SUPPORTING MATH TEACHERS IN THEIR TRANSITION TO THE COMMON CORE: IMPLEMENTATION AND EFFECTS OF MATHEMATICS DESIGN COLLABORATIVE

Deborah La Torre Matrundola
CRESST/University of California, Los Angeles

Abstract

This paper shares the results of an experimental study examining the relationship between changes in student understandings in mathematics concepts as evidence by an analysis of student artifacts and measures of teacher effectiveness in the implementation of MDC in ninth-grade Algebra I classes in Kentucky. This study included the development of rubrics to address two positive dimensions of learning, content accuracy and quality of mathematical explanations, as well as a third rubric examining evidence of common misconceptions. The study is framed in the literature on instructional change and it highlights the challenges in reaching the Common Core State Standards (CCSS) deeper learning goals (Heritage, 2013; Herman, 2013).

Study Overview and Purpose

According to surveys conducted by the Center on Education Policy of both state departments of education and school districts (Renter, 2013; Renter & Kober, 2014), the identification or development of Common Core aligned curricula is a challenge. Not only are the mathematics standards seen as more rigorous than previous state standards (Carmichael, Wilson, Porter-Magee, & Martino, 2010), but many deputy superintendents believe that successful implementation requires “fundamental changes in instruction” (Rentner, 2013, pg. 5).

The Mathematics Design Collaborative (MDC) initiative was designed to help teachers in this transition by integrating Classroom Challenges into their on-going curricula and instruction. Anchored in the Common Core standards and practices for mathematics, these two- to three-day formative assessment lessons (FALs) were designed to help teachers monitor and assess their students’ development of key skills and concepts. The Challenges are also intended to model and help teachers incorporate deeper mathematical reasoning and thinking into their practice. Towards this end, there are two primary types of Challenges, one focusing on conceptual understanding and the other on problem solving. At the time of the study, participating teachers could choose from among 40 Challenges at the high school level, as well as 61 Challenges geared towards middle school mathematics.

Designed by the Shell Center for Mathematical Education at the University of Nottingham in collaboration with the University of California, Berkeley, each Challenge follows the same general structure: 1) Students complete a pre-assessment to assess skills and reasoning; 2) Teachers review student responses to determine existing understandings and misconceptions; 3) Students are engaged in whole class and small group activities in order to provide feedback and discuss alternative approaches; 4) Students complete a post-assessment where they can revise their initial responses and reflect on their new understandings (see http://map.mathshell.org for more information about the Challenges).

This paper reports on a study funded by the Bill and Melinda Gates Foundation that examined the implementation of MDC in ninth-grade Algebra 1 classrooms in Kentucky. More specifically, the aspect of the study reported here examines the relationship between changes in student understandings and misconceptions of mathematics concepts as evidenced by an analysis of student artifacts and measures of teacher effectiveness in the implementation of MDC. This aspect of the study aimed to answer the following research question: What conditions and contexts, including implementation quality, influence MDC effectiveness?

Background

We locate our research in the literature on instructional change. Although classroom practice is notoriously impervious to reform (Cuban, 1982; Lortie, 1975), an emerging body of research has documented the relationship between student achievement and specific instructional practices that create "opportunities-to-learn" (see Bryk, 2010; Rowan & Correnti, 2009; Winters & Herman, 2011). Our implementation focus is classroom instruction, while recognizing that multiple factors influence and inhibit teacher innovation and instructional change (see, for example, Desimone, Porter, Garet, Yoon, & Birman, 2002; Fullan, 2007; Owston, 2003). Our ultimate goal is to leverage the findings from our analysis of student artifacts to inform efforts required for teachers to implement rigorous Common Core demands successfully.

Data and Methodology

The overall sample included 46 Algebra 1 teachers and their students in six districts and 17 schools across the state of Kentucky who participated in Phase I and/or II implementation of MDC. Of these teachers, 28 submitted student artifacts as part of their participation in the implementation study. Participating teachers were asked to focus their study submissions on one classroom, which was selected by the research team using a random number generator.

Measures in the analyses presented include the following.

MDC Classroom Artifacts

Each teacher participating in the implementation study was asked to submit four to six web-based logs over the course of the school year. Each log focused on the fidelity and quality of implementation of a Challenge. As part of the log process, teachers were also asked to submit electronic or hard copies of students' pre- and post-assessment responses for each Challenge.

Students' pre-to-post improvement on the assessments was treated as a measure of teachers' quality of MDC implementation. Towards this end, three rubrics were developed to examine each Challenge. These included rubrics to address two positive dimensions of learning, content accuracy and quality of mathematical explanations, while a third rubric examined evidence of misconceptions. Accuracy items and misconceptions were generally scored on a scale of 0–1. In contrast, the more cognitively complex explanations were scored on a scale of 0–3, with 0 meaning no response, 1 indicating that there was no evidence of conceptual understanding, 2 indicating partial understanding, and 3 indicating there was evidence of full understanding.

