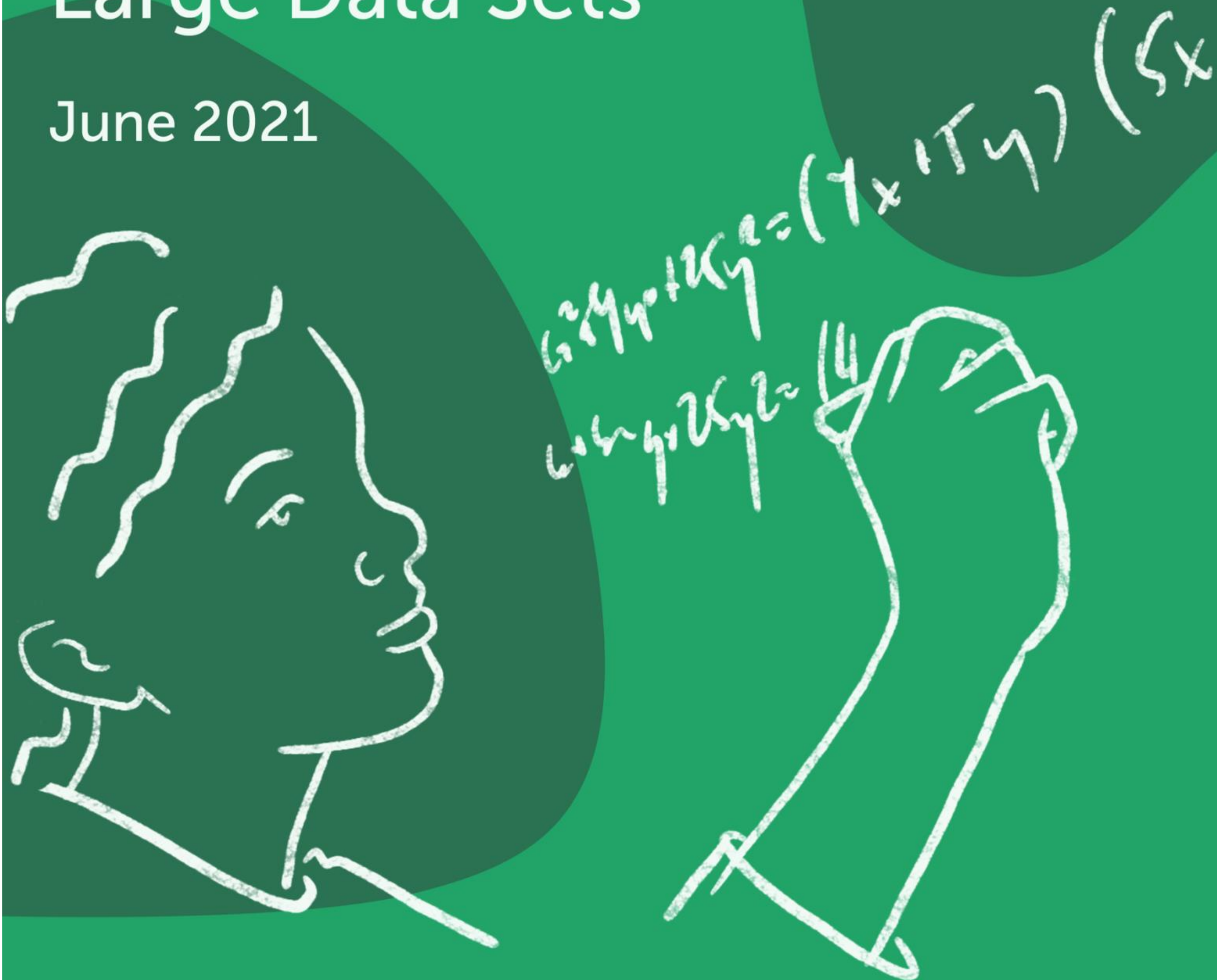
STUDENT
ACHIEVEMENT
PARTNERS

Preparation for Success in Algebra:

Exploring Math Education Relationships by Analyzing Large Data Sets

June 2021



About Student Achievement Partners

Student Achievement Partners believes challenging K-12 academic standards, in the context of culturally relevant teaching and equitable classroom practice, are non-negotiable when it comes to improving student outcomes. We believe that effective educational leadership and instruction persistently takes action to eradicate racist systems and policies so that all students can thrive academically.
In partnership with other passionate change-makers, we design tools and resources, professional learning, and other supports, grounded in research and the realities of the classroom.

Student Achievement Partners. (2021). Preparation for Success in Algebra: Exploring Math Education Relationships by Analyzing Large Data Sets (EMERALDS).

Table of Contents

Acknowledgements

This report was made possible through the generous support of the Bill & Melinda Gates Foundation and benefitted greatly from the thought partnership of Karen Johnson, Chaya Jones, and Rachel Leifer.

We gratefully acknowledge the contributions of the many individuals who made this report possible. David C. Geary, Ph.D., Curators' Distinguished Professor at the University of Missouri, led the development of the research design and is the primary author of the report. Shelbi Cole, Ph.D., Jessica Eadie, John Young, Ph.D., and Jason Zimba, Ph.D. from Student Achievement Partners provided key research and content expertise and overall management of the project along with project manager Ruth McKenna. Phil Daro developed the research questions and assisted in identifying content clusters for the study.

The refinement of the methodology and data analyses of Idaho, Washington, and Carnegie Learning MATHia data were led by Paulo Carvalho, Ph.D., Hui Cheng, Ph.D., and Ken Koedinger, Ph.D. from the Human-Computer Interaction Institute at Carnegie Mellon University (CMU). Educational Testing Service research scientist Jonathan Weeks, Ph.D. along with Patricia Baron, Ed.D., Terran Brown, Ph.D., Dan McCaffrey, Ph.D., and Max Wise helped to refine the methodology and analyzed the California state data.

Our report benefited greatly from thoughtful, probing feedback from The EMERALDS Research Advisory Committee members Robert Q. Berry III, Ph. D., David Conley, Ph.D., Linda Ruiz Davenport, Ph.D., Ray Hart, Ph.D., Neil Heffernan, Ph.D., Gerunda Hughes, Ph.D., Bill Schmidt, Ph. D., and Uri Treisman, Ph.D. We are grateful for the time and expertise they each have contributed to make the study stronger.

We also thank the Working Team Advisors, Derek Briggs, Ph.D. and Katherine Picho-Kiroga, Ph.D. Their constructive criticisms have been instrumental in refining the analyses and findings.

The study would not have been possible without the sponsorship of staff in the state departments of education who saw the potential for using longitudinal state data sets to discover findings that may support improvements in student achievement in early-grade mathematics for future success in Algebra. Shanine Coats, Johanna Harder, Jonathan Isler, Ph.D., Jonathan Lee, Constantino Silva, Mao Vang, Ed.D., and Eric Zilbert Ph.D. at the California Department of Education, and Karlynn Laraway, Kevin Whitman, and Kevin Chandler at the Idaho Department of Education, were instrumental in moving the data requests through the state processes. In Washington State, Deb Came, Ph.D., Julie Hoff, Kathe Taylor, Arlene Crum, Serena O'Neill, and Anton Jackson supported the process to participate in the study.

Lastly, the staff at the Smarter Balanced Assessment Consortium, Tony Alpert, Magda Chia, Ph.D., Rachel Kachchaf, Ph.D., Don Murray, Leslie Pearlman, Ph.D., and Carol Woods, Ph.D., helped with questions about the assessment design by sharing the Smarter Balanced item specifications, item metadata, and item parameters for the study. Similarly, Stephen Fancsali, Ph.D., Michael Sandbothe, and Steve Ritter, Ph.D. from Carnegie Learning and MATHia assisted CMU in understanding the design and components of the blended learning system.

The report is based on research funded by the Bill & Melinda Gates Foundation. The findings and conclusions contained within are those of the authors and do not necessarily reflect views, positions, or policies of the Bill & Melinda Gates Foundation, the state departments of education in any of the participating states, Smarter Balanced, nor the advisors.

Foreword

The doing of mathematics has always been inseparable from the teaching of it. Some of the oldest surviving written works of mathematics, created some four thousand years ago, appear to be problem-books intended for scribes and scholars. Some of those problems are still in the school mathematics curriculum today, four thousand years later. So are some of the instructional methods. Meanwhile, the frontiers of mathematical discovery have expanded at an ever-accelerating pace, allowing us to stand on the surface of the moon, to model the global climate, and soon perhaps to decrypt codes once thought unbreakable. I sometimes wonder what it would be like if our knowledge about the teaching and learning of mathematics could advance as quickly, as profoundly, as our knowledge of mathematics itself has.

It is a pleasure to introduce this EMERALDS report, which synthesizes the work of many contributors. Grateful thanks are due to the team of researchers, advisors, and educators who lent their abundant talents to the project. Please see page 5 for acknowledgement of the individuals and institutions without whom this report would not have been possible. The value of their efforts will be to help the field promote student success and equitable outcomes in mathematics, as detailed in the recommendations beginning on page 76.

This EMERALDS study used anonymized assessment data for some of its findings, though not for purposes of measuring differences between students or between groups of students. Indeed, the study was designed to look critically not at students, teachers, or schools, but rather at school mathematics itself. What did that critical look reveal? First, the results as I interpret them tend to validate previous smaller-scale studies pointing to the importance of investing in arithmetic in the elementary grades to get students ready for success in middle-grades algebra. Second, the design of the study also revealed a finding that strikes me as new: namely, that these patterns of mathematical progress from arithmetic to algebra seem to hold across student demographics and locations.

The study also made use of another type of data, anonymized data arising from usage of a digital Algebra curriculum in the middle grades. The analysis of this data source could support a particular claim of curricular efficacy and may also suggest, in a larger sense, a continuation of the story begun by the two findings about arithmetic described above. That is, if the report's findings about arithmetic emphasize the importance of getting a good start in a mathematical education—for every student—then the report's findings from the middle grades strengthen my optimism that our students can thrive in their formative studies of algebra. Our students deserve all the passion and support we can dedicate toward helping them do so.

Jason Zimba
Founding Partner of Student Achievement Partners
CEO and Cofounder of Math Milestones, LLC

Executive Summary

This final report on the Exploring Math Education Relationships by Analyzing Large Data Sets (EMERALDS) study shares both results and recommendations from the project. EMERALDS involved large-scale, four-year longitudinal student assessment data sets from the states of Idaho (n = 42,474), Washington (n= 216,595), California (n = 849,775), as well as more detailed assessments of middle school students' engagement with algebraic material in the MATHia Tutor (formerly Cognitive Tutor) as related to their Algebra I End-of-Course (EOC) performance (n = 1,304). The project was designed to explore ways to better support students who are Black, Latino, English learner-designated, experiencing poverty, and/or female to achieve better results in Algebra. The latter is the language of higher-level mathematics, a passport for expanding postsecondary opportunities, and a tool for confidently navigating the quantitative demands of daily life. To our knowledge, the result was the largest longitudinal assessments of what mathematics prepares students for success in Algebra. The study looked at student data from more than a million students representing a broad range of racial/ethnic backgrounds, income levels, and geographic regions.

The initial goal was to identify the core mathematics competencies at the end of the elementary school years that best predict students' later success in core Algebra topics, above and beyond overall mathematics competence. The measure used was the state-administered Smarter Balanced Assessment Consortium (SBAC) assessment, which is based on Common Core State Standards for Mathematics (CCSSM; hereafter CCSS). Content experts identified core Prealgebra knowledge domains and Algebra content areas with the goal of establishing clusters—groups of related items—for use in the analyses. The design of the SBAC assessment makes it well suited to providing an overall estimate of mathematics competencies, but our analyses suggest that measures such as this are not as well suited to identifying fine-grained subsets of competencies such as those explored in this study.

Despite this limitation, there were some competencies toward the end of the elementary school years that appeared to be more important than others in predicting Algebra outcomes four years later, but the effects were small compared to overall mathematics competencies. The latter was a substantive predictor of later performance in core Algebra areas and indicates that readiness for Algebra is dependent on a solid foundation in the elementary school CCSS. The measure for this study—the SBAC assessment—follows the CCSS in upper elementary grades, which emphasize the concepts, procedural fluencies, and applications of arithmetic (such as base-10 and fractions knowledge), as well as CCSS's practice standards for complex problem solving, reasoning & communicating, and the ability to use modeling to solve real-world problems. Accordingly, the results suggest a validation of the balanced rigor framed by the CCSS for arithmetic as the foundation for Algebra. The more specific predictors of later Algebra outcomes suggested effects that cut across content topics: multi-step problems were more predictive than one-step problems, and Mathematical Reasoning & Communication added the most predictive value above and beyond the total mathematics score. These are aspects of the CCSS Mathematical Practices, which were incorporated into the construct assessed by SBAC.

Middle school students' engagement and success with the MATHia system substantively improved their Algebra I EOC performance, above and beyond their end of elementary school mathematics competencies. The MATHia system includes translating word problems into equations and then correctly solving those equations and thus converges with the above noted results for multi-step problems. Moreover, extensive and successful engagement with MATHia reduced the Algebra I EOC performance differences comparing students with higher and lower prior mathematics scores. More fine-grain analyses suggest that, in addition to skill at translating word problems into algebra equations, an understanding of the different ways in which functions can be represented (e.g., equation, graphically) was uniquely related to later Algebra I EOC performance, controlling for prior mathematics scores. There were also several frequently engaged CCSS topics, such as basic work with functions, that did not predict later

Algebra I EOC performance. These conclusions should be interpreted cautiously, however, given the relatively small sample on which they are based.

A related goal was to determine if the strength of the relation between earlier mathematics competencies and later Algebra outcomes varied across subpopulations of students; note that this did not involve comparing the scores of students in different demographic groups. We addressed this goal using the large and diverse sample from California. There were several statistically significant differences due to the large sample, but none of these was practically important, and thus the study did not find group differences in the relation between earlier mathematics competencies, either overall or for specific skills (e.g., Reasoning & Communicating), and later Algebra scores. The results do show that a strong mathematical foundation in CCSS arithmetic topics at the end of the elementary school years is critical preparation for later Algebra, independent of student gender, ethnicity, race, disability status, English learner designation, eligibility for reduced or free lunches, or students' prior English language arts competencies. This is not to say, however, that students in all of these groups have had the same opportunities to learn this foundational material, but rather those who have solid skills by the end of the elementary school years (independent of demographic group) are on track for later success in Algebra. The results from the MATHia analyses were consistent with this conclusion but also revealed significant differences across these groups in engagement with MATHia. Lower engagement with MATHia, or a similar Algebra curriculum, will likely contribute to later differences in Algebra I EOC performance.

Another goal was to determine if there were substantive district, school, and classroom differences in Algebra outcomes, controlling for students' prior mathematics competencies. The data set did not allow assessment of classroom effects. In California, the results for districts and schools indicated that they were more similar than different, once prior achievement was considered. This does not mean that district- or school-wide reforms cannot substantively improve the mathematical development of their students; indeed, Carver-Thomas and Podolsky's (2019) assessment of individual districts in California indicates that systematic and wide-scale (e.g., involving teacher training, higher expectations for student performance, rigorous standards) reforms can have substantive and positive effects on the mathematical development of students who have been historically marginalized in educational systems. The MATHia analyses provided some evidence for this in middle schools in Florida. In terms of later Algebra I EOC performance, students in nearly all schools gained from engagement with and mastery of MATHia workspaces. The gains were, however, more variable for schools with more than 90% students of color, with the largest and smallest gains emerging for these schools. The reasons for the variation in these outcomes are not clear but merit follow-up study.

The analyses also sparked additional ideas for follow-up projects that could build upon these results. The first is to explore in greater detail the core components of elementary school mathematics. We were unable to do this with the data set, but it could be achievable with a modified replication of this study, using a fixed-form assessment with sufficient items in key predictor content in other states with different geographic and student demographic profiles. Such a study would benefit from close partnership with state departments of education and districts to understand aspects of curricular implementation, instruction, and other contextual factors, such as student experience, for mathematics success. The second is to follow up on the promising MATHia results; specifically, that intensive engagement with computer adaptive tutoring during the middle school years results in important gains in later Algebra I performance, controlling for prior mathematics competencies, and can help accelerate learning for historically marginalized students. Such a study would require close collaboration with providers of computer adaptive tutoring and with the schools and districts that use these programs. The current findings are based on a relatively small longitudinal sample with minimal contextual information. Follow-up with larger samples as well as greater understanding of the usage and student experience of the programs would help to verify our findings and enable a more fine-grain assessment of how computer adaptive tutoring supports mathematical academic identity if there are experience and usage differences for students who are Black, Latino, English learner-designated, experiencing poverty, and/or female.

The recommendations weave together the findings of this study, research outside the study, and the discussions of the Research Advisory Committee. Building on the findings that confirm a strong mathematical foundation in arithmetic as richly construed by CCSS at the end of elementary school leads to success in Algebra, the recommendations outline steps that states, districts, curriculum developers, and professional development providers can take to help students attain this mathematical foundation. Additionally, given the promising initial results of the Algebra tutoring program, the recommendations also outline steps that states, districts, curriculum developers, and professional development providers can take to help students during the middle school years to make gains in later Algebra I performance.

Recommendations

What should states and districts do to help every student leave elementary school with a solid foundation in elementary school CCSS and support middle school mathematics growth?

1. Communicate to teachers, students, families and caregivers, and the community the importance of a strong mathematical foundation in elementary grades for later success in Algebra. The messaging must not lead to middle-grade students being denied opportunities to learn pre-algebra and algebra on the basis of their opportunities in elementary grades.
2. Adopt an integrated, arithmetic-focused curriculum for the entire elementary grade span. Structure adoption processes to ensure the curriculum is designed to explicitly support teachers to facilitate the learning of students who have been historically marginalized by ensuring their unique identities, culture and needs are honored.
3. Support student transitions from elementary to middle school and middle school to high school by maintaining coherence of the K–12 mathematics learning pathway.
4. Provide professional learning opportunities that help teachers develop their own strong mathematical identity and a solid understanding of the key mathematical threads of their curricular programs.
5. Consider providing students with supplemental grade-level practice for content with significant evidential support for improving Algebra performance

What should curriculum developers do to support states, districts, teachers, and students to succeed in teaching and learning a coherent mathematics curriculum?

6. Invest in designing materials and explicit support for teachers in order to focus on students who have been historically marginalized by ensuring their unique identities, culture, and needs are honored.
7. Design curricular materials and programs—including supplementals—in the elementary school grades that emphasize the concepts, procedural fluencies, and applications of arithmetic as well as CCSS's practice standards for complex problem solving, communicating reasoning, and the ability to use modeling to solve real-world problems.
8. Attend to the content and coherence of the curricula materials, but also their mathematical fidelity and the quality of the mathematical tasks with which students are asked to engage.

What should designers of professional development and teacher preparation programs do to support teachers in helping historically marginalized students succeed in learning a solid foundation in elementary mathematics and have success in Algebra?

9. Design professional learning which helps teachers develop their own strong mathematical identity in order to positively impact their teaching of mathematics.
10. Design professional learning to support K–5 teachers to develop a solid understanding of the key threads of their curricular programs, specifically how knowledge at earlier grades provides the foundation for later learning and is not only a steppingstone but also is conceptually related to later material.
11. Require preservice teachers to take one or more courses aimed at helping teachers develop a solid mathematical understanding of fundamental mathematical concepts and the conceptual connections.

What should researchers do to support students, teachers, states, districts, curriculum developers and designers of professional development and teacher preparation programs?

12. Explore in greater detail the core components of a strong elementary school mathematics foundation through a modified replication of this study.
13. Follow up on the promising middle school MATHia results to understand the extent to which engagement with computer adaptive tutoring during the middle school years results in gains in later Algebra I performance.
14. Look inside upper elementary classrooms to learn about key curricular and instructional factors that make a difference for students who are Black, Latino, English learner-designated, experiencing poverty, and/or female and who are successful in upper elementary grades mathematics.
15. Examine how targeted professional learning for upper elementary teachers may affect the performance of students.

Introduction

Competence with algebra is the foundation for learning the more complex mathematics demanded in science, technology, engineering, and mathematics (STEM) fields (NMAP; National Mathematics Advisory Panel, 2008). Improving students' understanding of algebra has been a long-term educational priority in the United States; however, achieving this goal has been elusive (Stein et al., 2011), especially for students who have been historically underrepresented in STEM fields. For example, the eighth-grade mathematics section of the 2019 National Assessment of Educational Progress (NAEP) defines basic skills as including conceptual and procedural competence with whole and rational numbers. Although these are critical mathematics competencies (Siegler & Braithwaite, 2017), it is unlikely that eighth graders with only these basic skills are on track for successfully completing a rigorous high school Algebra course (NMAP, 2008). Overall, only ten percent of eighth-grade students have achieved the advanced competencies that position them well for rigorous high school mathematics coursework.

The content coverage of the NAEP mathematics assessment—a long-standing benchmark of U.S. students' educational progress—and the more recent Common Core State Standards (CCSS) are not completely aligned in terms of relative content emphases (e.g., emphasis on fractions). As a result, the percentage of students from the same population identified as adequately prepared could vary across assessments (Daro et al., 2015; Hughes et al., 2019). Even so, both assessment approaches reveal that a majority of U.S. students are not fully prepared for a rigorous course in high school Algebra. A key goal of the CCSS was to focus standards on the most critical procedures, concepts, and problem-solving skills that best prepare students for the high school mathematics curriculum and, through this, lay the foundation for college mathematics and entry into the workforce (Zimba, 2014). The focus of the CCSS, in turn, was based on mathematics standards from countries that consistently produce students who are well educated in mathematics and on recommendations for how U.S. students might achieve the same (e.g., NMAP, 2008; Schmidt & Houang, 2012). Whereas most mathematics content covered in elementary and middle school, and highlighted in CCSS, will have utility in some contexts, it is not likely that all this content is equally critical in terms of preparation for high school Algebra.

The current project is an attempt to identify specific CCSS procedural, conceptual, and problem-solving competencies in earlier grades that provide the most critical foundation for success in algebraic areas in later grades. This endeavor is part of an overall effort to improve outcomes and better support all students, but especially for those who are Black, Latino, English learner-designated, experiencing poverty, and/or female and have been historically underserved. Gaining algebraic competence is undergirded by strong procedural and conceptual competencies in key areas like fractions (Mou et al., 2016; Hurst & Cordes, 2018; Siegler et al., 2012). If these key procedural, conceptual, and problem-solving competencies can be identified, it will be an important step toward better preparing all U.S. students for success in high school Algebra, although identification alone is insufficient. Rather, the quantity and quality of the opportunities to learn this content, in addition to other factors like family background and students' engagement in classroom settings, must be taken into consideration (Bailey et al., 2014; Lee & Bull, 2016; Geary et al., 2017).

The available data did not allow us to control directly for all these myriad influences. To indirectly control for them, and because of the nature of the available data at the state level, we focused in our first approach on the relative strengths of individual students. This approach involved use of data from the Smarter Balanced computer adaptive test (below) to estimate how students' performance differed from expectations (based on their overall mathematics scores) on key Prealgebra and Algebra areas. The approach controls for factors that broadly influence mathematics achievement, while at the same time allowing an assessment of whether relative strengths in one domain or another are predictive of later above-expectations performance in key Algebra domains. The second approach included the same prealgebra predictors and algebra outcomes as the first approach but controls for general influences on school performance by statistically adjusting for overall mathematics competence and overall performance in English Language Arts (ELA). These state-level analyses were augmented with assessments of the relation between middle school students' engagement with and mastery of prealgebra and algebra CCSS standards, using a computer adaptive tutoring system (MATHia), and later Algebra I End-of-Course (EOC) performance.

Smarter Balanced Test

The Smarter Balanced Computer Adaptive Test was constructed using an underlying blueprint developed by mathematics content experts (for the mathematics section) and guided by the CCSS for Mathematics; see Appendix A. As noted, the goal of the latter was to provide a more focused and coherent mathematics education that is comparable to that found in nations with consistently high-achieving students.

Among other things, the blueprint designates mathematics content as being major, supporting, or additional. The designations reflect emphases in the standards and help to promote alignment by prioritizing instructional time and the foci of educational assessments at each grade level. For instance, fifth-grade students are expected to spend most of their time learning about place value, solving complex whole number and decimal arithmetic problems, as well as extending their conceptual understanding and operation skills with fractions. Supporting and additional instruction would include graphing quantitative relations in the coordinate plane and measurement.

The distribution of items on the Smarter Balanced Assessment Consortium assessment follows these instructional priorities, with about 75% of the elementary-grade items focused on topics that are considered most critical in the progression toward algebra. The latter items are designated as major in the blueprint, such as base-10 and fractions competencies in fifth grade. The major items largely assess fundamental conceptual and procedural knowledge for the content area, but also include items that assess complex problem solving, communicating reasoning, and the ability to use modeling to solve real-world problems. The remaining items assess some of the additional and supporting clusters, but not all students receive a significant number of items in these areas due to the adaptive nature of the assessment.

The focus of the items selected for the study is on major elementary-grade topics, such as fractions concepts, that are thought to be foundational for later algebra learning in later grades. Competencies in other areas, such as measurement and geometry, were also examined, allowing for an evaluation of the discriminant validity of the hypothesized major prealgebra competencies. Discriminant validity would be demonstrated, for instance, if fifth-grade fractions competencies were a stronger predictor of later algebra outcomes than fifth-grade geometry competencies.

Current Project

In sum, the project is designed to identify opportunity gaps that limit students' access to knowledge and skills so that students who are Black, Latino, English learner-designated, experiencing poverty, and/or female can be supported to achieve better results in Algebra, a keystone subject for future educational and career opportunities. Table 1 shows the core research questions of the EMERALDS study that will be addressed in this report.

