Improved Mathematical Teaching Practices and Student Learning Using Complex Performance Assessment Tasks

Pamela L. Paek

Center for Assessment

David Foster

Silicon Valley's Mathematics Assessment Collaborative

Paper presented April 15, 2012 at the annual meeting of the National Council on Measurement in Education (NCME), Vancouver, Canada.

Abstract

This longitudinal study examines the use of complex performance tasks, which appears to improve both teachers' practices and students' mathematics performance on these tasks as well as high-stakes large-scale assessments. Teachers are better equipped for understanding what students know and how to shape and reshape their instruction based on student needs. In addition, the longer students were involved with teachers who had intensive coaching professional development, the higher their mathematics performance, which was consistent across grades 3-10 and in all 35 districts participating in the study. The results are significant not only for the districts in this particular study but also potentially for districts nationwide, as most states are transitioning to the Common Core State Standards, and teachers will need the types of intensive support and resources described in this paper to make the transition successful.

Key words: Coaching, Mathematics Learning, Performance Assessments, Teaching Practices

Historically in the United States, each individual state has independently developed its own academic mathematics standards and assessments. Since 2001, the state standards and assessments have had to be compliant with the federal No Child Left Behind Act, which required all states to assess all students in grades 3-8 and once in high school in both mathematics and English Language Arts. However, since each state developed their own standards and assessments, it was impossible to compare how students across states were mastering content. For instance, a concept in one state's sixth-grade mathematics standards might be taught and tested in another state's fourth-grade mathematics classrooms. The Obama administration wanted all students to be expected to meet the same rigorous mathematics content at the same grade level and depth of complexity, across any state, district, and school. Thus, the Common Core State Standards (CCSS) were developed in 2010.

The CCSS in mathematics is a set of mathematics standards that have been adopted by 45 of the 50 states, the District of Columbia, and one of the five territories in the United States. The 47 entities that have signed on have committed to teaching and assessing the same knowledge, skills, and abilities, and for the first time, the U.S. will have a very nearly national set of standards. The CCSS call for a substantial increase in the breadth and depth of mathematical knowledge students must acquire to be college and career ready. Most states that have compared their own state mathematics standards to the CCSS have found that they need to move what is taught to earlier grades and teach the material with more depth. More modeling and mathematical practices are being included and are expected to be tested on large-scale assessments, and two assessment consortia, each consisting of a large number of states, have been formed to create common assessments to be used across states.

Given the upcoming transition to the CCSS, it is important to identify effective resources that provide guidance to teachers on next steps in their teaching practices. The use of complex performance tasks is touted as critical in the two assessment consortia, and as a result, there is high need to find ways to transition teachers' use of these new assessment types as well as ways to build their capacity to use that information to truly improve their teaching and thus student learning. This work cannot happen without effective and strong professional development and support for teachers. It is critical that we find a way to develop and support a high-capacity cadre of teachers to ensure students are being taught the depth of the mathematics expected in the CCSS. This requires investment in professional development beyond what we have proven time and again not to work. The field needs to find programs that have shown success in improving mathematical understanding for both teachers and students. We have a long way to go to scale and sustain programs that will steadily raise the bar supporting teachers to be as successful as they can.

Theoretical Framework

The Common Core State Standards (CCSS) call for a substantial increase in the breadth and depth of mathematical knowledge students must acquire to be college and career ready. Unfortunately, few school districts have the capacity to help their students meet these rigorous requirements. National and state-level reports document critical shortages and high attrition in the overall supply of appropriately trained and certified mathematics teachers (National Science Board, 2006). Additionally, the majority of secondary mathematics teachers lack deep knowledge of the content they are expected to teach (Barth & Haycock, 2004; Massell, 1998).

The most significant variable for improving student learning has been shown to be the teacher (Boaler, 1998; Sanders, Horn, 1994; Schmidt, McKnight, Valverde, Houang & Wiley, 1997; Wright, Horn & Sanders, 1997), yet research shows inconsistencies in instruction across classrooms within the same district and even within the same school. Teachers interpret the same instructional ideas in various ways (Marzano, 2003; Stigler & Hiebert, 1998; 1999), and accordingly make independent decisions about whether and/or how to ignore, adapt, or adopt policymakers' recommendations for instruction (Spillane, Reiser, & Reimer, 2002). As such, mathematics instruction has proven very difficult to improve (Bamburg, 1994; Beck-Winchatz & Barge, 2003) and students are not given adequate opportunity to succeed in challenging mathematics (National Science Board, 2006).

In urban districts faced with these and other difficult issues—including heavy turnover among administrators, administrators who do not understand what is needed to support a high level of mathematics learning, and low expectations from both teachers and administrators for the performance of their students—mathematics instruction has proven very difficult to improve (Bamburg, 1994; Beck-Winchatz & Barge, 2003; Tauber, 1997). As a result, all too often, students in urban school districts are not given adequate opportunity to enroll—and succeed—in challenging mathematics (National Science Board, 2006).

