Examining the Fidelity of Implementation of Early Algebra Intervention and Student Learning

Michael Cassidy
TERC
Michael_Cassidy@terc.edu

Rena Stroud
TERC
rena_stroud@terc.edu

Despina Stylianou
The City College of New York
dstylianou@ccny.cuny.edu

Maria Blanton
TERC
Maria_Blanton@terc.edu

Angela Gardiner
TERC
angela_gardiner@terc.edu

Eric Knuth
University of Wisconsin
knuth@education.wisc.edu

Ana Stephens
University of Wisconsin
acstephens@wisc.edu

Citation: Cassidy, M., Stroud, R., Stylianou, D., Blanton, M., Gardiner, A., Knuth, E. and Stephens, A. (2017). Examining the fidelity of implementation of an early algebra intervention and student learning. In Galindo, E., & Newton, J., (Eds.), Proceedings of the thirty-ninth annual meeting of the North American Chapter of the International Group for the Psychology of Mathematics Education (pp. 299–302). Hoosier Association of Mathematics Teacher Educators.

Through a theoretical framework emphasizing the importance of fidelity of implementation (FOI), this paper explores how 3rd and 4th grade teachers implemented an early algebra intervention, and the extent to which the FOI related to student learning. The data for this report are taken from the first two years of an experimental research project. Videotaped classroom observations, our primary measure of FOI, were coded by adding to and adapting the Mathematical Quality of Instruction (MQI) instrument, and student performance was measured by overall score (correctness) on an algebra assessment. Results revealed a significant positive relationship between teachers' implementation and their students' performance.

Keywords: Algebra and Algebraic Thinking, Elementary School Education, Assessment and Evaluation

This paper reports on the fidelity of implementation (FOI) of 3rd and 4th grade teachers as they implemented an early algebra intervention, and the relationship between FOI and student learning. The data from this study are taken from the first two years of an experimental research project (Project LEAP: Learning through an Early Algebra Progression) that tests the hypothesis that children who receive comprehensive, longitudinal early algebra instruction during the elementary grades are better prepared for algebra in middle school than children who have only arithmetic-based experiences during elementary grades. Through a focus on classroom observations and student assessment data, this paper explores differential patterns of implementation and student learning.

Theoretical Framework

The treatment of algebra in school mathematics has changed dramatically over the past two decades. The *Common Core State Standards for Mathematics* (National Governors Association Center for Best Practices and Council of Chief State School Officers [NGA Center & CCSSO], 2010) calls for algebraic reasoning to start in Grade K and span across the grades. In response to this challenge we initiated a study to examine the effectiveness of an early algebra intervention in Grades 3-5. The intervention consisted of 18 lessons per grade level that were taught throughout the school year. Teachers also attended ongoing professional development to support their implementation of the intervention.

The focus in this paper is on the relationship between student performance outcomes and teachers' FOI of the early algebra intervention. Our goal was to measure the fidelity with which teachers in diverse demographic settings implemented the intervention and how this intervention affected student learning outcomes. The degree of FOI has provided key insights into the viability of the implementation and, thus, helped us interpret findings about student performance. Measuring FOI revealed differential patterns of implementation and their relationship to the intervention (Boruch & Gomez, 1977; Mowbray et al., 2003), a critical factor in evaluating an intervention's effectiveness (NRC, 2004; Summerfelt & Meltzer, 1998) and promoting external validity (O'Donnell, 2008). While it is often assumed that students in experimental conditions receive comparable treatment, large variations in implementation might actually exist (Harachi et al., 1999).

Our efforts reflect a general interest for research in mathematics education to have an impact on practice at a large and measurable scale. Indeed, recently, the editors of the Journal for Research in Mathematics Education reiterated the question that underlies all our work: How can educational research have a larger impact on practice? (Cai, et al., 2017). Similarly, the National Council of Teachers of Mathematics Research Committee (Herbel-Eisenmann et al., 2016) recently deliberated the ways research can influence actual instruction and learning. Thus, our research report is framed around the question: To what extent does teachers' fidelity of implementation influence student learning?

Methodology

Three school districts (including approximately 240 classrooms and 3,400 children) participated in the cluster randomized trial, with entire schools being assigned to either the experimental or control condition. During the first year of implementation, we focused on third grade classrooms, and subsequently followed these children into fourth grade.

Data sources for the study reported here are classroom observations and student assessment data. Classroom observations (videotaped LEAP lessons) were conducted with a subsample of experimental teachers (grade 3: n=50, grade 4: n=45). The majority of teachers (n=78) were observed twice and the remaining teachers (n=17) were observed once, for a total of 173 observations.