Table 1: Core Research Questions

Question A: Can we identify the factors in K–8 achievement data most predictive of success in Algebra? Can we find factors more specific than "mathematics" and more useful than broad topics like "number" and "geometry"?
Question B: Are we spending too much time on some less important standards and not enough on some more important standards? Where should more time and effort be invested in mathematics instruction, and where less?
Question C: How do clusters of students classified according to their profiles across assessment items fare over time? (Note: this classification is not possible with the Smarter Balanced Assessment but may be possible with other approaches.) Do the achievement gaps widen for some clusters (controlling for background factors) but not others? Students with different profiles may benefit differently from different interventions. Some topics (see Question A) may be more difficult for some profiles, while other topics are more difficult for others. Are there some schools outperforming the expectations based on students' demographic profiles for some clusters?
Question D: Can the factors (Question A), emphasis (Question B), or student profiles (Question C), or trajectories in achievement differences among sub-populations be associated with the proportion of the variance in mathematics achievement among districts compared to among schools within districts compared to among classrooms within schools compared to among students within classrooms?

Methods

Data Sets

State Data

The data used to assess Questions A, C, and D were from the above-described Smarter Balanced Assessment Consortium (SBAC) assessment. The analyses capitalized on large longitudinal data sets from Idaho, Washington, and California; a description of the process of obtaining these data is provided in Appendix B.

The SBAC assessments include grades 3 to 8 as well as a high school assessment administered in tenth or eleventh grade. The grades 6 and 7 assessments include major early algebra items (e.g., expressions) and the grade 8 assessment includes major items (e.g., solving linear equations) that are traditionally taught in an Algebra I course in high school. The relatively recent adoption of the SBAC assessments resulted in the availability of only four years of longitudinal data, which limited the range of included grades. As such, some students might have assessment results in third through sixth grades while others have results in fifth through eighth grades. Regardless, they all received at least some items in the major prealgebra and algebra content domains (below). The computer adaptive design of the test presented an additional challenge for the analysis, however, because by design not all students received identical items.

A core requirement for the analyses was that student performance in earlier grades was directly linked to their algebra performance in later grades. In other words, only students with four years of longitudinal data were eligible for the study. The initial goal was to examine performance in key prealgebra areas (e.g., fractions concepts) in fifth grade and key algebra outcomes (e.g., expression evaluation) in eighth grade. However, the number of students with enough items to create the desired predictor and outcome variables, with defensible reliability, was insufficient. Thus, we expanded the grade ranges for both Idaho, Washington, and California and combined results across two or more grades.

As shown in Table 2, we included three cohorts of students from Idaho and Washington and created predictors based on performance in third to fifth grades, inclusive, and outcomes based on performance in sixth to eighth grades, inclusive. There were more students in California; hence, we were able to create predictors and outcomes using only two cohorts, resulting in combined fourth- and fifth-grade predictors and seventh- and eighth-grade outcomes (see also Weeks & Baron, 2021).

These constraints resulted in usable data for 42,474 students from Idaho, from 2016 to 2019, and 216,595 students from Washington. The first of the two cohorts from California, hereafter referred to as the 5/8 cohort, included 420,089 students with contiguous scores in grades 5, 6, 7, and 8, from 2016 to 2019, respectively. The second cohort, hereafter referred to as the 4/7 cohort, included 429,968 students with contiguous scores in grades 4, 5, 6, and 7, from 2016 to 2019, respectively. The cohorts did not include students who had missing scores, skipped a grade level, or were retained in the same grade during the interval. Table 3, Table 4, Table 5, and Table 6 summarize the demographic makeup of the students from Idaho, Washington, and the 5/8 and 4/7 cohorts from California, respectively. Relative to Idaho and Washington, the students from California were more demographically diverse, more likely to be English learner-designated, and more likely to experience economic disadvantage. State data sets from both states included student responses from all students who participated in the state testing program in each year.

Table 2: Longitudinal Samples from Idaho, Washington, and California

Grade						
	Prealgebra Predictors			Algebra Outcomes		
	3	4	5	6	7	8
Idaho and Washington						
Cohort 1	X	X	X	X		
Cohort 2		X	X	X	X	
Cohort 3			X	X	X	X
California						
Cohort 1		X	X		X	
Cohort 2			X		X	X

Table 3: Student Demographic Characteristics for Idaho

Demographic Group	Male	Female	Total	% of Total
White	15761	16551	32312	76
Hispanic	1302	1350	2652	6
Black	221	243	464	1
Asian	279	251	530	1
Pacific Islander	66	63	129	0.3
American Indian	271	241	512	1
Mixed race	2824	3052	5876	14
Economic disadvantage	3383	3555	6938	16
English second language	1436	1681	3117	7
Individual Education Plan	1344	2301	3645	9
Migrant status	236	241	447	1

Table 4: Student Demographic Characteristics for Washington

Demographic Group	Male	Female	Total	% of Total
White	60150	56547	116697	54%
Hispanic	25871	25130	51001	24%
Black	4484	4254	8738	4%
Asian	8291	8304	16595	8%
Pacific Islander	1208	1085	2293	1%
American Indian	1506	1385	2891	1%
Mixed race	8552	8352	16904	8%
Not specified	880	596	1476	1%
Economic disadvantaged	60740	57968	118708	55%
English second language	17961	15441	33402	15%
Individual Education Plan	21202	11598	32800	15%
Migrant status	2950	2891	5841	3%

Table 5: Student Demographic Characteristics for California – 5/8 Cohort

Demographic Student Group	Male	Female	Total	% of Total
American Indian or Alaska Native	1049	1032	2081	< .01
Asian	19632	19000	38632	9
Native Hawaiian or Other Pacific Islander	923	973	1896	< .01
Filipino	4909	4766	9675	2
Hispanic or Latino	119994	117115	237109	56
Black or African American	10637	10372	21009	5
White	47770	44910	92680	22
Two or more races	7400	7502	14902	4
Ethnicity not reported	902	898	1800	< .01
English only	116277	111908	228185	54
Initial fluent English proficient	7806	9203	17009	4
English learner	27032	19018	46050	11
Reclassified fluent English proficient	62090	66420	128510	31
English proficiency to be determined	8	9	17	< .01
English proficiency unknown	3	10	13	< .01
No special education services	182222	190592	372814	89
Special education services	30994	15976	46970	11
Not economically disadvantaged	82364	79886	162250	39
Economically disadvantaged	130852	126682	257534	61
Migrant	1582	1564	3146	< .01
Not migrant	211634	205004	416638	99
No available demographic information			305	< .01

Table 6: Student Demographic Characteristics for California – 4/7 Cohort

Demographic Student Group	Male	Female	Total	% of Total
American Indian or Alaska Native	1055	1059	2114	< .01
Asian	19465	18678	38143	9
Native Hawaiian or Other Pacific Islander	953	921	1874	< .01
Filipino	4814	4512	9326	2
Hispanic or Latino	124658	120940	245598	58
Black or African American	10795	10635	21430	5
White	48014	45180	93194	22
Two or more races	8108	7765	15873	4
Ethnicity not reported	1102	1077	2179	< .01
English only	121774	115864	237638	57
Initial fluent English proficient	7796	9296	17092	4
English learner	32325	24365	56690	13
Reclassified fluent English proficient	57043	61222	118265	28

English proficiency to be determined	13	12	25	< .01
English proficiency unknown	13	8	21	< .01
No special education services	186996	193530	380526	88
Special education services	31968	17237	49205	12
Not economically disadvantaged	83209	79713	162922	38
Economically disadvantaged	135755	131054	266809	62
Migrant	1558	1523	3081	< .01
Not migrant	217406	209244	426650	99
No available demographic information			237	< .01

MATHia Data

The information used to address Question B included data from the curriculum provider and the district. Carnegie Learning provided student response data from MATHia software used in a large Florida school district. The district provided demographic data and state test scores for students in the Carnegie Learning program in 2017–2020. Carnegie Learning's MATHia, (formerly known as Cognitive Tutor; Ritter et al., 2007) is part of Carnegie Learning's blended curriculum for middle school and high school mathematics. The blended learning context is designed to split the student's time between collaborative work using Carnegie Learning's print-based work-texts (60% of math classroom time) and self-paced study using intelligent tutoring software (i.e., MATHia), using a mastery learning approach (40% of math classroom time).

Extensive research, including a large-scale randomized-controlled trial (Pane et al., 2014), confirms the effectiveness of this type of blended learning approach in diverse, real-world learning contexts. MATHia's content is divided into major topic modules, comprised of multiple units, which in turn are divided into workspaces for student problem-solving. Following Anderson's ACT-R cognitive architecture (Anderson et al., 1998; Anderson et al., 1995), each workspace is comprised of complex, multi-step math problems, as shown in Figure 1, mapped to a set of fine-grained skills (or knowledge components).

As students work through steps in each problem, progress toward problem completion is tracked by model tracing. Progress toward skill mastery is tracked using Bayesian Knowledge Tracing (BKT; Corbett & Anderson, 1995). Student mastery is determined when MATHia's probability estimate reaches 95% that a student has mastered each skill within a workspace (mastered workspaces).

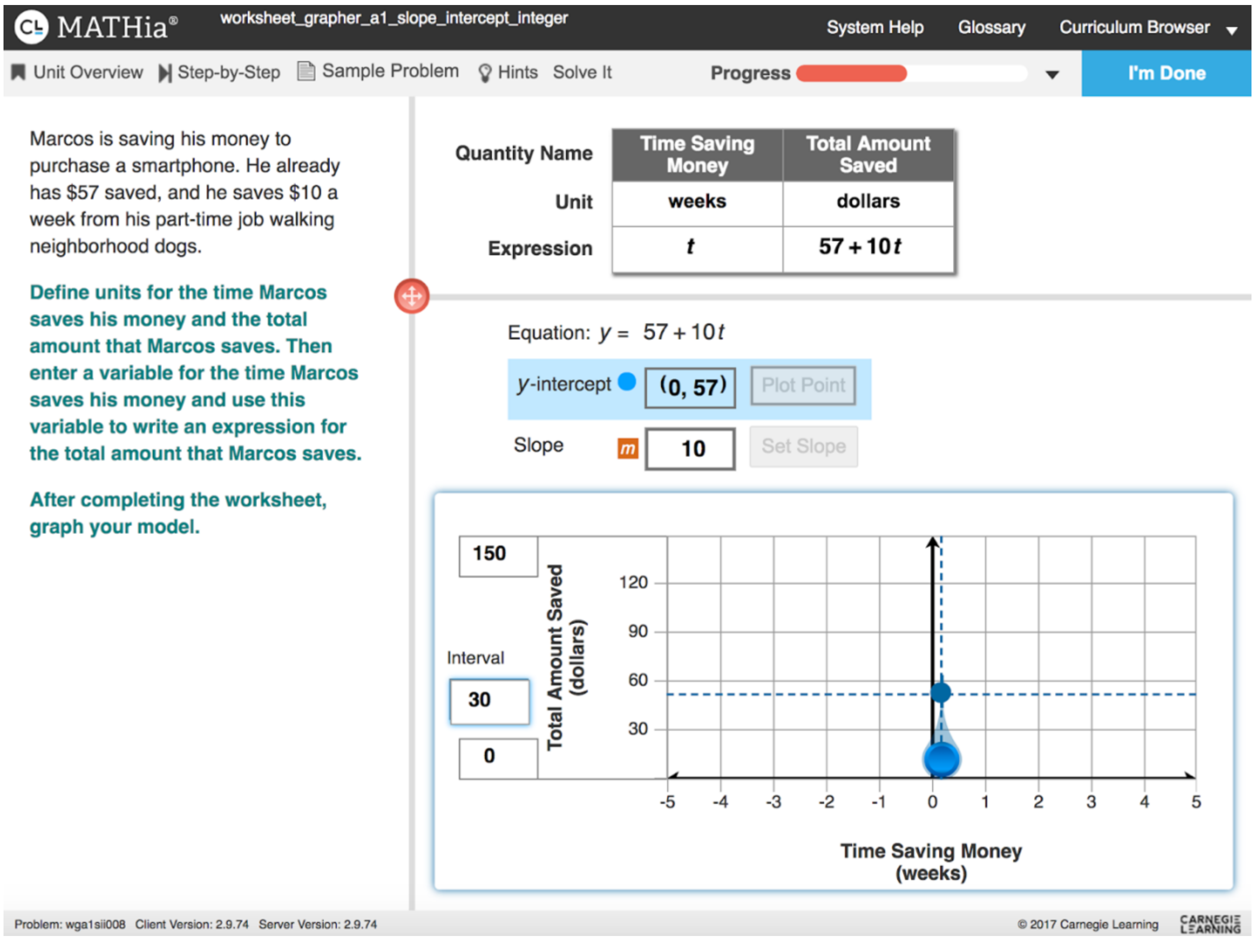
Estimates of student skill mastery are used to adaptively drive problem selection within a workspace; roughly, problems are selected that emphasize skills that a student has yet to master in each workspace. Once the student is determined to have mastered the workspace, they are moved to the next workspace within a unit (or to the first workspace in the next unit or module, depending on where they are in their assigned curricula).

Students may complete a workspace without mastery. If a student reaches a preset number of problems (typically 25) without having mastered all the skills in a workspace, the student is moved to the next workspace without mastery (non-mastered workspaces).

Key to MATHia's data-driven adaptive problem selection and mastery progression is the probabilistic BKT framework. BKT models student knowledge of each skill in a workspace as a latent variable using a two-state Hidden Markov Model. For each skill, the student is either in "un-mastered" or "mastered" state, and Bayesian Knowledge Tracing is used to infer which of the two states a student is in based on their sequence of performance on problem-solving steps requiring that skill.

Each skill in BKT is modeled by four parameters: the initial probability of knowing a skill *a priori* (i.e., prior skill knowledge), p(Init); the probability of a student transitioning from the unknown to the known state for a skill at a particular opportunity to practice the skill, p(Learn); the probability of a student answering incorrectly when the skill is in fact known (i.e., slipping when applying a mastered skill), p(Slip); and the probability of a student correctly applying an un-mastered skill, p(Guess).

Figure 1: Example of Multi-Step Problem-Solving in MATHia.



Workspaces are further organized into larger units of related topics. Each workspace is tagged with one or more Common Core State Standards (CCSS) that are covered by the associated activities. Multiple workspaces might be tagged with the same or overlapping CCSS. MATHia is used as part of regular mathematics and Algebra classes but can also be used as part of advanced and remedial activities and classes.

In addition to the MATHia data, student information was also available for the Florida Standards Assessment (FSA) that includes ELA in grades 3-10 and mathematics in grades 3-8. Florida also requires End-of-Course (EOC) assessments for Algebra 1 and Geometry. Table 7 includes the number of students in the sample for each grade and academic year who completed each type of standardized test (Mathematics FSA or Algebra 1 EOC).

Table 7: Number of Students in the Sample Who Completed Each Type of Test

Academic Year	Grade	Math FSA	Algebra 1 EOC
2016–17	6	100	0
	7	14683	16
	8	11200	1789
	Total	25983	1805
2017–18	6	16873	21
	7	14374	2200
	8	8802	4244
	Total	40049	6465
2018–19	7	14453	31
	8	13791	3011
	9	8697	5763
	Total	36941	8805

Overall, there was usage data from 36,010 students in grades 6–8 for the academic years 2017–18, 2018–19, and 2019–20, corresponding to over 1,000,000 total hours spent using the tutoring software. In addition, the dataset included Florida state standardized testing results for the same students for 2016–17, 2017–18, 2018–19; Florida state standardized testing was cancelled for 2019–20 because of the COVID-19 pandemic.

Table 8 shows the number of students and workspaces completed for each academic year and grade. A subset of students had MATHia usage data for more than one academic year (two years: 5050; three years: 1482).

Table 8: Number of Students and Workspaces Completed for Each Academic Year and Grade

Academic Year	Grade	Number of Students	Number of Workspaces	% Students in Advanced Classes	% Students Enrolled in Pre-Algebra/ Algebra
2017–18	6	5579	267	51.1%	0.0%
	7	4838	383	0.1%	0.4%
	8	3537	388	10.9%	0.0%
	Total	13954	1038		
2018–19	6	6147	356	51.8%	0.1%
	7	6401	393	28.5%	9.7%
	8	4649	443	1.1%	90.4%
	Total	17197	1192		
2019–20	7	4872	447	31.3%	13.2%
	8	4096	440	0.0%	98.9%
	9	87	200	1.1%	96.6%
	Total	9055	1087		

Table 9 shows the level of engagement with MATHia; all sections and distinct sections are separated because some students see the same workspace more than once either for review or as a requirement. When students master a workspace (as per BKT), they have "graduated." When the teacher advances students from a workspace they have not mastered, they have not mastered that workspace. When the student has not finished a workspace and has not started a new one (e.g., at the end a school year), that workspace is "incomplete." Table 10 shows the average level of engagement per covered CCSS.

Table 9: Average Number of Workspaces Mastered, Non-Mastered, or Incomplete Per Student/Year

All Sections	Distinct Sections	Mastered	Mastered Distinct	Non-Mastered	Non- Mastered Distinct	Incomplete	Incomplete Distinct
68.3801	64.1462	58.6262	57.1683	6.9185	6.8281	4.1198	4.0902

Table 10: Average Number of Workspaces Mastered, Non-Mastered, or Incomplete Per Student/Year/CCSS

All Sections	Mastered	Non- Mastered	Incomplete
2.2778	2.0645	1.5396	1.1691

To examine how opportunities to practice different CCSS are related to later Algebra scores, we limited our sample to students who completed the Algebra I EOC test and used MATHia for at least one year. The sample for these analyses is shown in Table 11, and their demographic characteristics are in Tables 12 and 13 (2017–18 cohort) and Tables 14 and 15 (2018–19 cohort).

Table 11: Descriptive Information for Students Included in the Analyses to Predict Algebra I EOC Performance

Academic Year	Grade	Number of Students	Number of Workspaces	% Students in Advanced Classes	% Students Enrolled in Pre-Algebra/ Algebra
2017–18	7	1775	381	14.48%	46.42%
	8	820	307	0.00%	74.02%
	Total	2595	688		
2018–19	7	428	222	0.00%	0.00%
	8	876	360	0.00%	95.32%
	Total	1304	582		

The 2017–18 cohort of students included in the study is approximately 5% of the total district enrollment of students in grades 7 and 8. The demographic characteristics of the study group are roughly comparable to the student groups reported at the district level.

Table 12: 2017–2018 Comparison of Study Group Cohort to the Total District Enrollment

Demographic Study Group vs. Total District[1]	Male	Female	All District	Study Group	% of Total District
All students Grade 7			26,218	1,775	6
All students Grade 8			26,522	820	3
Total Students	51%	49%	52,740	2,595	5
Ethnic Codes	**Grade 7**	**Grade 8**	**Total**	**Study Group**	**% of Total District**
Asian					
Black or African American	4896	4959	9,855	592	6
Hispanic or Latino	18,923	19,126	38,049	1,811	5
Native Hawaiian or Pacific Islander					
Two or more races					
White	1791	1944	3,735	123	3
ESE Codes	**Grade 7**	**Grade 8**	**Study Group**	**% of Total Study Group**	**% of Total [2]District**
Gifted	561	213	774	30	24
Lunch Codes	**Grade 7**	**Grade 8**	**Study Group**	**% of Total Study Group**	**% of Total [3]District**
The student is eligible for free lunch.	1408	653	2,061	79	69%
The student is eligible for reduced-price lunch.	86	32	118	4	

[1] Public Schools Statistical Highlights 2017–18
[2] Gifted all grades 39,781 + 43,529 / 354,172 = 24%
[3] Combined Free/Reduced lunch for Middle Schools

Table 13: Demographic Characteristics of the 2017–18 Cohort Included in the Analyses to Predict Algebra I EOC Performance

Demographic Student Group	Male	Female	Total	% of Total
All students Grade 7	857	918	1,775	68
All students Grade 8	383	437	820	32
Total Students	1240	1355	2,595	100
Ethnic Codes	**Grade 7**	**Grade 8**	**Total**	**% of Total**
Asian	29	8	37	1
Black or African American	404	188	592	23
Hispanic or Latino	1228	583	1,811	70
Native Hawaiian or Pacific Islander	2	0	2	.01
Two or more races	24	6	30	1
White	88	35	123	5
Total	1,775	820	2.595	100
Limited English Proficient Code	**Grade 7**	**Grade 8**	**Total**	**% of Total**
Distinct Student Count (DSC)	958	386	1,344	52
LY – Currently enrolled in class for English learner status students	52	8	60	23
LF – Exited English learner status and being followed	765	426	1,191	46
ESE Codes	**Grade 7**	**Grade 8**	**Total**	**% of Total**
Distinct Student Count (DSC)	1,165	586	1,751	67
Speech Impaired	9	1	10	0.30
Language Impaired	2	1	3	0.10
Visually Impaired	1	0	1	0.03
Emotional/Behavioral Disability	1	3	4	0.20
Specific Learning Disability	17	10	27	1
Gifted	561	213	774	30
Autism Spectrum Disorder	12	3	15	0.50
Other Health Impaired	7	3	10	0.30
Lunch Codes for free and reduced-price	**Grade 7**	**Grade 8**	**Total**	**% of Total**
Did not apply	247	122	369	14
Applied but not eligible	34	13	47	2
Eligible for free lunch	1408	653	2,061	79
Eligible for reduced-price lunch	86	32	118	4
Total	1,775	820	2,595	100

The 2018–19 cohort of students in grades 7 and 8 comprised approximately 2.4% of the total enrollment for those grades.

Table 14: 2018–2019 Comparison of Study Group Cohort to the Total District Enrollment

Demographic Study Group vs. Total District[4]	Male	Female	All District	Study Group	% of Total District
All students Grade 7			27,535	428	1.6
All students Grade 8			26,218	876	3
Total Students	51%	49%	53,736	1,304	2.4
Ethnic Codes	**Grade 7**	**Grade 8**	**Total**	**Study Group**	**% of Total District**
A – Asian					
B – Black or African American	5,288	4,821	9,569	592	6
H – Hispanic or Latino	19,833	19,045	38,878	1,811	4.6
I – Native Hawaiian or Pacific Islander					
M – Two or more races					
W – White	1,861	1,780	3,641	123	3
ESE Codes	**Grade 7**	**Grade 8**	**Study Group**	**% of Total Study Group**	**% of Total [5]District**
L – Gifted	561	213	774	30	24%
Lunch Codes	**Grade 7**	**Grade 8**	**Study Group**	**% of Total Study Group**	**% of Total [6]District**
2 – The student is eligible for free lunch.	1408	653	2,061	79	74.4%
3 – The student is eligible for reduced-price lunch.	86	32	118	4	

[4] Public Schools Statistical Highlights 2018–19
[5] Gifted all grades 40,416 + 43,990 / 350,040 = 24%
[6] Combined Free/Reduced lunch for Middle Schools

Table 15: Demographic Characteristics of the 2018–19 Cohort Included in the Analyses to Predict Algebra I EOC Performance

Demographic Student Group	Male	Female	Total	% of Total
All students Grade 7	196	232	428	33
All students Grade 8	459	417	876	67
Total Students	655	649	1,304	100
Ethnic Codes	**Grade 7**	**Grade 8**	**Total**	**% of Total**
A – Asian	19	14	33	3
B – Black or African American	72	201	273	21
H – Hispanic or Latino	293	613	906	69
I – Native Hawaiian or Pacific Islander	2	1	3	0.02
M – Two or more races	6	11	17	1
N – American Indian or Alaska Native	0	2	2	0.01
W – White	36	34	70	5
Total	428	876	1,304	100
Limited English Proficient Code – (English learner status)	**Grade 7**	**Grade 8**	**Total**	**% of Total**
Distinct Student Count – (DSC)	239	470	709	54
LF – Exited English learner status and being followed	189	377	566	43
LY – Currently enrolled in class for English learner status students	0	29	29	2
ESE Codes	**Grade 7**	**Grade 8**	**Total**	**% of Total**
Distinct Student Count (DSC)	139	608	747	57
F – Speech Impaired	0	0	0	0
G – Language Impaired	0	0	0	0
H – Deaf or Hard of Hearing	1	0	1	0.01
I – Visually Impaired	1	2	3	0.02
J – Emotional/Behavioral Disability	0	2	2	0.01
K – Specific Learning Disability	3	21	24	2
L – Gifted	279	227	506	39
P – Autism Spectrum Disorder	4	8	12	0.9
V – Other Health Impaired	1	8	9	0.6
Lunch Codes	**Grade 7**	**Grade 8**	**Total**	**% of Total**
0 – Did not apply	150	162	312	24
1 – Applied but not eligible	20	26	46	4
2 – Eligible for free lunch	217	607	824	63
3 – Eligible for reduced-price lunch	41	81	122	9
Total	428	876	1,304	100

Analytical Approach

State Data

The first task was to identify core prealgebra and core algebra items using the available items in the SBAC assessment. The associated procedures and item clusters are described in the first section below. Two approaches were then used to create core predictor and outcome variables. The first involved using Item Response Theory (IRT) residuals, that is, deviations from expected performance based on overall grade-level mathematics achievement (e.g., the residual of a fifth-grade item based on overall fifth-grade mathematics achievement). The second task involved the creation of new IRT subscore variables using the clusters of items that define major prealgebra and algebra content areas. The IRT residual approach was first used for the Idaho sample and then replicated with the California and Washington samples, whereas the IRT subscore approach was first used for the California sample and then replicated with the Idaho and Washington samples.