Since the specific Challenges implemented varied by school and often by teacher, it was decided to score linked pre- and post-assessments from the four most commonly implemented by the study sample. A minimum of 10% of responses for each assessment item were double-coded by specially trained members of the research staff, reaching reliability of $\geq .80$ per item and $\geq .90$ per Challenge. Our a priori hypothesis was that greater pre-to-post learning progress could serve as an indicator of higher quality MDC implementation.

ACT Plan Assessment Data

All Kentucky students take ACT's Plan test in the fall of tenth grade. The mathematics component of the Plan is a 40-minute multiple-choice test, including 22 items addressing algebra and pre-algebra content and 18 items addressing geometry. The test yields subscores for each of these areas. Score reliabilities for the subscores and total scores ranged from .65 to .86, with a reliability of above .80 for algebra, the primary area of interest in the study of MDC (ACT, 2013). Tenth-grade scores from the fall of 2013 served as an outcome for a larger quasi-experimental design (QED) analysis of MDC effects, and were also used to examine the relationship between pre-post learning progress and teacher effectiveness (see Herman et al., 2015 for results of the larger QED study).

Results and Conclusions

As was previously noted, student work was collected as part of the log process to create an indicator of implementation quality. While teachers submitted class sets for 18 of the Challenges, our analyses focused solely on four of the most commonly administered by participating ninth-grade Algebra 1 teachers. Furthermore, of the 28 teachers, only 29–57% submitted matched pre-post class sets for each of these Challenges (see Table 1). Because of this lack of representativeness within and across Challenges, results should be interpreted with caution.

Table 1

Number of Classroom Sets of Student Work Analyzed by Classroom Challenge (n = 28)

Classroom Challenge	Pre and/or post		Matched pre-post	
	Class sets (*n*)	Teachers (%)	Class sets (*n*)	Teachers (%)
Solving Linear Equations in Two Variables (FAL03)	17	60.7	16	57.1
Sorting Equations and Identities (FAL13)	16	57.1	16	57.1
Interpreting Algebraic Expressions (FAL16)	13	46.4	13	46.4
Finding Equations of Parallel and Perpendicular Lines (FAL22)	11	39.3	8	28.6
Total (one or more common Challenges)	25	89.3	25	89.3

Table 2

Mean Score Percentages for Individual Classroom Challenges by Rubric Type and Assessment

	n	Accuracy	Explanation	Misconceptions
Solving Linear Equations in Two Variables (FAL03)				
Pre-assessment	307	0.45 (0.28)	0.23 (0.18)	0.69 (0.19)
Post-assessment	307	0.53 (0.29)	0.29 (0.21)	0.72 (0.18)
Sorting Equations and Identities (FAL13)				
Pre-assessment	96	0.37 (0.18)	0.30 (0.18)	0.81 (0.15)
Post-assessment	96	0.58 (0.19)	0.43 (0.20)	0.89 (0.11)
Interpreting Algebraic Expressions (FAL16)				
Pre-assessment	96	0.63 (0.18)	0.35 (0.27)	0.50 (0.25)
Post-assessment	96	0.72 (0.16)	0.41 (0.27)	0.61 (0.26)
Finding Equations of Parallel and Perpendicular Lines (FAL22)				
Pre-assessment	56	0.21 (0.13)	0.23 (0.21)	0.43 (0.28)
Post-assessment	56	0.27 (0.17)	0.33 (0.23)	0.38 (0.28)

Note. For this analysis, misconceptions were reverse coded so that higher means indicate a lower presence of misconceptions. Standard deviations presented in parentheses.

Descriptive Results

Table 2 presents descriptive statistics for the pre- and post-assessments that were matched for individual students. Although results were generally low, considerable variation was found.

On the pre-assessments, the average student only earned 23–35% of possible points on the explanation items and 21–63% of possible points on the accuracy items. Furthermore, students received an average misconceptions score of 43–81% on their pre-assessments. It should be noted that misconceptions scores were reverse coded so that a higher score indicates a lower presence of misconceptions. Post-assessment performance also showed variation across Challenges. More specifically, average scores on the post-assessments ranged from 29–43% on the explanation items, from 27–72% on the accuracy items, and from 38–89% on the misconceptions scores.

Part of the explanation for this variability in performance at both time points may involve differences in the focus for each of these Challenges (see Appendix). For example, FAL16, on which the average student tended to score higher, focused on algebraic expressions, a content area initially introduced in middle school, while FAL22, with which students tended to most struggle, focused on high school standards involving geometric theorems and formal representations of functions such as $y = mx + b$ and graphs. In contrast, both FAL03 and FAL13 focused on linear equations, content that has been traditionally taught in Algebra 1, but with which many students still struggle.

