org/20063/pdf/ednext20063_28.pdf. Retrieved 23 July 2006.

[Schoenfeld 2002] A. H. Schoenfeld, "Making mathematics work for all children: Issues of standards, testing and equity", *Educational Researcher* **31**:1 (2002), 13–25.

[Siegler 2003] R. S. Siegler, “Implications of cognitive science research for mathematics education”, pp. 219–233 in *A research companion to Principles and Standards for School Mathematics*, edited by J. Kilpatrick et al., Reston, VA: National Council of Teachers of Mathematics, 2003.

[Wilson 2003] S. Wilson, *California dreaming: Reforming mathematics education*, New Haven: Yale University Press, 2003.