Creating Predictor and Outcome Measures for SBAC

A critical step in the analyses was the construction of core predictor and outcome variables that represent major procedural, conceptual, and problem-solving competencies in core prealgebra and algebra domains. To construct these variables, mathematics content experts—who were very familiar with the SBAC assessment—identified relevant clusters of items that varied by mathematical content and complexity. The first group of items **(A1)** was based on **SBAC Claim 1 (Concepts and Procedures)**. These are the number sense predictors shown in Table 16; the Claims and Standards used to identify items are shown in Appendix C. As can be seen, the number sense items typically require students to execute mathematical procedures (e.g., solve multi-digit multiplication problems), as well as explain and apply basic concepts (e.g., understanding place value in the base-10 system).

There are three other SBAC Claims; detailed descriptions for the Claims are available through SBAC. We focused the second group of items **(A2)** on **Claim 2 (Problem Solving)** and the third group **(A3)** on **Claim 4 (Modeling and Data Analysis)**, which largely assessed students' ability to solve basic and applied word problems, and their ability to take real-world problems and construct and use mathematical models to analyze these problems. The fourth group **(A4)** focused on **Claim 3 (Communicating Reasoning**), the construction of arguments to support reasoning about mathematics. Examples of these types of items are shown in Table 16 under problem solving and reasoning & communicating predictors. Within each of these broader categories (e.g., number sense), items were further differentiated based on arithmetical content and problem-solving complexity. The creation of subdomains of items enabled a more fine-grain prediction of later algebra outcomes than would be otherwise possible.

As a contrast, the content experts identified items (generally additional and supporting items) that were hypothesized to not be as strongly related to later algebra performance as the number sense, problem solving, and reasoning items. As can be seen in Table 16, these items largely included geometry and measurement. It was not possible to identify enough items that did not include some arithmetic operations, and as a result, this contrast variable was not as pure a contrast as we would have liked it to be.

IRT Residual Variables

As described earlier, the SBAC assessment uses an adaptive testing algorithm based on the use of IRT. The assessment provides an estimate of students' overall grade-level mathematics competence based on their performance on a fixed number of items; the specific items vary from person to person. The nature of the assessment thus complicates the determination of student strengths and weaknesses for particular subskills. Stated differently, each student has relatively few item responses for a given skill. This makes it difficult to directly examine the psychometric properties of the collection of items for each competency. One potential solution is to use students' deviation from expected performance on each item (based on overall grade-level mathematics scores) to determine their relative strengths in core prealgebra areas and in later core algebra outcomes for the SBAC item standards identified by content experts (Table 16).

Calculating and compiling item residuals requires a multi-step approach (described in more detail below). As a first step, we used the SBAC item parameters and overall estimates of student mathematics competence to compute the expected probability of a correct response (or for items scored in more than two categories, the probability of scoring in a particular category). These probabilities were computed for all items in the dataset. The deviation between the scored response and the expected probability for each item, for each student, is referred to as the item residual. These residuals are then standardized and averaged across the items associated with each competency for the predictor and outcome variables, respectively. These residual indices were subsequently used in a series of linear regressions to examine the relationship between relative strengths in early prealgebra domains and relative strengths in later core algebra outcomes.

Table 16: Core Predictor and Algebra Outcome Variables

Potential Predictive Factors		
Content Area	**Target**	**Examples**
Predictor: Number Sense		
A1a - Whole Numbers	Operations with and conceptual understanding of place value with whole numbers	Released Items
A1b - Fractions	Operations with and conceptual understanding of proper fractions and mixed numbers	Released Item
A1c - Decimals and Place Value	Read, write, and compare decimals to the thousands	Released Item
Predictor: Problem Solving		
A2a - Basic Problem Solving: Whole Numbers	One-step word problems involving whole numbers	Released Item
A2b - Basic Problem Solving: Fractions	One-step word problems involving fractions	Released Item
A3a - Complex Problem Solving: Whole Numbers	Multi-step or higher complexity word problems with whole numbers	Released Item
A3b - Complex Problem Solving: Rational Numbers	Multi-step or higher complexity word problems with fractions or decimals	Released Item
Predictor: Reasoning & Communicating		
A4- Mathematical Reasoning & Communication	Construction of arguments to support mathematical reasoning or to critique the reasoning of others	Released Item
Contrast: Geometry and Measurement		
AG - Understand shapes, volume and measurement	Convert like measurements within a measurement system; Identify properties of shapes; volumes of solids	Released Item
Outcomes: Critical Components of Algebra I		
B1 - Quantitative Literacy	Identify and graph relationships between quantities	Released Item
B2 - Algebra as Generalized Arithmetic	Read, write, and transform expressions and equations using arithmetic operations	Released Item
B3 - Algebra as Functional Thinking	Formulating and interpretating, linear, quadratic, and exponential relations between quantities	Released Item
B4- Algebra in Constraint Equations	Solve constraint equations involving linear, quadratic, and exponential relations between quantities	Released Item
Contrast: Geometry and Statistics		
B5- Understand the properties of geometric figures and the basics of sampling, distributions, and inferences	Construct and describe the features of geometrical figures and the relations between them; use random sampling to draw inferences about a population	Geometry Released Item Statistics Released Item

Expected performance. To determine the expected performance for each item, we used unidimensional IRT models for selected response and constructed response (polytomous) items. For selected-response items (e.g., multiple-choice, true/false), we used the two-parameter logistic (2PL) model (Birnbaum, 1968). The 2PL model is given by:

$$P_i(\theta_j) = exp[Da_i(\theta_j - b_i)] \;/\; (1 + exp[Da_i(\theta_j - b_i)]), \qquad \text{Eq 1}$$

where $P_i(\theta_j)$ is the probability of a correct response to item *i* by a student with an overall mathematics competence θ_j, and a_i is the item discrimination parameter and b_i is the item difficulty parameter. The student (θ_j) parameters were provided as part of the data report for each student and each academic year. The item parameters (a_i and b_i) were provided by Smarter Balanced for each item. D is a constant that puts the θ competence scale into the same metric as a normal ogive model (D = 1.7).

For constructed-response items, that is, items scored in categories ranging from 0–5, we used the generalized partial credit model (GPCM; Muraki, 1992) to estimate performance for each item. GPCM is given by:

$$P_{ih}(\theta_j) = \frac{exp\sum_{v=1}^{h}[Da_i(\theta_j - b_i + d_{iv})]}{\sum_{h=1}^{n_i} exp\sum_{v=1}^{h}[Da_i(\theta_j - b_i + d_{iv})]}, \qquad \text{Eq 2}$$

where, $P_{ih}(\theta_j)$ is the probability of student *j* with overall mathematics competence θ_j obtaining a score of *h* on item *i*, *n* is the number of item categories, b_i is the item location parameter, d_{iv} is the category parameter for item *i* for category *v*, and *D* is a scaling constant given previously.

We used custom-built scripts written in Python to estimate for each item, student, and year, the expected performance using one of the models mentioned above (Python Software Foundation, 2020).[7] For the selected-response items, we used the parameters provided by SBAC to estimate expected performance. For constructed-response items, we used the generalized partial credit model to determine the probability of achieving each category-score and then summed all those probabilities to achieve an overall single performance score.

Residual calculation. To calculate the deviation between the student's actual score and the model predicted expected performance (calculated in the previous step), we subtracted the observed score by the predicted score and divided the resulting score by the number of categories - 1. In this way, all residuals were in the same scale and varied between -1 and 1. Note that positive residuals are observed for correct responses (i.e., a correct response scored as 1 will always be greater than the expected probability of a correct response). Conversely, negative residuals are observed for incorrect responses. For polytomous items, scored responses in higher categories will generally correspond to higher residuals, and vice versa.

Residual standardization. To facilitate comparison across years, grades, and students, we standardized all residuals. We subtracted each item residual by the average residual on that item for the same grade and academic year and divided by the standard deviation of residuals for that item for the same grade and academic year. This standardization process transforms each residual in standard deviation units from the average residual for that item among students of the same grade and who took that item at the same time.

Averaged standardized residuals by group of items. Finally, for each student, we calculated a single standardized residual score for each key algebra predictor and each algebra outcome of interest (see Table 16). This was done by averaging the standardized residuals of all items that belonged to each predictor or outcome variable.

[7] All analytic scripts are available through Github: https://github.com/LearnSphere/EMERALDS-II

For Idaho, not all students completed items for all predictor variables (see Appendix D), and thus we used k-nearest neighbor (k-NN) models to impute missing data. If a student did not complete any item for a given predictor variable (e.g., A1a - Whole Numbers), we used k-NN models to impute this value. That is, taking as the full sample all students who had completed at least one item for the main outcome variable B (Table 16), we imputed data for predictor variables if the student did not complete at least one item for that predictor. Using k-NN imputation fills in missing values with the distance-weighted averaged of the k closest neighbor observations based on Euclidean distance (Altman, 1992). Following this process, we derived scores for all students for all predictor and outcome variables.

To achieve our goal of longitudinally following students, an item had to have been classified as an early prealgebra predictor and had to have been completed when the student was in grades 3 to 5, inclusive (Table 16). Similarly, for an item to be considered as part of an Algebra variable, it had to be among the Algebra outcomes of interest and had to have been completed when the student was in grades 6 to 8, inclusive.

Finally, to evaluate our approach we compared the residuals (before standardization) to the student's overall mathematics scores (θ_j). The residuals should capture variation above and beyond the students' overall mathematics competence; therefore, correlations between residuals and mathematics competence should be approximately 0. The approach, as expected, resulted in a trivial correlation between students' θ_j and their scores on the residualized items, as shown in Figure 2 for Idaho (r = .039). Figures 3 and 4 show the same pattern for the 5/8 cohort for students from California in grades 5 and 8 respectively; the results are the same for the 4/7 cohort.

The residual approach is premised on the notion that student performance for a given cluster can be accurately distinguished from overall performance. To the extent that the construct is essentially unidimensional, average item residuals should be small and insignificant. On the other hand, if there is substantive variability in the clusters that is not explained by the overall score, one should expect to see larger average residuals for those clusters. If the average residuals by cluster are small, identifying relationships between proximal clusters (predictor clusters in the earlier grades) and distal clusters (outcome clusters in the later grades) may not be very informative.

As an alternative to the residual approach, item responses could be modeled separately for each cluster of items (see Appendix C). With the residual approach, the goal was to partial out the variability associated with a subset of items from the overall performance score. Conversely, by only including a particular subset of items when creating an IRT subscale, only the associated variability in those specific items is considered. While this is a subtle difference, the approach may provide more interpretable results with respect to examinations of subscore reliability and the practical significance of any associated effects in the regression analyses.

Figure 2. Correlations Between Residual Scores and Overall Mathematical Competence for Idaho

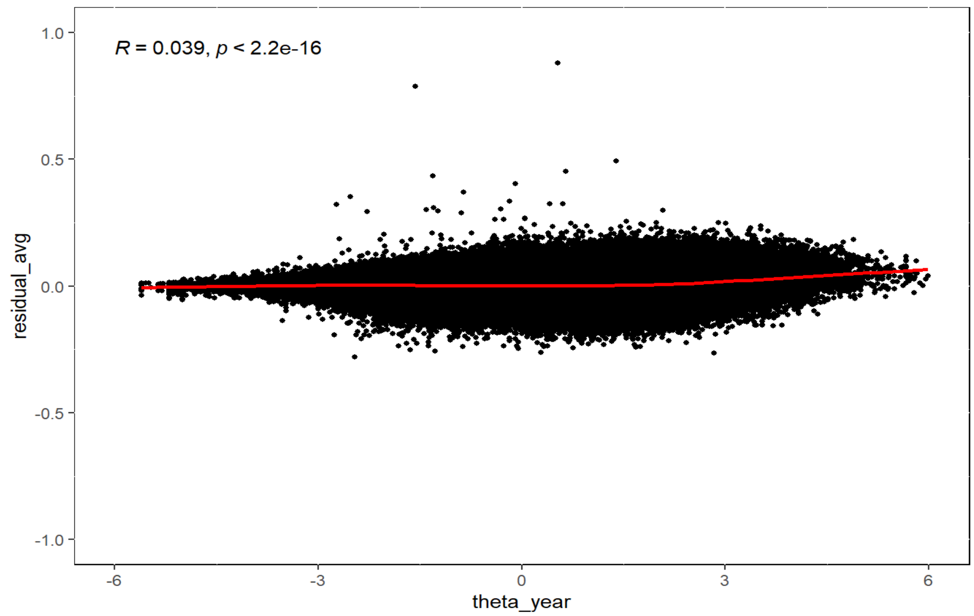


Figure 3. Grade 5 Correlations Between Residual Scores and Overall Mathematical Competence for California

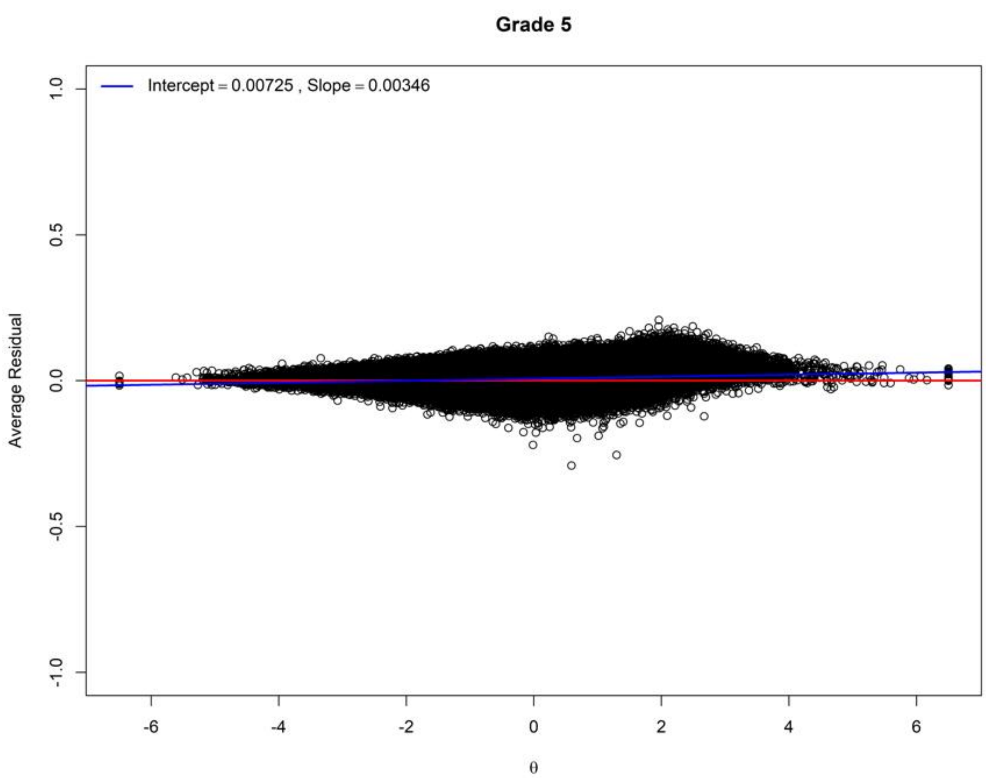
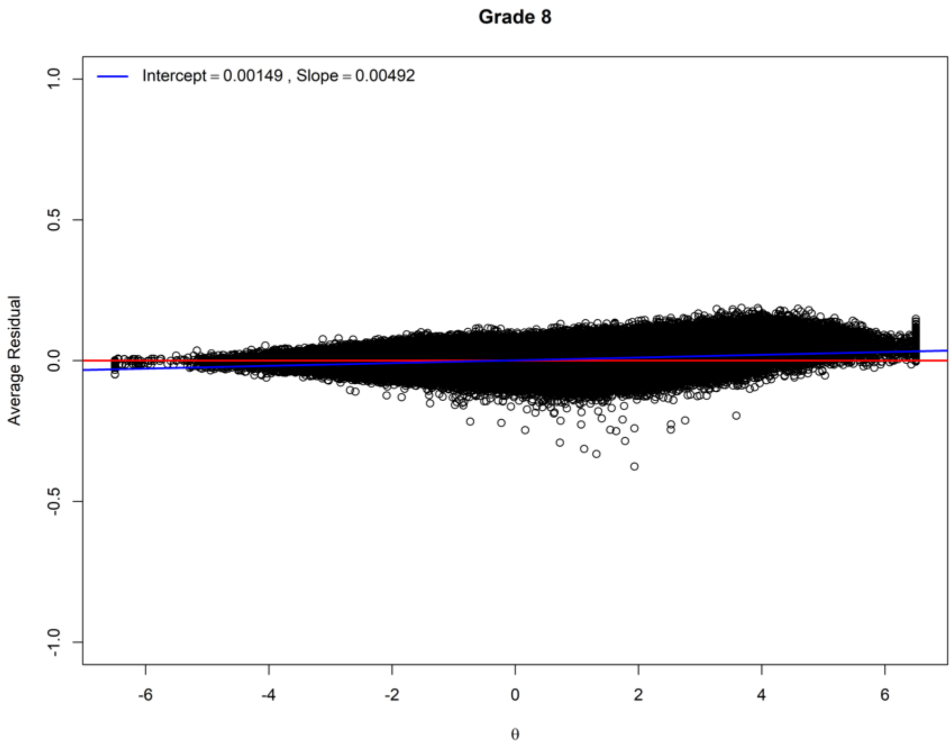


Figure 4. Grade 8 Correlations Between Residual Scores and Overall Mathematical Competence for California



IRT Subscore Variables

The development of subscores for each of the item clusters described in Table 16 involved several stages. New scales were created for each cluster using the same general IRT approach used to establish the overall SBAC mathematics competence (i.e., theta) score.

Item parameters (provided by SBAC) and student abilities were estimated separately for each cluster (i.e., separate unidimensional scales were created); a multidimensional model was not employed. In general, all available data were used to estimate the item and person parameters; however, items with fewer than 100 responses and students with fewer than three responses, for a given cluster, were excluded from the estimation. The output from the IRT subscore estimation was reviewed with particular attention to convergence, out-of-range item parameters (items with very low or negative slopes and items with absolute difficulty/location estimates greater than 6), item fit, and marginal reliability. Several items, across scales, had negative discrimination slopes; these items were excluded. Parameters for the retained items were re-estimated.

Based on the marginal reliabilities, scores could not be created for several of the variables, primarily due to students having taken too few items in these clusters. As such, the design was expanded to include students from the 4/7 cohort. Due to staggered grades across the two cohorts, predictor clusters were established using grades 4 and 5 items across the cohorts; outcome clusters were established using grades 7 and 8 items across the cohorts.

To combine the responses from the 5/8 and 4/7 cohorts for a given cluster, the set of all administered items was identified for fifth-grade students from the 5/8 cluster and fourth- and fifth-grade students from the 4/7 cluster. The resulting dataset included one row per student (a total of 850,057 rows) and columns for each unique item in the cluster. The grades 4 and 5 responses for the students from the 4/7 cohort were combined. That is, they were treated as if they were collected as part of a single test administration. In the rare instances where a student received the same item in both years, the higher scored response was retained. The grade 7 and grade 8 data from the 5/8 cohort were combined with the grade 7 data from the 4/7 cohort using the same approach.

It is important to note that the defensibility of concatenating the item responses for the grades 4 and 5 students and the grades 7 and 8 students is based on the assumptions that the same construct is measured at both grades and that student performance across grades is highly correlated. The scores in grades 4 and 5 from the 4/7 cohort are correlated at $r = .86$; the scores in grades 7 and 8 from the 5/8 cohort are correlated at $r = .87$. While these correlations are high, the shared variance between grades is only around 75%. It is unclear how much of the unexplained variance is due to random error, systematic differences associated with student learning, and/or potential changes in the construct. As such, concatenating the responses may be reasonable to allow for the inclusion of a more complete set of subscores in the regression analyses; however, the approach is not ideal.

The numbers of students, available items, and reliabilities for the subscore variables across the two cohorts are shown in Appendix E. With respect to the marginal reliabilities, the presumption is that each student has responses for the same number of items. While this is generally true, there are some clusters with notably greater variability in the number of responses. As such, the reported marginal reliabilities should be interpreted as rough indicators of stability. Note that all the predictor and outcome clusters were included in the regression analyses, with the exception of Basic Problem Solving with Whole Numbers (A2a, Table 16) or Algebra in Constraint Equations (B4, Table 16). These were excluded because there was not a sufficient number of items to construct reliable variables.

Item parameters and student abilities were scaled separately for each item cluster. That is, each cluster was treated as a separate, unidimensional scale. Dichotomously scored items were fitted using the two-parameter logistic model (2PL); polytomously scored items were fitted using the generalized partial credit model (GPCM). A scaling constant of 1.7 was used to place the estimates on a normal metric. Item parameters were estimated via marginal maximum likelihood using the program MDLTM (von Davier, 2017) based on a single group design. That is, the combined data across cohorts were treated as a single population. For the purpose of identification, the item slopes were constrained to have a mean of unity; the item difficulty/location parameters were constrained to have a mean of zero. Expected *a posteriori* estimates of student ability were compiled and standardized. Convergence of the estimation algorithm was assured before the resulting parameter estimates were used in further analyses.

After each estimation run, the item parameters were reviewed. Items with negative slopes and/or items with difficulties with an absolute value greater than six were excluded. The estimation for the scales in these instances were rerun. Marginal reliabilities were compiled based on the final ability estimates (see Appendix E; Table E1, Table E2).

Regression Analyses

Initial regression models for the IRT residuals were developed based on the data from Idaho and then replicated with the data from California and Washington. The initial models for the IRT subscale scores were first run based on the data from California and then replicated with data from Idaho and Washington.

IRT residual variables. For these analyses, we only included data from students from whom we had both early prealgebra predictors and later algebra outcomes. We compared three models to estimate the unique influence of each prealgebra predictor on algebra outcomes: (1) a model predicting all Algebra outcomes; (2) a model predicting Geometry and Measurement outcomes in grades 6 to 8, inclusive; and (3) a model predicting English proficiency in grades 6 to 8, inclusive. In all models, we used simultaneous linear regression including each of the prealgebra variables as well as the Geometry and Measurement contrast variable, and students' overall English proficiency (as estimated by SBAC IRT models, ELA θj) as predictors.

The basic model provides information about the relative importance of prealgebra variables in the prediction of core Algebra outcomes. The models predicting later Geometry and Statistics and ELA outcomes provide important contrasts by determining if the same or different prealgebra variables predict later non-Algebra outcomes. The contrasts allow us to determine if some prealgebra variables are uniquely related to later Algebra outcomes.

We also ran separate models predicting each specific Algebra subscore, that is, Algebra: Quantitative Literacy, Algebra: Generalized Arithmetic, Algebra: Functional Thinking, and Algebra: Constraint Equations. We compared each one of these with the results of models for Geometry and Statistics and ELA to extract unique predictors of each of these critical components of Algebra.

For Idaho and Washington, we completed a second step of repeating all models including school and district as nested random variables to account for natural variation that is likely to occur across districts and schools inside districts across the state. This type of mixed model is more complicated for the data from California. This is due, in part, to single charter schools that are also designated as a single district.

All regression models were run in R using the base linear modeling function (R Core Team, 2017). Because all residuals were standardized previously to create variables, the regression estimates can be directly compared and interpreted as effect sizes, allowing us to compare the relative impact that 1 standard deviation increase/decrease in each predictor has on the outcome variable.

IRT subscore variables. By combining the data across cohorts, scores were available for a large number of California students for all but two of the predictor and outcome clusters. To maximize the information available in the regression analyses, a multiple imputation approach was used to fill in the missing scores for the predictor variables. The R package, “Amelia,” was used for the imputation (Honaker et al., 2011). This approach uses the EM algorithm to replicate the observed covariance matrix. Five imputed datasets were created. The pairwise correlations between the predictor and outcome variables, including overall Math and ELA scores, before and after the imputation are shown in Appendix E; Table E1 and Table E2, respectively. The median absolute difference between the correlations is zero; the root mean squared difference is 0.008. In short, there are essentially no differences.

Five separate regression models, as shown in Table 17, were fit for each of the three retained Algebra outcome measures (B1, B2, B3; see Table 15), the overall measure of Algebra performance (including all available items from B1, B2, B3, and B4), and the contrast Geometry and Statistics outcome. The model specification for each of the models was the same for each outcome measure (i.e., the only difference for a given model is the outcome variable). Below is an overview of the five models.

Table 17: Basic Regression Models

Model 1	Main effects only for each of the predictor variables. ○ A1a, A1b, A1c, A2b, A3a, A3b, A4, AG
Model 2	Main effects only for each of the predictor variables with ELA score as a covariate. ○ A1a, A1b, A1c, A2b, A3a, A3b, A4, AG, ELA
Model 3	Main effects only for each of the predictor variables with overall Math score as a covariate. ○ A1a, A1b, A1c, A2b, A3a, A3b, A4, AG, Math
Model 4	Main effects only for each of the predictor variables with ELA and Math scores as covariates. ○ A1a, A1b, A1c, A2b, A3a, A3b, A4, AG, ELA, Math
Model 5	Model 5A: Main effects for A4 (the strongest predictor of algebra, based on the previous models), ELA score, and a range of demographic variables; additionally, interaction effects between A4 and the demographic variables (the interaction effects are denoted by the colons, e.g., A4: Female is the effect for Females with a given A4 score) Model 5B: Identical to Model 5A, except earlier overall Mathematics competence (theta score) was substituted for A4

MATHia

All analyses were done at the workspace level instead of the problem level because, with the mastery and adaptive nature of the software, analyses would be confounded with initial knowledge and ability.

After inspection of the data to identify possible missing data or variable data by school and course to spot any potential issues, we conducted basic data cleaning. We removed students for whom we did not have outcome data; we imputed grade-attended data where missing based on previous years' grade attended, and we removed students for whom we did not have both early FSA math and Algebra I EOC scores. In the clean data, we conducted two sets of analyses. The first focused on CCSS topics that were practiced by students. We did this at two levels: first, considering the entire dataset and second, considering only the sample of students for whom we also had Algebra I EOC scores.