Table 3

Mean Percentage Change in Performance for Individual Classroom Challenges by Rubric Type

	n	Accuracy	Explanation	Misconceptions
Solving Linear Equations in Two Variables (FAL03)	307	0.08 (0.35)	0.06 (0.22)	0.02 (0.24)
Sorting Equations and Identities (FAL13)	96	0.21 (0.21)	0.14 (0.24)	0.08 (0.17)
Interpreting Algebraic Expressions (FAL16)	96	0.09 (0.22)	0.06 (0.31)	0.11 (0.34)
Finding Equations of Parallel and Perpendicular Lines (FAL22)	56	0.06 (0.17)	0.09 (0.23)	-0.05 (0.29)

Less variability was found when examining pre- to post-assessment changes in performance (Table 3). More specifically, average student accuracy and explanation scores increased by only 6–9% for three of the Challenges (i.e., FAL03, FAL16, and FAL22). In contrast, FAL13, which was the only Challenge to include multiple-choice items, showed an average improvement of 21% on the accuracy items and 14% on the explanation items. At the same time, FAL13 showed little improvement (8%) in regards to evidence of misconceptions, which may indicate improvement in computation rather than in conceptual understanding of algebra. It also should be noted that FAL22 was the only Challenge to show a slight average increase in misconceptions pre-to-post.

Relationship between the Artifact Implementation Measure and Teacher Effectiveness

Using the results of the artifact analysis, the research team created indicators of implementation quality. These indicators were based on students' pre- to post-assessment improvement in content accuracy (i.e., short answer items, multiple-choice items), quality of students' explanations, and the presence of misconceptions that are common for the targeted standards. Although recognizing serious limitations in the data, including the small sample size, we then conducted exploratory analyses of the relationship between these measures of quality and teacher effectiveness.

A first step in our exploratory analysis was to run a hierarchical linear model (HLM) with the purpose of identifying individual MDC teacher's effectiveness on improving student performance on Classroom Challenge assessments. As was previously noted, because teachers submitted artifacts for different Challenges, it was important to control for potential differences in the Challenges by limiting our ratings and estimates to four of the most common ones submitted. A two-level HLM model was used to estimate performance on the post-assessments by controlling for the effect of each Challenge as well as pre-assessment scores at both the student and teacher levels. Separate models were initially run for each rating and then, based on the results, another model was run that combined the mean scores for both the accuracy and explanation ratings. Bayes estimates for each teacher's effectiveness at improving student performance on the Challenge assessments for each rating were saved.

Table 4

Pearson Correlations for Teacher Effectiveness on Classroom Challenge ratings and ACT Plan scores

	1	2	3	4	5
1. Accuracy & explanation (combined)	---				
2. Accuracy	.839**	---			
3. Explanation	.775**	.487	---		
4. Misconceptions	.248	.493	-.198	---	
5. ACT Plan Algebra	.746**	.581*	.527*	.251	---

Note. Aggregate scores for item types control for student pre-assessment scores.
* $p \leq 0.05$ level. ** $p \leq 0.01$ level.

A second step was to examine the relationship between teacher effectiveness on each of the Challenge assessment ratings and teacher effectiveness on state ACT Plan scores. During this stage of the analysis, we explored correlations for the effectiveness scores for the individual Challenge ratings as well as for the combined accuracy-explanation score. Table 4 presents the correlations between these various effectiveness scores. As can be seen, both the accuracy and explanation scores showed strong positive relationships with teacher effectiveness on the ACT Plan, and the combined accuracy-explanation score showed a very strong, significant relationship. That is, teachers who better implemented the Challenges, as evidenced by their students' improvement from pre- to post- on Challenge assessments, also were more effective in affecting student scores on the ACT Plan state assessment.

Conclusions

While the QED analysis conducted for the larger study of MDC did find a positive, statistically significant effect on students' ACT Plan performance in algebra, descriptive findings from the artifact analysis highlight the challenge of moving to standards that are more rigorous and performance tasks that align with math practices involving greater cognitive complexity (e.g., assessment items requiring explanations). We find evidence for this in students' low mean scores on the assessments from the common Classroom Challenges, the generally small changes found pre-to-post, and the general lack of improvement in the prevalence of common misconceptions. Furthermore, results of the HLM analyses showed a strong positive, albeit highly tentative relationship between students' learning as evidenced by their performance on the tenth-grade ACT Plan state assessment and the quality of MDC implementation suggested by pre-post changes in the quality of student explanations and the accuracy of their short answer and multiple-choice responses.