Student participants were given a one-hour, written algebra assessment as a pre/post measure in Grade 3 and a related assessment as a post measure in Grade 4. The assessments were designed by the project team (Blanton et al, 2015) and measured students' understanding of early algebra along four main components: (1) Generalized Arithmetic; (2) Equivalence, Expressions, Equations, and Inequalities; (3) Functional Thinking; and (4) Variable. We have assessment data from approximately 800 students in the observed classrooms. For the purposes of the analyses reported here, student assessment data were coded according to item correctness.

We coded classroom observation data with a specific focus on the degree to which teachers implemented the early algebra materials with fidelity as well as the quality of mathematics instruction. Teachers were rated on 5-point Likert scales on each of the three cognitive demand variables created by the project team: *justify an answer*, *generalize a mathematical relationship*, and *represent with variables*. Separate codes were given for whole class work and individual/group work. Observations were also coded for six items adapted from the Mathematical Quality of Instruction (MQI) instrument (Hill et al, 2008).

- *Efficiency:* extent to which lesson time is used efficiently and class is on task.
- *Distorted:* extent to which the mathematics of the lesson is clear and correct and not distorted.
- *Engaged:* extent to which the classroom environment is characterized by student engagement.
- *Student Difficulty:* extent to which the teacher attends to students' struggles and challenges when working through the material.
- *Uses Student Ideas:* extent to which the teacher uses student ideas and solutions to move the lesson forward.
- *Imprecision:* extent to which there are incorrect uses of mathematical language or notation.

Approximately 15% of videos were double coded in order to assess inter-rater reliability. Factor analysis was then employed in order to create composite variables that could be used as teacher-level predictors of student outcomes (i.e., student performance on the early algebra assessments) in a multilevel analysis.

Results

As a first step, we looked at the implementation of the intervention. Lessons in this early algebra intervention consist of two parts – the Jumpstart (a review and warm up activity related to the objectives of the lesson) and the main early algebra activity. In Grades 3 and 4 combined, the jumpstart activity was completed in 96% of observed classes (see Table 1). Of those that did, 100% of the observed classes included whole class discussion led by the teacher, 66% included individual student work, 42% included group work, and 34% included student-led presentations. Jumpstarts activities lasted, on average, 16 minutes and 13 seconds (*SD* = 07:03).

Table 1: Summary of Jumpstart Activities

	Grade 3	Grade 4	Overall
Jumpstart	95%	97%	96%
Whole class discussion	100%	100%	100%
Individual student work	67%	64%	66%
Group work	32%	54%	42%
Student-led presentations	35%	32%	34%
Average time	15m 32s (7m54s)	16m 01s (6m 52s)	16m 13s (7m 3s)

After completing the Jumpstart, lessons moved on to the early algebra task (see Table 2). Overall, in 88% of observed classrooms teachers read the problem aloud or had a student read the problem. In 47% of classrooms, teachers ensured students understood any terms or concepts that might be unfamiliar to them. In 36% of classrooms, teachers demonstrated methods of presenting numerical information that might be unfamiliar to students, and of those who did so, 98% used information from the worksheet itself.

After the lesson was introduced, in 74% of the observed classrooms students worked on their own (either independently or in groups) on the bulk of the remainder of the activity. During individual/group work, teachers were rated on whether they were active or passive. An active teacher would actively visit individuals/groups to help students with questions, but also to challenge their thinking. A passive teacher would go around to groups only when asked, and would be largely reactive to students' needs rather than being proactive and challenging mathematical thinking. In 72% of observed classrooms, the teacher was coded as "active."

Table 2: Summary of Early Algebra Activities

	Grade 3	Grade 4	Overall
Read aloud	90%	85%	88%
Define	52%	41%	47%
Demo	42%	28%	36%
Demo info from worksheet	98%	100%	98%
Post-intro work	71%	78%	74%
Active teacher	69%	77%	72%

LEAP Cognitive Demand Variables: Justify, Generalize, Represent

A rating of 1 indicates that a teacher did not ask students to justify, generalize, or represent at all, while a rate of 5 indicates that the teacher asked students to justify, generalize, or represent in a way that went beyond the lesson expectations (see Table 3). Because elementary school teachers were participating in algebraic content professional development for the first time, we conjectured teachers to rarely go beyond the lesson expectations.

