American Educational Research Journal
Month XXXX, Vol. XX, No. X, pp. 1–43
DOI: 10.3102/0002831219832301
Article reuse guidelines: sagepub.com/journals-permissions

Does Early Algebra Matter? The Effectiveness of an Early Algebra Intervention in Grades 3 to 5

Maria Blanton
TERC
Rena Stroud
Merrimack College
Ana Stephens
University of Wisconson-Madison
Angela Murphy Gardiner
TERC
Despina A. Stylianou
The City College of New York
Eric Knuth
University of Texas at Austin
Isil Isler-Baykal
Middle East Technical University
Susanne Strachota
University of Wisconson-Madison

A cluster randomized trial design was used to examine the effectiveness of a Grades 3 to 5 early algebra intervention with a diverse student population. Forty-six schools in three school districts participated. Students in treatment schools were taught the intervention by classroom teachers during regular mathematics instruction. Students in control schools received only regular mathematics instruction. Using a three-level longitudinal piecewise hierarchical linear model, the study explored the impact of the intervention in terms of both performance (correctness) and strategy use in students' responses to written algebra assessments. Results show that during Grade 3, treatment students, including those in at-risk settings, improved at a significantly faster rate than control students on both outcome measures and maintained their advantage throughout the intervention.

KEYWORDS: cluster randomized trial, early algebra, early math intervention, effectiveness study, elementary grades

> Algebra has become an academic passport for passage into virtually every avenue of the job market and every street of schooling. With too few exceptions, students who do not study algebra are therefore relegated to menial jobs and are unable often to even undertake training programs for jobs in which they might be interested. They are sorted out of the opportunities to become productive citizens in our society. (Schoenfeld 1995, pp. 11–12)

MARIA BLANTON is a senior scientist at TERC, 2067 Massachusetts Avenue, Cambridge, MA 02140; e-mail: *maria_blanton@terc.edu*. Her primary interests include understanding the foundations of children's algebraic thinking in elementary grades and designing inclusive innovations that improve children's algebra readiness for middle grades.

RENA STROUD is an assistant professor in the Department of Education at Merrimack College; e-mail: *stroudr@merrimack.edu*. She specializes in quantitative methods and analysis, including advanced regression techniques such as hierarchical linear modeling and structural equation modeling. Her research interests include teaching and learning in early mathematics and the role of technology in education.

ANA STEPHENS is an associate researcher at the Wisconsin Center for Education Research, University of Wisconsin-Madison; email: *acstephens@wisc.edu*. She is interested in studying the development of elementary students' algebraic reasoning and tasks that support this development.

ANGELA MURPHY GARDINER is a senior research associate at TERC; email: *angela_gardiner@terc.edu*. Her primary research interests include students' algebraic thinking and understanding of functions, methods for delivering effective teacher professional development, and understanding the relationship between characteristics of professional development and student outcomes.

DESPINA A. STYLIANOU is a professor of Mathematics Education in the Department of Curriculum and Instruction at the City College of New York; email: *dstylianou@ccny.cuny.edu*. She studies the development of students' mathematical practices across the grades, including argumentation and representation.

ERIC KNUTH is an Elizabeth Shatto Massey Endowed Chair in Education and Director of the STEM Center at the University of Texas at Austin; *e-mail: eric.knuth@austin.utexas.edu*. His research focuses on the meaningful engagement of students in mathematical practices and their development of increasingly more sophisticated ways of engaging in those practices, focusing in particular on practices related to algebraic reasoning and learning to prove.

ISIL ISLER-BAYKAL is an assistant professor in Mathematics Education in the Department of Mathematics and Science Education at the Middle East Technical University; email: *iisler@metu.edu.tr*. She studies early algebra, reasoning, and proof in elementary and middle grades. She is also interested in preservice and in-service teacher education.

SUSANNE STRACHOTA is an assistant researcher at the Wisconsin Center for Education Research, University of Wisconsin Madison; *e-mail: sstrachota@wisc.edu*. Her research focuses on algebraic reasoning, specifically how students generalize functional relationships and justify those generalizations.

Introduction

Currently, one of the most significant questions in mathematics education reform is "Does early algebra[1] matter?" That is, will a comprehensive, sustained effort to develop children's algebraic thinking as part of their mathematics learning in elementary grades improve their readiness for a more formal study of algebra as they enter middle grades? The answer to this question has deep implications for school mathematics.

Historically, teaching and learning algebra in the United States emphasized computational work in arithmetic in the elementary and middle grades, followed by a superficial treatment of algebra in secondary grades (Kaput, 2008). This approach resulted in widespread student failure in school mathematics (e.g., Kaput, 1999; Stigler, Gonzales, Kawanaka, Knoll, & Serrano, 1999) that positioned algebra as a gatekeeper whereby those who were not successful were "sorted out of the opportunities to become productive citizens in our society" (Schoenfeld, 1995, p. 12). The resulting marginalization of students particularly affected those in underrepresented groups (e.g., Moses & Cobb, 2001) in ways that have propagated to their underrepresentation in STEM fields in general (Museus, Palmer, Davis, & Maramba, 2011).

In recent decades, however, algebra's status as a gateway to academic and economic success (Moses & Cobb, 2001) has led to calls for new approaches to algebra education (e.g., National Council of Teachers of Mathematics [NCTM] & Mathematical Sciences Education Board, 1998; U.S. Dept. of Education, 2008). One key recommendation that gained widespread acceptance was to reconceptualize teaching and learning algebra from a K–12 perspective whereby students would have long-term, sustained algebra experiences, beginning in elementary grades (e.g., Kilpatrick, Swafford, & Findell, 2001; NCTM, 1989, 2000). In theory, such an approach would allow children's algebraic thinking to develop more organically by leveraging their natural intuitions about structure and relationships (Mason, 2008) from the start of formal schooling. In theory, too, the development of algebraic thinking in this way would increase children's success with more formal mathematics, particularly algebra, as they progressed into middle grades and beyond.

While this approach promises to address the widespread failure in algebra in the United States, it also entails significant costs. As Kaput (2008) argues, it "involves deep curriculum restructuring, changes in classroom practice and assessment, and changes in teacher-education—each a major task" (p. 6). These costs highlight the need for carefully constructed models of early algebra instruction that can provide a curricular road map for systematically developing children's algebraic thinking. However, such tools are typically lacking in educational practice in elementary grades. Instead, we often see fragmented efforts to infuse algebraic ideas into arithmetic-focused curricula. This is evidenced by instructional materials that might address more ubiquitous algebraic tasks (e.g., solving simple linear equations) but may not reflect an underlying, comprehensive treatment of algebra rooted in empirical research on children's algebraic understandings. Moreover, the recent adoption by many U.S. states of the Common Core State Standards (National Governors Association Center for Best Practices and Council of Chief State School Officers [NGA Center & CCSSO], 2010)—which rightly reiterates the place of early algebra in school mathematics beginning in kindergarten—has at least implicitly elevated the role of algebra and, thus, potentially increased its gatekeeper status, leaving students vulnerable to a persistent marginalization in school. As such, research-based models of comprehensive early algebra instruction are urgently needed to clarify and deepen the role of early algebra in elementary grades mathematics.

Such models would provide not only an instructional blueprint for teachers but also a critical means to rigorously study early algebra's impact. And while the study of early algebra's impact alone has substantial merit in the national discourse on teaching and learning algebra, it simultaneously addresses a broader concern. The National Mathematics Advisory Panel Report (U.S. Dept. of Education, 2008), which emphasizes the importance of student success in algebra, calls broadly for experimental research to "investigate the effects of programs, practices, and approaches on students' mathematics achievement" (p. 84). Other scholars have also noted the deficit of research in evaluating the effectiveness of curricular innovations, particularly for innovations aimed at young populations of learners (Clements & Sarama, 2008; National Research Council, 2004). Moreover, they point specifically to the need for evaluation studies that employ rigorous, experimental designs such as randomized field trials (Clements, 2002) and that are situated within at-risk populations, including in schools with high percentages of students categorized as low socioeconomic status (SES; Clements, 2007).

In short, a rigorously tested model of early algebra instruction would provide not only a much-needed curricular innovation by which elementary teachers could implement an ambitious early algebra agenda, but also a careful measure of its risks and rewards from which we could better understand early algebra's impact. The research reported here addresses this need. We used a cluster randomized trial design in a large-scale, longitudinal study to examine the effectiveness of a Grades 3 to 5 early algebra intervention developed in our prior work. The intervention was taught by classroom teachers in demographically diverse school settings that included a high percentage of low-SES populations. In the study reported here, we compare the mathematics achievement of students who participated in the teacher-led intervention as part of their regular classroom instruction to that of students who had only their regular, arithmetic-centered instruction. We focus on the following questions:

Research Question 1: What are differences in students' performance on measures of algebraic knowledge?

Research Question 2: What are differences in students' use of structural (algebraic) strategies to solve nonroutine problems on measures of algebraic knowledge?

Research Question 3: What are differences in students' performance and use of structural (algebraic) strategies among at-risk populations (e.g., low-SES)?

Framework for the Development of the Intervention

We see the design and evaluation of our early algebra intervention as aligned with Clements's (2007) Curriculum Research Framework (CRF; see also Clements & Sarama, 2008). The CRF provides a template for the development of research-based curricula that utilizes three categories: *a priori foundations, learning model*, and *evaluation*. The first category, *a priori foundations*, includes identifying appropriate subject matter around which the curriculum will be designed and analyzing relevant research in teaching and learning, both of which inform the design of the curriculum. *Learning model* involves the design and sequencing of activities that align with empirical models of children's thinking around the content of the curriculum. Finally, *evaluation* involves the use of multiple methodologies to evaluate the appeal, usability, and effectiveness of the curriculum. In the discussion that follows, we highlight how our conceptual approach to (early) algebra, the design of our intervention based on this approach, and our evaluation process for the intervention, aligns with the three dimensions of the CRF.

Our conceptual approach to early algebra is based on Kaput's (2008) content analysis of algebra as a set of *core aspects* across several mathematical *content strands* (see Blanton, Brizuela, et al., 2018, for a more detailed treatment of our framework). A vital contribution of Kaput's analysis to our work is its identification of valid subject-matter content and practices (i.e., ways of reasoning algebraically). This, along with our synthesis of empirical research on teaching and learning algebra, as well as the canonical development of algebra as a mathematical discipline (Battista, 2004), served as an *a priori foundation* (Clements, 2007) for the design of our intervention. From Kaput's core aspects—namely, making and expressing generalizations in increasingly formal and conventional symbol systems and acting on symbols within an organized symbolic system through an established syntax—we derived four essential algebraic thinking practices that served as organizing principles for the intervention: *generalizing, representing, justifying*, and *reasoning with* mathematical structure and relationships (see, e.g., Blanton, Brizuela et al., 2018; Blanton, Levi, Crites, & Dougherty, 2011).

Generalizing and representing are deeply symbiotic processes central to algebraic thinking (Cooper & Warren, 2011). Generalizing as the mental activity by which one compresses multiple instances into a single unitary form (Kaput, Blanton, & Moreno 2008) is conveyed through the action of representing—that is, symbolizing—the resulting unitary form using an appropriate notational system (e.g., natural language, variable notation, graphs, tables, pictures). For example, as students operate on particular whole numbers, they might notice—either spontaneously or through teacher scaffolding—that the action of adding two odd numbers results in an even number. This compression of their observation about multiple instances of adding two particular odd numbers can be represented in a unitary or generalized form through natural language (e.g., "The sum of two odd numbers is even"). In turn, the action of representing a generalization is a socially mediated process whereby one's thinking about symbol and referent are iteratively refined (Kaput et al., 2008), leading to a mediation of the generalization itself.