References

ACT. (2013). *ACT Plan technical manual 2013-2014*. Iowa City, IA: Author. Retrieved from https://www.act.org/content/dam/act/unsecured/documents/PlanTechnicalManual.pdf

Bryk, A. (2010). Organizing schools for improvement. *Phi Delta Kappan, 91(7),* 23-34.

Carmichael, S.B., Wilson, W.S., Porter-Magee, K., & Martino, G. (2010). *The state of state standards and the Common Core—in 2010*. Washington, DC: Thomas B. Fordham Institute.

Cuban, L. (1982). Persistent instruction: The high school classroom, 1900-1980. *Phi Delta Kappan,64*(2), 113-118.

Desimone, L., Porter, A. C., Garet, M. S., Yoon, K. S., & Birman, B. F. (2002). Effects of professional development on teachers' instruction: Results from a three-year longitudinal study. *Educational Evaluation and Policy Analysis*, 24(2), 81–112.

Fullan, M. (2007). *The new meaning of educational change*. New York, NY: Teachers College Press.

Heritage, M. (2013). *Formative assessment in practice*. Boston, MA: Harvard Education Press.

Herman, J. (2013). *Formative assessment for Next Generation Science Standards*. ETS Invitational Research Symposium. Washington, DC.

Herman, J. L., La Torre Matrundola, D., Epstein, S., Leon, S., Dai, Y., Reber, S., & Choi, K. (2015). *The implementation and effects of the Mathematics Design Collaborative (MDC): Early findings from Kentucky ninth-grade Algebra 1 courses* (CRESST Report 845). Los Angeles, CA: University of California, National Center for Research on Evaluation, Standards, and Student Testing (CRESST).

Lortie, D. (1975). *Schoolteacher: A sociological study*. Chicago, IL: University of Chicago Press.

Owston, R.D. (2003) School context, sustainability and transferability of innovation. In R. Kozma, *Ed.Technology, innovation and educational change: A global perspective*. Eugene, OR: ISTE.

Rentner, D.S. (2013). *Year 3 of implementing the Common Core State Standards: An overview of sates' progress and challenges*. Washington, DC: Center on Education Policy.

Rentner, D.S., & Kober, N. (2014). *Common Core State Standards in 2014: Curriculum and professional development at the district level*. Washington, DC: Center on Education Policy.

Rowan, B., & Correnti, R. (2009). Studying reading instruction with teacher logs: Lessons from the study of instructional improvement. *Educational Researcher*, *28(2),* pp. 120-131.

Winters, L., & Herman, J. (2011). *Turn-around toolkit.* Thousand Oaks, CA: Corwin Press.

Appendix:
Classroom Challenges

Table 1

Overview of the Common Classroom Challenges

Challenge	Goals	Standards	Practices	Structure
FAL03	• Solving a problem using two linear equations with two variables. • Interpreting the meaning of algebraic expressions.	Algebra: • CED: Create equations that describe numbers or relationships. • REI: Solve systems of equations.	2. Reason abstractly and quantitatively. 3. Construct viable arguments and critique the reasoning of others.	Accuracy: 3 Explanation: 4 Misconceptions: 4
FAL13	• Recognize the differences between equations and identities. • Substitute numbers into algebraic statements in order to test their validity in special cases. • Resist common errors when manipulating expressions such as $2(x - 3) = 2x - 3$; $(x + 3)^2 = x^2 + 3^2$. • Carry out correct algebraic manipulations.	Algebra: • SSE: Interpret the structure of expressions. Write expressions in equivalent forms to solve problems. • REI: Solve equations and inequalities in one variable.	3. Construct viable arguments and critique the reasoning of others. 7. Look for and make use of structure.	Accuracy: 10 Explanation: 1 Misconceptions: 6
FAL16	• Recognizing the order of algebraic operations. • Recognizing equivalent expressions. • Understanding the distributive laws of multiplication and division over addition (expansion of parentheses).	Algebra: • SSE: Interpret the structure of expressions. • APR: Rewrite rational expressions.	2. Reason abstractly and quantitatively. 7. Look for and make use of structure.	Accuracy: 9 Explanation: 3 Misconceptions: 1
FAL22	• Find, from their equations, lines that are parallel and perpendicular. • Identify and use intercepts. • It also aims to encourage discussion on some common misconceptions about equations of lines.	Geometry: • PE: Use coordinates to prove simple geometric theorems algebraically. Functions: • IF: Analyze functions using different representations.	1. Make sense of problems and persevere in solving them. 3. Construct viable arguments and critique the reasoning of others. 7. Look for and make use of structure.	Accuracy: 2 Explanation: 1 Misconceptions: 3