Table 3: Mean Justify, Generalize, and Represent Ratings

		Grade 3 (SD)	Grade 4 (SD)	Overall (SD)
Justify				
	Whole Class	3.75 (1.02)	3.62 (0.89)	3.69 (0.97)
	Individual/Group	2.91 (1.39)	2.66 (1.20)	2.79 (1.32)
Generalize				
	Whole Class	3.51 (1.08)	3.54 (1.04)	3.52 (1.06)
	Individual/Group	2.55 (1.33)	2.32 (1.23)	2.45 (1.29)
Represent				
	Whole Class	3.00 (1.25)	3.12 (1.15)	3.05 (1.20)
	Individual/Group	2.10 (1.34)	1.93 (1.10)	2.02 (1.24)

Adapted MQI Codes

With the exception of *imprecision*, which was coded on a 4-point scale, all items were coded using a 5-point Likert scale. For all items, 1 is the most "negative" rating, indicating inefficient use of class time, severely distorted mathematics, total lack of student engagement with the lesson, student difficulty without any teacher remediation, no substantive use of student ideas, or imprecision that obscured the mathematics of the lesson (see Table 4).

Table 4: Mean Adapted MQI Ratings

	Grade 3 (SD)	Grade 4 (SD)	Overall (SD)
Efficiency	3.33 (1.11)	3.43 (1.05)	3.38 (1.08)
Distorted	4.24 (.96)	4.51 (0.81)	4.36 (0.91)
Engaged	3.80 (1.00)	3.61 (1.06)	3.72 (1.03)
Student Difficulty	3.56 (.99)	3.59 (1.11)	3.57 (1.04)
Uses Student Ideas	3.85 (.93)	4.05 (1.12)	3.94 (1.02)
Imprecision	3.31 (.83)	3.51 (0.69)	3.40 (0.78)

Inter-rater reliability was assessed for the cognitive demand and MQI data using weighted Cohen's kappa. The analysis suggested that raters had acceptable levels of agreement (κ_w>.60) for all MQI variables, for the three individual/group cognitive demand variables, and for the whole class represent variable. However, there was only moderate agreement for two of the whole class cognitive demand variables: whole class generalize (κ_w =.54) and whole class justify (κ_w =.53). For this reason, subsequent analyses do not include the whole class variables.

Relationship Between FOI and Student Performance

We hypothesized that several of our observed variables would be correlated due to their association to latent (unobserved) variables. In order to identify these underlying latent variables, factor analysis using principal components analysis was utilized. For teachers who were observed twice, codes were first averaged across to create one set of variables per teacher. Separate analyses were then conducted for Grades 3 and 4.

The six MQI variables and the three individual/group cognitive demand variables were entered. In Grade 3, a two-factor solution, which explained 71% of the variance, was preferred (see Table 5), while in Grade 4 a three-factor solution, which explained 82% of the variance, was preferred (see Table 6).

Table 5: Grade 3 Factor Loadings Based on a PCA with Varimax Rotation

	Factor 1	Factor 2
Justify – Ind/Group		0.86
Generalize – Ind/Group		0.84
Represent – Ind/Group		0.78
Efficiency	0.75	
Distorted	0.84	

Engaged	0.72	
Student Difficulty	0.74	0.50
Uses Student Ideas	0.70	0.48
Imprecision	0.83	

Note. Factor loadings < .4 are suppressed.

Table 6: Grade 4 Factor Loadings Based on a PCA with Varimax Rotation

	Factor 1	Factor 2	Factor 3
Justify – Ind/Group		0.75	
Generalize – Ind/Group		0.88	
Represent – Ind/Group		0.87	
Efficiency	0.79		
Distorted			0.92
Engaged	0.87		
Student Difficulty	0.83	0.50	
Uses Student Ideas	0.77	0.48	
Imprecision			0.91

Note. Factor loadings < .4 are suppressed.

In Grade 3, two factors emerged: The three individual/group cognitive demand codes (justify, generalize, represent) were added together to create a composite variable ("cognitive demand"), and the six MQI variables were added together to create a composite variable ("MQI") (see Table 7).

In Grade 4, three factors emerged: 1) the three individual/group cognitive demand codes (justify, generalize, represent) were added together to create a composite variable ("cognitive demand"), 2) "Imprecision" and "Distorted" were added together to create a composite variable ("teacher precision"), and the remaining four MQI variables ("Efficiency," "Engaged," "Student Difficulty," and "Uses Student Ideas") were added together to create a composite variable ("student-focused") (see Table 7).

Table 7: Descriptive statistics for composite variables in Grades 3 and 4.

	Mean (*SD*)	Minimum	Maximum
Grade 3			
cognitive demand	7.55 (*2.82*)	3	14
MQI	22.25 (*3.79*)	14.5	29
Grade 4			
cognitive demand	6.97 (*2.91*)	3	14
teacher precision	7.94 (1.33)	4	9
student-focused	14.49 (*3.52*)	4	20

Using these composite variables as level 2 (teacher-level) predictors, we conducted separate multilevel regression analyses at Grades 3 and 4 to explore the relationship between teacher-level FOI variables and student performance on the LEAP assessment. Baseline measures of performance were included as student-level (level 1) predictors, and a measure of school SES (percentage of students with free or reduced lunch) was included as a school-level (level 3) factor.