The practices of justifying and reasoning with mathematical structure and relationships are themselves actions on a unitary form (i.e., a generalization), where those actions are governed by an established syntax. In the case of justifying, one builds an argument about the validity of a generalization within a given representational system. For example, consider a representation-based argument (Schifter, 2009) that students might construct through either physical objects, such as cubes, or a drawing that depicts such objects. They might reason that since an odd-numbered set of cubes can be separated into pairs of cubes with one cube left over, the combination of two odd-numbered sets of cubes results in no cube without a "partner" cube. That is, since the leftover cube in each of the two sets combines to form a new pair, the resulting sum is even. In later years, students might use algebraic syntax to reason formally on symbolic (variable) representations of odd numbers to construct their arguments. In the case of reasoning with a generalization, one acts on generalizations as mathematical objects (Sfard, 1991) themselves in novel situations. For example, students might reason inductively with the generalization "the sum of two odd numbers is even" to examine the parity of the sum of three odd numbers. These practices, too, are socially mediated processes that refine the scope of the generalization and "drive the symbolization process" (Kaput et al., 2008, p. 46).

In designing our intervention, we were interested in the occurrence of these practices in two of Kaput's (2008, p. 11) three content strands ("the study of structures and systems abstracted from computations and relations" and the "study of functions, relations, and joint variation") because of their close alignment with empirical research on children's algebraic thinking. As reported elsewhere (Fonger et al., 2018), we organized key early algebraic concepts and practices relative to these strands under the "Big Ideas" (Shin, Stevens, Short, & Krajcik, 2009) of *generalized arithmetic; equivalence, expressions, equations, and inequalities*; and *functional thinking* (see Blanton, Brizuela et al., 2018, for an elaboration of these Big Ideas).

Using this conceptual approach to algebra organized around essential algebraic thinking practices within content-based Big Ideas, we drew from learning progressions research (e.g., Battista, 2004; Clements & Sarama, 2004; Maloney, Confrey, & Nguyen, 2011; Shin et al., 2009; Simon, 1995) to develop an *early algebra learning progression* for Grades 3 to 5 that includes the following four components (Clements & Sarama, 2004): (1) a curricular framework and associated learning goals that identify core algebraic concepts within the Big Ideas and that are organized around the four algebraic thinking practices, (2) a Grades 3 to 5 instructional sequence (referred to here as the *intervention*) designed to address the learning goals, (3) validated assessments to measure student learning in response to the intervention, and (4) a specification of the increasingly sophisticated levels of algebraic thinking students exhibit about algebraic concepts and practices as they progress through the intervention (see Fonger et al., 2018, for an extensive treatment of the development of these components). Components 1 to 3 are the basis for the effectiveness study reported here. Table 1 provides an illustration of these components using the Big Idea of *generalized arithmetic* for Grade 3.

As alluded to earlier, Kaput's (2008) content analysis of algebra, along with empirically based research on children's engagement in core algebraic thinking practices within the Big Ideas, provided the *a priori foundation* (Clements & Sarama, 2008) for the design of the curricular framework, learning goals, intervention, and assessments. Consistent with the *learning model* dimension of the CRF (Clements, 2007), the intervention was designed as a conjectured route whose sequencing was based on known or hypothesized progressions in children's thinking about core algebraic concepts and practices (our targeted subject matter domain). The sequencing of activities, or lessons, within the intervention was intended to advance students' understanding of a concept or practice. For example, the significant body of empirical research on the development of children's relational understanding of the equal sign—which entails interpreting the equal sign as an equivalence relation indicating two mathematical objects are equivalent (Jones, Inglis, Gilmore, & Dowens, 2012)—and the type and sequencing of tasks that support this (Rittle-Johnson, Matthews, Taylor, & McEldoon, 2011) were used to design lesson activities for the intervention that would advance children's relational thinking about this symbol.

Moreover, the treatment of core algebraic concepts and practices was interwoven throughout the 3-year intervention so that their treatment would not be isolated in instruction. Clements and Sarama (2008) attribute several reasons to the importance of a distributed approach in the learning model, including that students' mathematical learning is naturally incremental, learning progressions themselves reflect a multiyear process that cannot be appropriately compressed into units of instruction, and such an approach improves recall and retention and can mutually reinforce common ideas

Table 1

Selected Components of the Early Algebra Learning Progression for the Big Idea "Generalized Arithmetic" in Grade 3

Selected Generalized Arithmetic Constructs[a] for the Curricular Framework
The Fundamental Properties of Number and Operation (e.g., the Commutative Property of Addition) represent the underlying relationships that govern how operations behave and relate to each other. Operations are inversely related to each other: • Addition and subtraction have an inverse relationship. • Multiplication and division have an inverse relationship. Generalizations in arithmetic other than the Fundamental Properties can be derived from the Fundamental Properties. These include relationships in classes of numbers (e.g., even and odd numbers) and outcomes of calculations. The Fundamental Properties are relationships that are true for all values of the variables in a specified number domain.

Learning Goals	Associated Algebraic Thinking Practices
• Analyze information to develop a conjecture (generalization) about an arithmetic relationship	*Generalize* the relationship
• Represent the generalization using words and/or variables • Examine the meaning of variables used in a representation • Examine the meaning of repeated variables or different variables in the same equation	*Represent* the relationship
• Develop a justification or an argument to support the conjecture's validity • Identify values for which the conjecture is true • Explore different types of arguments, including empirical arguments, representation-based arguments, and arguments based on the algebraic use of number • Examine the characteristic that the generalization is true for all values of the variable in a given domain	*Justify* the relationship
• Identify the generalization (e.g., property) in use when doing computational work • Use the generalization to reason about the validity of new conjectures	*Reason with* the relationship

(continued)

Table 1 **(continued)**

Instructional Sequence (Intervention)	
Focus and Sequence of 5 Lessons on Generalized Arithmetic	Overall Aims of Grade 3 Lessons on Generalized Arithmetic
• Lesson 3: Additive Identity and Additive Inverse Properties • Lesson 4: Commutative Property of Addition • Lessons 6 and 7: Generalizations About the Sums of Evens and Odds • Lesson 11: Multiplicative Identity Property and Multiplication Property of Zero ($a \times 0 = 0$ for all a)	Students begin generalizing, representing, justifying, and reasoning with Fundamental Properties of Number and Operation, focusing initially on properties involving addition and subtraction. These are followed later in Grade 3 by multiplicative properties, once multiplication is introduced. The order in which properties are introduced progresses from those that contain a single, repeated variable to those with multiple variables. Students also begin to explore other arithmetic generalizations about classes of numbers, particularly those about sums of even numbers and odd numbers. Although arithmetic generalizations in Grade 3 are initially expressed in words, the use of variables to express Fundamental Properties is introduced. Students also begin to use Fundamental Properties to simplify computational work and to identify these properties in use in computational work. Students begin to explore ways to justify arithmetic generalizations, beginning with empirical arguments and moving towards more general arguments based on reasoning with representations.

Grade 3 Assessment Items for Generalized Arithmetic
• Marcy's teacher asks her to solve "23 + 15." She adds the two numbers and gets 38. The teacher then asks her to solve "15 + 23." Marcy says she already knows the answer is 38 because the numbers are just "turned around." a. Do you think Marcy's idea will work for any two numbers? Why or why not? b. Write an equation using variables (letters) to represent the idea that you can add two numbers in any order and get the same result. • Brian knows that if you add any three odd numbers, you will get an odd number. Explain why this is true.

[a]Adapted from Blanton, Levi, Crites, and Dougherty (2011).

across different learning trajectories. For example, in our intervention generalizing was viewed not as a mental activity that students learned in a given instructional unit but as a practice in which they were expected to engage across all the Big Ideas throughout the 3-year intervention.

The *evaluation* phase (Clements, 2007) of our work has progressed from the use of classroom-based teaching experiments from which we could examine the meanings students made of concepts and practices addressed in the intervention, to more formal small-scale, quasi-experimental studies that examined the usability and potential efficacy of the intervention when implemented by a member of our team with expertise in the intervention, to the current large-scale, randomized study reported here. The prior small-scale studies included a 1-year, cross-sectional study at each of Grades 3 to 5 and a 3-year, longitudinal study across Grades 3 to 5 that examined the potential efficacy of the intervention under the most favorable conditions (O'Donnell, 2008). These studies showed statistically significant differences favoring students who were taught the intervention in comparison to their peers who received only regular instruction (see, e.g., Blanton et al., 2015; Blanton, Isler, et al., 2018). Early indications of the intervention's potential positioned us for the longitudinal, large-scale, randomized effectiveness study (e.g., Mihalic, 2002; Raudenbush, 2007) reported here, where the intervention was implemented in diverse demographic settings using an authentic approach involving classroom teachers who were expected to vary in their fidelity of implementation (FOI; Clements, 2007). Figure 1 summarizes the alignment between the CRF and the development of our intervention.

Method

Setting and Participants

The study took place in three school districts in one state within the southeastern United States. All schools within the districts were invited to participate, although a few schools opted not to do so. For example, some schools followed a year-round academic calendar and were unable to coordinate this with the implementation timetable.

The districts represented diverse settings with urban, suburban, and rural populations. Forty-six schools within the districts participated, with 23 schools randomly assigned to the treatment condition and 23 to the control condition in the summer prior to the September commencement of the study in Grade 3. All 46 schools participated throughout the study (i.e., there was no school attrition). Although the randomization occurred at the school level, we blocked by district since schools were demographically similar within districts but not between districts. This resulted in an equal number of treatment and control schools within districts (see Table 2 for overall

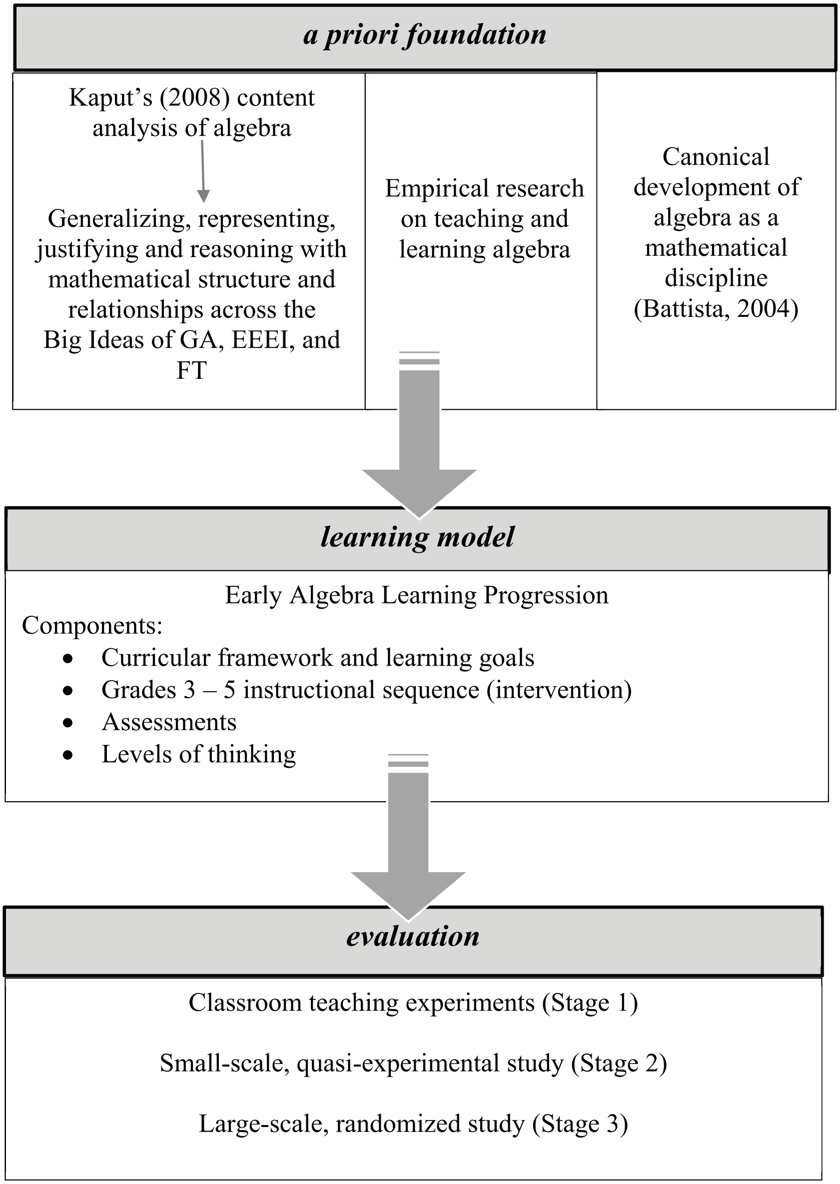


Figure 1. **Alignment between the CRF (Clements, 2007) and our design and evaluation process.**

Note. CRF = Curriculum Research Framework; GA = generalized arithmetic; EEEI = equivalence, expressions, equations, and inequalities; FT = functional thinking.