Second, mixed-effect regression models were used to predict Algebra I EOC scores using the number of CCSS-linked workspaces completed as predictors. The models included random effects for schools and type of class in which the student was enrolled while using MATHia (Regular Math, Advanced Math, Remedial Math, Algebra I). The grade enrolled and the standardized FSA Math test score were also added as covariates. Inspection of the data revealed that the distribution of the number of workspaces completed was skewed right. To address this potential issue, we transformed the variable by taking the log of the number of workspaces. Analyses with and without logging showed the same pattern of results. For ease of interpretability, we are reporting analyses without the log transformation.

All numeric predictor and outcome variables were standardized in all regression analyses at the academic year, school, and class level by converting the raw values to z-scores. This transformation makes it easier to compare variables as standard deviations of change and interpret coefficients as standardized effect sizes.

Results

The findings for each of the four questions are presented in succession. For readers who do not wish to read through the detailed analyses, the core findings associated with each set of analyses are previewed at the beginning of each of these sections.

QUESTION A

Question A: Can we identify the factors in K–8 achievement data most predictive of success in Algebra? Can we find factors more specific than "mathematics" and more useful than broad topics like "number" and "geometry"?

In preview, for the IRT residuals, several earlier mathematics competencies (e.g., Problem Solving with Whole Numbers) predicted later overall Algebra scores for students in Idaho, controlling for earlier overall mathematics scores. However, these effects were not specific to Algebra (some also predicted later Geometry or ELA scores), and none of them were practically significant. There were few significant results for students in California.

For the IRT factor scores, there are a few specific earlier mathematics competencies (e.g., Mathematical Reasoning & Communication) that predict later overall performance in Algebra for students in California, but these effects are small and not practically significant once earlier overall mathematical performance is controlled. The latter is a substantive predictor of later overall Algebra outcomes and reveals that performance in elementary grade CCSS, which emphasizes the concepts, procedural fluencies, and applications of arithmetic, is an early indicator of students' preparation for learning algebraic material in later grades. This finding was also confirmed for students in Idaho and Washington.

The first of two results sections present findings for the IRT residuals and the second for the IRT subscores.

IRT Residuals

Idaho and Washington

An overall summary of the results for Idaho is presented in Table 18 without estimation of random effects for districts and schools; the individual regression results for overall Algebra, Geometry and Statistics, and ELA outcomes are in Appendix G. The initial model predicted overall Algebra outcomes using each of the critical prealgebra variables, as well as overall ELA ability and the contrast Geometry and Measurement score (also in grades 3–5). The model was significant and accounted for 4% of the variance in later Algebra outcomes ($F(10,42461) = 156.3, p < .0001$).

As it can be seen in Table 18, multiple factors were positive predictors of overall Algebra performance, including Number Sense: Fractions, Basic Problem Solving with Whole Numbers, and Complex Problem Solving with Whole Numbers. The strongest of these predictors is Basic Problem Solving with Whole Numbers ($\beta = 0.098$), followed by Fractions ($\beta = 0.009$) and Complex Problem Solving with Whole Numbers ($\beta = 0.003$). Basic Problem Solving with Fractions was a negative predictor of overall Algebra knowledge ($\beta = -0.004$), controlling for other factors in the model. Both ELA and Geometry and Measurement were positive predictors of later Algebra as well. However, it is important to note that both were weaker predictors than Basic Problem Solving with Whole Numbers ($\beta = 0.007$and $\beta = 0.011$, respectively).

As mentioned before, to fully characterize the unique predictors of Algebra, it is important to compare the basic model (Algebra) with models predicting Geometry and Statistics and ELA. Both of the latter models were significant, but only explained 0.33% ($F(10{,}42461) = 14.62, p < .0001$), and 0.49% ($F(9{,}42461) = 24.28, p < .0001$) of the variance in Geometry and Statistics and ELA, respectively.

The key outcomes in Table 18 for Idaho and Table 19 for Washington are later overall Algebra scores, and the core contrast outcomes are later Geometry and Statistics and ELA scores. Earlier Fractions performance predicted all three outcomes, whereas Basic Problem Solving with Whole Numbers significantly predicted overall Algebra and Geometry and Statistics but was negatively related to later ELA scores. The pattern suggests that early Fractions knowledge and Whole Number Problem Solving were predictive of later math outcomes broadly and not algebra specifically, above and beyond earlier overall mathematics competence. Overall, however, the magnitude of these effects is not large and may not be practically important.

Moreover, earlier Geometry and Measurement was as important as Fractions in predicting later Algebra and thus the specificity of the relation between early Fractions knowledge and later Algebra outcomes is uncertain. Complex Problem Solving with Whole Numbers was the one factor that predicted later Algebra but not Statistics and Geometry outcomes. This predictor did, however, predict later ELA scores and thus likely assesses reading and language comprehension, on top of early mathematical content knowledge.

Looking at each critical aspect of Algebra, we see a pattern that is similar to the results for overall Algebra, but with the following exceptions. Although performance in Number Sense: Fractions was a positive predictor for most critical components of Algebra, it is a negative predictor of Algebra in Constraint Equations (B4). Similarly, Basic Problem Solving with Whole Numbers and Complex Problem Solving with Whole numbers were negative predictors of Quantitative Literacy only, while remaining positive predictors of all the other critical components of Algebra.

Finally, to account for variability that might exist across districts and schools within districts, we re-ran all models with school nested in district as a random effect, which allows for variation across students from different schools in the same district and across districts. The summary results are presented below in Table 20 (Idaho) and Table 21 (Washington). As it can be seen, the pattern is similar to that shown in Tables 18 and 19, respectively.

Table 18: Summary of Standardized Estimates Across all Regression Models for Idaho (without School and District as Random Effects)

Outcomes	*Overall Algebra*	*Quantitative Literacy*	*Generalized Arithmetic*	*Functional Thinking*	*Constraint Equations*	*Geometry and Statistics*	*ELA*
Predictor							
A1a - Whole Numbers	-0.002	0.004	0.006	-0.023*	0.045	-0.003	-0.055*
A1b - Fractions	0.009*	0.005*	0.010*	0.014*	-0.018*	0.013*	0.050*
A1c - Decimals and Place Value	0.003	0.012*	-0.013*	-0.001	0.000	-0.011*	0.015
A2a - Basic Problem Solving: Whole Numbers	0.098*	-0.040*	0.469*	0.056*	0.248*	0.081*	-0.078*
A2b - Basic Problem Solving: Fractions	-0.004*	-0.004*	-0.024*	0.009*	0.025*	-0.006	-0.005
A3a - Complex Problem Solving: Whole Numbers	0.003*	-0.005*	0.018*	0.007	0.026*	-0.002	0.050*
A3b - Complex Problem Solving: Rational Numbers	0.002	-0.002	0.003*	0.006	0.041*	0.002	-0.012
A4 - Mathematical Reasoning & Communication	-0.001	0.008*	-0.021*	-0.008	-0.011	-0.002	0.097*
ELA	0.007*	-0.001	0.023*	0.007*	0.079*	0.011*	---
Geometry and Measurement	0.011*	0.004*	0.032*	0.013*	-0.058*	0.010*	0.044*

*Statistically significant, $p < .05$.

Table 19: Summary of Standardized Estimates Across all Regression Models for Washington (without School and District as Random Effects)

Outcomes	*Overall Algebra*	*Quantitative Literacy*	*Generalized Arithmetic*	*Functional Thinking*	*Constraint Equations*	*Geometry*	*ELA*
Predictor							
A1a - Whole Numbers	-0.001	0.002*	-0.007*	-0.003	0.003	-0.003	-0.018*
A1b - Fractions	0.002*	0.008*	-0.015*	0.003	0.000	0.010*	-0.021*
A1c - Decimals and Place Value	0.006*	0.010*	0.007*	-0.001	-0.002	-0.003	-0.030*
A2a - Basic Problem Solving: Whole Numbers	-0.009*	0.000	-0.020	-0.017*	0.009*	-0.002	0.088*
A2b - Basic Problem Solving: Fractions	0.003*	0.005*	0.001	0.000	-0.003	-0.004*	-0.029*
A3a - Complex Problem Solving: Whole Numbers	-0.001	-0.001	-0.004	0.001	-0.001	0.004	0.047*
A3b - Complex Problem Solving: Rational Numbers	-0.001	-0.001	-0.008*	0.004*	-0.008*	0.004*	-0.020*
A4 - Mathematical Reasoning & Communication	0.005*	0.010*	0.000	0.001	-0.005	0.007*	0.163*
ELA	0.007*	0.000	0.022*	0.011*	-0.013*	0.006*	--
Geometry and Measurement	0.000	0.004*	-0.006*	-0.001	-0.012*	0.012*	0.077*

*Statistically significant, $p < .05$.

Table 20: Summary of Standardized Estimates Across all Regression Models for Idaho (with School and District as Random Effects)

Outcome Variables	*Overall Algebra*	*Quantitative Literacy*	*Generalized Arithmetic*	*Functional Thinking*	*Constraint Equations*	*Geometry and Statistics*	*ELA*
Predictor Variables							
A1a - Whole Numbers	-0.002	0.004	0.003	-0.021*	0.046*	-0.003	-0.058*
A1b - Fractions	0.009*	0.005*	0.011*	0.013*	-0.017*	0.013*	-0.063*
A1c - Decimals and Place Value	0.003*	0.011*	-0.012*	0.000	0.000	-0.011*	0.011
A2a - Basic Problem Solving: Whole Numbers	0.096*	-0.039*	0.453*	0.057*	0.248*	0.082*	-0.087*
A2b - Basic Problem Solving: Fractions	-0.004*	-0.005*	-0.021*	0.009*	0.025*	-0.007	-0.009
A3a - Complex Problem Solving: Whole Numbers	0.003*	-0.005	0.016*	0.007	0.026*	-0.002	0.051*
A3b - Complex Problem Solving: Rational Numbers	0.002	-0.002	0.004	0.005	0.041*	0.001	-0.009
A4 - Mathematical Reasoning & Communication	-0.001	0.007*	-0.017*	-0.008	-0.011	-0.002	0.089*
ELA	0.007*	-0.001	0.026*	0.007	0.078*	0.009*	---
Geometry and Measurement	0.010*	0.003*	0.024*	0.013*	-0.054*	0.012*	0.047*

*Statistically significant, $p < .05$.

Table 21: Summary of Standardized Estimates Across all Regression Models for Washington (with School and District as Random Effects)

Outcome Variables	*Overall Algebra*	*Quantitative Literacy*	*Generalized Arithmetic*	*Functional Thinking*	*Constraint Equations*	*Geometry and Statistics*	*ELA*
Predictor Variables							
A1a - Whole Numbers	0.000	0.003*	-0.004*	-0.003	0.004	-0.004*	-0.021*
A1b - Fractions	0.002*	0.008*	-0.013*	0.003	0.000	0.010*	-0.031*
A1c - Decimals and Place Value	0.005*	0.010*	0.006*	-0.001	-0.002	-0.002	-0.012*
A2a - Basic Problem Solving: Whole Numbers	-0.008*	0.000	-0.019*	-0.017*	0.009*	-0.002	0.085*
A2b - Basic Problem Solving: Fractions	0.003*	0.005*	0.001	0.000	-0.004	-0.004*	-0.027*
A3a - Complex Problem Solving: Whole Numbers	-0.001	-0.001	-0.004	0.002	-0.002	0.004	0.033*
A3b - Complex Problem Solving: Rational Numbers	-0.001	-0.001	-0.006*	0.004*	-0.008*	0.004*	-0.025*
A4 - Mathematical Reasoning & Communication	0.005*	0.009*	0.002	0.000	-0.005	0.007*	0.138*
Geometry and Measurement	0.000	0.004*	-0.006*	-0.001	-0.012*	0.011*	0.057*
ELA	0.008*	0.000	0.025*	0.012*	-0.015*	0.006*	--

California

The same analyses, as noted, were conducted for California as in Idaho and Washington, with the exception of Algebra: Constraint Equations as an outcome and Problem Solving with Whole Numbers as a predictor. This is because reliable estimates of these variables could not be obtained for the IRT subscore analyses (below), and thus they were dropped for all California analyses. As with the Idaho and Washington analyses, later Geometry and Statistics and ELA scores were used as contrasts. As will be seen, the most consistent results emerge for the overall later Algebra measure, likely because it was more reliable than the more specific Algebra outcome measures.

The most basic results, without control of early mathematics competence and ELA scores, are shown in Table 22. To provide an added control, the same predictors along with control of early ELA and overall mathematics competencies are shown in Table 23; the results for models with only control of ELA and early overall mathematics are in Appendix H. Note that cluster-adjusted standard errors are presented in these tables; district was used as the clustering variable (Bell & McCaffrey, 2002).

Unlike the findings for Idaho and Washington, there are few significant results for California; overall R^2s for all models was < 1% (Appendix I). One potential reason is the inclusion of three grade levels for predictors and outcomes in Idaho and Washington and only two grade levels in California. The inclusion of only two grade levels had the benefit of focusing on the end of the elementary school years but may have come with the cost of less variation among students for both predictors and outcomes.

Table 22: Model 1 Standardized Regression Coefficients for IRT Residuals for California

Outcome Variables	*Overall Algebra Beta*	*Overall Algebra SE*	*Quantitative Literacy Beta*	*Quantitative Literacy SE*	*Generalized Arithmetic Beta*	*Generalized Arithmetic SE*	*Functional Thinking Beta*	*Functional Thinking SE*	*Geometry & Statistics Beta*	*Geometry & Statistics SE*
Predictor Variables										
A1a - Whole Numbers	-0.0005	0.0016	0.0012	0.0017	-0.0005	0.0016	-0.0027*	0.0013	-0.0009	0.0019
A1b - Fractions	0.0012	0.0010	0.0004	0.0010	0.0015	0.0011	-0.0001	0.0011	-0.0015	0.0011
A1c - Decimals	-0.0010	0.0011	-0.0003	0.0011	-0.0014	0.0010	-0.0013	0.0011	0.0008	0.0010
A2b – Basic Problem Solving: Fractions	0.0002	0.0012	0.0004	0.0012	-0.0013	0.0014	0.0019	0.0013	-0.0009	0.0014
A3a – Complex Problem Solving: Whole Numbers	-0.0002	0.0015	-0.0007	0.0014	-0.0004	0.0014	0.0014	0.0014	-0.0005	0.0012
A3b – Complex Problem Solving: Fractions	0.0004	0.0013	-0.0006	0.0012	0.0021	0.0012	0.0001	0.0011	-0.0001	0.0010
A4 – Mathematical Reasoning & Communication	0.0005	0.0011	0.0007	0.0011	0.0003	0.0011	-0.0009	0.0011	-0.0004	0.0011
AG - Geometry & Measurement	0.0001	0.0014	-0.0014	0.0014	0.0016	0.0011	0.0010	0.0013	0.0004	0.0011

*Statistically significant, $p < .05$.

Table 23: Model 4 Standardized Regression Coefficients for IRT Residuals for California Controlling Overall Mathematics and ELA Scores

Outcome Variables	*Overall Algebra Beta*	*Overall Algebra SE*	*Quantitative Literacy Beta*	*Quantitative Literacy SE*	*Generalized Arithmetic Beta*	*Generalized Arithmetic SE*	*Functional Thinking Beta*	*Functional Thinking SE*	*Geometry & Statistics Beta*	*Geometry & Statistics SE*
Predictor Variables										
A1a - Whole Numbers	-0.0005	0.0016	0.0012	0.0017	-0.0005	0.0016	-0.0027*	0.0013	-0.0009	0.0019
A1b – Fractions	0.0012	0.0010	0.0004	0.0010	0.0015	0.0011	-0.0001	0.0011	-0.0015	0.0011
A1c – Decimals	-0.0010	0.0011	-0.0003	0.0011	-0.0014	0.0010	-0.0013	0.0011	0.0008	0.0010
A2b – Basic Problem Solving: Fractions	0.0002	0.0012	0.0004	0.0012	-0.0013	0.0014	0.0020	0.0013	-0.0009	0.0014
A3a – Complex Problem Solving: Whole Numbers	-0.0001	0.0015	-0.0009	0.0014	-0.0005	0.0014	0.0025	0.0014	-0.0012	0.0012
A3b – Complex Problem Solving: Fractions	0.0004	0.0013	-0.0006	0.0012	0.0021	0.0012	0.0001	0.0011	-0.0001	0.0010
A4 – Mathematical Reasoning & Communication	0.0005	0.0011	0.0007	0.0011	0.0003	0.0011	-0.0009	0.0011	-0.0004	0.0011
AG – Geometry & Measurement	0.0001	0.0014	-0.0014	0.0014	0.0016	0.0011	0.0010	0.0013	0.0004	0.0011
ELA	-0.0027	0.0019	-0.0011	0.0018	-0.0019	0.0018	-0.0015	0.0019	0.0018	0.0017
Math	0.0023	0.0020	0.0013	0.0020	0.0019	0.0018	-0.0002	0.0018	-0.0005	0.0017

*Statistically significant, $p < .05$.

IRT Subscore Variables

California

The overall percentage of variance in the outcome measures explained by the predictor models are shown in Table 24; reliabilities of predictor and outcome variables are in Appendix E. The standardized coefficients for Model 1 and Model 4 are shown in Table 25 and Table 26, respectively; results from Model 2 and Model 3 are in Appendix J.

As noted, Model 1 includes the simultaneous estimation of all of the reliable prealgebra variables in the prediction of later overall Algebra scores (across all B items, Table 16) and (separately) for each of the specific later outcomes (Table 25); the one exception was Algebra: Constraint Equations, which could not be reliably estimated. These results are only presented for readers who might be interested in seeing the magnitude of the relations between earlier specific math competencies and later math outcomes before the control of earlier overall mathematics competence and ELA scores.

The key results are in Model 4, which includes the same predictors as in Model 1, along with overall ELA scores and overall mathematics scores (Table 26). A summary of the overall findings for the prediction of later ELA scores (as a contrast to the prediction of mathematics scores) across all models is provided in Table 27.

Table 24: Overall Variance Explained (R^2) for IRT Subscores for California

Outcome Variables	*Overall Algebra*	*Quantitative Literacy*	*Generalized Arithmetic*	*Functional Thinking*	*Geometry & Statistics*	*ELA*
Model 1	0.618	0.564	0.164	0.162	0.156	0.522
Model 2	0.641	0.581	0.176	0.177	0.168	0.659
Model 3	0.656	0.597	0.179	0.175	0.164	0.551
Model 4	0.665	0.603	0.184	0.183	0.171	0.662

All models are statistically significant, $p < 05$.

As shown in Table 24, all of the models were predictors of later overall Algebra and Quantitative Literacy; the results for these two outcomes are similar due to the high correlation between them ($r = .96$, Table F1). The prediction of Generalized Arithmetic and Functional Thinking, as well as Geometry and Statistics, is considerably lower but still statistically significant.

Table 25: Model 1 Standardized Regression Coefficients for IRT Subscores for California

Outcome Variables	*Overall Algebra Beta*	*Overall Algebra SE*	*Quantitative Literacy Beta*	*Quantitative Literacy SE*	*Generalized Arithmetic Beta*	*Generalized Arithmetic SE*	*Functional Thinking Beta*	*Functional Thinking SE*	*Geometry & Statistics Beta*	*Geometry & Statistics SE*
Predictor Variables										
A1a - Whole Numbers	0.059*	0.001	0.056*	0.001	0.037*	0.002	0.023*	0.002	0.025*	0.002
A1b - Fractions	0.159*	0.001	0.158*	0.001	0.058*	0.002	0.074*	0.002	0.084*	0.002
A1c - Decimals	0.057*	0.001	0.053*	0.001	0.038*	0.001	0.027*	0.001	0.021*	0.001
A2b - Basic Problem Solving: Fractions	0.071*	0.001	0.069*	0.001	0.033*	0.002	0.028*	0.001	0.042*	0.002
A3a - Complex Problem Solving: Whole Numbers	0.095*	0.001	0.087*	0.001	0.049*	0.002	0.046*	0.002	0.049*	0.002
A3b - Complex Problem Solving: Fractions	0.198*	0.001	0.192*	0.001	0.093*	0.002	0.103*	0.002	0.103*	0.002
A4 - Mathematical Reasoning & Communication	0.386*	0.002	0.367*	0.002	0.208*	0.004	0.207*	0.003	0.188*	0.003
AG - Geometry & Measurement	0.080*	0.001	0.069*	0.001	0.068*	0.002	0.051*	0.001	0.042*	0.001

*Statistically significant, $p < .05$.

Table 26: Model 4 Standardized Regression Coefficients for IRT Subscores for California

Outcome Variables	*Overall Algebra Beta*	*Overall Algebra SE*	*Quantitative Literacy Beta*	*Quantitative Literacy SE*	*Generalized Arithmetic Beta*	*Generalized Arithmetic SE*	*Functional Thinking Beta*	*Functional Thinking SE*	*Geometry & Statistics Beta*	*Geometry & Statistics SE*
Predictor Variables										
A1a - Whole Numbers	0.037*	0.001	0.036*	0.001	0.022*	0.002	0.009*	0.002	0.013*	0.002
A1b - Fractions	0.023*	0.001	0.032*	0.001	-0.027*	0.002	-0.004*	0.002	0.023*	0.001
A1c - Decimals	0.013*	0.001	0.011*	0.001	0.010*	0.001	0.002*	0.001	0.002	0.001
A2b - Basic Problem Solving: Fractions	0.017*	0.001	0.019*	0.001	-0.001	0.002	-0.002	0.001	0.019*	0.002
A3a - Complex Problem Solving: Whole Numbers	0.035*	0.001	0.034*	0.001	0.009*	0.002	0.006*	0.002	0.015*	0.002
A3b - Complex Problem Solving: Fractions	0.053*	0.001	0.060*	0.002	0.000	0.002	0.015*	0.002	0.033*	0.002
A4 - Mathematical Reasoning & Communication	0.130*	0.002	0.135*	0.002	0.043*	0.003	0.048*	0.003	0.059*	0.002
AG - Geometry & Measurement	0.010*	0.001	0.006*	0.001	0.023*	0.001	0.009*	0.001	0.008*	0.001
ELA	0.165*	0.002	0.130*	0.002	0.129*	0.003	0.160*	0.002	0.151*	0.002
Math	0.454*	0.005	0.430*	0.005	0.269*	0.004	0.225*	0.003	0.163*	0.004

*Statistically significant, $p < .05$.

Given the high correlation between overall Algebra and Quantitative Literacy and the lower overall variance explained in the prediction of Generalized Arithmetic and Functional Thinking, we focused on a contrast of overall Algebra and Geometry and Statistics and later ELA scores. The estimates for the prealgebra prediction of overall Algebra are about double those for the prediction of Geometry and Statistics, suggesting earlier competencies are better predictors of later Algebra than Geometry and Statistics. However, early Geometry and Measurement is also a better predictor of overall Algebra than later Geometry and Statistics, and thus it cannot be stated with certainty that the stronger relation to later Algebra is related to the identified content of the prealgebra and overall Algebra measures; the relations could be stronger because overall Algebra is a more reliable measure than Geometry and Statistics (see Appendix E), or because the latter items require arithmetic.

Most of the prealgebra variables were more strongly related to later Algebra than to later ELA, indicating the prealgebra measures were better indicators of domain-specific mathematical knowledge than reading and language comprehension. There were, however, two exceptions, Complex Problem Solving with Whole Numbers and Geometry and Measurement, that were more strongly related to later ELA than to later Algebra. The implication is that these measures have a relatively strong reading and language comprehension component to them. Complex Problem Solving with Fractions and Reasoning & Communicating were also good predictors of later ELA, suggesting reading and language comprehension contribute to performance on these measures.

As shown in Table 26, controlling for earlier mathematics and English Language Arts scores lowers the magnitude of the relations between early prealgebra predictors and later Algebra outcomes. Nevertheless, nearly all the prealgebra variables remain significant predictors of later overall Algebra but also of later Geometry and Statistics, but in both cases the magnitude of these effects was small. The most important early predictors of later Algebra are those that result in larger estimates in the prediction of Algebra than Geometry and Statistics and larger estimates than were found for the relation between Geometry and Measurement and overall Algebra (i.e., 0.010 in column 2 of Table 26).

Using these criteria, better earlier performance in Whole Numbers, Complex Problem Solving with Whole Numbers, and Reasoning & Communication better predicted Algebra than Geometry and Statistics competencies in later grades. Importantly, these relations emerge controlling for earlier overall mathematics and ELA performance and are stronger than the relation between earlier Geometry and Measurement and later Algebra. There is also some evidence for the importance of Fractions and Complex Problem Solving with Fractions, as the strength of the relation between these and later Algebra is at least double the strength of the relation between Geometry and Measurement and later Algebra, but both of these variables are also predictors of later Geometry and Statistics. Again, the magnitude of these effects was small and much lower than the effect of earlier overall mathematics competence (beta = 0.454 for prediction of overall Algebra).