In Grade 3, after controlling for baseline performance (Grade 3 pre-test) and school-level SES, the cognitive demand composite variable was found to be a significant predictor of students' score on the Grade 3 LEAP post-assessment ($\gamma = .015$, $t(30) = 3.093$, $p < .001$). A one-unit increase in cognitive demand score was associated with a 1.5% increase (.015 points) in post-test score. The MQI composite variable was not a significant predictor. In other words, the extent to which teachers emphasized the core algebraic practices of generalizing, representing and justifying generalizations in LEAP lessons had a significant, positive impact on student performance.

In Grade 4, after controlling for baseline performance (Grade 3 pre- and post-test) and school-level SES, the student-focused composite variable was found to be a significant predictor of students' score on the Grade 4 LEAP post-assessment ($\gamma = .010$, $t(22) = 3.940$, $p < .001$). A one-unit increase in student-focused score was associated with a 1.0% increase (.010 points) in post-test score. The cognitive demand and teacher precision composite variables were not significant predictors.

Discussion

Understanding the ways in which teachers implemented the early algebra intervention and how it impacted student learning can have important implications for this particular study and, more broadly, how we as a community understand the complexity of finding ways for educational research to influence actual instruction and have an impact on the mathematics learning of large numbers of students. This is critical if we are to take educational innovations to scale.

In both Grades 3 and 4, aspects of teachers' implementation were significantly positively related to their students' performance on the early algebra assessments. In Grade 3, given the range of the cognitive demand composite variable (3 to 14), students in classrooms where the teacher received the highest rating outperformed their peers in the classroom of the lowest rated teachers by an average of 16.5%. Similarly, in Grade 4, given the range of the student-focused composite variable (4 to 20), students in classrooms where the teacher received the highest rating outperformed their peers in the classroom of the lowest rated teachers by an average of 16%. Therefore, students of teachers who implemented the intervention with higher fidelity had higher mean scores on the early algebra assessment. Which leads us to believe that these students are better prepared for algebra in middle school than children whose teachers implement the intervention with lower fidelity.

Acknowledgements

The research reported here was supported by the US Department of Education under IES Award #R305A140092. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the US Dept of Education.

References

Blanton, M., Stephens, A., Knuth, E., Gardiner, A., Isler, I., & Kim, J. (2015). The Development of Children's Algebraic Thinking: The Impact of a Comprehensive Early Algebra Intervention in Third Grade. *Journal for Research in Mathematics Education, 46*(1), 39-87.

Boruch, R. F., & Gomez, H., (1977). Sensitivity, bias, and theory in impact evaluation. *Professional Psychology*, *8*(4), 411–433.

Cai, J., Morris, A., Hwang, S., Hohensee, C., Robinson, V. & Hiebert, J. (2017), Improving the impact of educational research. *Journal for Research in Mathematics Education*, *48*(1), 2-6.

Common Core State Standards Initiative. (2010). National Governors Association Center for Best Practices and Council of Chief State School Officers.

Harachi, T. W., Abbott, R. D., Catalano, R. F., Haggerty, K. P., & Fleming, C. B. (1999). Opening the black box: Process evaluation measures to assess implementation and theory building. *American Journal of Community Psychology*, *27*(5), 711–731.

Herbel-Eisenmann, B., Sinclair, N., Chval, K. B., Clements, D. H., Civil, M., Pape, S. J., . . . Wilkerson, T. L. (2016). Positioning mathematics education researchers to influence storylines. *Journal for Research in Mathematics Education*, *47*(2), 102–117.

Hill, H. C., Blunk, M. L., Charalambous, C. Y., Lewis, J. M., Phelps, G. C., Sleep, L., & Ball, D. L. (2008). Mathematical knowledge for teaching and the mathematical quality of instruction: An exploratory study. *Cognition and instruction*, *26*(4), 430-511.

Mowbray, C., Holter, M. C., Teague, G. B., & Bybee, D. (2003). Fidelity criteria: Development, measurement, and validation. *American Journal of Evaluation*, *24*(3), 315–340

National Research Council. (2004). *On curricular effectiveness: Judging the quality of K-12 mathematics evaluations.* Washington, DC: National Academy Press.

O'Donnell, C. (2008). Defining, conceptualizing, and measuring fidelity of implementation and its relationship to outcomes in K-12 curriculum intervention research. *Review of Educational Research 78*(1), 33–84.

Summerfelt, W. T., & Meltzer, H. Y. (1998). Efficacy vs. effectiveness in psychiatric research. *Psychiatric Services, 49*(6)*,* 834–835.