Table 2
Demographics for Districts Participating in the Study

District	Total No. of Elementary Schools	Distribution of Condition for Participating Schools	Low SES (Free/Reduced-Price Lunch), %	English Language Learners, %	Minority (Non-White), %
A	7	Treatment: 3; Control: 3	20	6	36
B	22	Treatment: 6; Control: 7	54	10	40
C	30	Treatment: 14; Control: 13	62	20	82

demographics for the participating districts across the study's 3-year implementation, as well as the distribution of treatment and control schools within districts). Data regarding the percent of students receiving free or reduced-price lunch were collected at the school level and used as a proxy for SES. Across all participating schools, an average of 63% of students received free or reduced-price lunch.

Students in treatment schools were taught the early algebra intervention by their classroom teachers as part of regular mathematics instruction. Because of early algebra's deep connections to arithmetic, the intervention provided opportunities to develop not only students' algebraic understanding but also their knowledge of important arithmetic concepts and skills. As such, although concepts addressed in the regular curriculum might have been realigned at the discretion of the teacher or other school authority, they did not need to be eliminated to accommodate the intervention. Students in control schools received only regular instruction.

In Year 1, we collected assessment data from 3,085 treatment and control students at Grade 3 pretest. The pretest was administered at the start of the school year (September) prior to the intervention. This student cohort was followed across Grades 3 to 5. As shown in Table 3, the number of students assessed varied from year to year, a common characteristic of longitudinal studies. Since we could not test the missing data directly, we looked for patterns in the data regarding missingness. A review of the data suggested that student absences during the administration of the assessment and students moving out of district were the two primary causes of missingness. Given that missingness was not found to be related to our outcome measures (student performance and strategy use on the assessment), the missing data were presumed to be missing at random (Little & Rubin, 1987). Full-information maximum likelihood was used in the multilevel analyses so as to include

Table 3
Number of Algebra Assessments Collected by Grade and Treatment Condition

Grade	Treatment	Control	Total
Grade 3 Pretest	1,637	1,448	3,085
Grade 3 Posttest	1,495	1,343	2,838
Grade 4 Posttest	1,341	1,245	2,586
Grade 5 Posttest	1,087	1,079	2,166

all available data in model estimation. Additionally, all multilevel models were run twice, once with the full data set and once with only the subset of students who completed all four testing points. The pattern of results for the subsample paralleled that of the full sample for both outcome measures. Accordingly, in subsequent analyses, only results from the full sample are reported.

The Intervention

The intervention consisted of 18 one-hour lessons at each of Grades 3 to 5, with lessons taught throughout the school year (approximately September through March). As noted earlier, lessons were designed to engage students in the algebraic thinking practices of generalizing, representing, justifying, and reasoning with mathematical structure and relationships within the Big Ideas of generalized arithmetic; equivalence, expressions, equations, and inequalities; and functional thinking.

Lessons began with a 15-minute "Jumpstart" constructed to review previous concepts or prompt students' thinking about the concept to be addressed in the given lesson. They then transitioned into an investigative activity or set of activities in which students explored the particular lesson focus through small group work. Finally, lessons concluded with a whole-group discussion of students' findings, followed by a brief "Review and Discuss" that served as a formative assessment. All lessons emphasized developing meaning for mathematical ideas by engaging students in explaining their thinking, both orally and in writing. Table 4 illustrates the lesson structure with Lesson 4 (Grade 3) on the Commutative Property of Addition.

The study was implemented with grade-level teams across participating schools. Grade 3 teachers from all schools participated in Year 1, Grade 4 teachers from all schools participated in Year 2, and Grade 5 teachers from all schools participated in Year 3. During each year of the study, teachers implementing the intervention were provided professional development (PD) to support their FOI. PD, which included a 1-day training session prior to the start of school and a ½-day session each month thereafter, had three primary goals: (1) developing teachers' knowledge of algebraic thinking practices and core concepts by engaging teachers in these practices across the Big Ideas, (2) developing teachers' understanding of students' algebraic

Table 4
Lesson 4 (Grade 3, Generalized Arithmetic)

Lesson 4: Exploring Fundamental Properties (Commutative Property of Addition)	
Jumpstart	1. Which of the following equations are true? Explain. $14 - 14 = 0$ $394 + 0 = 394$ $17 + 5 = 23 + 5$ $30 + (10 + 19) = (30 + 10) + 19$ 2. Marta has 6 pieces of candy. Her friend, Sarah, has 9 pieces of candy. How would you represent the relationship between the numbers of pieces of candy they have? Using the same numbers, can you represent the relationship in a different way?
Small-group investigation	A. Which of the following equations are true? Use numbers, pictures, cubes, or words to explain your reasoning. $17 + 5 = 5 + 17$ $20 + 15 = 15 + 20$ $148 + 93 = 93 + 148$ B. What numbers make the following equations true? $25 + 10 =$ ____ $+ 25$ ___ $+ 237 = 237 + 395$ $38 +$ ___ $=$ ____ $+ 38$ C. What do you notice about these problems? Write a conjecture about what you notice in your own words. D. Write your conjecture as an equation (with variables). What do your variables represent? E. Can you write your equation in a different way? F. For what numbers is your conjecture true? Use numbers, cubes, pictures, or words to explain your thinking. G. Find the following. Think about how you might use the properties you have learned in Lessons 3 and 4. $95 + 39 - 39 + 12$ $68 + 27 + 32 - 27$
Review and discuss	1. Is $23 + 17 = 17 + 23$ true or false? What is a different way you can write this equation, using only these numbers, so that the equation is still true? 2. ___ $+ 0 = 0 +$ ____. What numbers will make this equation true? 3. Kara said that you could use any number in (2) and that she could represent "any number" with a variable. She represented this idea in the following way: $b + 0 = 0 + b$. Marcus agreed but wrote $c + 0 = 0 + b$. Do you agree with how Marcus represented the idea? Explain.

thinking—as identified in both research and in teachers' own classroom data collected during their implementation of the intervention—and how to build on students' thinking in instruction, and (3) strengthening teaching practices (such as teacher questioning strategies) that could increase students' engagement with core algebraic concepts and practices.

Algebra Assessments

Grade-level algebra assessments were designed and validated to measure within-grade and across-grade (longitudinal) growth in students' understanding of algebraic thinking practices (e.g., generalizing) and core algebraic concepts (e.g., relational understanding of the equal sign), as well as growth in students' use of structural (algebraic) strategies.

Each grade-level assessment was designed as a 1-hour, written assessment containing 12 to 14 items. Nine of these items were used across the assessments in order to measure longitudinal growth, with these common items being most difficult for Grade 3 (see Appendix for common items). Eight of the common items were open response, with five items containing multiple subparts. Each of the subparts was coded separately for performance and structural strategy use, for a total of 21 subparts across the common items used to measure longitudinal growth.

In validating the assessments, items went through multiple cycles of internal reviews by the research team and external reviews by the project's advisory board (content experts) and an independent evaluator. Items were then revised to create grade-level assessments that were administered at grade level to about 100 students per grade. Student responses were scored using a coding scheme developed by the project team. The coding scheme was designed to capture both correctness of student responses as well as the types of strategies students used. Mean initial agreement was 86%. After negotiation of the coding scheme, mean agreement after a second round of coding was 89%.

Under the direction of the project's quantitative methodologist, the assessments were then tested for psychometric soundness in several ways. First, internal consistency estimates of reliability (i.e., Cronbach's alpha) were calculated for each assessment. Individual items within each assessment were evaluated with respect to their contributions to test reliability, as well as other characteristics such as item difficulty, variance, and item-total correlations. Items that failed to demonstrate positive contributions to the test (e.g., proportion correct > 0.95 or < 0.05; variance < 0.05; item-total correlations $< .10$) were removed or modified. For construct validity, we compared our assessments to external and established assessments such as standardized state tests based on, for example, Pearson correlations. Using these data to refine the items, final versions of the Grades 3 to 5 assessments were constructed.

Data Collection and Analysis of Student Responses on Algebra Assessments

In Grade 3, participating students in treatment and control schools were administered the Grade 3 algebra assessment both as a pretest (baseline) prior to the start of the intervention and as a posttest. The Grade 4 and Grade 5 algebra assessments were administered as a posttest in Grades 4 and 5, respectively.

Coding Student Responses

Student responses to algebra assessment items were coded using coding schemes developed through multiple refinements in our prior work (e.g., see Blanton et al., 2015). Items were coded by trained coders, without knowledge of students' treatment condition, in terms of performance (i.e., correctness) and students' strategy use. Strategy codes were developed to try to capture the variety of ways in which students approached our assessment items and included a wide range of sophistication, some of which reflected correct thinking and some of which did not. Some strategy codes indicated evidence of common misconceptions, some indicated engagement in arithmetic approaches, and others indicated more sophisticated algebraic thinking.

In analyzing students' strategies, we were particularly interested in the degree to which students used strategies we refer to as structural (Kieran, 2007). Structural strategies involve recognizing and acting on underlying mathematical relationships. This might occur when representing a relationship between two quantities using variables, when making a general argument that does not rely on specific values, or when reasoning about equations. For example, on the assessment item for which students were asked to find the missing value in the equation $7 + 3 = \_\_\_ + 4$, students using a structural strategy might argue that the missing value must be one less than 7 (i.e., 6) because 4 is one more than 3. This approach indicates an understanding of the relationships among the quantities in the equation and an ability to view the equation as a whole object. A student using a computational approach, on the other hand, might find the missing value by adding 7 and 3 and subtracting 4 from the result to find the missing value (i.e., 6). While both are correct, the first approach (known as a *compensation strategy*) demonstrates a type of structural reasoning that is important in algebraic thinking (Kieran, 2007; Knuth, Stephens, McNeil, & Alibali, 2006). A third, albeit incorrect, strategy for this item is an *operational* strategy, in which students perform the operation to the left of the equal sign (writing 10 in the blank) or add all given numbers in the equation (writing 14 in the blank). The full coding scheme for this item is included in the Appendix to illustrate the variation of strategies captured from student responses.

Finally, to assess interrater reliability of the coding process, a random sample of 15% to 20% of assessments at each grade level were double-coded by a pair of trained coders. Training involved having coders discuss individual items and coding schemes and then independently code student assessments from previous research as well as hypothetical responses generated by noncoding members of the research team. Some items were less complex and thus more straightforward to code than others. Oftentimes, multiple rounds of training and discussion among coders and other research team

Table 5
Interrater Reliability (*k*) for Common Items Across Grade Levels

	Correctness				Structural Strategy Use			
Item	Grade 3 Pretest	Grade 3 Posttest	Grade 4	Grade 5	Grade 3 Pretest	Grade 3 Posttest	Grade 4	Grade 5
1	.98	.99	.98	.96	.90	.95	.96	.90
2a	.98	.99	.98	.95	.84	.91	.89	.86
2b	.95	.97	.96	.96	.85	.89	.84	.86
2c	.95	.97	.97	.97	.83	.87	.85	.85
3a	.84	.89	.79	.76	.76	.84	.74	.72
3b	.94	.96	.97	.95	.89	.86	.94	.91
4	.94	.88	.89	.81	.90	.86	.83	.76
5a	.97	.93	.90	.89	.94	.89	.86	.87
5b	.97	.94	.92	.88	.94	.86	.86	.80
5c1	.98	.97	.93	.92	.95	.90	.84	.78
5c2	.96	.99	.97	.91	.90	.85	.79	.78
8	.93	.96	.94	.96	.88	.87	.88	.84
9a	.92	.95	.93	.98	n/a	n/a	n/a	n/a
9b	.95	.93	.88	.95	n/a	n/a	n/a	n/a
9c1	.96	.98	.93	.96	.88	.86	.89	.87
9c2	.97	.97	.89	.97	.91	.83	.84	.84
9d	.97	.97	.97	.92	n/a	n/a	n/a	n/a
10a	.99	.96	.97	.92	n/a	n/a	n/a	n/a
10b	.98	.98	.97	.98	n/a	n/a	n/a	n/a
10c	.97	.97	.96	.95	n/a	n/a	n/a	n/a

Note. Structural strategies were not applicable to some items (indicated by n/a).

members were needed to achieve consistency on more complex items before the coding of the present study's data could begin. Such discussions enriched the coding manual by leading to more refined definitions and examples of particular strategy codes. Results indicated substantial agreement between raters, with all Cohen's kappa (*k*) statistics .70 or greater across each grade level (see Table 5).