To address this problem, school districts are pouring enormous quantities of resources to improve their mathematics programs' capacity to deliver a consistent and rigorous curriculum, aligned with state (and now national) standards and assessments, that prepares students for success in college and entry into high-quality workplaces. Yet despite these substantial investments, evidence of the effectiveness of district and school reform efforts varies greatly in quality and usefulness by district. Fullan (1991, p. 315) once summed up this mismatch: "Nothing has promised so much and has been so frustratingly wasteful as the thousands of workshops and conferences that led to no significant change in practice when teachers returned to their classrooms."

Description of the Practice

This study looks at the use of complex performance tasks for informing teaching practices. The Silicon Valley Mathematics Assessment Collaborative (MAC) was created to provide richer assessment information for teachers, schools, and districts to use to inform instruction, given that statewide assessment data did not provide the type of information practitioners needed to make changes in the classroom. Using the recommendations from the National Council of Teachers of Mathematics (NCTM), MAC outlined five core topics at each grade level they thought were worthy of teachers' efforts, of sufficient scope to allow for deep student thinking, and able to be assessed on an exam that lasted a single class period.

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) to design the exams. MARS is a project funded by the National Science Foundation, with roots in three universities: the University of California, Berkeley; Michigan State University; and the Shell Centre, Nottingham, England (http://www.nottingham.ac.uk/~ttzedweb/MARS/). The MARS exam is a performance assessment consisting of five tasks that assess concepts and skills at each grade level in addition to problem solving, reasoning, and communication skills. The tasks require students to evaluate, optimize, design, plan, model, transform, generalize, justify, interpret, represent, estimate, and calculate their solutions.

The MARS tasks 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 of points 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. Teachers are involved in scoring these performance assessments and talking about the depth of student learning as part of their professional development.

The combination of constructed-response tasks and weighted rubrics provides a detailed picture of student performance and 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. The reports also provide 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 wealth of resources provides a structure for teachers to study student understanding and misconceptions as evidence to help inform and adjust future instruction. By focusing on these detailed and structured analyses of student work on complex tasks, teachers can get a better sense of what students understand or what skills still need to be developed.

Questions

The larger project beyond this paper investigates an intensive coaching program and its impact on changes to teachers' understanding of mathematical content and their pedagogy as well as on students' mathematics performance, on both complex performance assessments and high-stakes large-scale assessments. The two main research questions ask how teachers' involvement in using complex performance tasks – involvement in trained scoring sessions, use of the data for formative feedback and discussions with colleagues and coaches – impact students' performance within a single year and across years:

1. What is the relationship with students' mathematics performance within a single year for teachers using complex performance assessments?
2. What is the relationship in performance of students who work with teachers using complex performance tasks across several years?

Data Sources and Methods

Quantitative data includes thirteen years of data, matching both teachers and students across the years since 2004. This data includes the number of years teachers have been involved in using MARS and the performance of their students on the MARS tasks and statewide assessment results, including demographic information for teachers (years teaching, gender, ethnicity, type of school) and students (gender, ethnicity). For one year, access to a comparable demographic of schools was used to compare teachers who working with MARS data to those who were not involved at all. Qualitative data includes feedback and rating of teachers by the coaches across the years.

The purpose of this study was to analyze the potential of using complex performance tasks on both teachers' pedagogy and students' mathematics performance. The results of these analyses across the 13 years and the accompanying trends were studied to see how the experience of teachers using complex tasks may be related with student performance within and across years. We began with exploratory analyses using analysis of variance, where student performance on the MARS and statewide assessment were used as dependent variables, disaggregated by teacher experience using MARS data, years teaching, and other demographic information within a single year. We then conducted some preliminary analyses looking across years to see how students' mathematical performance (as seen on both MARS and the statewide assessment) changed when they were taught by teachers using the MARS resources for examining student learning.

Results

Use of the MARS data shows promise for improving teachers' pedagogy and students' mathematical learning. Qualitative results show that teachers are better equipped for understanding what their students know and how to shape and reshape their instruction accordingly. Teachers spend more time talking about student work and find evidence of what students have learned rather than use anecdotal information to gauge students' understanding (Paek, 2008).

Not only do there appear to be changes in teachers' practices, but there also appear related to be improvements in student learning. For instance, Table 1 shows the pass rates on a large-scale mathematics assessment. Teachers who engaged in MARS tasks had a higher percentage of students passing the statewide assessment than those who did not. The differences in results are significant for all four grades.