Table 27: Standardized Regression Coefficients for IRT Subscores Predicting ELA for California

Regression Models	***Model 1 Beta***	***Model 1 SE***	***Model 2 Beta***	***Model 2 SE***	***Model 3 Beta***	***Model 3 SE***	***Model 4 Beta***	***Model 4 SE***
Predictor Variables								
A1a – Whole Numbers	0.041*	0.002	0.018*	0.002	0.024*	0.002	0.013*	0.002
A1b – Fractions	0.108*	0.001	0.031*	0.001	-0.012*	0.002	-0.004*	0.001
A1c – Decimals	0.029*	0.003	0.010*	0.002	-0.012*	0.003	-0.002*	0.002
A2b – Basic Problem Solving: Fractions	0.043*	0.003	0.016*	0.003	-0.005*	0.003	0.001*	0.003
A3a – Complex Problem Solving: Whole Numbers	0.120*	0.003	0.040*	0.003	0.080*	0.003	0.031*	0.003
A3b – Complex Problem Solving: Fractions	0.185*	0.003	0.061*	0.002	0.070*	0.003	0.029*	0.002
A4 – Mathematical Reasoning & Communication	0.371*	0.003	0.124*	0.002	0.175*	0.003	0.072*	0.003
AG – Geometry & Measurement	0.086*	0.003	0.027*	0.003	0.030*	0.003	0.012*	0.003
ELA			0.608*	0.001			0.579*	0.001
Math					0.471*	0.002	0.154*	0.002

*Statistically significant, $p < .05$.

Controlling for earlier ELA, the relation between the prealgebra measures and later ELA is substantively reduced. Whole Numbers stands out as the only predictor with estimates at least twice as large in the prediction of Algebra than Geometry and Statistics or ELA, and more than three times larger than early Geometry and Measurement in the prediction of Algebra. Fractions and Complex Problem Solving with Fractions predict later Algebra and Geometry and Statistics but are effectively unrelated to later ELA, suggesting Fractions knowledge is important across mathematical domains.

Finally, both Complex Problem Solving and the Reasoning & Communicating measures predict later ELA, even with control of earlier ELA, in keeping with a reading and language comprehension component to these measures. Even so, Complex Problem Solving with Fractions and Reasoning & Communication are more strongly related to later Algebra than Geometry and Statistics or ELA. Nevertheless, the relation between overall early mathematics competence and later Algebra is much stronger than the relation between any specific prealgebra predictor and later Algebra.

Idaho and Washington

The overall percentage of variance in the outcome measures explained by the predictor models are shown in Table 28 and Table 29 for Idaho and Washington, respectively. The standardized coefficients for Model 1 are shown in Table 30 and Table 31 for Idaho and Washington, respectively; respective results for Model 4 are in Table 32 and Table 33. As noted for California, Model 1 includes the simultaneous estimation of all of the reliable prealgebra variables in the prediction of later overall Algebra scores (across all B items, Table 16) and (separately) for each of the specific later outcomes. Model 4 includes the same predictors, along with overall ELA scores and overall mathematics scores.

Table 28: Overall Variance Explained (*R^2*) for IRT Subscores for Idaho

Outcome Variables	*Overall Algebra*	*Quantitative Literacy*	*Generalized Arithmetic*	*Functional Thinking*	*Constraint Equations*	*Geometry & Statistics*	*ELA*
Model 1	0.58	0.45	0.20	0.30	0.19	0.16	0.45
Model 2	0.62	0.50	0.22	0.33	0.20	0.18	0.64
Model 3	0.65	0.50	0.23	0.33	0.20	0.18	0.51
Model 4	0.66	0.51	0.23	0.34	0.20	0.19	0.64

Table 29: Overall Variance Explained (*R^2*) for IRT Subscores for Washington

Outcome Variables	*Overall Algebra*	*Quantitative Literacy*	*Generalized Arithmetic*	*Functional Thinking*	*Constraint Equations*	*Geometry & Statistics*	*ELA*
Model 1	0.629	0.518	0.022	0.359	0.192	0.211	0.501
Model 2	0.663	0.540	0.236	0.389	0.207	0.231	0.664
Model 3	0.690	0.567	0.248	0.397	0.211	0.232	0.553
Model 4	0.690	0.578	0.252	0.415	0.213	0.231	0.668

As in California, all of the models were predictors of later overall Algebra and Quantitative Literacy (see Table 28 and Table 29). The prediction of Generalized Arithmetic, Functional Thinking, Constraint Equations, as well as Geometry and Statistics is considerably lower, but is still statistically significant.

Table 30: Model 1 Standardized Regression Coefficients for IRT Subscores for Idaho

Outcome Variables	*Overall Algebra Beta*	*Overall Algebra SE*	*Quantitative Literacy Beta*	*Quant-itative Literacy SE*	*Generalized Arithmetic Beta*	*Generalized Arithmetic SE*	*Functional Thinking Beta*	*Functional Thinking SE*	*Constraint Equations Beta*	*Constraint Equations SE*	*Geometry & Statistics Beta*	*Geometry & Statistics SE*
Predictor Variables												
A1a - Whole Numbers	0.139	0.004	0.116	0.004	0.097	0.005	0.118	0.005	0.000	0.004	0.099	0.005
A1b - Fractions	0.181	0.004	0.162	0.004	0.074	0.005	0.137	0.005	0.085	0.005	0.114	0.005
A1c - Decimals	0.071	0.003	0.068	0.004	0.043	0.004	0.039	0.004	0.084	0.005	0.023	0.005
A2a - Basic Problem Solving: Whole Numbers	0.014	0.003	-0.020	0.004	0.033	0.005	0.043	0.004	0.033	0.004	-0.005	0.005
A2b - Basic Problem Solving: Fractions	0.125	0.003	0.107	0.004	0.085	0.005	0.077	0.004	0.036	0.005	0.077	0.005
A3a – Complex Problem Solving: Whole Numbers	0.154	0.004	0.131	0.004	0.127	0.005	0.133	0.005	0.114	0.005	0.064	0.005
A3b – Complex Problem Solving: Fractions	0.207	0.004	0.183	0.004	0.101	0.005	0.165	0.005	0.054	0.005	0.114	0.005
A4 – Mathematical Reasoning & Communication	0.232	0.004	0.202	0.005	0.129	0.006	0.122	0.005	0.131	0.005	0.086	0.006
Geometry & Measurement	0.082	0.003	0.077	0.004	0.051	0.005	0.056	0.004	0.129	0.006	0.048	0.005

Table 31: Model 1 Standardized Regression Coefficients for IRT Subscores for Washington

Outcome Variables	*Overall Algebra Beta*	*Overall Algebra SE*	*Quantitative Literacy Beta*	*Quant-itative Literacy SE*	*Generalized Arithmetic Beta*	*Generalized Arithmetic SE*	*Functional Thinking Beta*	*Functional Thinking SE*	*Constraint Equations Beta*	*Constraint Equations SE*	*Geometry & Statistics Beta*	*Geometry & Statistics SE*
Predictor Variables												
A1a - Whole Numbers	0.098	0.002	0.092	0.002	0.056	0.002	0.075	0.002	0.045	0.002	0.055	0.002
A1b - Fractions	0.194	0.002	0.181	0.002	0.091	0.002	0.156	0.002	0.130	0.002	0.130	0.002
A1c - Decimals	0.059	0.001	0.058	0.002	0.045	0.002	0.032	0.002	0.023	0.002	0.027	0.002
A2a - Basic Problem Solving: Whole Numbers	0.042	0.001	0.039	0.002	0.029	0.002	0.039	0.002	-0.005	0.002	0.005	0.002
A2b - Basic Problem Solving: Fractions	0.101	0.001	0.091	0.002	0.075	0.002	0.053	0.002	0.085	0.002	0.056	0.002
A3a – Complex Problem Solving: Whole Numbers	0.121	0.002	0.104	0.002	0.080	0.002	0.096	0.002	0.030	0.002	0.060	0.002
A3b – Complex Problem Solving: Fractions	0.199	0.002	0.179	0.002	0.113	0.002	0.161	0.002	0.127	0.002	0.127	0.002
A4 – Mathematical Reasoning & Communication	0.294	0.002	0.265	0.002	0.176	0.003	0.212	0.002	0.173	0.003	0.165	0.003
Geometry & Measurement	0.076	0.001	0.069	0.002	0.056	0.002	0.058	0.002	0.014	0.002	0.047	0.002

As with California, all of the prealgebra variables are significant predictors of all later outcomes (Model 1), although the most weight should be given to Number Sense: Fractions and Reasoning & Communication because these were the most reliable predictors (Table E3). Again, we focused on a contrast of overall Algebra and Geometry and Statistics and focused on the key results found for California. As shown in Table 30 and Table 31, the results are consistent with the core findings for California in that Number Sense: Whole Numbers, Number Sense: Fractions, Complex Problem Solving with Fractions, and Reasoning & Communication are all stronger predictors of later Algebra than later Geometry and Statistics. Moreover, each of these variables is more strongly related to later Algebra than is earlier Geometry and Measurement. One additional finding is that Problem Solving with Fractions is a stronger predictor of later Algebra performance in Idaho than in California.

Table 32: Model 4 Standardized Regression Coefficients for IRT Subscores for Idaho

Outcome Variables	***Overall Algebra Beta***	***Overall Algebra SE***	***Quantitative Literacy Beta***	***Quant-itative Literacy SE***	***Generalized Arithmetic Beta***	***Generalized Arithmetic SE***	***Functional Thinking Beta***	***Functional Thinking SE***	***Constraint Equations Beta***	***Constraint Equations SE***	***Geometry & Statistics Beta***	***Geometry & Statistics SE***
Predictor Variables												
A1a – Whole Numbers	0.080	0.003	0.066	0.004	0.061	0.005	0.076	0.005	0.058	0.005	0.000	0.004
A1b – Fractions	-0.007	0.004	-0.006	0.005	-0.041	0.006	0.014	0.006	-0.001	0.006	0.066	0.005
A1c – Decimals	0.011	0.003	0.014	0.004	0.007	0.004	0.002	0.004	0.007	0.005	0.015	0.006
A2a – Basic Problem Solving: Whole Numbers	-0.019	0.003	-0.049	0.004	0.013	0.004	0.020	0.004	0.020	0.005	-0.007	0.005
A2b – Basic Problem Solving: Fractions	0.037	0.003	0.029	0.004	0.031	0.005	0.018	0.004	0.074	0.005	-0.023	0.005
A3a – Complex Problem Solving: Whole Numbers	0.050	0.004	0.043	0.004	0.063	0.005	0.056	0.005	0.005	0.006	0.030	0.005
A3b - Complex Problem Solving: Fractions	0.024	0.004	0.023	0.005	-0.011	0.006	0.041	0.005	0.048	0.006	0.005	0.006
A4 – Mathematical Reasoning & Communication	0.048	0.004	0.042	0.005	0.016	0.006	-0.006	0.006	0.044	0.006	0.016	0.006
Geometry & Measurement	0.002	0.003	0.006	0.004	0.002	0.005	0.001	0.004	0.017	0.005	-0.015	0.006
ELA	0.157	0.005	0.098	0.006	0.106	0.007	0.181	0.007	0.087	0.007	0.004	0.005
Math	0.533	0.007	0.500	0.009	0.319	0.011	0.307	0.010	0.233	0.011	0.126	0.007

As in California, controlling for earlier mathematics and ELA scores lowers the magnitude of the relations between early prealgebra predictors and later Algebra outcomes in both Idaho and Washington (Table 32 and Table 33). As an example, for Idaho, the relation between Fractions and later Algebra is no longer significant. For Idaho, the effects for Whole Numbers and Complex Problem Solving with Fractions remain significant but predict later Geometry and Statistics almost as well as they predict later Algebra. In other words, early competence with Whole Numbers and Problem Solving with Fractions predicts better performance in later Algebra and later Geometry and Statistics, controlling for overall mathematics competence. In contrast, Reasoning & Communicating not only remains a significant predictor of later Algebra, but it is also a stronger predictor of later Algebra than later Geometry and Statistics in Idaho. This pattern is confirmed for the data from Washington.

Table 33: Model 4 Standardized Regression Coefficients for IRT Subscores for Washington

Outcome Variables	*Overall Algebra Beta*	*Overall Algebra SE*	*Quantitative Literacy Beta*	*Quant-itative Literacy SE*	*Generalized Arithmetic Beta*	*Generalized Arithmetic SE*	*Functional Thinking Beta*	*Functional Thinking SE*	*Constraint Equations Beta*	*Constraint Equations SE*	*Geometry & Statistics Beta*	*Geometry & Statistics SE*
Predictor Variables												
A1a – Whole Numbers	0.046	0.001	0.046	0.002	0.024	0.002	0.032	0.002	0.016	0.002	0.022	0.002
A1b – Fractions	0.018	0.002	0.023	0.002	-0.020	0.003	0.019	0.002	0.033	0.003	0.030	0.003
A1c – Decimals	0.012	0.001	0.016	0.001	0.013	0.002	-0.003	0.002	0.000	0.002	0.004	0.002
A2a – Basic Problem Solving: Whole Numbers	0.012	0.001	0.013	0.001	0.010	0.002	0.013	0.002	-0.021	0.002	-0.015	0.002
A2b – Basic Problem Solving: Fractions	0.026	0.001	0.024	0.002	0.026	0.002	-0.004	0.002	0.041	0.002	0.014	0.002
A3a – Complex Problem Solving: Whole Numbers	0.034	0.001	0.030	0.002	0.026	0.002	0.023	0.002	-0.021	0.002	0.003	0.002
A3b – Complex Problem Solving: Fractions	0.039	0.002	0.037	0.002	0.013	0.003	0.033	0.002	0.035	0.003	0.031	0.003
A4 – Mathematical Reasoning & Communication	0.102	0.002	0.097	0.002	0.054	0.003	0.056	0.003	0.062	0.003	0.049	0.003
Geometry & Measurement	0.003	0.001	0.004	0.002	0.008	0.002	-0.001	0.002	-0.026	0.002	0.003	0.002
ELA	0.152	0.002	0.105	0.002	0.098	0.003	0.174	0.003	0.110	0.003	0.145	0.003
Math	0.508	0.003	0.471	0.004	0.331	0.005	0.366	0.004	0.253	0.005	0.254	0.005

As shown in Table 34 and Table 35, most of the early prealgebra measures also predict later ELA scores. In fact, Reasoning & Communicating is a better predictor of later ELA than later Algebra (Model 1) in Idaho and about as good in Washington, in keeping with a strong language and reading component to the former measure.

Table 34: Standardized Regression Coefficients for IRT Subscores Predicting ELA for Idaho

Regression Models	***Model 1 Beta***	***Model 1 SE***	***Model 2 Beta***	***Model 2 SE***	***Model 3 Beta***	***Model 3 SE***	***Model 4 Beta***	***Model 4 SE***
Predictor Variables								
A1a – Whole Numbers	0.107	0.004	0.044	0.004	0.061	0.004	0.035	0.004
A1b – Fractions	0.133	0.004	0.026	0.004	-0.042	0.005	-0.018	0.004
A1c – Decimals	0.040	0.004	0.018	0.003	-0.018	0.004	0.002	0.003
A2a – Basic Problem Solving: Whole Numbers	0.042	0.004	0.014	0.003	0.013	0.004	0.008	0.003
A2b – Basic Problem Solving: Fractions	0.086	0.004	0.025	0.003	0.007	0.004	0.006	0.003
A3a – Complex Problem Solving: Whole Numbers	0.185	0.004	0.064	0.004	0.105	0.004	0.048	0.004
A3b – Complex Problem Solving: Fractions	0.187	0.004	0.049	0.004	0.026	0.005	0.011	0.004
A4 – Mathematical Reasoning & Communication	0.217	0.005	0.057	0.004	0.060	0.005	0.022	0.004
Geometry & Measurement	0.077	0.004	0.018	0.003	0.005	0.004	0.001	0.003
ELA	--	--	0.645	0.004	--	--	0.600	0.005
Math	--	--	--	--	0.577	0.008	0.170	0.008

Table 35: Standardized Regression Coefficients for IRT Subscores Predicting ELA for Washington

Regression Models	*Model 1 Beta*	*Model 1 SE*	*Model 2 Beta*	*Model 2 SE*	*Model 3 Beta*	*Model 3 SE*	*Model 4 Beta*	*Model 4 SE*
Predictor Variables								
A1a – Whole Numbers	0.079	0.002	0.033	0.001	0.035	0.002	0.022	0.001
A1b – Fractions	0.144	0.002	0.038	0.002	-0.022	0.002	-0.007	0.002
A1c – Decimals	0.021	0.002	0.004	0.001	-0.031	0.002	-0.011	0.001
A2a – Basic Problem Solving: Whole Numbers	0.079	0.002	0.035	0.001	0.057	0.002	0.032	0.001
A2b – Basic Problem Solving: Fractions	0.063	0.002	0.019	0.001	-0.007	0.002	0.000	0.001
A3a – Complex Problem Solving: Whole Numbers	0.150	0.002	0.044	0.002	0.085	0.002	0.032	0.002
A3b – Complex Problem Solving: Fractions	0.178	0.002	0.051	0.002	0.038	0.002	0.016	0.002
A4 – Mathematical Reasoning & Communication	0.255	0.002	0.081	0.002	0.094	0.002	0.043	0.002
Geometry & Measurement	0.082	0.002	0.024	0.001	0.017	0.002	0.008	0.001
ELA	--	--	0.633	0.002	--	--	0.586	0.002
Math	--	--	--	--	0.563	0.004	0.179	0.003

QUESTION B

Question B: Are we spending too much time on some less important standards and not enough on some more important standards? Where should more time and effort be invested in mathematics instruction, and where less?

In preview, engagement with and mastery of algebra-related and CCSS-aligned workspaces during the middle school years was a substantive predictor of later Algebra I EOC performance. Across students, a 1 standard deviation increase in the number of mastered workspaces was associated with a 0.406 standard deviation increase in later Algebra I EOC performance, controlling for prior mathematics competence. Moreover, students who scored lower on prior mathematics assessments appeared to gain more than their peers with higher prior mathematics scores by mastering these workspaces, which was associated by a smaller difference in later Algebra I EOC performance. However, the relations between mastering specific CCSS-aligned workspaces and later Algebra I performance was much weaker than the overall number of mastered workspaces.

Student Engagement with Different CCSS

There was high variability in the level of practice of different CCSS, across both the entire dataset and for the students who completed the Algebra I EOC exam. For some CCSS, students completed on average nine workspaces. Table 36 shows the most frequently and infrequently practiced CCSS for the entire dataset, only the analyses sample, and separately for seventh and eighth grade students.

Table 36: Most Frequently and Least Frequently Practiced CCSS (means and confidence intervals in parentheses)

		Most Frequently Practiced	Least Frequently Practiced
Entire Dataset	**All grades**	HSF.BF.B.3 (9.62, [5.73,17.39]) 6.EE.B.7 (9.48, [5.91,16.62]) 6.EE.A.3 (7.94, [3.18,17.48]) 7.EE.B.4.a (7.79, [2.93,18.79])	HSF.TF.B.5 (1.02, [0.89,1.27]) HSS.CP.A.5 (1.01, [0.90,1.23]) HSA.SSE.A.1.b (1.00, [0.94,1.14]) HSS.CP.A.3 (1.00, [0.95,1.11])
Analyses Sample	**All grades**	7.EE.B.4.a (12.59, [8.58,16.93]) 6.EE.B.7 (11, [11,11]) HSF.BF.B.3 (10.11, [6.66,13.56]) 6.EE.A.3 (7, [1.48,12.51])	8.SP.A.1 (1.02, [0.88,1.14]) 8.SP.A.3 (1.02, [0.88,1.14]) HSG.CO.A.4 (1.01, [0.90,1.22]) HSG.GPE.B.6 (1.00, [0.93,1.08])
	Grade 7	7.EE.B.4.a HSF.BF.B.3 6.EE.B.7 HSA.APR.A.1	HSS.ID.A.1 HSF.BF.A.1.b HSF.LE.A.1.c 8.SP.A.3
	Grade 8	6.EE.A.4 6.EE.B.7 6.EE.A.3 HSF.BF.B.3	HSF.LE.A.1.c HSG.CO.A.5 HSG.CO.A.4 HSG.GPE.B.6

Note. The terms refer to grade-level and specific CCSS. For instance, 6.EE.B.7 refers to grade 6 standard, Expressions and Equations (*Solve real-world and mathematical problems by writing and solving equations of the form $x + p = q$ and $px = q$ for cases in which p, q and x are all nonnegative rational numbers).*

CCSS Predictors of Algebra I EOC Exam Performance

The first task was to establish if practice in the MATHia Tutor was related to later Algebra I EOC exam performance. To do this, we predicted the Algebra score using the number of workspaces mastered, number of workspaces not mastered (moved forward before mastery), and number of workspaces that were incomplete. If practice in the system is important, we should find that higher Algebra I EOC scores are predicted by the number of workspaces mastered, as opposed to those not mastered or incomplete. For this analysis, the earliest available FSA Math standardized test score and enrolled grade were covariates, and school and type of class were random effects. All scores were standardized with a mean of 0 and standard deviation of 1.

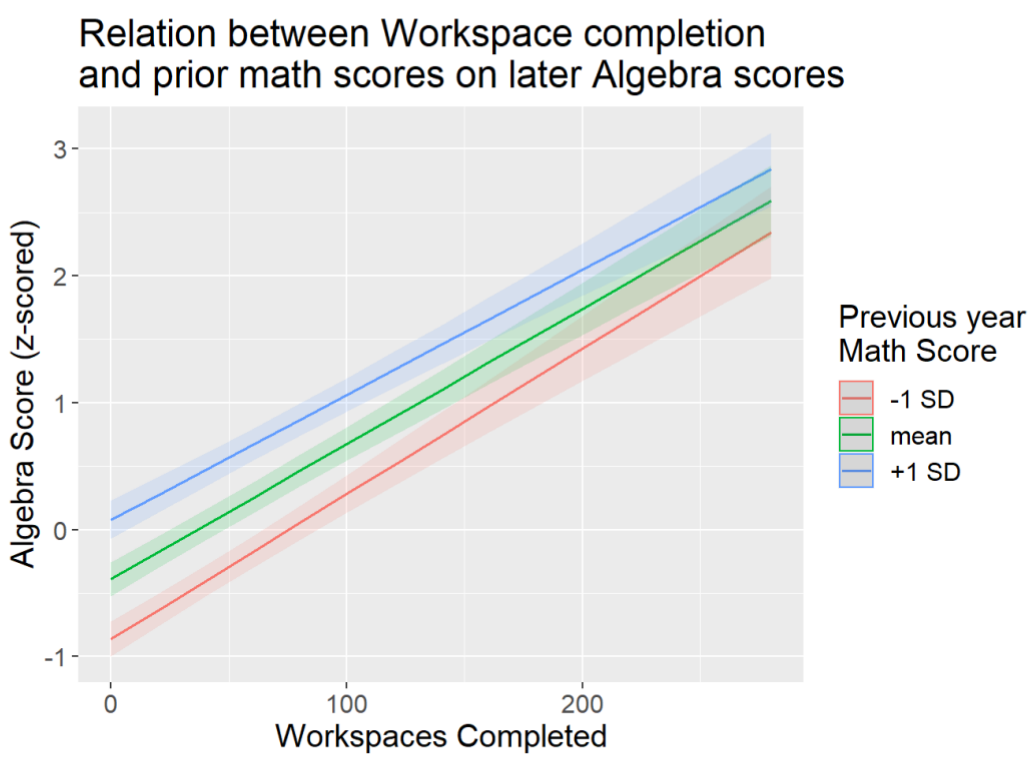
The results show that the previous mathematics score was the best predictor of Algebra I EOC scores, followed by the number of mastered workspaces (Table 37). The latter effect is substantive: a 1 standard deviation (*SD*) increase in number of mastered workspaces is related to a 0.406 *SD* increase in later Algebra 1 scores; overall $R^2 = .66$. The numbers of non-mastered and incomplete workspaces were both negative predictors of Algebra scores. The interaction between number of mastered workspaces and previous mathematics scores was significant, indicating that gains associated with mastered workspaces varied by prior mathematics competence.

Figure 5 shows the interaction effect (based on parameter estimates) and suggests that students with lower scores gain more than ones with higher scores by mastering additional workspaces. In other words, the gap between higher- and lower-scoring students in later Algebra I EOC scores becomes smaller with increases in the number of mastered workspaces.

Table 37: Results of Regression Predicting Algebra I EOC Exam Scores Based on Different Types of Tutor Usage Frequency

Predictor	Estimate	Std Error	*t*	*p*
(Intercept)	-1.747	0.225	-7.77	0.000
Workspaces Mastered	0.406	0.024	17.28	< .001
Workspaces Not Completed	-0.029	0.011	-2.65	0.008
Workspaces Not Mastered	-0.167	0.012	-13.45	< .001
Previous Mathematics Score	0.707	0.020	35.66	< .001
Mathematics Course Grade (grades 6 or 7)	0.232	0.027	8.57	< .001
Interaction: Mastered by Previous Mathematics Score	-0.049	0.016	-3.14	0.002

Figure 5: Interaction Between Prior Mathematics Score and Workspaces Mastered in the Prediction of Later Algebra I EOC Scores



To explore the relation between specific CCSS engagement and later Algebra I EOC scores, we selected all the CCSS with variability in usage (regardless of mastered, non-mastered, or incomplete status), using the frequency analyses from the previous section (Table 37). CCSS mastered by all students were removed from the analyses.(See Appendix K for the list of these standards.) If a particular student did not complete any workspaces for a given CCSS, we entered 0 for that CCSS. We then repeated the same regression as above using the number of workspaces for each CCSS as predictors. We included in the analyses only CCSS where more than 25% of the students saw more than 0 workspaces. Using these criteria, the model included 50 CCSS.

To account for the large number of predictors (CCSS, grade, and previous math score), we used a significance criterion of $p < .01$. The model was significant and explained 64% of the variance. As shown in Table 38, two CCSS were significant predictors of later Algebra I EOC success, controlling for grade and prior mathematics performance; engagement in one was inversely related to engagement in the other ($r = -.44$). The absolute magnitude of both of these effects was modest. For instance, a 1 *SD* increase in the mastery of workspaces related to 7.EE.B.4.a was associated with a 0.149 *SD* increase in later Algebra I EOC scores, controlling for prior math scores.