Analysis of Performance and Structural Strategy Use

We calculated individual student scores for each outcome measure (i.e., performance and structural strategy use) based on the percentage of items coded as correct (as a measure of performance) or that used a structural strategy. Using HLM Version 7.01 (Raudenbush & Bryk, 2002), we modeled performance and structural strategy use using separate three-level longitudinal piecewise growth models, with repeated assessments nested within individuals nested within schools. Two linear slope factors were modeled. The first growth factor examined the initial changes in performance and structural strategy use during the first year of the intervention (Time 1, Grade 3 pretest to Grade 3 posttest), whereas the second examined the growth that occurred from the end of Grade 3 through Grade 5 (Time 2, Grade 3 posttest to Grade 5 posttest). These time periods were identified based on descriptive statistics that suggested a differential effect in the first year of the intervention as compared to subsequent years. At the school level, we modeled the effect that the intervention had on student growth and between-school variability, controlling for SES. The interaction between SES and treatment condition was also included as a predictor at the school level to determine if the impact of the intervention was equitable across the range of SES.

Results

Before we discuss findings on student performance and strategy use, we briefly address two issues that will provide further context for our findings: the fidelity with which teachers implemented the intervention and the regular mathematics curriculum used by teachers in both treatment and control classrooms.

All districts followed the state's common standard course of study, based on the Common Core State Standards for Mathematics (NGA Center & CCSSO, 2010), which prescribed specific content to address at each grade level. The districts addressed these content standards through different curricular materials. One district used a well-known "reform-based" curriculum that encouraged inquiry and investigation, one district used a more mainstream curriculum that encouraged computational skill, and one used a curriculum that was a hybrid of the first two. Regardless, when examined using our conceptual framework for algebra, all three curricula were found to address algebraic topics but with limited emphasis and frequency. For example, the "mainstream" curriculum contained activities such as finding the missing value in equations (including equations containing a variable), identifying properties of operations informally using natural language (e.g., framing the Commutative Property of Addition as "adding in any order"), finding the recursive rule in a sequence of values, and examining function tables to look for relationships. Moreover, the curriculum introduced variables as letters representing fixed unknowns in Grade 3 (through equation-solving tasks such as $7 = n - 8$) and variables as varying quantities in Grade 5 through functional relationships. By Grade 5, the curriculum included topics such as substituting the value of a letter in expressions and equations, finding a (functional) relationship between two quantities in a table, and interpreting graphs by constructing a story that could be represented by a given graph. Not surprisingly, the treatment of algebra in the curriculum increased across Grades 3 to 5. Given this, we anticipated that control students would exhibit some level of success on our algebra assessment, particularly in Grade 5.

Elsewhere, Stylianou et al. (in press) conducted a full FOI study and found that while there was variation in fidelity, teachers generally implemented the intervention faithfully. In particular, they found that, overall, teachers followed the lesson structure: 94% of teachers used the Jumpstart in their instruction and 73% of teachers placed students in group or individual formats to investigate the lesson activity, with 77% of these teachers coded as "actively" interacting with students during this investigative component by clarifying ideas and asking questions to challenge students' thinking. Furthermore, they found that teachers scored well on their use of time, clear presentation of mathematics, engagement of students in instruction, attention to students' difficulties, use of students' ideas in instruction, and use of precise mathematical language or notation. Moreover, they found that teachers' engagement of students in the algebraic thinking practices significantly predicted student outcomes at posttest, whereby students in classrooms where teachers were rated more highly on their implementation of core algebraic practices outperformed students whose teachers received lower ratings. Given that it can take several years of PD for teachers to develop the necessary expertise to implement a curriculum with high fidelity (see e.g., Jacobs, Lamb, & Philipp, 2010; Superfine, 2008), we view results of these teachers' first-year attempts at implementing the intervention as positive and reasonable.

Results for Students' Responses on Algebra Assessments

Descriptive statistics for both performance and structural strategy use are provided in Table 6. In the sections that follow, we summarize the hierarchical linear models tested for each outcome measure.

Performance

Figure 2 compares the average student performance (i.e., percentage correct) on common items on the algebra assessment for the given conditions, across the four testing points. An initial exploratory examination of the data suggested a sizeable effect in the first year of the intervention, followed by a potential leveling off of that effect in subsequent years. Following a procedure used by Frank and Seltzer (1990) and described in Raudenbush and Bryk (2002), an inspection of growth plots for a random subsample of 100 students confirmed this pattern, suggesting a differential effect in the first year of the study as compared to subsequent years. According to Raudenbush and Bryk (2002), a piecewise linear growth model is appropriate when an analysis of the data suggests nonlinearity such that growth trajectories can be broken into two linear components. Therefore, given the nature of our data, we employed a piecewise linear growth model

Table 6
Percentages for Correctness and Structural Strategy Use on Algebra Assessments

	Treatment				Control			
Outcome Measure	Grade 3 Pretest	Grade 3 Posttest	Grade 4 Posttest	Grade 5 Posttest	Grade 3 Pretest	Grade 3 Posttest	Grade 4 Posttest	Grade 5 Posttest
Correctness								
M	13.64%	40.42%	52.82%	64.67%	14.90%	27.31%	38.84%	47.80%
SD	11.08%	21.56%	22.91%	22.04%	11.54%	15.88%	19.37%	20.31%
n	1,637	1,495	1,341	1,087	1,448	1,343	1,245	1,079
Strategy								
M	3.68%	18.39%	27.94%	37.75%	4.17%	7.13%	13.40%	19.09%
SD	5.34%	16.66%	19.67%	22.11%	5.45%	8.00%	13.78%	16.90%
n	1,637	1,495	1,341	1,087	1,448	1,343	1,245	1,079

to explore the impact of the intervention on performance at each distinct time period.

An unconditional piecewise linear growth model was first fit in order to assess the variability at each level, as well as to provide a baseline from which to compare subsequent models. The Level 1 (repeated assessments within students) model is the following:

$$\mathrm{Y}_{tij}=\pi_{0ij}+\pi_{1ij}(\mathrm{Time}\ 1_{tij})+\pi_{2ij}(\mathrm{Time}\ 2_{tij})+e_{tij},$$

where Y_{tij} is the observed performance (correctness) score at time t for individual i in school j, π_{0ij} is the performance of student i in school j at baseline (Grade 3 pretest), π_{1ij} is the growth rate for student i in school j during the first year of the intervention (Time 1), π_{2ij} is the growth rate of student i in school j over the following 2 years of the intervention (Time 2), and e_{tij} is the Level 1 residual.

Baseline performance and growth rate during the first year of the intervention (Time 1) and the following 2 years (Time 2) were subsequently modeled at Level 2 (student level) and Level 3 (school level). The unconditional model at Levels 2 and 3 is the following:

Level 2

$$\pi_{0ij}=\beta_{00j}+r_{0ij},$$
$$\pi_{1ij}=\beta_{10j}+r_{1ij},$$
$$\pi_{2ij}=\beta_{20j}+r_{2ij},$$

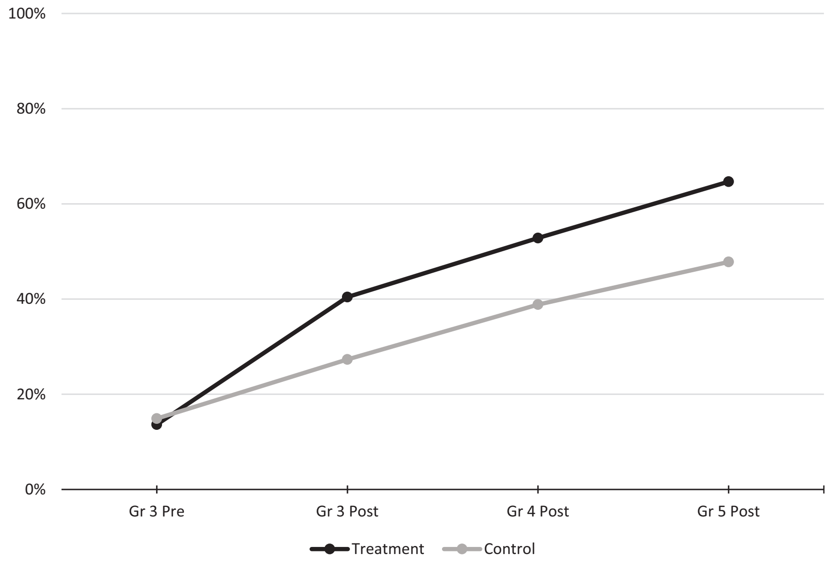


Figure 2. **Comparison of overall correctness for common items on algebra assessments.**

Level 3

$$\beta_{00j} = \gamma_{00j} + u_{00j},$$
$$\beta_{10j} = \gamma_{100} + u_{10j},$$
$$\beta_{20j} = \gamma_{200} + u_{20j},$$

where β_{00j} is the average performance within school j at pretest; β_{10j} and β_{20j} are the average growth parameters at Time 1 and Time 2, respectively, within school j; r_{0ij}, r_{1ij}, and r_{2ij} represent variability between individuals for pretest performance and growth rates at Times 1 and 2, respectively; γ_{000} is the overall average performance across all schools at pretest; γ_{100} and γ_{200} are the average growth parameters at Time 1 and Time 2, respectively, across schools; and u_{00j}, u_{10j}, and u_{20j} represent variability between schools on pretest performance and each growth parameter, respectively.

Across all schools, including both treatment and control, the average performance at Grade 3 pretest (γ_{000}) was 14.02% (see Table 7). The average growth rate in the first year of the intervention was 19.06% (γ_{100}), and 9.93% (γ_{200}) in subsequent years. The results of the unconditional model suggest significant variability to be explained in pretest scores (π_{0ij}) and growth rates (π_{1ij}, and π_{2ij}) at both the student level (Level 2) and school level (Level 3). Furthermore, we found that 20.3% of the variance in pretest scores, 46.1% of

Table 7

Parameter Estimates for the Unconditional, Full, and Final Piecewise Linear Growth Models for Performance (Correctness)

	Unconditional Model	Full Model	Final Model
Fixed effects			
Model for baseline, π_{0ij}			
Intercept, γ_{000}	14.02 (0.52)***	14.48 (0.59)***	13.96 (0.43)***
Treatment, γ_{001}		1.57 (2.15)	
Socioeconomic status, γ_{002}		−0.06 (0.02)*	−0.09 (0.02)***
Treatment × socioeconomic status, γ_{003}		−0.04 (0.03)	
Model for Time 1, π_{1ij}			
Intercept, γ_{100}	19.06 (1.31)***	12.29 (0.94)***	12.14 (0.94)***
Treatment, γ_{101}		20.81 (3.45)***	20.74 (3.43)***
Socioeconomic status, γ_{102}		−0.07 (0.04)	−0.08 (0.04)*
Treatment × socioeconomic status, γ_{103}		−0.11 (0.05)*	−0.11 (0.05)*
Model for Time 2, π_{2ij}			
Intercept, γ_{200}	9.93 (0.40)***	9.25 (0.54)***	9.95 (0.40)***
Treatment, γ_{201}		−0.54 (1.97)	
Socioeconomic status, γ_{202}		−0.04 (0.02)	
Treatment × socioeconomic status, γ_{203}		0.03 (0.03)	
Random effects			
Level 1 (within students)			
Temporal variation, e_{tij}	92.45	92.30	92.28
Level 2 (between students)			
Initial status, r_{0ij}	39.45***	39.68***	39.66***
Time 1 growth rate, r_{1ij}	86.89***	87.06***	87.09***
Time 2 growth rate, r_{2ij}	12.68***	13.19***	13.21***
Level 3 (between schools)			
Initial status, u_{00j}	10.05***	5.48***	6.05***
Time 1 growth rate, u_{10j}	74.42***	15.12***	15.25***
Time 2 growth rate, u_{20j}	5.85***	5.04***	5.29***

Note. Standard errors are in parentheses.
$^*p < .05$. $^{**}p < .01$. $^{***}p < .001$.

variance in growth during the first year of the intervention, and 31.6% of variance in growth in subsequent years lies between schools, which is the level of our intervention. The results for pretest performance and both growth factors are on par with findings from cross-sectional research where variability between schools is approximately 10% to 30% (Raudenbush & Bryk, 2002).