[Insert Table 1 here]

In general, an increasing percentage of students taught by teachers using MARS tasks and their accompanying data are classified as proficient on the statewide mathematics assessment that has been in place since 2005 (note: only grades 3, 5, and 7 were assessed in 2005). Table 2 shows these increases from 2005 to 2011. However, in the middle school grades (grades 6-8) the percentage of proficient students vary across years without any clear pattern of improvement that is clearly seen in the elementary grade levels (grades 2-5) or in Algebra 1. The most drastic case is eighth grade, where the majority of students are not deemed proficient on the statewide assessment.

[Insert Table 2 here]

Since the MARS and statewide assessments require different types of mathematics' understanding – MARS requires students to explain and justify their answers on performance-based tasks while the statewide assessment is comprised of all multiple-choice items––we compared the percentage of proficient students on MARS to the percentage proficient on the statewide assessment. Even though the assessments measure different mathematics skills, one would expect mathematics proficiency to show a higher match of students meeting standards on both of these assessments. For grades 2-5, the majority of students who were proficient on the statewide assessment were also deemed proficient on the MARS exam. This finding is opposite for grades 6-8 and Algebra 1, as the majority of students are not proficient on both assessments. This finding is most striking for grade 8 students, as the highest percentage of proficient eighth graders on both assessments is one-quarter of the population across the years. It is clear that even with the continued increases of proficient students across the years and grades on the statewide exam, as students move through middle school, students are challenged in understanding both the basic concepts and higher-order thinking, especially with problem solving tasks.

[Insert Table 3 here]

Using performance on the MARS tasks, we were able to track both teachers and students over an eight-year period. The longer students were involved in classrooms with MAC-experienced teachers, the more likely they were to meet the performance standards on MARS. An example of the performance of a cohort of students in grades 4–7 on MARS taught by MAC teachers shows that before teachers had MAC coaching, only 30% of students met the standards, compared with 58% after one year, 66% after two years, and almost 100% of students with MAC teachers after three years.

Within MAC, there is more intensive formative assessment coaching for a subset of teachers. Results on the large-scale assessment and the MARS show even higher percentages of proficiency. For all grades, students with formative assessment coached teachers reached the proficiency cut-score for both MARS and the statewide multiple-choice assessment more than those who did not have this intensive training, but still utilized MARS data. Table 4 show that students are gaining conceptual understanding in more complex tasks where they need to justify and explain their reasoning (as on the MARS exams) and on the standardized assessment that span the content range of a given grade.

[Insert Table 4 here]

These findings are examples of what can be found in the many years of data we have yet to explore. Given the limitations of the length of this paper, the findings reported here represent what has been consistently found across the years, even with an increasing number of students and teachers participating. For instance, in 1999, 21 school districts, 462 teachers, and 23,128 students were involved in the study. The project reached the largest number of students in 2004 with 81,075 students and 1622 teachers in 28 districts. When funding for the project needed to be supported solely by district funds, the number of participating districts, teachers, and schools decreased. In 2011, 28 districts participated with 38,538 students.

Conclusions

Teachers need intensive support and resources to help them successfully transition to the new standards of the CCSS. This transition includes a shift in instruction and types of assessments, as the CCSS stress a new set of problem solving and skills more explicitly than most state assessments to date. As a result, the field needs to understand what teacher support around studying complex performance tasks has proven successful for promoting positive changes in teaching and learning. Additionally, the MARS tasks are in line with the two assessment consortia's descriptions of must-needed performance tasks, so better understanding the data from these assessments and their relationship to improved teaching practices is paramount for informing the pragmatic decisions that will be made in the upcoming years around what types of mathematics tasks assess deeper student learning and reflect improved teaching practices.

The Mathematics Assessment Collaborative has been able to demonstrate that utilizing complex performance tasks significantly enhances student achievement, both on state standardized achievement multiple-choice tests and on more complex performance-based assessments. An even more focused intervention around formative assessment using these tasks and data shows that students with such teachers have even further gains in both types of assessments. This finding is even true with the ever-increasing number of students and teachers involved every year, indicating that the process is scalable and sustainable, and not just privy to a privileged few. We need to find ways to capitalize on gains and findings like these to show what strong professional development looks like, and how such investments can truly improve mathematics teaching and learning.

Limitations And Next Steps

This paper is the first in a series studying the usefulness of MARS resources for teachers and their students, and thus provides only a sneak peek at the many layers of data to be studied.

It also does not include all of the results we have to date, which would demonstrate the consistency of findings that have been sampled and summarized in this paper.

This work used simple exploratory analyses and comparisons, and thus does not utilize the types of methodology for studying longitudinal nested data, which will provide a much clearer view of longer-term impact, following students and teachers over the years, controlling for previous performance and other trends that cannot be seen by simple year-to-year splicing of data. The plan is to use hierarchical linear modeling and potentially other nested models to provide information on how different sites may have varying results and to verify findings cross-sectionally across different grade spans. Qualitative case studies will be conducted to better understand practices, which will then be incorporated into improvements in the professional development resources for future use.