Table 38: Practiced Workspaces that Predict Later Algebra I EOC Exam Scores

Predictor (CCSS)	Estimate	Std Error	*t*	*p*	Non-Zero Values	Mean Use in Sample
Grade 7: Expressions & Equations. Solve word problems leading to equations of the form px + q = r and p(x + q) = r, where p, q, and r are specific rational numbers. Solve equations of these forms fluently. (7.EE.B.4.a)	0.149	0.039	3.87	<.0001	1369	11.33
High School: Interpreting Functions. Compare properties of two functions each represented in a different way (algebraically, graphically, numerically in tables, or by verbal descriptions). (HSF.IF.C.9)	0.165	0.054	3.05	<.0001	2288	2.31
Grade	0.334	0.030	11.11	<.0001	---	---
Previous Mathematics Score	0.836	0.020	41.96	<.0001	---	---

Mismatches Between Predictors of Algebra I Success and Practice Emphasis

Finally, to investigate the degree of practice time that is being spent on CCSS topics that do *not* predict later Algebra I EOC scores, we examined the CCSS that were practiced most frequently and least frequently. The goal was to determine which of these were significantly related to later Algebra success.

Table 39 includes six CCSS topics practiced least frequently (top) and six CCSS topics practiced most frequently (bottom); these were assessed in terms of workspaces seen for the students with Algebra I EOC score and included only CCSS with at least 25% non-zero values. As can be seen in the Table, none of the least frequently practiced items predicted later Algebra performance. The only significant frequently practiced CCSS was also identified in Table 38 (e.g., 7.EE.B.4.a).

The five remaining CCSS were weakly and non-significantly related to later Algebra I EOC performance. For instance, build new functions from existing functions (HSF.BF.B.3) is frequently practiced, but this practice is not related to later Algebra I EOC scores. The reasons for this (e.g., the students already had sufficient competencies in these areas, or the areas were not emphasized in the Algebra I EOC exam) cannot be determined from these data.

Table 39: Least and Most Practiced CCSS As Related to Later Algebra I EOC Scores

CCSS	Estimate	*p*	Mean Use in Sample
Low Frequency of Practice			
High School: Functions, Linear, Quadratic, & Exponential Models. Construct and compare linear, quadratic, and exponential models and solve problems. (HSF.LE.A.1.c)	0.061	0.263	1.01
High School: Reasoning with Equations & Inequalities. Solve equations and inequalities in one variable. (HSA.REI.B.4.a)	-0.022	0.592	1.02
High School: Reasoning with Equations & Inequalities. Represent and solve equations and inequalities graphically. (HSA.REI.D.11)	-0.054	0.206	1.02
High School: Statistics & Probability. Interpreting Categorical & Quantitative Data » Summarize, represent, and interpret data on a single count or measurement variable. (HSS.ID.A.1+HSS.ID.A.2+HSS.ID.A.3)	0.035	0.403	1.03
High School: Interpreting Functions. Analyze functions using different representations. (HSF.IF.C.8.a)	-0.060	0.102	1.03
High School: Interpreting Functions. Interpret functions that arise in applications in terms of the context. (HSF.IF.B.5)	0.078	0.086	1.03
High Frequency of Practice			
Grade 7: Expressions & Equations. Solve word problems leading to equations of the form px + q = r and p(x + q) = r, where p, q, and r are specific rational numbers. Solve equations of these forms fluently. (7.EE.B.4.a)	0.149	<.0001	11.33
High School: Building Functions. Build new functions from existing functions. (HSF.BF.B.3)	0.083	0.161	10.05
High School: Arithmetic with Polynomials & Rational Expressions. Perform arithmetic operations on polynomials. (HSA.APR.A.1)	-0.032	0.409	6.60
Grade 8: Functions. Use functions to model relationships between quantities. (8.F.B.4)	0.026	0.169	4.84
High School: Arithmetic with Polynomials & Rational Expressions. Rewrite rational expressions. (HSA.APR.D.6)	0.036	0.443	3.92
High School: Creating Equations. Create equations that describe numbers or relationships. (HSA.CED.A.1)	-0.066	0.116	3.20

QUESTION C

***Question C**: How do clusters of students classified according to their profiles across assessment items fare over time? Do the achievement gaps widen for some clusters (controlling for background factors) but not others? Students with different profiles may benefit differently from different interventions. Some topics (see Question A) may be more difficult for some profiles, while other topics are more difficult for others. Are there some schools outperforming the expectations based on students' demographic profiles for some clusters?*

In preview, using IRT subscores from California, we assessed whether there were differences across demographic groups in the *relation* between earlier performance in Mathematical Reasoning & Communication and overall mathematics competence and later overall Algebra scores. These analyses revealed several statistically significant effects, but these were all small and of little practical importance. In all, strong performance in elementary mathematics CCSS, which emphasizes the concepts, procedural fluencies, and applications of arithmetic, was important for the later Algebra performance of all students. The results were similar for the MATHia analyses; that is, the benefits of engagement with and mastery of CCSS workspaces in middle school is associated with similar gains in later Algebra I EOC performance across most demographic groups. The benefits may have differed, however, because the level of engagement differed across demographic groups.

California

As noted before, the SBAC data did allow not us to reliably assess of many of the earlier prealgebra abilities (e.g., fractions vs. overall math), and thus we focused on the Mathematical Reasoning & Communication. The latter was the strongest predictor among the A variables and had the highest marginal reliability (0.74). We also assessed earlier overall mathematics competence (i.e., the theta score; Model 5B); the theta score was from fourth grade for the 4/7 cohort and fifth grade from the 5/8 cohort.

We used these as predictors (IRT subscores) of later performance in Algebra (overall IRT subscore). The Algebra outcomes were the same as those reported for Question A (for completeness) and included Geometry and Statistics as a contrast, but the results for overall Algebra should be considered the most reliable. As described in the methods, the associated regression model is the same as Model 4 but with the inclusion of an interaction between the Mathematical Reasoning & Communication measure and demographic groups (including English learner status, having a 504 Plan, and eligibility for subsidized lunches) in the prediction of later Algebra outcomes, and in a separate model, the interaction between overall Mathematical Competence and demographic group in the prediction of later Algebra outcomes.

The main effects for the Mathematical Reasoning & Communication model are shown in Table 40, and the associated interaction effects are shown in Table 41. The corresponding results for earlier overall mathematics competence are shown in Tables 42 and 43, respectively. Statistically significant effects are in bold, although these do not necessarily indicate practically important effects.

As an illustration of these effects, the coefficient for earlier mathematics competence (for all students) in the prediction of later overall Algebra (Table 42) is 0.637. This indicates that being 1 standard deviation above average in math skills at the end of the elementary school years (i.e., end of grades 4 or 5) is associated with being 0.637 standard deviations above average in later overall Algebra, controlling for all other factors in the model. This result is not surprising since we generally expect higher math performance in earlier grades to be associated with higher math performance in later grades.

The interactions in Table 41 and Table 43 indicate the extent to which the strength of the relation between earlier competencies (Mathematical Reasoning & Communication and overall mathematics) and later Algebra outcomes differ across demographic groups. As can be seen in these tables, there are many statistically significant effects, but all of them are very small and of no practical significance. As an example, the interaction effect (Table 41) for economic disadvantage (eligible for free or reduced lunch) is -0.007, which is statistically significant. However, controlling for all other variables in the model, the strength of the relation between earlier Reasoning & Communication scores and later overall Algebra is 0.435, meaning that a 1 standard deviation increase in Reasoning & Communication scores is associated with a 0.435 standard deviation increase in later overall Algebra scores for students who are not eligible for free or reduced lunch. For students who are eligible for free or reduced lunch, the relation between earlier Reasoning & Communication scores and later overall Algebra is 0.428 (0.435 – 0.007).

The overall results mean that strong earlier competencies in mathematics are similarly predictive of strong later outcomes in Algebra across gender, ethnic, and racial groups, English learner status, students with and without a 504 Plan, and students with and without economic disadvantage.

Table 40: Model 5A IRT Scores Main Effects by Student Group with Predictor A4 by Outcome Variables—Reasoning & Communication

Predictor Variables	*Overall Algebra Beta*	*Overall Algebra SE*	*Quantitative Literacy Beta*	*Quantitative Literacy SE*	*Generalized Arithmetic Beta*	*Generalized Arithmetic SE*	*Functional Thinking Beta*	*Functional Thinking SE*	*Geometry & Statistics Beta*	*Geometry & Statistics SE*
A4 – Reasoning & Communication	0.435*	0.003	0.442*	0.003	0.142*	0.006	0.175*	0.003	0.224*	0.004
Female	-0.050*	0.001	-0.054*	0.002	-0.004*	0.002	0.007*	0.002	-0.043*	0.002
Hispanic or Latino	-0.031*	0.005	-0.040*	0.005	0.036*	0.007	0.007	0.005	-0.064*	0.005
American Indian or Alaska Native	-0.008*	0.012	-0.007*	0.012	-0.002*	0.018	-0.006*	0.016	-0.005*	0.016
Asian	0.090*	0.014	0.084*	0.014	0.065*	0.011	0.047*	0.010	0.031*	0.007
Black or African American	-0.040*	0.008	-0.046*	0.007	0.004*	0.011	-0.006*	0.009	-0.034*	0.008
Native Hawaiian or Other Pacific Islander	-0.001*	0.013	-0.002	0.013	0.004	0.018	0.001	0.017	-0.004*	0.015
Filipino	0.017*	0.009	0.016*	0.009	0.014	0.012	0.018*	0.011	-0.009*	0.008
English learner status	-0.038*	0.008	-0.051*	0.008	0.033	0.008	-0.020*	0.007	-0.042*	0.004
504 Plan	-0.006*	0.008	-0.004*	0.008	-0.006*	0.010	-0.004*	0.010	-0.001	0.011
Economic Disadvantage	-0.051*	0.006	-0.059*	0.006	0.013*	0.007	-0.001*	0.005	-0.063*	0.005
ELA	0.363*	0.002	0.325*	0.002	0.220*	0.004	0.242*	0.003	0.223*	0.002

*Statistically significant, $p < .05$.

Table 41: Model 5A Factor Scores Interaction Effects by Student Group with Predictor A4 by Outcome Variables—Reasoning & Communication

Predictor Variables	*Overall Algebra Beta*	*Overall Algebra SE*	*Quantitative Literacy Beta*	*Quantitative Literacy SE*	*Generalized Arithmetic Beta*	*Generalized Arithmetic SE*	*Functional Thinking Beta*	*Functional Thinking SE*	*Geometry & Statistics Beta*	*Geometry & Statistics SE*
A4: Female	-0.034*	0.001	-0.026*	0.002	-0.029*	0.002	-0.017*	0.002	-0.033*	0.002
A4: Hispanic or Latino	-0.007*	0.003	-0.016*	0.003	0.042*	0.004	0.006*	0.004	-0.040*	0.003
A4: American Indian or Alaska Native	-0.002	0.010	-0.003*	0.011	0.003	0.015	-0.001	0.017	-0.004	0.017
A4: Asian	0.001	0.007	0.001	0.007	-0.001	0.008	-0.004	0.007	0.015*	0.007
A4: Black or African American	-0.011*	0.006	-0.023*	0.006	0.029*	0.007	0.009*	0.006	-0.015*	0.006
A4: Native Hawaiian or Other Pacific Islander	-0.001	0.010	-0.003*	0.010	0.003	0.016	0.001	0.016	0.000	0.018
A4: Filipino	-0.003*	0.006	-0.003*	0.006	-0.005*	0.008	0.005*	0.008	-0.001	0.008
A4: English learner status	-0.028*	0.005	-0.044*	0.006	0.064*	0.004	-0.018*	0.004	-0.028*	0.003
A4: 504 Plan	-0.001	0.007	-0.001	0.007	0.000	0.012	-0.002	0.011	-0.002	0.011
A4: Economic Disadvantage	-0.007*	0.003	-0.020*	0.003	0.046*	0.004	0.020*	0.004	-0.047*	0.003

*Statistically significant, $p < .05$.

Table 42: Model 5A Factor Scores Main Effects by Student Group with Predictor Overall Mathematics

	Overall Algebra		*Quantitative Literacy*		*Generalized Arithmetic*		*Functional Thinking*		*Geometry & Statistics*	
	Beta	**SE**	**Beta**	**SE**	**Beta**	**SE**	**Beta**	**SE**	**Beta**	**SE**
Math – Overall	0.637*	0.001	0.640*	0.001	0.238*	0.001	0.264*	0.001	0.306*	0.000
Female	0.002*	0.001	-0.007*	0.001	0.028*	0.001	0.032*	0.001	-0.013*	0.000
Hispanic or Latino	-0.015*	0.001	-0.022*	0.001	0.023*	0.001	0.010*	0.001	-0.041*	0.001
American Indian or Alaska Native	-0.006*	0.008	-0.005*	0.009	-0.003*	0.008	-0.004*	0.005	-0.003*	0.003
Asian	0.071*	0.002	0.064*	0.002	0.057*	0.002	0.042*	0.001	0.009*	0.001
Black or African American	-0.028*	0.003	-0.029*	0.003	-0.005*	0.003	-0.006*	0.002	-0.025*	0.001
Native Hawaiian or Other Pacific Islander	0.000	0.008	0.000	0.008	0.003*	0.008	0.000	0.005	-0.004*	0.003
Filipino	0.018*	0.003	0.017*	0.004	0.018*	0.003	0.014*	0.002	-0.008*	0.001
English learner status	-0.032*	0.002	-0.040*	0.002	0.005*	0.002	-0.012*	0.001	-0.033*	0.001
504 Plan	-0.005*	0.005	-0.004*	0.005	-0.006*	0.005	-0.002*	0.003	0.001	0.002
Economic Disadvantage	-0.034*	0.001	-0.037*	0.001	-0.002	0.001	-0.004*	0.001	-0.037*	0.001
ELA	0.184*	0.001	0.152*	0.001	0.126*	0.001	0.157*	0.001	0.154*	0.000

*Statistically significant, $p < .05$.

Table 43: Model 5A Factor Scores Interaction Effects by Student Group with Predictor Overall Mathematics

	Overall Algebra		*Quantitative Literacy*		*Generalized Arithmetic*		*Functional Thinking*		*Geometry & Statistics*	
	Beta	**SE**	**Beta**	**SE**	**Beta**	**SE**	**Beta**	**SE**	**Beta**	**SE**
Math: Female	-0.002*	0.001	0.004*	0.001	-0.012*	0.001	-0.002	0.001	-0.017*	0.000
Math: Hispanic or Latino	-0.011*	0.001	-0.019*	0.001	0.035*	0.001	0.006*	0.001	-0.032*	0.001
Math: American Indian or Alaska Native	-0.004*	0.006	-0.005*	0.007	0.003*	0.006	0.000	0.004	-0.003*	0.003
Math: Asian	-0.001	0.002	-0.001	0.002	0.001	0.002	-0.004*	0.001	0.017*	0.001
Math: Black or African American	-0.019*	0.002	-0.028*	0.002	0.017*	0.002	0.003*	0.002	-0.016*	0.001
Math: Native Hawaiian or Other Pacific Islander	-0.002*	0.007	-0.003*	0.007	0.001	0.007	0.000	0.005	-0.001	0.003
Math: Filipino	0.000	0.003	0.000	0.003	-0.004*	0.003	0.004*	0.002	0.001	0.001
Math: English learner status	-0.031*	0.001	-0.044*	0.001	0.038*	0.001	-0.014*	0.001	-0.022*	0.001
Math: 504 Plan	-0.002*	0.004	-0.002*	0.004	-0.001	0.004	-0.002	0.003	-0.002	0.002
Math: Economic Disadvantage	-0.009*	0.001	-0.020*	0.001	0.036*	0.001	0.017*	0.001	-0.041*	0.000

*Statistically significant, $p < .05$.

MATHia

The MATHia data were also used to determine if the relation between completion (i.e., mastered) of CCSS workspaces and later Algebra I EOC scores varied across demographic groups. To do this, the interaction between a demographic group variable (e.g., male, female) and workspaces completed in the prediction of Algebra I EOC scores was estimated, as with the California analyses. These regression analyses also included the main effects of the demographic group, workspaces completed (mastered), workspaces not completed, workspaces not mastered, prior mathematics achievement, and grade level. For all groups, Algebra I EOC scores increased with increases in the number of CCSS workspaces that were mastered.

One significant difference to emerge was for gender ($p = .009$). Here, there was no gender difference in later Algebra I EOC scores for students with lower levels of workspace mastery, but boys had higher Algebra I scores than did girls with higher levels of workspace mastery. Another significant effect was the contrast of non-gifted-designated and gifted-designated students, whereby the gap in later Algebra I EOC scores became smaller as the number of completed workspaces increased ($p = .002$).

The interactions between number of workspaces completed (mastered) and later Algebra I EOC scores were not significant for English learner status ($p = .132$), ethnic status (ps > .28), average-scoring students with a disability as compared to average students ($p = .66$), or family economic status (as indexed by eligibility for free lunch; $p = .471$).

At the same time, there were differences in the numbers of workspaces that were attempted and mastered ($p = .003$), but no differences in the number of workspaces not completed or not mastered (ps > .786). Asian students attempted ($n = 58.25$) and mastered ($n = 52.53$) more workspaces than White ($n = 52.73$, 46.73), Hispanic ($n = 49.68$, 43.13) and Black or African American ($n = 47.46$, 40.23) students.

In all, the results are largely consistent with those found for California; that is, the benefits of engagement with and mastery of CCSS workspaces in middle school are associated with similar gains in later Algebra I EOC performance across most demographic groups (with the exceptions of gender and gifted status). However, this result is qualified by differences in engagement with MATHia across racial/ethnic groups that will likely contribute to later differences in Algebra I EOC performance.

QUESTION D

***Question D**: Can the factors (Question A), emphasis (Question B), or student profiles (Question C), or trajectories in achievement differences among sub-populations be associated with the proportion of the variance in mathematics achievement among districts compared to among schools within districts compared to among classrooms within schools compared to among students within classrooms?*

In preview, for California there were no substantive differences across districts or schools within districts in preparing students for success in Algebra once students' prior mathematics performance was taken into consideration. However, prior case studies of individual districts indicate that district-level reforms can promote the mathematics performance of historically underserved students. Students in all schools benefited from mastery of MATHia algebra workspaces and the magnitude of these benefits may have varied across the demographic composition of the schools, but there was too little school-level variation in demographics to draw strong conclusions.

California

On the basis of the results for Question A and Question C, we examined overall (across demographic groups) differences in Algebra outcomes across districts, schools nested within districts, and students nested within schools. The approach involved using the regression model (Model 4) presented for Question A, but with intercepts at the school and district level modeled as random effects; in the previously presented results for Model 4, the regression models were based on an assumption that the intercepts and slopes are the same for all schools and districts. The basic model is:

- Model 4A—Main effects only for each of the predictor variables with earlier overall ELA and mathematics competence (theta) scores as covariates
  - A1a, A1b, A1c, A2b, A3a, A3b, A4, AG, ELA, Math
  - Random intercepts for schools and districts

Intraclass correlations (ICCs) were computed to evaluate the utility of fitting random intercepts, relative to the previously specified fixed-effects models. Table 44 shows the ICCs for each of the models. The within-school correlations and the within-district, within-school correlations are all small, controlling for individual differences in the predictor variables and the earlier ELA and mathematics competence scores.

Table 44: Model 4A Factor Score Intraclass Correlations by Outcome Variables

Intraclass Correlation Levels	*Overall Algebra*	*Quantitative Literacy*	*Generalized Arithmetic*	*Functional Thinking*	*Geometry & Statistics*
Within school	0.104	0.071	0.040	0.027	0.012
Within district, within school	0.067	0.028	0.017	0.011	0.006

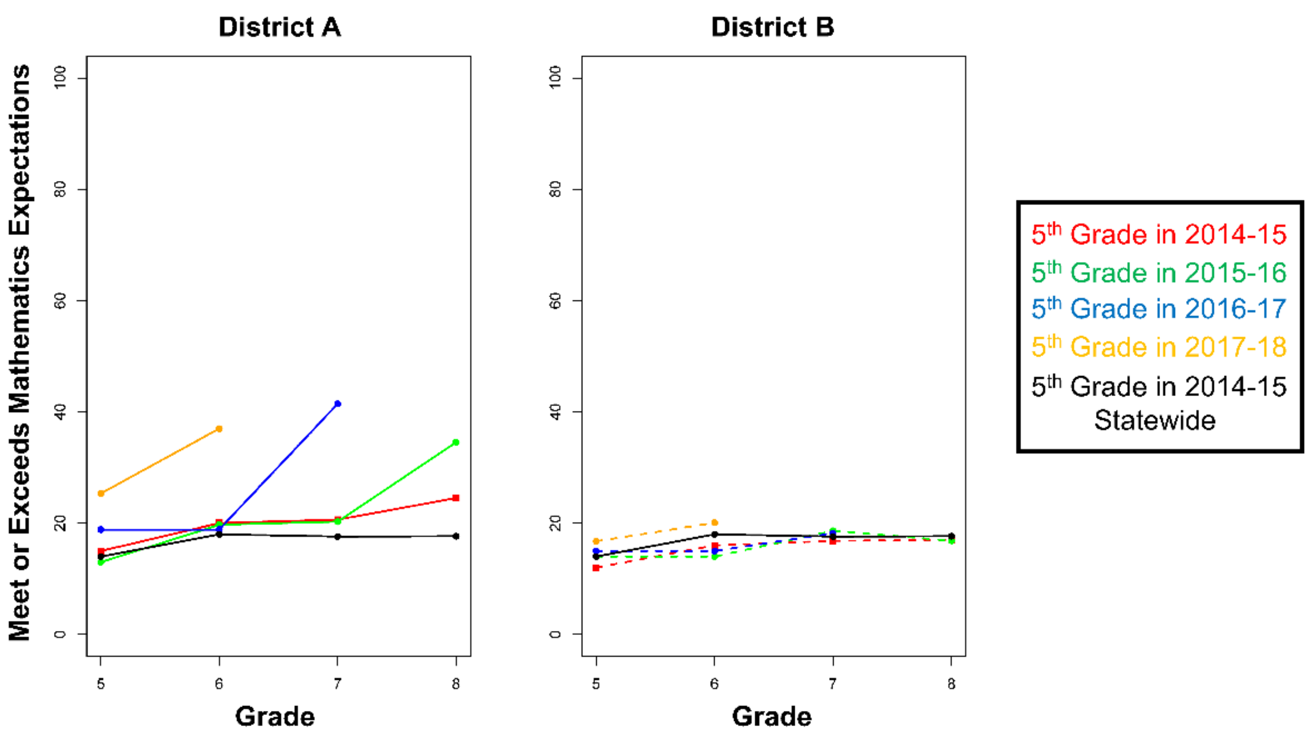
These results suggest that there is not enough variability in the effects at the school and district levels to justify fitting a series of multilevel models. This does not mean that there are no differences across districts or schools in Algebra outcomes, but rather any such effects are small once earlier prealgebra, ELA, and overall mathematics competencies are taken into consideration. *However*, this is not the same as concluding that districtwide policies and instructional practices do not matter. In fact, there is evidence that such policies and practices can make a substantive difference in achievement outcomes, even if most districts and schools appear to be more similar than different.

To illustrate the point, Figure 6 shows mathematics achievement changes for Black or African American students in two geographically close districts in California; the districts serve predominantly Hispanic students but enroll large numbers of Black or African American students. The figure shows the percentage of Black or African American students who met or exceeded expectations for Common

Core State Standards from fifth to eighth grade, inclusive, across four cohorts of students. As an example, the red lines indicate districtwide performance for students who were in fifth grade during the 2014–15 academic year. The black line shows statewide performance for Black or African American students in this same cohort.

District A began updating its policies and practices in anticipation of the CCSS assessments that began in 2014–15 (Carver-Thomas & Podolsky, 2019). The changes included high and clearly articulated achievement expectations, frequent professional development centered on the mathematics standards, and increased availability of eighth-grade Algebra courses, among other changes. As can be seen in Figure 6, the percentage of fifth-grade Black or African American students who met or exceeded expectations in mathematics was similar across districts for the 2014–15 and 2015–16 academic years, and both were similar to statewide results. Within two years, the districts began to diverge. In District B and statewide, there is little change in the percentage of Black or African American students who met or exceeded expectations for any of the grades or cohorts. The pattern is clearly different for District A several years after the changes in policies and practices.