We next considered an explanatory model, with predictor variables added at the school level. The Level 1and Level 2 models remained the same as in the unconditional model, while school-level predictors were added at Level 3 for pretest performance and each growth rate. Specifically, we were interested in the impact of the treatment on performance, controlling for school-level SES (% of students receiving free or reduced-price lunch), as well as the interaction between the treatment and SES. Treatment was dummy coded (0 = control, 1 = treatment), while SES was grand mean centered. The full level 3 model is specified as follows:

$$\beta_{00j} = \gamma_{000} + \gamma_{001}(\text{Treatment}) + \gamma_{002}(\text{SES}) + \gamma_{003}(\text{Treatment} \times \text{SES}) + u_{00j},$$

$$\beta_{10j} = \gamma_{100} + \gamma_{101}(\text{Treatment}) + \gamma_{102}(\text{SES}) + \gamma_{103}(\text{Treatment} \times \text{SES}) + u_{10j},$$

$$\beta_{20j} = \gamma_{200} + \gamma_{201}(\text{Treatment}) + \gamma_{202}(\text{SES}) + \gamma_{203}(\text{Treatment} \times \text{SES}) + u_{20j},$$

where γ_{001}, γ_{002}, γ_{003}, and so on, represent the effect of the specified predictor variable (treatment, SES, or their interaction) on pretest performance; γ_{101}, γ_{102}, and γ_{103} represent the effect of the predictors on growth during the first year of the intervention; and γ_{201}, γ_{202}, and γ_{203} represent the effect of the predictors on growth in Grades 4 and 5.

The results of the unconditional piecewise growth model, the full fitted conditional model, and the final model, with nonsignificant effects removed, can be found in Table 7. As shown in the final model, students' performance score at baseline (Grade 3 pretest) was statistically significantly predicted by SES, such that students from more affluent schools significantly outperformed their peers from less affluent schools prior to the start of the intervention. Every 1% increase in the number of students receiving free or reduced-price lunch was associated with a 0.09% reduction in performance at Grade 3 pretest. Given that free or reduced-priced lunch values ranged from 14% to 100% across schools, this equated to a 7.6% difference in pretest scores between schools at the highest and lowest levels of SES. Baseline performance was significantly predicted neither by treatment condition nor by the treatment by SES interaction effect, suggesting that our randomization was effective. Accordingly, these nonsignificant predictors were removed from the final model.

During the first year of the intervention, holding SES constant, students in our treatment condition improved at a significantly faster rate than students in the control condition, with students in the treatment condition gaining a 21% advantage over those in the control condition by the end of Grade 3. We also found a significant main effect of SES, suggesting that students at higher SES schools outperformed their peers at lower SES schools.[2] Overall, growth rates in performance decreased by 0.08% for every 1% increase in the number of students receiving free or reduced-priced lunch. For control schools, this equates to a 6.9% gap between the highest and lowest SES schools, an effect similar in magnitude to the one found at baseline.

However, the impact of the treatment was attenuated by SES. The significant treatment by SES interaction term suggests that students in our treatment condition from low-SES schools grew at a significantly slower rate than students in our treatment condition from relatively higher SES schools. Specifically, for every 1% increase in the number of students receiving free or reduced-price lunch, there was a 0.11% decrease in the growth rate. In combination with the main effect for SES described above, this equates to a 16.4% difference in growth rate between our highest and lowest SES schools in the treatment condition.

Importantly, however, regardless of SES, all students benefited from the intervention. To understand this more fully, consider that 9 out of our 46 schools had 100% of their participating students characterized as having low-SES backgrounds. Focusing solely on students in these nine lowest SES treatment and control schools, we found that students in the treatment condition improved significantly faster than control students, with treatment students outperforming controls by 9.2% at the end of Grade 3. In the latter 2 years of the intervention, there was a marginally significant difference in the rate of growth between treatment and control students, with treatment students gaining an additional 2.0% above and beyond the growth seen by control students. Though overall scores are lower and gains are attenuated, there remains clear evidence of improvement among these treatment students, relative to their control peers, due to the intervention (see Figure 3).

When combined, treatment condition, SES, and their interaction explained 80% of the between-school variability in growth in performance between Grade 3 pretest and Grade 3 posttest. Though significant between-school variation remains, the results suggest that the treatment condition explains a substantial amount of school-level variability in performance.

In the latter 2 years of the intervention (Time 2 of our model, between Grade 3 posttest and Grade 5 posttest), there was no significant difference in the rate of growth between treatment and control students. Importantly, however, the treatment effect seen in the initial year was maintained. That is, on average, control students did not gain on their peers in the treatment condition. The SES effect and the treatment by SES interaction were also no longer significant, suggesting that growth between the end of Grades 3 and 5 for performance on algebraic items was equitable across the range of SES.

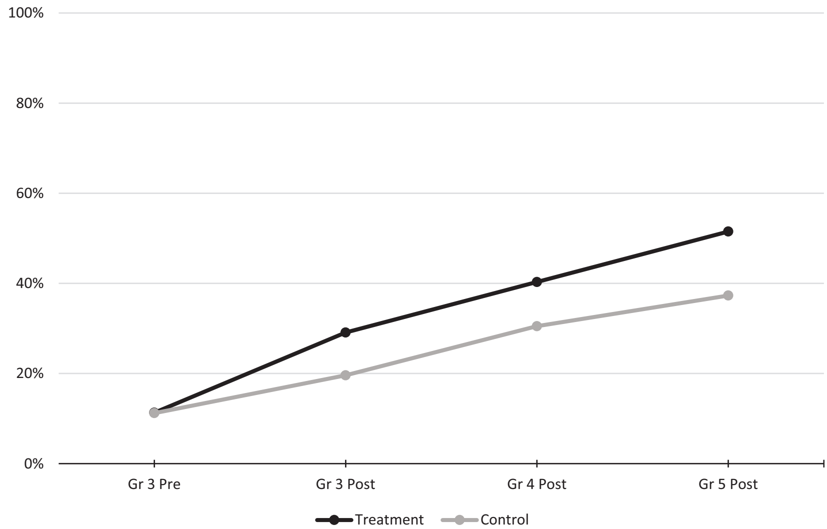


Figure 3. **Comparison of overall correctness for lowest socioeconomic status schools on algebra assessments.**

Structural Strategies

Figure 4 compares the average student use of structural strategies on common items on the algebra assessment for the given conditions, across the four testing points. Similar to the findings for performance, we employed a piecewise linear growth model based on the descriptive data coupled with an inspection of growth plots.

The unconditional piecewise growth model for structural strategy use was identical to that described above for performance. The results of the unconditional model revealed significant variability in pretest performance (π_{0ij}) and both growth factors (π_{1ij} and π_{2ij}) across schools (Level 3). However, at Level 2, only the growth factors (π_{1ij} and π_{2ij}) showed significant variability across individuals. Therefore, we treated baseline performance (π_{0ij}) as fixed at Level 2. With the random effect removed, the unconditional Level 1 and Level 3 models remained identical to the performance unconditional model, while the Level 2 model is altered as follows:

$$\pi_{0ij} = \beta_{00j}.$$
$$\pi_{1ij} = \beta_{10j} + r_{1ij}.$$
$$\pi_{2ij} = \beta_{20j} + r_{2ij}.$$

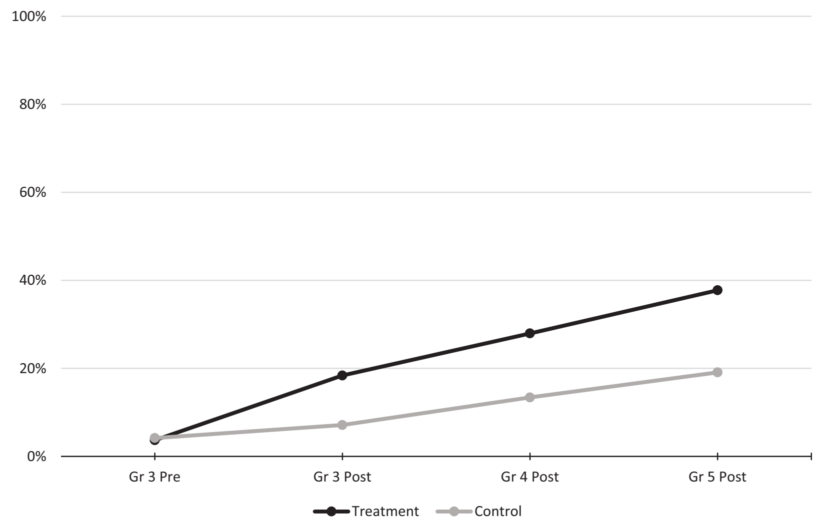


Figure 4. **Comparison of structural strategy use for common items on algebra assessments.**

The results of the unconditional piecewise growth model, the full fitted conditional model, and the final model, with nonsignificant effects removed, can be found in Table 8.

Across the entire population of student participants, the average pretest performance (γ_{000}) was 3.87% for structural strategy use. The average growth rate was 8.34% (γ_{100}) in the first year of the intervention and 6.62% (γ_{200}) in subsequent years. The results of the unconditional model revealed that 31.8% of the variability in growth during the first year of the intervention was between schools, while 18.7% of the variability in growth in subsequent years was between schools. These results are, again, on par with what is typically found in the literature with respect to between-school variability (Raudenbush & Bryk, 2002).

We subsequently fitted a conditional model that included all school-level predictors. Other than the fixed random effect at the student level, the full model was identical to that described previously for performance. Nonsignificant effects were removed, resulting in the final model.

As shown in the full model, students' structural strategy use score at baseline (Grade 3 pretest) was not statistically significantly predicted by any of the school-level factors. While it is not surprising that the treatment and interaction terms are nonsignificant at pretest given random assignment to intervention condition, the lack of an SES effect for strategy use departs from the findings for performance. This is likely due to the fact that use of

Table 8

Parameter Estimates for the Unconditional, Full, and Final Piecewise Linear Growth Models for Structural Strategy Use

	Unconditional Model	Full Model	Final Model
Fixed effects			
Model for baseline, π_{0ij}			
Intercept, γ_{000}	3.87 (0.18)***	4.10 (0.23)***	3.91 (0.18)***
Treatment, γ_{001}		1.14 (0.76)	
Socioeconomic status, γ_{002}		−0.01 (.01)	
Treatment × socioeconomic status, γ_{003}		−0.02 (0.01)	
Model for Time 1, π_{1ij}			
Intercept, γ_{100}	8.34 (1.07)***	2.86 (0.72)***	2.82 (0.73)***
Treatment, γ_{101}		19.38 (2.59)***	19.48 (2.63)***
Socioeconomic status, γ_{102}		−0.03 (0.03)	−0.03 (0.03)
Treatment × socioeconomic status, γ_{103}		−0.13 (0.04)**	−0.13 (0.04)**
Model for Time 2, π_{2ij}			
Intercept, γ_{200}	6.62 (0.44)***	5.24 (0.49)***	5.25 (0.49)***
Treatment, γ_{201}		3.13 (1.47)*	2.64 (0.69)***
Socioeconomic status, γ_{202}		−0.05 (0.02)*	−0.06 (0.01)***
Treatment × socioeconomic status, γ_{203}		−0.01 (0.03)	
Random effects			
Level 1 (within students)			
Temporal variation, e_{tij}	58.38	58.32	58.36
Level 2 (between students)			
Time 1 growth rate, r_{1ij}	105.05***	105.06***	105.05***
Time 2 growth rate, r_{2ij}	31.39***	31.42***	31.41***
Level 3 (between schools)			
Initial status, u_{00j}	0.56**	0.42	0.60
Time 1 growth rate, u_{10j}	49.03***	7.70***	8.00***
Time 2 growth rate, u_{20j}	7.23***	3.92***	3.96***

Note. Standard errors are in parentheses.
*$p < .05$. **$p < .01$. ***$p < .001$.

structural strategies was quite low across all students at the outset of the intervention (approximately 4%), suppressing any potential impact of SES.