References

Bamburg, J. (1994). *Raising expectations to improve student learning*. Oak Brook, Illinois: North Central Regional Educational Laboratory.

Barth, P., & Haycock, K. (2004). A core curriculum for all students. In R. Kazis, J. Vargas, & N. Hoffman (Eds.). *Double the numbers: Increasing postsecondary credentials for underrepresented youth* (pp. 35-45). Cambridge, MA: Harvard Education Press.

Beck-Winchatz, B. & Barge, J. (2003). A new graduate space science course for urban elementary and middle school teachers at DePaul University in Chicago. *The Astronomy Education Review*, *1*(2), 120–128.

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

Fullan, M. (1991). *The new meaning of educational change*. London, UK: Cassell.

Fuhrman, S. (Ed.). (2001). From the capitol to the classroom: Standards-based reform in the states. Chicago, IL: University of Chicago Press.

Marzano, R. J. (2003). *What works in schools: Translating research into action*. Alexandria, VA: Association for Supervision and Curriculum Development.

Massell, D. (1998). *State strategies for building local capacity: Addressing the needs of standards-based reforms*. Philadelphia, PA: Center for Policy Research in Education, University of Pennsylvania.

National Science Board. (2006, January). *America's pressing challenge: Building a stronger foundation.* NSB 06-02. Washington, DC: Author.

Paek, P. L. (2008, January). Mathematics coaching: Silicon Valley Mathematics Initiative. Case study from *Practices worthy of attention: Local innovations in strengthening secondary mathematics*. Austin, TX: Charles A. Dana Center at The University of Texas at Austin.

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.

Spillane, J., Reiser, B., & Reimer, T. (2002). Policy implementation and cognition: Reframing and refocusing implementation research. *Review of Educational Research, 72,* 387–431.

Stigler, J. W., & Hiebert, J. (1998, Winter). Teaching is a cultural activity. *American Educator*. Retrieved March 14, 2007, from http://www.aft.org/pubs-reports/american_ educator/winter98/index.html.

Stigler, J. W., & Hiebert, J. (1999). *The teaching gap: Best ideas from the world's teachers for improving education in the classroom*. New York: Free Press.

Tauber, R. (1997). Self-fulfilling prophecy: A practical guide to its use in education. Westport, CT: Praeger.

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.

Table 1. Percentage of Proficient Students on the Mathematics Statewide Assessment

Grade/ Course	Percentage of Proficient Students with Non-MAC Teachers	Percentage Proficient Students with MAC Teachers
6	42	64
7	29	59
8	15	25
Algebra I	52	70

Table 2. Percentage of Students Above Proficient on the Statewide Assessment

	2005	2006	2007	2008	2009	2010	2011
Grade 2		71.9	72.8	72.0	75.5	76.4	79.8
Grade 3	70.0	69.9	70.3	71.0	74.4	76.6	79.0
Grade 4		66.8	67.8	75.9	77.0	75.9	78.7
Grade 5	56.0	59.8	62.1	63.2	65.4	70.1	76.8
Grade 6		60.0	58.7	53.6	55.3	62.8	61.3
Grade 7	51.0	52.6	54.3	53.7	52.7	62.1	61.6
Grade 8		37.9	31.6	40.8	36.9	41.5	42.0
Algebra 1		45.3	49.7	52.6	58.4	63.3	63.5

Table 3. Percentage of Students Above Proficient on MARS and the Mathematics Statewide Assessment

	2005	2006	2007	2008	2009	2010	2011
Grade 2		65.6	67.7	63.6	64.8	72.1	72.5
Grade 3	57.0	63.6	62.8	57.3	66.1	61.3	66.3
Grade 4		57.9	56.0	58.4	67.4	57.6	65.8
Grade 5	49.0	51.6	52.1	51.1	53.6	47.9	64.4
Grade 6		52.5	49.0	47.6	46.7	36.3	39.6
Grade 7	42.0	32.6	42.7	43.9	33.7	48.6	23.5
Grade 8		25.5	18.1	12.1	6.0	25.5	17.0
Algebra 1		35.2	36.7	37.9	35.6	43.2	42.0

Table 4. Percentage of Proficient Students on MARS and Statewide Assessments:

Intensive Vs. General Coaching and Support to Use of Complex Performance Assessments, 2009

Coaching and Support	Grade 6 MARS	Grade 7 MARS	Grade 8 MARS	Grade 6 statewide assessment	Grade 7 statewide assessment	Grade 8 statewide assessment
Intensive Formative Assessment	69	45	38	65	59	48
General MAC	45	28	20	50	48	35