Figure 6. Changes in Percentage of Black or African American Students Who Meet or Exceed Mathematics Expectations in Two California School Districts

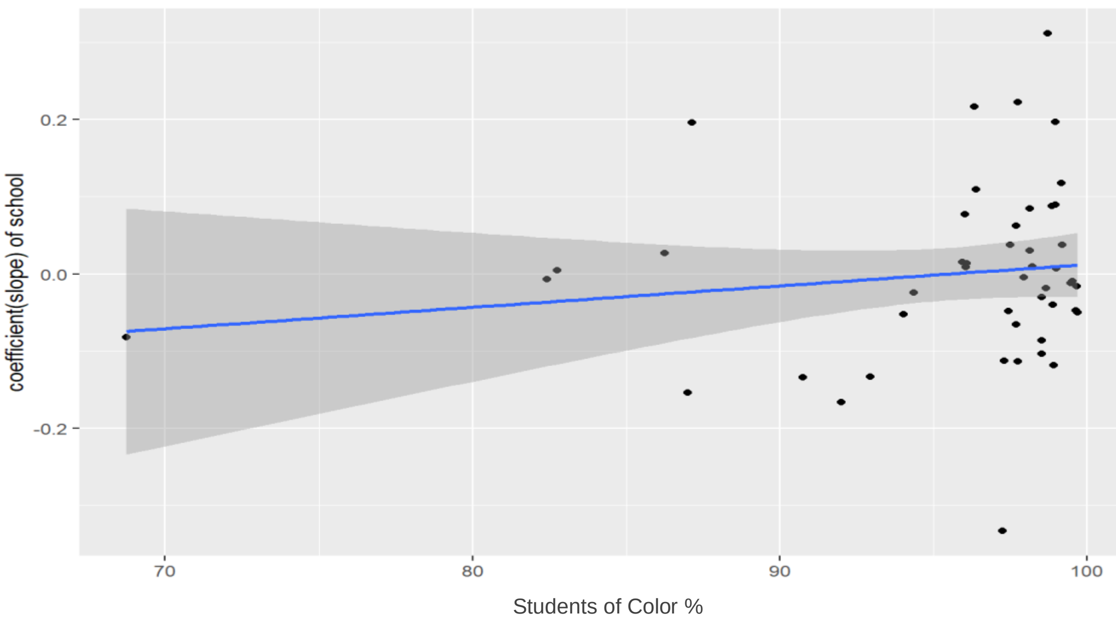


MATHia

The MATHia data were used to assess whether engagement with and mastery of CCSS workspaces in the prediction of later Algebra I EOC scores varied with the racial/ethnic composition of individual middle schools. First, the relation between the number of workspaces mastered and later Algebra I EOC scores was estimated for each school. To do this, we developed a model predicting Algebra I EOC using number of workspaces mastered and allowed the slope of this model to vary across schools. This slope indicates the impact that mastering more workspaces has on Algebra I EOC for each school. Then we relate the impact of mastering more workspaces on Algebra outcomes (y-axis in Figure 7) for each school with the percentage of students of color in the school (x-axis in Figure 7).

As shown in the figure, nearly all of the effects are positive, meaning that students in nearly all middle schools benefitted from mastery of CCSS workspaces. Figure 7 also shows that there is more variation in outcomes among schools with a majority of students of color. Schools showing the least and most gains from engagement with and mastery of CCSS workspaces are those with more than 90% students of color. There is also a small positive trend, suggesting that students in middle schools with larger numbers of students of color may gain more, in terms of later Algebra I EOC performance, by engagement and mastery of CCSS workspaces, but these results need to be interpreted with caution because there was not a lot of school-level variation in the percentage of students of color.

Figure 7. The School-Level Gains in Algebra I EOC Performance as Related to the Percentage of Students of Color in the School



In follow-up analyses, we examined whether the impact of mastering more workspaces on later Algebra I EOC scores varied across school characteristics. Figure 8 shows that the relation between workspaces mastered and later Algebra I score is stronger for schools with a larger percentage of students receiving free or reduced lunches (p = .041). Figure 9 shows that the number of mastered workspaces did not vary across these schools (p = .87), suggesting that students in schools with a high percentage of students receiving free or reduced lunches show more gains for each mastered workspace than do students in other schools.

Figure 8. The School-Level Relations between Workspaces Mastered and Later Algebra I EOC Scores as Related to Percentage of Students Receiving Free or Reduced Lunches

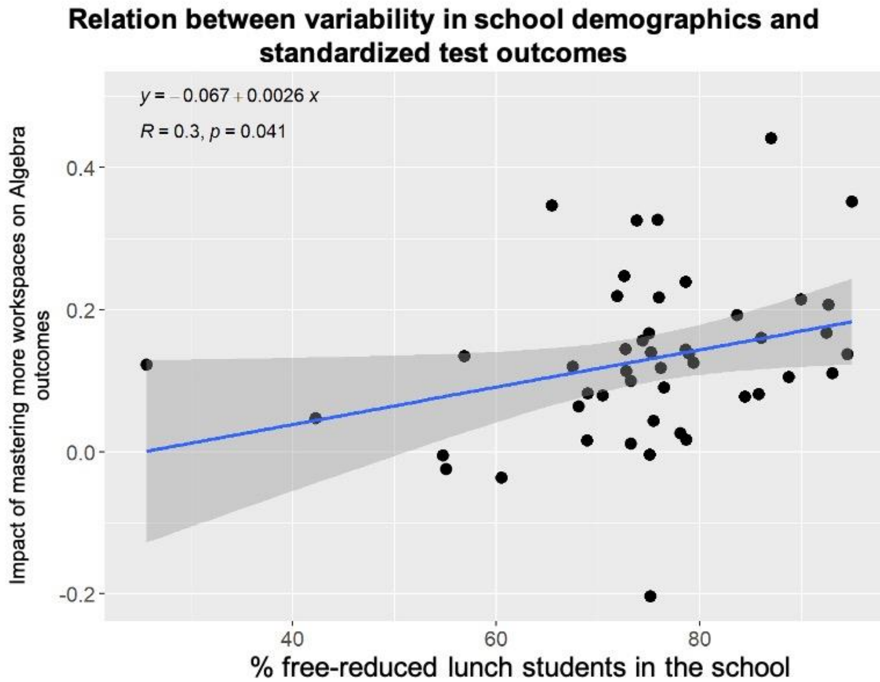


Figure 9. The Number of Workspaces Mastered and the Percentage of Students Receiving Free or Reduced Lunches

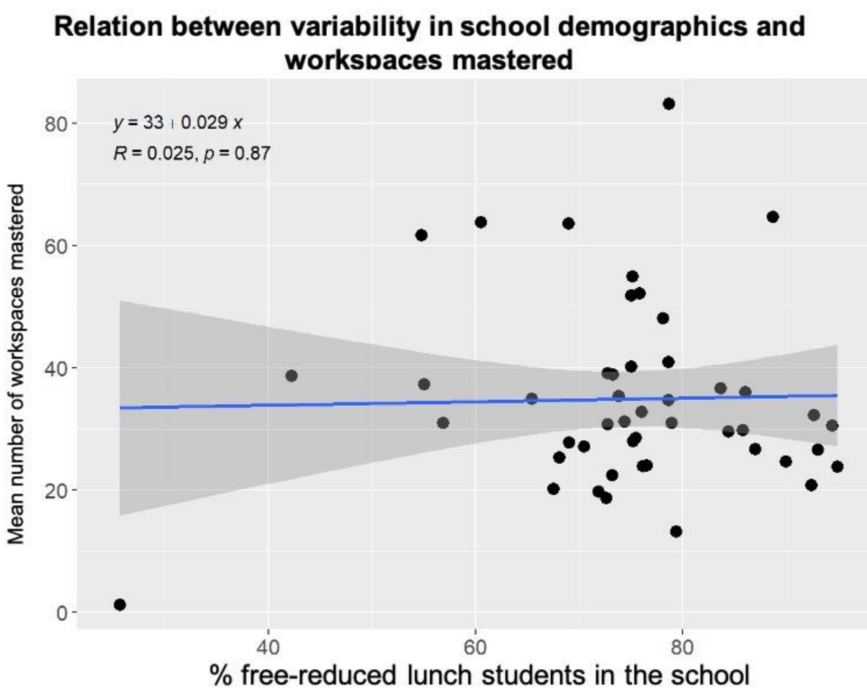
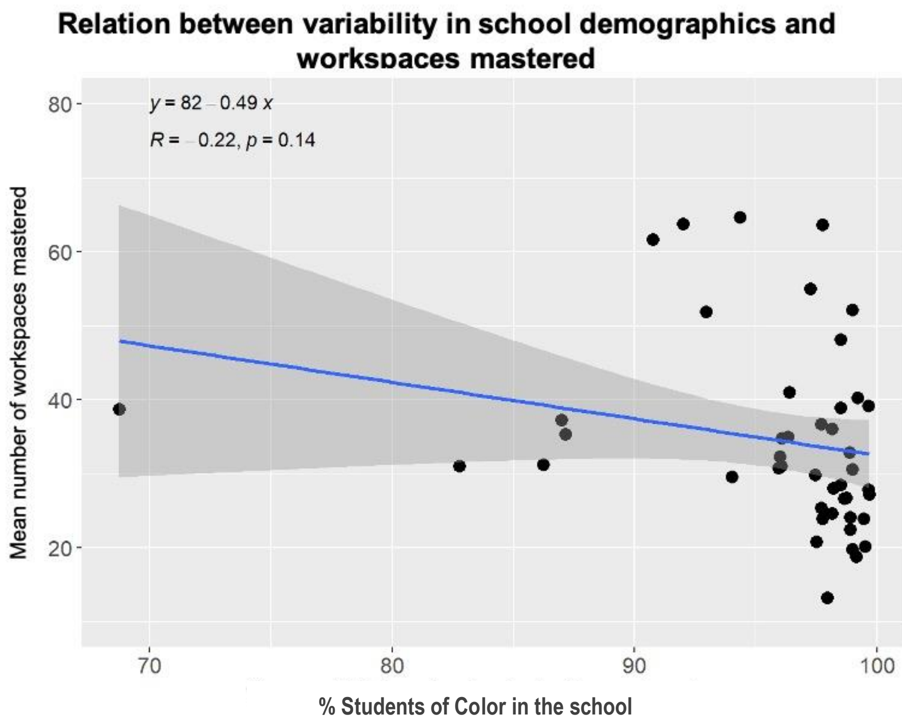


Figure 10 shows the same analyses across school that vary in the percentage of students of color. As can be seen, the relation was not significant (ps > .14), indicating that the strength of the relationship between mastered workspaces and later Algebra I EOC performance is similar across schools with different demographic makeups.

Figure 10. The Number of Workspaces Mastered and the Percentage of Students of Color



Discussion

The purpose of this paper is to provide a final report of the EMERALDS study. The study involved large-scale, longitudinal student assessment data sets from Idaho, California, and Washington as well as more detailed assessments of middle school students' engagement with pre-algebraic and algebraic material in the computer adaptive tutor MATHia (formerly the Cognitive Tutor), as related to their Algebra I end-of-course performance.

The initial goal was to identify the core mathematics competencies at the end of the elementary school years that best predict students' later success in core algebra topics, above and beyond overall mathematics competence. There were some earlier competencies that appeared to be more important than others in predicting later algebra (Question A, below), but the effects were small and overwhelmed by the influence of overall mathematics competencies. That said, it is clear that students with a strong foundation in the elementary school CCSS at the end of the elementary school years tend to do better in later algebra, or conversely students who are not well prepared by the end of elementary school are likely to struggle with algebra. Because the measure for this study—the SBAC assessment—follows the CCSS in upper elementary grades by emphasizing the concepts, procedural fluencies, and applications of arithmetic (such as base-10 and fractions knowledge but also assessing complex problem solving, communicating reasoning, and the ability to use modeling to solve real-world problems), this result suggests a validation of the importance of arithmetic for algebra. The results for the MATHia component of the project indicate that intensive engagement with computer adaptive tutoring related to CCSS can significantly improve later Algebra I EOC performance (Question B, below).

The relation between earlier mathematics competencies and later algebra outcomes is largely the same across gender, ethnicity, race, disability status, English learner status, and economic disadvantage (Question C, below). Note that we did not compare groups in terms of absolute levels of performance on the earlier and later mathematics measures, as this was not the goal, but rather assessed whether the *strength of the relationship* between earlier mathematics competencies and later Algebra outcomes differed across groups; for the most part, it did not. With control of students' prior mathematics competencies, districts and schools are more similar than different in algebra outcomes, although case studies indicate that substantial improvements in outcomes can be achieved with wide-scale reforms (Question D, below).

QUESTION A

Implications for Better Preparing Students for Success in Algebra

The decomposition of SBAC algebra items into the subdomains of Quantitative Literacy, Generalized Arithmetic, Functional Relations, and Constraint Equations makes sense in terms of students' progression through algebraic material. However, there were not enough item responses to construct a Constraint Equations measure, and the reliabilities for Generalized Arithmetic (reliability = .48) and Functional Relations (reliability = .33) were significantly lower in comparison to Quantitative Literacy (reliability = .80) and overall Algebra (reliability = .86; see Table E2). The two latter measures were highly correlated (e.g., $r = .96$ in California) and thus essentially are measuring the same competencies. As a result, the best outcome was overall Algebra. The contrast outcome of Geometry and Statistics was not particularly reliable (reliability = .25) due to too few items; therefore, the associated results need to be interpreted with caution. Nevertheless, the outcome was retained to provide a non-algebra contrast.

In the prediction of later overall Algebra performance, modest effects in the IRT subscore analyses suggested that students who had relative advantages in Whole Numbers and Fractions performed better in later Algebra, controlling for overall mathematics competence and ELA scores. The former was uniquely related to later Algebra, whereas the latter was related to later Algebra and later Geometry and Statistics, but neither was related to later ELA scores. The results suggest that—above and beyond overall mathematics competence—mastery of conceptual knowledge and procedural skills in whole number arithmetic may contribute to later outcomes in Algebra, whereas mastery of conceptual knowledge and procedural skills in fractions may contribute to later mathematics more broadly. These patterns are consistent with recommendations of the National Mathematics Advisory Panel (2008), but the size of these effects was small, and it is not clear that they are practically significant.

Interpretations of the relations between earlier performance on the Problem Solving, Complex Problem Solving, and Reasoning & Communicating measures are not straightforward, because these measures are complex and go beyond the ability to use mathematical knowledge in a problem-solving context. Performance in problem-solving contexts, as typically measured by word problems, is influenced by reading and language comprehension (Fuchs et al., 2020), students' prior knowledge as related to the context of the problem (Thevenot, 2017), and domain-general abilities, such as working memory (Geary & Widaman, 1992) and visuospatial abilities (Casey et al., 1995).

The contrast of the relations between prealgebra problem solving and reasoning competencies and later Algebra and ELA scores will help to control for some of these confounds but will not likely control all of them (Bailey et al., 2014). Although the size of these effects was small, especially in comparison to the effect of overall mathematics competence, the current findings indicate that Complex Problem Solving with Fractions and Reasoning & Communicating may be more related to later Algebra than to later Geometry and Statistics or later ELA scores. The pattern suggests that the earlier ability to use fractions and other arithmetical knowledge in the context of complex problem solving (e.g., multistep word problems) and reasoning could be important for later performance in Algebra, but again the size of these effects was small, and it is not clear that they are practically significant.

Despite small effects for specific areas, such as fractions, there was a substantive and practically important relation between overall performance on the SBAC mathematics assessment at the end of elementary school and performance in algebra four years later. The assessment follows the CCSS in upper elementary grades by emphasizing the concepts, procedural fluencies, and applications of arithmetic. So when the results of this study point to the importance of quality preparation in upper elementary grades mathematics, that should be read as a validation of the importance of arithmetic for algebra. This is not to say that students should be denied algebra in middle grades contingent on this or that level of arithmetic competency, but rather to say that elementary educators should be aware of the stakes of what they teach for what comes next.

QUESTION B

The results from the MATHia component were promising and worthy of follow-up study. The core finding was that successful completion of algebraic workspaces, that is, the solving of multi-step problems related to CCSS, during the middle school years resulted in improved Algebra I EOC performance, controlling for prior mathematics competencies. More fine-grain analyses suggested that skill at translating word problems into algebra equations and an understanding of the different ways in which functions can be represented (e.g., equation, graphically) were uniquely related to later Algebra I EOC performance, controlling for prior mathematics competencies. There were also several frequently engaged CCSS topics, such as arithmetic with polynomials and basic work with functions, that did not predict later Algebra I EOC performance, and thus might be de-emphasized. Given the limitations of the sample and understanding of how the program was used, these latter results should be considered tentative and in need of replication.

The second core finding was that there was an interaction between prior mathematics competencies and successful engagement with MATHia. The interaction revealed that extensive and successful engagement with MATHia reduced the Algebra I EOC performance differences comparing students who had higher and lower scores on prior assessments. Follow-up studies with larger, more randomized samples are needed to confirm this interaction and, if possible, determine the key mathematical activities (e.g., practice with solving expressions or linear equations) that will have substantive and positive effects on the mathematical development of students who have been historically marginalized in educational systems, and whether these activities vary across gender, ethnicity, race, 504 status, English learner status, and economic disadvantage.

QUESTION C

The core question here was whether the relation between earlier mathematics competencies and later algebra outcomes varied across different groups; specifically, across gender, ethnicity, race, 504 status, English learner status, and economic disadvantage. We focused on the California data to address this question due to the large and very diverse sample afforded by these data.

The most straightforward way to assess this question is through an interaction between earlier overall mathematics competences, such as overall SBAC score in fifth grade, and group membership in the prediction of later algebra performance. For instance, the relation between earlier mathematics competencies and later algebra was strong (β = 0.637), and the interaction between earlier math and student gender was significant but very small (β = -0.002). This means that the strength of the relation between earlier math and later algebra was slighter smaller for girls than for boys, but the magnitude of this difference is of no practical importance. The same pattern emerged for all of the other group differences; some were statistically significant, due to the large sample size, but none of them is of practical significance.

The core finding here is that a strong mathematical foundation in elementary grade CCSS, which emphasizes the concepts, procedural fluencies, and applications of arithmetic, is critical to success in later algebra. This relation between earlier math and later algebra appears to be independent of student gender, ethnicity, race, disability status, English language proficiency, family income, or students' prior English language arts competencies. This is not to say that students in all demographic groups have had the same opportunities to acquire this early foundation or have done so, but rather that those students who have a solid foundation, independent of demographic group, are on track for success in later Algebra.

The results from the analyses of the MATHia data are largely consistent with the results for California. Specifically, all middle school students benefit from engagement with and mastery of CCSS workspaces in terms of later Algebra I EOC scores. These gains are similar across demographic groups, but there were significant differences across these groups in engagement with MATHia. Lower engagement with MATHia, or a similar algebra curriculum, will likely contribute to later differences in Algebra I EOC performance.

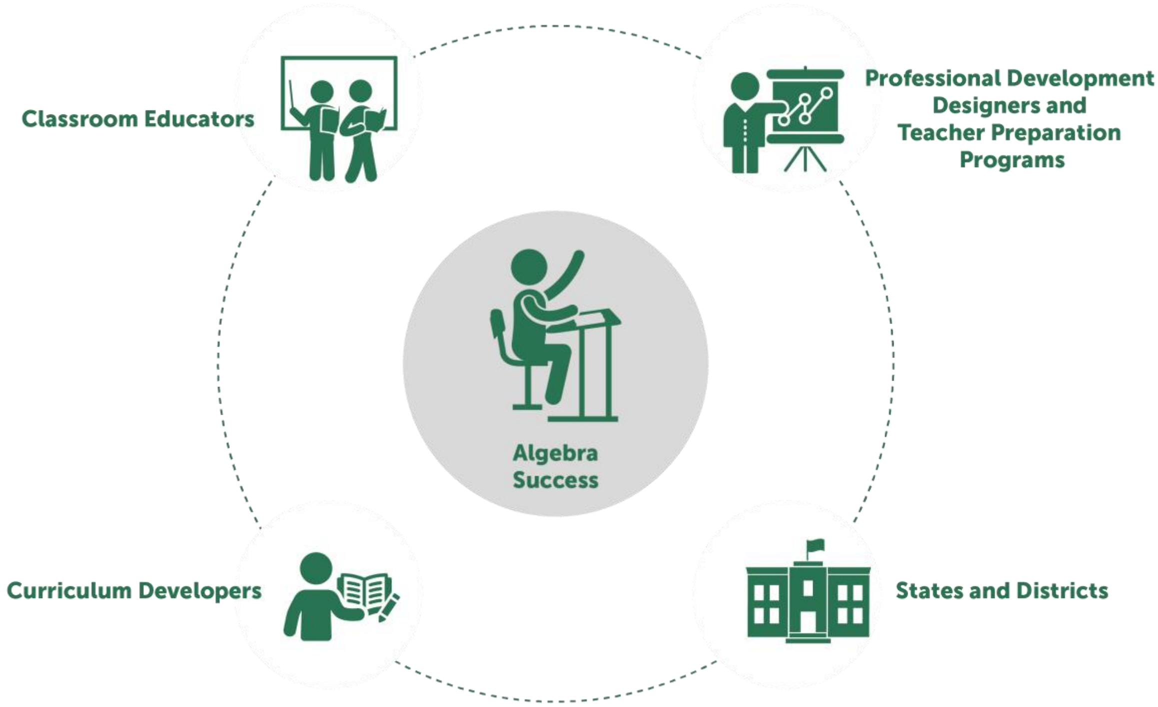
QUESTION D

The core question here was whether any of the above-described effects varied across districts, schools, or classrooms. Due to limitations of the data set, we were unable to assess classroom-level effects but were able to assess variation in algebra outcomes across districts and schools, controlling for students' prior mathematics competencies. Again, we focused on California due to the large and diverse sample. Preliminary results indicated that, once prior achievement was controlled, there was not substantive variation in algebra outcomes across districts or schools. This does not mean that there is not variation in algebra outcomes across districts and schools, but rather this variation is largely explained by the level of preparation of students entering the middle schools where the algebra outcomes were assessed.

In other words, most districts and schools are more similar than different once prior achievement is taken into account. This does not mean that districtwide or schoolwide reforms cannot substantively improve the mathematical development of their students, as was illustrated by Carver-Thomas and Podolsky's (2019) case studies of individual districts. Their analysis indicates that systematic and wide-scale (e.g., involving teacher training, higher expectations for student performance, rigorous standards) reforms can have substantive and positive effects on the mathematical development of students who have been historically marginalized in educational systems.

Carver-Thomas and Podolsky's (2019) case studies are consistent with the MATHia results for middle school students in Florida. In terms of later Algebra I EOC performance, students in nearly all schools gained from engagement with and mastery of MATHia workspaces. The gains were, however, more variable for schools with more than 90% students of color, with the largest and smallest gains emerging for these schools. The reasons for the variation in these outcomes are not clear but merits follow-up study.

Recommendations



The following recommendations weave together the findings of this study, research outside the study, and the discussions of the Research Advisory Committee.

S=study findings;
PR=prior research;
RAC=Research Advisory Committee discussions

Classroom-focused | Curriculum-focused | Professional learning focused | Future research

What should states and districts do to help every student leave elementary school with a solid foundation in elementary school CCSS and support middle school mathematics growth?

1. **Communicate to teachers, students, families and caregivers, and the community the importance of a strong mathematical foundation in elementary grades for later success in Algebra.** Since the findings in this report emphasize the progress that all students can make in middle-grade Algebra, messaging must not lead to middle-grade students being denied opportunities to learn pre-algebra and algebra on the basis of their opportunities in elementary grades. Thus, provide educators with resources and professional learning aimed at the goal of regularly engaging students in grade-level and challenging mathematics even in cases where the educational system has not provided them an adequate mathematical foundation (Balfanz, Mac Iver & Byrnes, 2006; Baker, Gersten & Lee, 2002; Burris, Heubert & Levin, 2006; Global Family Research Project, 2017; TNTP, 2021). **S, PR, RAC**

2. **Adopt an integrated, arithmetic-focused curriculum for the entire elementary grade span.** The curriculum should be coherently organized around key content threads (e.g., an understanding of numerical magnitude) that tie material together across grades in order to better prepare students for later success in Algebra. Structure adoption processes to ensure the curriculum is designed to explicitly support teachers to facilitate the learning of students who have been historically marginalized by ensuring their unique identities, culture and needs are honored (Brown, 2007; Howard-Hamilton, 2002; Santamaria, 2009). **PR, RAC**

3. **Support student transitions from elementary to middle school and middle school to high school by maintaining coherence of the K–12 mathematics learning pathway**. For example, adopt curricular materials that build coherently across the grades, and ensure that school-based staff understand the value of instructional coherence across the grades in their school and beyond (ACT, 2008; National Research Council, 1999). **PR, RAC**

4. **Provide professional learning opportunities that help teachers develop their own strong mathematical identity and a solid understanding of the key mathematical threads of their curricular programs**. (Feiman-Nemser, 2001; Gallagher, 2016; Schoenfeld, 2014). **PR, RAC**

5. **Consider providing students with supplemental grade-level practice for content with significant evidential support for improving Algebra performance** (e.g., fractions in upper elementary grades, Siegler et al., 2012; Algebra by Example during Algebra 1, Booth *et al*., 2015)**.** The supplementals should support and coherently reinforce the tier 1 instructional materials (Gersten, Beckmann, Clarke, Foegen, Marsh, Star & Witzel, 2009). **PR**

What should curriculum developers do to support states, districts, teachers, and students to succeed in teaching and learning a coherent mathematics curriculum?

6. **Invest in designing materials and explicit support for teachers in order to focus on students who have been historically marginalized by ensuring their unique identities, culture, and needs are honored**. For example, consider how the curricular materials cultivate or become a barrier to cultivation of positive mathematical identities for students who are Black or Latino, and how they engage students learning English. This is in service to the core goal of students acquiring a solid understanding of whole number and rational number arithmetic during the elementary school years and prealgebra and early algebra in the middle school years (Leonard, Knapp & Adeleke, 2009; Peoples, Islam & Davis, 2021; Ukpokodu, 2011). **PR, RAC**

7. **Design curricular materials and programs—including supplementals—in the elementary school grades that emphasize the concepts, procedural**

fluencies, and applications of arithmetic as well as CCSS's practice standards for complex problem solving, communicating reasoning, and the ability to use modeling to solve real-world problems. Curricular materials and programs should be coherently organized around key content threads and tie material together across grades. Specifically, these key threads would include number sense, that is, a developing understanding of numerical magnitudes (including fractions and later rational numbers) and the arithmetical operations that can be applied to them (Siegler & Braithwaite, 2017). Success at using this knowledge to better understand mathematical relationships as well as to apply it to problem-solving contexts, as in word problems, is a critical component of early mathematics education and preparation for later algebra. **S, RAC**

8. **Attend to the content and coherence of the curricula materials, but also their mathematical fidelity and the quality of the mathematical tasks with which students are asked to engage.** For instance, in pre-CCSS textbooks in the United States, arithmetic problems were typically presented overwhelmingly in a result-unknown format as $a + b = ?$ (e.g., $4 + 3 = ?$), an approach that results in many students inferring that the '=' sign means to operate on the numbers to the left rather than indicating the equality of the quantities to the left and right of it (McNeil et al., 2006). Textbooks must follow CCSS in this area (see, e.g., 1.OA.D.7) by coherently integrating the forms of number relationships (e.g., $c = b + a$) that express number decompositions and facilitate students' understanding of the '=' as a relational construct (McNeil et al., 2011). **PR, RAC**

What should designers of professional development and teacher preparation programs do to support teachers in helping historically marginalized students succeed in learning a solid foundation in elementary mathematics and have success in Algebra?