During the first year of the intervention, holding SES constant, students in the treatment condition again improved at a significantly faster rate than students in the control condition, gaining approximately 19% above and beyond gains made by control students. However, similar to results for performance, this effect was attenuated by SES, with students in low-SES schools gaining at a significantly slower rate than students in higher SES schools. For control students, there was a 0.03% reduction in the use of structural strategies for every 1% increase in the number of students receiving free or reduced-price lunch, equating to a 2.6% gap in growth rates between the highest and lowest SES schools.

Compared to performance findings, the interaction effect was slightly larger for structural strategy use. For students in the treatment condition, every 1% increase in the number of students receiving free or reduced-priced lunch was associated with a 0.13% reduction in growth rate in the use of structural strategies during the first year of the intervention. In addition to the more general SES effect described above, this equates to a 13.8% gap between schools of the highest and lowest SES in the treatment condition. However, similar to our findings for performance, an exploration of the structural strategy use of students at the nine lowest SES schools (100% free or reduced-price lunch) showed that treatment students from low-SES backgrounds did in fact benefit from the intervention, in comparison to control students from similar SES backgrounds.

Combined, treatment condition, SES, and their interaction explained 84% of the between-school variability in growth rate for structural strategy use in the first year of the intervention. This effect is similar to what was found for performance and, again, points to the substantial explanatory power of the predictor variables.

In subsequent years (Time 2), the treatment advantage seen in the initial year decreased but was still present. This is unlike findings for performance, where the effect of the intervention leveled off in Time 2. Students in the treatment condition gained approximately 3% per year above and beyond the gains made by control students in strategy use. There was also a significant effect of SES in the expected direction, with every 1% increase in the number of students receiving free or reduced-price lunch associated with a 0.06% decrease in the growth rate. However, there was no significant treatment by SES interaction, suggesting that the treatment was equitable across the range of SES for structural strategy use gains during this time period. Treatment and SES combined explained 45% of the between-school variability in growth in the latter 2 years of the intervention.

Table 9
Average Student Score on End-of-Grade Standardized Assessment at Each Grade Level

Grade Level	Treatment	Control
Grade 3	449.51	450.33
Grade 4	448.97	450.00
Grade 5	450.68	451.53

Results on State Accountability Assessments

Separate multilevel analyses of students' performance on the state's end-of-grade standardized assessment in mathematics showed no significant difference between treatment and control students' performance in any of Grades 3 to 5. As shown in Table 9, scores were nearly identical between students in the treatment and control conditions at each grade level. This finding is not surprising, given that the state assessment is not well aligned with the intervention and therefore would not be sufficiently sensitive to changes in students' understanding of the algebraic content addressed in the intervention. Further analyses revealed that students' state assessment scores and outcomes on the algebra assessment (both correctness and strategy use) were uncorrelated, with Pearson's r statistics ranging from .02 to .04. This finding provides additional evidence that the two assessments are measuring different knowledge and skills. However, the implication that the intervention "did no harm" is an important finding in light of both the significant gains made in treatment students' algebraic thinking and the concerns teachers expressed initially that treatment students might not have sufficient time with content in their regular curriculum. In other words, the intervention only added value to children's elementary grades experiences relative to their potential algebra readiness and did not impede their learning of other mathematical content.

Discussion

Overall, the strong effect of the intervention in Grade 3 as measured by student performance (i.e., correctness) on common algebra assessment items resulted in a significant advantage for treatment students that they maintained throughout the intervention in Grades 4 and 5, placing them significantly ahead of control students in their understanding of core algebraic concepts and practices just prior to entering middle grades. A similar growth pattern was found in students' use of structural strategies, where treatment students were able to achieve a significant advantage over control students in the growth of their use of structural strategies in the first year of the intervention and even added to those gains in Grades 4 and 5.

While results for both outcome measures studied here hold important implications for the impact of early algebra in developing children's algebra readiness for middle grades, we particularly emphasize the significant effect of the intervention on students' use of structural strategies. Scholars have long argued that noticing and reasoning with mathematical structure—activities that we see as implicit in the algebraic thinking practices identified here—are essential to thinking algebraically (e.g., Kieran, 2007; Knuth et al., 2006; Pimm, 1995), yet difficult for students to achieve (Linchevski & Livneh, 1999). To further compound this, arithmetic instruction in elementary grades has not adequately emphasized attention to structure (Arcavi, Drijvers, & Stacey, 2017), resulting in opportunities missed for young learners to engage in structural reasoning.

Our analysis of students' use of structural strategies points to the impact of the intervention on students' "structure sense" (Linchevski & Livneh, 1999) in particular ways. Evidence of this is found, for example, in students' use of the compensation strategy to reason relationally about the equal sign. In tasks that required students to find a missing value in an equation or determine if an equation was true or false, treatment students were more successful in using a compensation strategy to correctly solve these tasks than control students.

Some have argued that the ability to notice, represent, and reason with fundamental properties (e.g., the Commutative Property of Addition) is central to structural reasoning (Mason, Stephens, & Watson, 2009). This, coupled with the view that students' difficulties with algebraic structure are rooted in arithmetic, particularly in students' lack of understanding of the structural aspects of the number system (e.g., Booth, 1988; Linchevski & Livneh, 1999), underscores the importance of interventions that strengthen students' understanding of structural properties in arithmetic. We found that treatment students were more successful than control students in noticing a structural property of the number system and using it as the basis of an argument. Treatment students were also better able than control students at building general arguments to justify arithmetic relationships, particularly with classes of numbers. Moreover, treatment students were better able to solve equations using strategies such as "unwinding" (Knuth et al., 2006; Koedinger & MacLaren, 2002), a strategy whereby students find the missing value in an equation by working backwards through inverse operations, rather than by a numerical (arithmetic) strategy such as "guess and test," whereby they guess values for the unknown and work forward to see if the values satisfy the equation.

Linchevski and Vinner (1990) further argue that students' success in algebra is predicated on their noticing of "hidden structures" in algebraic representations and that instruction that includes functional situations can increase students' attention to structure. Elsewhere, Blanton, Brizuela, et al. (2018) found that young children's inability to represent algebraic quantities was rooted more in their lack of perceiving variable quantities in mathematical situations rather than in difficulties with symbolic systems. Collectively, these studies point to students' challenges in seeing structure conveyed not just explicitly, in symbolic forms, but also implicitly in mathematical situations. We found that treatment students were better able than control students to perceive variable quantities in mathematical situations and represent these quantities and the relationships between them with variable notation across contexts where variable plays different roles, including as a fixed unknown, as a varying quantity, and as a generalized pattern (Blanton et al., 2011). This conveys treatment students' advantage over control students by way not only of structure sense but also of "symbol sense" (Arcavi, 2005), that is, their ability to represent and reason with variable notation in different types of mathematical situations.

We note that control students' performance and strategy use improved over time as well. This is promising, given the decades-long effort in the United States to reconceptualize teaching and learning algebra and the resulting increase in attention to algebraic ideas in elementary grades curricula, including the curricula used by schools participating in the study reported here. We suggest that the pattern of gains made by control students across Grades 3 to 5, which were greater in Grades 4 and 5, can be attributed in part to the increased attention to early algebraic concepts in these grades in students' regular curriculum, as noted earlier. However, the significant advantage treatment students exhibited in this study highlights the opportunities missed in current curricular approaches. Moreover, research shows that *when* students learn core algebraic ideas—not just *that* they do—matters, with earlier gains in knowledge leading to more success later (Alibali, Knuth, Hattikudur, McNeil, & Stephens, 2007). This, coupled with the strong effect of the intervention in Grade 3, suggests that the earlier introduction of algebraic thinking has the potential for significant benefits in terms of students' mathematical success.

Finally, we highlight findings relative to performance and strategy use for socioeconomically disadvantaged students participating in this study (i.e., treatment and control schools with 100% of student participants designated as from low-SES backgrounds). We also note that other relevant demographic data for these particular schools include, on average, a highly diverse population of students (approximately 94% persons of color, with 42.4% Hispanic and 51.4% Black). Thus, while our findings regarding the effect of the intervention on students' capacity to think algebraically are framed here relative to SES, we suggest that they can be viewed with an eye toward these other at-risk populations as well.

Research establishes a strong link between SES and student achievement (Chudgar & Luschei, 2009). By Grade 5, for example, students from lower SES backgrounds are 2 times more likely to lack the proficiency needed in math than more advantaged students (U.S. Department of Education, National Center for Education Statistics, 2007), placing this population of students at particular risk for later difficulties in algebra. Moreover, Grades 3 to 5, the period addressed by our intervention, represent some of the years of highest risk for students from low-SES backgrounds (Schmidt, McKnight, Cogan, Jakwerth, & Houang, 1999), making interventions in this grade band that can mitigate the risk all the more critical.

Since its inception, a primary motivation for early algebra has been to democratize access to algebra for students who have historically had limited access to the STEM pipeline due to algebra's gatekeeper effect (Kaput, 2008). The long-term implications of the inequity of access for students from low-SES backgrounds to STEM-related careers, given their already existing underrepresentation in the STEM pipeline (Oscos-Sanchez, Oscos-Flores, & Burge, 2008), are significant. Students from lower SES backgrounds are less likely than their more advantaged peers to complete high school or attend college and postgraduate school (Halle, Kurtz-Costes, & Mahoney, 1997), sharply decreasing their access to such careers. Yet research suggests that underserved students who receive equitable learning opportunities achieve STEM outcomes comparable to mainstream students (Lee, 2011). Thus, it was reasonable to expect that disadvantaged participants in our study, too, might show relative gains in their understanding of core algebraic thinking practices and concepts.

In this regard, we found that students from economically disadvantaged (and ethnically diverse) backgrounds could successfully engage with core algebraic practices and concepts when given appropriate curricular and instructional supports. Importantly, the algebra assessments used in this study were largely composed of nonroutine, open-response items. Results of standardized testing (e.g., National Assessment of Educational Progress) suggest that such problems are particularly challenging for students from low-SES backgrounds, who tend to perform more poorly than students from high-SES backgrounds on nonroutine problems (Lubienski, 2007). While we did find that students from high-SES backgrounds outperformed those from low-SES backgrounds, our findings on students in the most disadvantaged cases (i.e., in schools for which 100% of students received free or reduced-price lunch) showed that treatment students' responses followed the same pattern in terms of performance (correctness) and structural strategy use as the overall sample. Specifically, treatment students in the lowest SES schools showed a clear advantage over their control peers in similar schools in their ability to solve nonroutine algebra tasks. During the first year of the intervention, control students increased their overall percentage of structural strategy use from 3% to 6%, while treatment students jumped from 3% to 10%. This marks a significant advantage for treatment students. In Grades 4 and 5, control students, on average, gained 3% per year, while treatment students gained 6% per year. By the end of the study, control students were using structural strategies 12% of the time, while treatment students were using them more often, at 22% of the time. These gains by students from low-SES schools are particularly encouraging if we are to change the story on algebra's gatekeeper effect.

Conclusion

Does early algebra matter? Our findings suggest not only that a systematic effort to engage upper elementary grades students in understanding core algebraic thinking practices and concepts does matter in terms of developing their algebra readiness for middle grades but also that its effect can be felt among our most vulnerable populations. Moreover, while our central aim was to understand the impact of early algebra, there are secondary outcomes of this study that are important to note as well. In particular, the intervention used in this study reflects a research-based model of a comprehensive approach to early algebra instruction whose effectiveness is established through findings presented here. In this, our design methodology reinforces the utility of the CRF (Clements, 2007) as a template for developing effective, research-based curricula. Moreover, the study itself contributes to the body of experimental research that evaluates the effectiveness of curricular innovations, particularly among young learners and with at-risk populations and, thereby, addresses the call to "investigate the effects of programs, practices, and approaches on students' mathematics achievement" (U.S. Dept. of Education, 2008, p. 84).

In spite of the positive findings from this study, however, we acknowledge that there was still room for growth in algebraic understanding among participants (e.g., by the end of Grade 5, overall performance for treatment students was 65%). We also recognize the need to understand reasons for the dampened effect of the treatment in Grades 4 and 5 (and, accordingly, how the intervention or instruction might be modified to accelerate the rate of growth in students' algebraic thinking during this phase). To this end, we close here with some factors that might help us understand limitations in our results and how we are currently addressing these.

First, we note that the significant growth we observed among treatment students—including economically disadvantaged and ethnically diverse students—in comparison to students in business-as-usual, arithmetic-focused settings resulted after only a relatively small amount of instructional time (in one sense, about 20 hours of instruction per year for the intervention). The brevity of the intervention raises the question of the extent to which students' algebraic thinking might grow in ideal situations, where algebraic concepts and practices are woven more organically and continuously into mathematics instruction across all of elementary grades by classroom teachers who have deep expertise in building rich algebraic discussions around tasks that foster higher level thinking.

Second, we recognize that teachers in treatment schools were teaching, largely for the first time, an intervention that had content that was noticeably different from their more familiar arithmetic-based curriculum. As noted earlier, this is an important factor because it can take several years of PD to develop the expertise needed to implement an intervention with the highest fidelity (Jacobs et al., 2010; Superfine, 2008). In our view, the limited amount of time for the intervention, coupled with teachers' limited experience in teaching early algebraic ideas, can help explain what might be viewed as only modest overall growth among students who received the intervention.

To this end, our study points to the need for further critical research. We think that part of the answer to how we might further increase student performance lies in better understanding how to support elementary teachers. Elementary teachers in the United States have been underserved in their preparation for early algebra instruction (Greenberg & Walsh, 2008), a problem that is compounded by the fact that a disproportionately large number of preservice and in-service elementary teachers have deeply rooted anxieties about mathematics (Battista, 1986; Haycock, 2001)—particularly algebra—that can impede the confidence with which they teach children (Bursal & Paznokas, 2006). To this end, we are currently examining instruction for the intervention by teachers who implemented the intervention with highest fidelity in order to identify more nuanced aspects of instruction that support students' engagement with algebraic thinking practices.

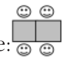
Part of the answer also lies in the depth to which early algebra instruction occurs, both in its duration across *all* of elementary grades and in its capacity to engage *all* learners. In response to this, we are extending our Grades 3 to 5 work in two important ways. First, we are currently expanding the intervention into Grades K–2 in order to develop a complete, comprehensive approach to early algebra across elementary grades, an approach that is consistent with calls by the Common Core State Standards for Mathematics (NGA Center & CCSSO, 2010) that the development of algebraic thinking begin at the start of formal schooling. Second, we are examining design principles by which our intervention can better support struggling learners (e.g., students with learning differences) and be more culturally sensitive to diverse populations. Finally, we are currently exploring how the performance and strategy use of treatment and control students in the study reported here change once they enter middle grades and are assimilated into common classrooms. Understanding retention in treatment students' knowledge of algebraic concepts and practices, as well as gains or losses for either group, will be important in identifying what should happen after elementary grades to keep students on a positive trajectory of development. In our view, working to address all of these factors can help us continue to find ways to mediate algebra's gatekeeper effect.

Designing effective, research-based approaches to the development of children's algebraic thinking in elementary grades that align with national learning standards is a critical need for STEM education, particularly if current college and career readiness standards and practices such as the Common Core Content Standards and Standards for Mathematical Practice (NGA Center & CCSSO, 2010) are to be implemented with fidelity. The early algebra innovation in the study reported here provides an effective curricular road map that can strengthen the implementation of these standards and practices relative to children's algebraic thinking. We are hopeful that such models can continue to change the national discourse on teaching and learning algebra and lead to more opportunities for success, for *all* students.

Appendix

Item No.	**Common Assessment Items**
1	Fill in the blank with the value that makes the number sentence true. Explain how you got your answer. $7 + 3 = \_\_\_\_ + 4$
 2-a 2-b 2-c	Circle True or False. Explain how you got your answer. a) 12 + 3 = 10 + 5 True False b) 57 + 22 = 58 + 21 True False c) 39 + 121 = 121 + 39 True False
   3-a 3-b	Marcy's teacher asks her to solve "23 + 15." She adds the two numbers and gets 38. The teacher then asks her to solve "15 + 23." Marcy already knows the answer without adding. a) Do you think Marcy's idea will work for any two numbers? Why or why not? b) Write an equation using variables (letters) to represent the idea that you can add two numbers in any order and get the same result.
4	Brian knows that anytime you add three odd numbers, you will always get an odd number. Explain why this is true.
    5-a 5-b 5-c1 5c2	Tim and Angela each have a piggy bank. They know that their piggy banks each contain the same number of pennies, but they don't know how many. Angela also has 8 pennies in her hand. a) How would you represent the number of pennies Tim has? b) How would you represent the total number of pennies Angela has? c) (c1) Angela and Tim combine all of their pennies. How would you represent the number of pennies they have all together? (c2) Suppose Angela and Tim now count their pennies and find they have 16 altogether. Write an equation that represents the relationship between this total and the expression you wrote above.
8	Find the value of n in the following equation. Show or explain how you got your answer. $10 \times n + 2 = 42$

Brady is celebrating his birthday at school. He wants to make sure he has a seat for everyone. He has square desks.

He can seat 2 people at one desk in the following way:  If he joins another desk to the first one, he can seat 4 people: 

If he joins another desk to the second one, he can seat 6 people: 

9-a a) Fill in the table below to show how many people Brady can seat at different numbers of desks.

Number of desks	Number of people
1	2
2	4
3	
4	
5	
6	
7	

9-b b) Do you see any patterns in the table from part a? If so, describe them.

c) Think about the relationship between the number of desks and the number of people.

9-c1 (c1) Use words to write the rule that describes this relationship.

9c2 (c2) Use variables (letters) to write the rule that describes this relationship.

9-d d) If Brady has 100 desks, how many people can he seat? Show how you got your answer.

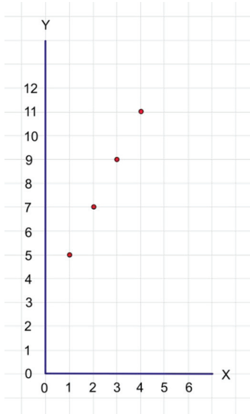
The table below shows the relationship between two variables, k and p. The rule $p = 2 \times k + 1$ describes their relationship.

10-a a) Some numbers in the table are missing. Use this rule to fill in the missing numbers.

k	p
1	3
2	
	9

10-b b) What is the value of p when $k = 21$? Show how you got your answer.

10-c c) What is the value of k when $p = 61$? Show how you got your answer.

<table>
<tr><td>12</td><td>Answer the following question for the graph given below:

Which of the following tables could have been used to construct the graph? Circle the table and explain why you think it is the correct table.

<table>
<tr><th>x</th><th>y</th><th>x</th><th>y</th></tr>
<tr><td>1</td><td>2</td><td>1</td><td>5</td></tr>
<tr><td>2</td><td>4</td><td>2</td><td>7</td></tr>
<tr><td>3</td><td>6</td><td>3</td><td>9</td></tr>
<tr><td>4</td><td>8</td><td>4</td><td>11</td></tr>
</table>
</td></tr>
</table>

Source. Assessment items were reproduced or adapted from Blanton et al. (2015).

Coding Scheme for Item 1:

Fill in the blank with the value that makes the number sentence true. Explain how you got your answer.

$7 + 3 = \_\_\_\_ + 4$

Code	Description	Sample Response
Structural (Compensation)	Student notices structure in the equation. Student may or may not additionally compute.	*4 is one more than 3, so the number in the blank must be one less than 7.*
Computational	Student adds the two numbers on the left side and then subtracts the number on the right side from the result or uses another computational strategy to equalize the two sides. Student states that he/she computed (with some detail about what was computed).	*7 + 3 = 10; 10 – 4 = 6* *7 + 3 = 10 and 6 + 4 = 10* *6. I added 7 + 3 and 6 + 4.*
Balance	Student makes vague statement about the equation balancing or the two sides being equal, but it is not clear how this conclusion was reached. (If correct, the strategy was likely S or C but there is just not enough information to categorize either way.)	*Both sides are balanced/not balanced.* *The sides have the same total/do not have the same total.* *The sides are equal/not equal.*
Operational	Student shows an operational understanding of the equal sign by indicating that the missing value is either the sum of the two numbers on the left side (7 and 3) or the sum of all numbers (7, 3, and 4). Also applies when student gives answer of 3 "because 4 plus 3 equals 7" (reversing equation and ignoring the 3 on the left side). An explanation is required in this case or the response is coded "answer only."	*7 + 3 = 10* *7 + 3 + 4 = 14* *4 + 3 = 7*
Other	Student uses strategy different from those above or strategy is not discernable.	
Answer only	Student gives numerical answer with no work shown. Includes "I guessed."	
No response	Student leaves the item blank or writes "I don't know" or "?" or any other marks not interpreted as a solution attempt.	

Notes

The research reported here was supported by the U.S. Department of Education under Institute of Education Sciences Award No. R305A140092. Any opinions, findings, and conclusions or recommendations expressed in this article are those of the authors and do not necessarily reflect the views of the U.S. Department of Education.

[1]We take *early algebra* to refer to algebraic thinking in elementary grades (i.e., Grades K–5). It should not be confused with "pre-algebra" courses offered in middle grades.

[2]We emphasize that, here, because the designation of high or low SES is made at the school level, the reason for lower mean school performance in schools designated as low-SES may be due to the school having fewer resources.

References

Alibali, M., Knuth, E., Hattikudur, S., McNeil, N., & Stephens, A. (2007). A longitudinal examination of middle school students' understandings of the equal sign and performance solving equivalent equations. *Mathematical Thinking and Learning*, *9*, 221–247.

Arcavi, A. (2005). Developing and using symbol sense in mathematics. *For the Learning of Mathematics*, *25*(2), 42–47.

Arcavi, A., Drijvers, P., & Stacey, K. (2017). *The learning and teaching of algebra: Ideas, insights, and activities*. London, England: Routledge.

Battista, M. (1986). The relationship of mathematical anxiety and mathematical knowledge to the learning of mathematical pedagogy by preservice elementary teachers. *School Science and Mathematics*, *86*, 10–19.

Battista, M. (2004). Applying cognition-based assessment to elementary school students' development of understanding of area and volume measurement. *Mathematical Thinking and Learning 6*, 185–204.

Blanton, M., Brizuela, B., Gardiner, A., & Sawrey, K. (2017). A progression in first-grade children's thinking about variable and variable notation in functional relationships. *Educational Studies in Mathematics, 95*, 181–202, doi:10.1007/s10649-016-9745-0

Blanton, M., Brizuela, B., Stephens, A., Knuth, E., Isler, I., Gardiner, A., . . . Stylianou, D. (2018). Implementing a framework for early algebra. In C. Kieran (Ed.), *Teaching and learning algebraic thinking with 5- to 12-year-olds: The global evolution of an emerging field of research and practice* (pp. 27–49). Cham, Switzerland: Springer International.

Blanton, M., Isler, I., Stroud, R., Stephens, A., Knuth, E., & Gardiner, A. (2018). *Growth in children's understanding of generalizing and representing mathematical structure and relationships*. Manuscript submitted for publication.

Blanton, M., Levi, L., Crites, T., & Dougherty, B. (2011). *Developing essential understanding of algebraic thinking for teaching mathematics in Grades 3-5* (Essential Understanding Series). Reston, VA: National Council of Teachers of Mathematics.

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*, 39–87.

Booth, L. R. (1988). Children's difficulties in beginning algebra. In A. F. Coxford & A. P. Schulte (Eds.), *The ideas of algebra, K–12* (pp. 20–32). Reston, VA: National Council of Teachers of Mathematics.

Bursal, M., & Paznokas, L. (2006). Mathematics anxiety and pre-service elementary teachers' confidence to teach mathematics and science. *School Science and Mathematics*, *106*, 173–179.

Chudgar, A., & Luschei, T. F. (2009). National income, income inequality, and the importance of schools: A hierarchical cross-national comparison. *American Educational Research Journal*, *46*, 626–658.

Clements, D. H. (2002). Linking research and curriculum development. In L. D. English (Ed.), *Handbook of international research in mathematics education* (pp. 599–636). Mahwah, NJ: Lawrence Erlbaum.

Clements, D. H. (2007). Curriculum research: Toward a framework for "research-based curricula." *Journal for Research in Mathematics Education*, *38*, 35–70.

Clements, D. H., & Sarama, J. (2004). Learning trajectories in mathematics education. *Mathematical Thinking and Learning*, *6*, 81–89.

Clements, D. H., & Sarama, J. (2008). Experimental evaluation of the effects of a research-based preschool mathematics curriculum. *American Educational Research Journal*, *45*, 443–494. doi:10.3102/0002831207312908

Cooper, T., & Warren, E. (2011). Years 2 to 6 students' ability to generalize: Models, representations, and theory for teaching and learning. In J. Cai & E. Knuth (Eds.), *Early algebraization: A global dialogue from multiple perspectives* (pp. 187–214). Heidelberg, Germany: Springer.

Fonger, N. L., Stephens, A., Blanton, M., Isler, I., Knuth, E., & Gardiner, A. (2018). Developing a learning progression for curriculum, instruction, and student learning: An example from mathematics education. *Cognition and Instruction*, *36*, 30–55. doi:10.1080/07370008.2017.1392965

Frank, K., & Seltzer, M. (1990, April). *Using the hierarchical linear model to model growth in reading achievement*. Paper presented at the Annual Meeting of the American Educational Research Association, Boston, MA.

Greenberg, J., & Walsh, K. (2008). *No common denominator: The preparation of elementary teachers in mathematics by America's education schools*. Washington, DC: National Council on Teacher Quality. Retrieved from www.nctq.org/publications/No-Common-Denominator:-The-Preparation-of-Elementary-Teachers-in-Mathematics-by-Americas-Education-Schools

Halle, T. G., Kurtz-Costes, B., & Mahoney, J. L. (1997), Family influences on school achievement in low-income, African American children. *Journal of Educational Psychology*, *89*, 527–537.

Haycock, K. (2001). Closing the achievement gap. *Educational Leadership*, *58*(6), 6–11.

Jacobs, V. R., Lamb, L. L. C., & Philipp, R. A. (2010). Professional noticing of children's mathematical thinking. *Journal for Research in Mathematics Education*, *41*, 169–202.

Jones, I., Inglis, M., Gilmore, C., & Dowens, M. (2012). Substitution and sameness: Two components of a relational conception of the equals sign. *Journal of Experimental Child Psychology*, *113*, 166–176.

Kaput, J. (1999). Teaching and learning a new algebra. In E. Fennema & T.A. Romberg (Eds.), *Mathematical classrooms that promote understanding* (pp. 133–155). Mahwah, NJ: Lawrence Erlbaum.

Kaput, J. (2008). What is algebra? What is algebraic reasoning? In J. Kaput, D. Carraher, & M. Blanton (Eds.), *Algebra in the early grades* (pp. 5–17). Mahwah, NJ: Lawrence Erlbaum.

Kaput, J., Blanton, M., & Moreno, L. (2008). Algebra from a symbolization point of view. In J. Kaput, D. Carraher, & M. Blanton (Eds.), *Algebra in the early grades* (pp. 19–55). Mahwah, NJ: Lawrence Erlbaum.

Kieran, C. (2007). Learning and teaching of algebra at the middle school through college levels: Building meaning for symbols and their manipulation. In F. K. Lester Jr. (Ed.), *Second handbook of research on mathematics teaching and learning* (Vol. *2*, pp. 707–762). Charlotte, NC: Information Age.

Kilpatrick, J., Swafford, J., & Findell, B. (Eds.). (2001). *Adding it up: Helping children learn mathematics*. Washington, DC: National Academies Press.

Knuth, E. J., Stephens, A. C., McNeil, N. M., & Alibali, M. W. (2006). Does understanding the equal sign matter? Evidence from solving equations. *Journal for Research in Mathematics Education*, *37*, 297–312.

Koedinger, K., & MacLaren, B. (2002). *Developing a pedagogical domain theory of early algebra problem solving*. Retrieved from http://pact.cs.cmu.edu/pubs/Koedinger,%20McLaren%20.pdf

Lee, O. (2011, May). *Effective STEM strategies for diverse and underserved learners*. Paper presented at the National Research Council's Workshop on Successful STEM Education in K–12 Schools, Washington, DC.

Linchevski, L., & Livneh, D. (1999). Structure sense: the relationship between algebraic and numerical contexts. *Educational Studies in Mathematics*, *40*, 173–196.

Linchevski, L., & Vinner, S. (1990). Embedded figures and structures of algebraic expressions. In G. Booker, P. Cobb, & T. N. de Mendicuti (Eds.), *Proceedings of the 14th conference of the International Group for the Psychology of Mathematics Education* (Vol. *2*, pp. 85–92). Oaxtepec, Mexico: PME.

Little, R., & Rubin, D. (1987). *Statistical analysis with missing data*. New York, NY: John Wiley.

Lubienski, S. (2007). What we can do about achievement disparities. *Educational Leadership*, *65*(3), 54–59.

Maloney, A.P., Confrey, J., & Nguyen, K. (Eds.). (2011). *Learning over time: Learning trajectories in mathematics education*. Charlotte, NC: Information Age.

Mason, J. (2008). Making use of children's powers to produce algebraic thinking. In J. Kaput, D. Carraher, & M. Blanton (Eds.), *Algebra in the early grades* (pp. 57–94). Mahwah, NJ: Lawrence Erlbaum/Taylor.

Mason, J., Stephens, M., & Watson, A. (2009). Appreciating mathematics structure for all. *Mathematics Education Research Journal*, *21*(2), 10–32.

Mihalic, S. (2002, April). *The importance of implementation fidelity*. Boulder, CO: Center for the Study and Prevention of Violence.

Moses, R. P., & Cobb, C. E. (2001). *Radical equations: Math literacy and civil rights*. Boston, MA: Beacon Press.

Museus, S., Palmer, R. T., Davis, R. J., & Maramba, D. C. (2011). *Racial and ethnic minority students' success in STEM education*. Hoboken, NJ: Jossey-Bass.

National Council of Teachers of Mathematics. (1989). *Curriculum and evaluation standards for school mathematics*. Reston, VA: Author.

National Council of Teachers of Mathematics. (2000). *Principle and standards for school mathematics*. Reston, VA: Author.

National Council of Teachers of Mathematics & Mathematical Sciences Education Board. (Eds.). (1998). *The nature and role of algebra in the K–14 curriculum: Proceedings of a national symposium*. Washington, DC: National Research Council, National Academies Press.

National Governors Association Center for Best Practices & Council of Chief State School Officers. (2010). *Common core state standards for mathematics*. Washington, DC: Council of Chief State School Officers. Retrieved from http://www.corestandards.org/assets/CCSSI_Math%20Standards.pdf

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

Sciences Education Board, Center for Education, Division of Behavioral and Social Sciences and Education, National Academies 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*, 33–84.

Oscos-Sanchez, M. A., Oscos-Flores, L. D., & Burge, S. K. (2008). The Teen Medical Academy: Using academic enhancement and instructional enrichment to address ethnic disparities in the American healthcare workforce. *Journal of Adolescent Health*, *42*, 284–293.

Pimm, D. (1995). *Symbols and meanings in school mathematics*. London, England: Routledge. doi:10.4324/9780203428610

Raudenbush, S. W. (2007). Designing field trials of educational innovations. In B. Schneider & S.-K. McDonald (Eds.), *Scale-up in education: Vol. 2. Ideas in practice* (pp. 23–40). Lanham, MD: Rowman & Littlefiel.

Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical linear models: Applications and data analysis methods* (2nd ed.). Thousand Oaks, CA: Sage.

Rittle-Johnson, B., Matthews, P. G., Taylor, R. S., & McEldoon, K. L. (2011). Assessing knowledge of mathematical equivalence: A construct-modeling approach. *Journal of Educational Psychology*, *103*, 85–104. doi:10.1037/a0021334

Schifter, D. (2009). Representation-based proof in the elementary grades. In D. Stylianou, M. Blanton, & E. Knuth (Eds.), *Teaching and learning proof across the grades: A K-16 perspective* (pp. 87–101). New York, NY: Taylor & Francis.

Schmidt, W. H., McKnight, C. C., Cogan, L. S., Jakwerth, P. M., & Houang, R. T. (1999). *Facing the consequences: Using TIMSS for a closer look at U.S. mathematics and science education*. Dordrecht, Netherlands: Kluwer Academic.

Schoenfeld, A. H. (1995). Report of Working Group 1. In C. Lacampagne, W. Blair & J. Kaput (Eds.), *The algebra colloquium: Vol. 2. Working group papers* (pp. 11–18). Washington, DC: U.S. Department of Education, Office of Educational Research and Improvement.

Sfard, A. (1991). On the dual nature of mathematical conceptions: Reflections on processes and objects as different sides of the same coin. *Educational Studies in Mathematics*, *22*(1), 1–36. doi:10.1007/BF00302715

Shin, N., Stevens, S. Y., Short, H., & Krajcik, J. S. (2009, June). *Learning progressions to support coherence curricula in instructional material, instruction, and assessment design*. Paper presented at the Learning Progressions in Science (LeaPS) conference, Iowa City, IA.

Simon, M. (1995). Reconstructing mathematics pedagogy from a constructivist perspective. *Journal for Research in Mathematics Education*, *26*, 114–145.

Stigler, J. W., Gonzales, P., Kawanaka, T., Knoll, S., & Serrano, A. (1999). *The TIMSS videotape classroom study: Methods and findings from an exploratory research project on eighth-grade mathematics instruction in Germany, Japan, and the United States* (NCES1999-074). Washington, DC: National Center for Education Statistics.

Stylianou, D., Stroud, R., Cassidy, M., Stephens, A., Knuth, E., Gardiner, A., & Demers, L. (in press). Putting early algebra in the hands of elementary school teachers: Examining fidelity of implementation and its relation to student performance. *Infancia y Aprendizaje/Journal for the Study of Education and Development*.

Superfine, A. S. (2008). Planning for mathematics instruction: A model of experienced teachers' planning processes in the context of a reform mathematics curriculum. *The Mathematics Educator*, *18*(2), 11–22.

U.S. Department of Education. (2008). *Foundations for success: The final report of the National Mathematics Advisory Panel*. Retrieved from www.ed.gov/about/bdscomm/list/mathpanel/index.html

U.S. Department of Education, National Center for Education Statistics. (2007). *The condition of education 2007* (NCES 2007-064). Washington, DC: U.S. Government Printing Office.

Manuscript received July 9, 2018
Final revision received December 17, 2018
Accepted January 24, 2019