9. **Design professional learning which helps teachers develop their own strong mathematical identity in order to positively impact their teaching of mathematics** (Ball & Forzani, 2011; Thompson, 1992; Wei, Darling-Hammond, Andree, Richardson & Orphanos, 2009). **PR, RAC**

10. **Design professional learning to support K–5 teachers to develop a solid understanding of the key threads of their curricular programs, specifically how knowledge at earlier grades provides the foundation for later learning and is not only a steppingstone but also is conceptually related to later material.** Professional learning should help teachers support students' unfinished learning by building on their understandings and assets to access the topic at hand, as opposed to re-teaching prior-grades material. Professional learning should also help teachers build their knowledge and ability to navigate decisions about when and how to modify the curriculum to make it stronger and more relevant for students, and not make changes that unravel the coherence and priorities of a strong curriculum (as described above). For example, improvising and skipping tasks in number and operations can cause incoherence in the curriculum leading to algebra. Teachers who are well equipped understand that working through these tasks provides additional practice on basic skills.

(Darling-Hammond, Hyler & Gardner, 2017; National Mathematics Advisory Panel, 2008). **PR, RAC**

11. **Require preservice teachers to take one or more courses aimed at helping teachers develop a solid mathematical understanding of fundamental mathematical concepts and the conceptual connections.** (Conference Board of the Mathematical Sciences, 2012; National Research Council, 2001). **PR, RAC**

What should researchers do to support students, teachers, states, districts, curriculum developers and designers of professional development and teacher preparation programs?

12. **Explore in greater detail the core components of a strong elementary school mathematics foundation through a modified replication of this study**, using a non-computer-adaptive assessment with sufficient items in key predictor content from other states with different geographic and student demographic profiles; the study may require the inclusion of additional assessment items to assess core math areas (e.g., fractions). Such a study would benefit from close partnership with state departments of education and districts. One goal of the latter should be to better understand aspects of curricular implementation and other contextual factors for mathematics success, such as student experience. **S**

13. **Follow up on the promising middle school MATHia results to understand the extent to which engagement with computer adaptive tutoring during the middle school years results in gains in later Algebra I performance.** A follow up study with larger samples as well as a greater understanding of the usage and student experience would help to verify these findings and enable a more fine-grained assessment of how computer adaptive tutoring supports mathematical development and identity, and if there are experience and usage differences across students who are Black, Latino, English learner-designated, experiencing poverty, and/or female. Of course, strong causal conclusions will also have to await randomized controlled trials. **S**

14. **Look inside upper elementary classrooms to learn about key curricular and instructional factors that make a difference for students who are Black, Latino, English learner-designated, experiencing poverty, and/or female and who are successful in upper elementary grades mathematics**. Studying the practices of teachers of students who are Black, Latino, English learner-designated, experiencing poverty, and/or female and are succeeding in learning upper elementary mathematics can help identify key factors that can support students on the path for future success in Algebra. Observations and interviews of students and teachers would document student-teacher relationships, classroom or school environment, curriculum implementation, and instructional practices that contribute to student success. **RAC**

15. **Examine how targeted professional learning for upper elementary teachers may affect the performance of students**. Using research-backed practices, design, deliver and evaluate the effectiveness of professional learning which integrates Mathematics Practice 1 (make sense of problems and persevere in

solving them), fractions, supporting student identity, and instruction for equity. A study such as this would require the active involvement of district staff, school leaders, teachers, caregivers, and students, including but not limited to interviews and surveys, observations, student work analysis, and shared interpretation of findings. **RAC**

References

ACT. (2008). *The forgotten middle: Ensuring that all students are on target for college and career readiness before high school.* Iowa City, IA: ACT.

Altman, N. S. (1992). An introduction to kernel and nearest-neighbor nonparametric regression. *The American Statistician, 46*, 175–185. https://doi.org/10.1080/00031305.1992.10475879

Anderson, J. R., Bothell, D., Lebiere, C., & Matessa, M. (1998). An integrated theory of list memory. *Journal of Memory and Language*, *38*, 341–380. https://doi.org/10.1006/jmla.1997.2553

Anderson, J. R., Corbett, A. T., Koedinger, K. R., & Pelletier, R. (1995). Cognitive tutors: Lessons learned. *The Journal of the Learning Sciences*, *4*, 167–207.

Bailey, D. H., Watts, T. W., Littlefield, A. K., & Geary, D. C. (2014). State and trait effects on individual differences in children's mathematical development. *Psychological Science, 25*, 2017–2026. https://doi.org/10.1177/0956797614547539

Baker, S., Gersten, R., & Lee, D.-S. (2002). A synthesis of empirical research on teaching mathematics to low-achieving students. *The Elementary School Journal, 103(1),* 51–73.

Balfanz, R., Mac Iver, D. J., & Byrnes, V. (2006). The implementation and impact of evidence based mathematics reforms in high poverty middle schools: Multi-school, multi-year studies. *Journal for Research in Mathematics Education, 37,* 33–64.

Ball, D. L., & Forzani, F. M. (2011). Building a Common Core for learning to teach: And connecting professional learning to practice. *American Educator, 35(2),* 17-39.

Birnbaum, A. (1968). Some latent trait models and their use in inferring an examinee's ability. In F. M. Lord & M. R. Novick (Eds.), Statistical theories of mental test scores (pp. 397–479). Reading, MA: Addison-Wesley.

Bell, R. M., & McCaffrey, D. F. (2002). Bias reduction in standard errors for linear regression with multi-stage samples. *Survey Methodology*, *28*, 169–182.

Brown, M. R. (2007). Educating All Students: Creating Culturally Responsive Teachers, Classrooms, and Schools. I*ntervention in School and Clinic*, *43*(1), 57–62. https://doi.org/10.1177/10534512070430010801

Burris, C. C., Heubert, J. P. & Levin, H. M. (2006). Accelerating Mathematics Achievement Using Heterogeneous Grouping. *American Educational Research Journal, 43*, 105-136.

Carver-Thomas, D., & Podolsky, A. (2019). *Long Beach Unified School District: Positive outliers case study*. Palo Alto, CA: Learning Policy Institute.

Casey, M. B., Nuttall, R., Pezaris, E., & Benbow, C. P. (1995). The influence of spatial ability on gender differences in mathematics college entrance test scores across diverse samples. *Developmental Psychology*, *31*, 697–705. https://doi.org/10.1037/0012-1649.31.4.697

Common Core State Standards (CCSS) for Mathematics. http://www.corestandards.org/Math/

Conference Board of the Mathematical Sciences. (2012). *The mathematical education of teachers.* Issues in mathematics education (Volume 17). Providence, RI: American Mathematical Society.

Corbett, A. T., & Anderson, J. R. (1995). Knowledge tracing: Modeling the acquisition of procedural knowledge. *User Modeling and User-Adapted Interaction, 4*, 253–278. https://doi.org/10.1007/BF01099821

Darling-Hammond, L., Hyler, M. E. & Gardner, M. (2017). *Effective Teacher Professional Development.* Palo Alto, CA: Learning Policy Institute.

Daro, P., Hughes, G. B., & Stancavage, F. (2015). *Study of the Alignment of the 2015 NAEP Mathematics Items at Grades 4 and 8 to the Common Core State Standards (CCSS) for Mathematics*. San Mateo, CA: American Institutes for Research. Study of the Alignment of the 2015 NAEP Mathematics Items at Grades 4 and 8 to the Common Core State Standards (CCSS) for Mathematics (air.org)

Feiman-Nemser, S. (2001). From preparation to practice: Designing a continuum to strengthen and sustain teaching. *Teachers College Record, 103,* 1013-1055.

Fuchs, L. S., Fuchs, D., Seethaler, P. M., & Craddock, C. (2020). Improving language comprehension to enhance word-problem solving. *Reading & Writing Quarterly*, *36*, 142–156. Improving Language Comprehension to Enhance Word-Problem Solving: Reading & Writing Quarterly: Vol 36, No 2 (tandfonline.com)

Gallagher, A. (2016). Professional development to support instructional improvement: Lessons from research (working paper). Menlo Park, CA: SRI International.

Geary, D. C., Nicholas, A., Li, Y., & Sun, J. (2017). Developmental change in the influence of domain-general abilities and domain-specific knowledge on mathematics achievement: An eight-year longitudinal study. *Journal of Educational Psychology*, *109*, 680–693. https://doi.org/10.1037/edu0000159

Geary, D. C., & Widaman, K. F. (1992). Numerical cognition: On the convergence of componential and psychometric models. *Intelligence, 16*, 47–80. https://doi.org/10.1016/0160-2896(92)90025-M

Gersten, R., Beckmann, S., Clarke, B., Foegen, A., Marsh, L., Star, J. R., & Witzel, B. (2009). *Assisting students struggling with mathematics: Response to Intervention for elementary and middle schools* (NCEE 2009-4060). Washington, DC: National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education. Retrieved from http://ies.ed.gov/ncee/wwc/publications/practiceguides/.

Global Family Research Project. (2017). *Formula for Success: Engaging Families in Early Math Learning.* Cambridge, MA: Harvard Graduate School of Education. Retrieved from https://globalfrp.org/content/download/83/561/file/Early+Math+FINE.pdf.

Honaker J., King G., & Blackwell M. (2011). Amelia II: A program for missing data. *Journal of Statistical Software*, *45*, 1–47. http://www.jstatsoft.org/v45/i07/

Howard-Hamilton, M. F. (2002). Creating a culturally responsive learning environment for African American students. doi:10.1002/tl.8205.

Hughes, G., Behuniak, P., Norton, S., Kitmitto, S., & Buckley, J. (2019). *NAEP Validity Studies Panel Responses to the Reanalysis of TUDA Mathematics Scores*. San Mateo, CA: American Institutes for Research. https://eric.ed.gov/?id=ED599642

Hurst, M. A., & Cordes, S. (2018). Children's understanding of fraction and decimal symbols and the notation-specific relation to pre-algebra ability. *Journal of Experimental Child Psychology*, *168*, 32–48. https://doi.org/10.1016/j.jecp.2017.12.003

Lee, K., & Bull, R. (2016). Developmental changes in working memory, updating, and math achievement. *Journal of Educational Psychology, 108*(6), 869–882. https://doi.org/10.1037/edu0000090

Leonard, J., Napp, C., & Adeleke, S. (2009). The Complexities of Culturally Relevant Pedagogy: A Case Study of Two Secondary Mathematics Teachers and Their ESOL Students. *High School Journal, 93(1),* 3–22.Mathematics: Focus by Grade Level. https://achievethecore.org/category/774/mathematics-focus-by-grade-level

McNeil, N. M., Fyfe, E. R., Petersen, L. A., Dunwiddie, A. E., & Brletic-Shipley, H. (2011). Benefits of practicing 4= 2+ 2: Nontraditional problem formats facilitate children's understanding of mathematical equivalence. *Child Development, 82*, 1620–1633. https://doi.org/10.1111/j.1467-8624.2011.01622.x

McNeil, N. M., Grandau, L., Knuth, E. J., Alibali, M. W., Stephens, A. C., Hattikudur, S., & Krill, D. E. (2006). Middle-school students' understanding of the equal sign: The books they read can't help. *Cognition and instruction*, *24*(3), 367–385.

Mou, Y., Li, Y., Hoard, M. K., Nugent, L., Chu, F., Rouder, J., & Geary, D. C. (2016). Developmental foundations of children's fraction magnitude knowledge. *Cognitive Development, 39*, 141–153. https://doi.org/10.1016/j.cogdev.2016.05.002

Muraki, E. (1992). A generalized partial credit model: Application of an EM algorithm. *Applied Psychological Measurement, 16*, 159–176.

National Research Council. (1999). *Designing Mathematics or Science Curriculum Programs: A Guide for Using Mathematics and Science Education Standards.* Washington, DC: The National Academies Press. https://doi.org/10.17226/9658.

National Research Council. 2009. Mathematics Learning in Early Childhood: Paths Toward Excellence and Equity. Washington, DC: The National Academies Press. https://doi.org/10.17226/12519.

National Research Council. (2001). *Adding It Up: Helping Children Learn Mathematics.* Washington, DC: The National Academies Press. https://doi.org/10.17226/9822.

National Mathematics Advisory Panel. (2008). *Foundations for Success: Final Report of the National Mathematics Advisory Panel*. Washington, DC: United States Department of Education. http://www.ed.gov/about/bdscomm/list/mathpanel/report/final-report.pdf

National Assessment of Educational Progress (NAEP). https://nces.ed.gov/nationsreportcard/mathematics/achieve.aspx#grade8

Pane, J. F., Griffin, B. A., McCaffrey, D. F., & Karam, R. (2014). Effectiveness of cognitive tutor algebra I at scale. *Educational Evaluation and Policy Analysis, 36*, 127–144. https://doi.org/10.3102/0162373713507480

Peoples, L.Q., Islam, T., & Davis, T. (2021). The culturally responsive-sustaining STEAM curriculum scorecard. New York: Metropolitan Center for Research on Equity and the Transformation of Schools, New York University.

Python Software Foundation. (2020). *Python Language Reference*, version 3.7. Available at http://www.python.org

R Core Team. (2017). *R: A language and environment for statistical computing*. Retrieved from https://www.R-project.org/

Race to the Top Grant. https://portal.smarterbalanced.org/library/en/v1.0/end-of-grant-report.pdf

Ritter, S., Anderson, J. R., Koedinger, K. R., & Corbett, A. (2007). Cognitive Tutor: Applied research in mathematics education. *Psychonomic Bulletin & Review, 14*, 249–255. https://doi.org/10.3758/BF03194060

Santamaría, L.J. (2009). Culturally Responsive Differentiated Instruction: Narrowing Gaps between Best Pedagogical Practices Benefiting All Learners. *Teachers College Record, 111*, 214-247.

Schoenfeld, A. H. (2014). What Makes for Powerful Classrooms, and How Can We Support Teachers in Creating Them? A Story of Research and Practice, Productively Intertwined. *Educational Researcher, 43*, 404-412.

Schmidt, W. H., & Houang, R. T. (2012). Curricular coherence and the common core state standards for mathematics. *Educational Researcher*, *41*, 294–308. https://doi.org/10.3102/0013189X12464517

Siegler, R. S., & Braithwaite, D. W. (2017). Numerical development. *Annual Review of Psychology*, *68*, 187–213. https://doi.org/10.1146/annurev-psych-010416-044101

Siegler, R. S., Duncan, G. J., Davis-Kean, P. E., Duckworth, K., Claessens, A., Engel, M., Susperreguy, M. I. & Chen, M. (2012). Early predictors of high school mathematics achievement. *Psychological Science*, *23*, 691–697. https://doi.org/10.1177/0956797612440101

Smarter Balanced Computer Adaptive Test. https://portal.smarterbalanced.org/library/en/strategic-plan.pdf

Smarter Balanced Content Specifications. https://contentexplorer.smarterbalanced.org/test-development#content

Stein, M. K., Kaufman, J. H., Sherman, M., & Hillen, A. F. (2011). Algebra: A challenge at the crossroads of policy and practice. *Review of Educational Research*, *81*, 453–492. https://doi.org/10.3102/0034654311423025

Thevenot, C. (2017). Arithmetic word problem solving: The role of prior knowledge. In D. C. Geary, D. B. Berch, R. Ochsendorf, & K. Man Koepke (Eds), *Acquisition of complex arithmetic skills and higher-order mathematics concepts* (pp. 47–66). San Diego: Elsevier Academic Press.

Thompson, A. G. (1992). Teachers' beliefs and conceptions: A synthesis of the research. In D. A. Grouws (Ed.), *Handbook of research on mathematics teaching and learning: A project of the National Council of Teachers of Mathematics* (p. 127–146). Macmillan Publishing Co, Inc.

TNTP. (2021). *Accelerate, don't remediate: New evidence from elementary math classrooms.* New York: TNTP, Inc. Retrieved from https://tntp.org/assets/documents/TNTP_Accelerate_Dont_Remediate_FINAL.pdf.

Ukpokodu, O. N. (2011). How Do I Teach Mathematics in a Culturally Responsive Way?: Identifying Empowering Teaching Practices. *Multicultural Education, 19(3),* 47-56.

von Davier, M. (2017). *mdltm: Software for the general diagnostic model and for estimating mixtures of multidimensional discrete latent traits models* [Computer software]. Princeton, NJ: ETS.

Weeks, J. & Baron, P. (2021). *Exploring math education relations by analyzing large data sets II*. [ETS-RM-21-02]. Princeton, NJ: ETS.

Wei, R. C., Darling-Hammond, L., Andree, A., Richardson, N., Orphanos, S. (2009). Professional learning in the learning profession: A status report on teacher development in the United States and abroad. Dallas, TX: National Staff Development Council.

Zimba, J. (2014). The development and design of the Common Core State Standards for Mathematics. *New England Journal of Public Policy, 26*(1), 10.

Appendices

Appendix A: The Development of the Smarter Balanced Assessment System

2010–2014

The Smarter Balanced Assessment System was developed by a consortium of states in response to a federal Race To The Top Grant to develop a summative assessment that would assess students' learning and their readiness for college and careers after graduation from high school. The consortium of states developed the application for funds and was awarded the four-year grant to develop the summative assessment system aligned to the Common Core State Standards (CCSS). The Consortium was made up of two groups of states that joined together around the goals of 1) technology innovation (Smarter) to improve the accessibility and opportunity for students to demonstrate their knowledge in a variety of ways and 2) commitment to a balanced assessment system (Balanced) with the formative assessment process, interim, and summative assessments as three legs of the stool. This balanced system supports students and teachers in understanding what has been learned and what needs to continue to be improved for students to perform at a level that demonstrates college and career readiness.

The federal grant provided resources and guidance that encouraged innovation in the development of state summative assessments. Smarter Balanced incorporated evidence-based design to ground the assessment in a tight alignment to the implementation of the CCSS in the classroom. The assessment is built around Claims (what knowledge can students demonstrate) and Targets (the evidence students can provide to show they have a deep understanding). Related CCSS standards are grouped within Targets, which are organized under Claims described in the Smarter Balanced Content Specifications. Educators participated in an extensive review and feedback process to finalize the Content Specifications. Once the Consortium adopted the Content Specifications in 2012, they became the foundation for the development of item specifications, rubrics, and sample items to guide item writing. Updates to the Content Specifications continued throughout the test development, with the latest update in 2015.

A foundational theory of action was to include educators at all steps along the way to ensure their feedback impacted the item development process. Educators from participating states developed guidelines for item writing based on best practices for classroom instruction. The technology available through online testing opened up the opportunity for all students, including students with special needs, to respond to different types of items. All items were reviewed by teams of practicing teachers who compared the evidence expected by the item writer with their own knowledge of student competencies. There were strict guidelines for readability, accessibility, and language density that made the inclusion of all students a priority at the item development stage. Whole groups of students traditionally excluded from state testing were included through accessibility tools and supports. The Consortium engaged in cognitive labs in 2012–2013 to answer key questions about how the students would respond to different types of items and be able to use technology tools. The first field test of items was held in 2013–2014. The assessment went operational during the 2014–2015 school year.

Another innovation in the operational summative assessment was the use of computer adaptive technology. This approach to assessment had not been widely used in K–12 state assessments previously. The advantage of the computer adaptive technology was the ability to elicit sufficient evidence to evaluate student learning with fewer test questions and still cover the depth and breadth of the CCSS. The Consortium developed the blueprint for the assessment over two years as they engaged the member states in discussions of the benefits and constraints of a computer adaptive assessment for large-scale state assessment. The blueprint was finalized with two parts to the summative assessment: a computer adaptive test (CAT) and a performance task. The two tests are combined to make a complete test for a student with evidence from both tests contributing to the claim scores and the overall score. The result is a "constrained CAT" where at least 60% of the assessment uses grade-level items. The CAT algorithm may pull off-grade items (up to 2 grades) if the performance of students at the very high end or low end of the performance scale has not been adequately measured. The performance task is written at grade level with multiple items that build toward the completion of a task. The blueprint specifies the percentage of items in each Claim that must be at Depth of Knowledge (DOK) Levels 2 and 3 to ensure the student provides evidence of higher order thinking and deep understanding necessary for college and career readiness through applications and interpretations.

Appendix B: State Data Requests and Data Agreements

The Exploring Math Education Relationships by Analyzing Large Data Sets (EMERALDS) Phase II project analyzed longitudinally linked student-level responses to state summative assessment items from 2015–16, 2016–17, 2017–18, and 2018–19 to determine if student knowledge in earlier grades predicted success in high school algebra. Although states have developed policies and procedures to facilitate research aligned with state goals to improve student learning, analysis of individual student responses is novel and has great potential to inform basic research on students' mathematical development and to inform statewide educational policies and mathematics standards.

Approaching States in the Smarter Balanced Assessment Consortium

Student Achievement Partners (SAP) approached the Smarter Balanced Assessment Consortium (SBAC) with the proposed research in April of 2019. The SBAC leadership staff and an SAP representative discussed the research proposal with the representatives from each member state. In June 2019, an interest survey was sent out to the member states to volunteer for the research study. Seven states—California, Connecticut, Idaho, Nevada, Oregon, South Dakota, and Washington—indicated an interest in participating. Accordingly, EMERALDS project staff submitted data requests through the state websites or by direct communication with the state agency representative in all seven states in July 2019. Once the data requests were received by the designated state agency staff, each state followed a data request review process to determine whether the data request would meet the minimum state criteria (e.g., to maintain student confidentiality).

Once the applications were accepted, they were sent to Data Request Review Committees in three of the states (Nevada, Oregon, and Washington) while four other states (California, Connecticut, Idaho, and South Dakota) completed the review by state agency staff. Of the seven original states, two (Nevada and Oregon) declined to participate due to concerns about releasing student-level responses; two (Connecticut and South Dakota) declined due to concerns with the complications and time required to complete the state contracting process, and three states (Idaho, Washington, and California) agreed to move the request to the next level. At this stage, the negotiation of the data-sharing agreements took place between the states and the EMERALDS data analysis partner, Carnegie Mellon University (CMU).

Data Agreements with Idaho, Washington, and California

The established data-agreement templates went through multiple revisions by state agencies and CMU to meet state and university requirements. The first agreement was executed between Idaho and CMU on 1/10/20. It followed the most straightforward process, managed by the Idaho State Director of Assessment. The data-agreement template provided by CMU was adjusted by the state agency to meet the state laws related to data privacy and the acceptable uses of student assessment data for research to benefit the state. The agreement language was approved by several departments in the state agency and CMU. The negotiation and signature process between CMU and Idaho took 161 days from 7/20/19–1/10/20.

Washington's data request and data-agreement process required the data analysis partner to agree to the state data agreement prior to submitting the request. For this reason, the process started with CMU on 8/7/19 with the evaluation of the state data agreement prior to finalizing the data request. CMU requested modifications in the data agreement that Washington state legal staff countered with proposed alternative language. Once the staff in both agencies had confidence the issues could be resolved, the state agency in Washington advised the EMERALDS staff to submit the data request to be reviewed by the Data Request Committee at their November 2019 meeting. The request was submitted and approved on 11/22/19 with the requirement that a more detailed data agreement be executed with CMU. The data-agreement negotiations began again as of 12/15/19 with several revisions on both sides. The agreement was executed on 1/5/21.

The data agreement with California experienced the most complicated path. State agency staff were concerned that the long data-agreement process would impede the progress of the research such that the findings would not be available in time to inform the California Mathematics Framework

Committee during the summer of 2020. The first approach was to seek approval from SBAC to act as an agent of the state to share the data. By October 2019, they determined the approach would be unsuccessful due to the limited scope of the existing agreements between SBAC and the state agency. State agency staff began the traditional data-agreement process with CMU, requiring the university to accept the terms of the CA data agreement “as is” and be approved for an IRB with the California Committee for Protection of Human Subjects (CPHS), also known as the “State IRB.” CMU proposed several revisions to the state agency, which were not accepted by the California legal staff.

In January 2020, EMERALDS started a parallel process facilitated by California state agency staff to discuss the research project with the state assessment contractor, Educational Testing Service (ETS). As the state's testing contractor, ETS stores California state assessment data and thus no data transfer was necessary. ETS was still required to complete the data request and data-agreement process including approval for the IRB. However, since ETS had already accepted the California data-agreement terms, there were no apparent impediments to executing the agreement.

Appendix C: Matching Smarter Balanced Correlates to Item Metadata

What kinds of problems do students need to be successful on in earlier grades (Table C1) to be successful in algebra (Table C2)?

To identify appropriate items, the item metadata was filtered by Grade + Claim + CCSS + Target Model based on the metadata criteria on the table, *Matching Smarter Balanced Correlates to Item Metadata.* The corresponding item categories were used to create IRT residual and IRT subscore predictor and outcome variables.

Table C1: CCSS and Metadata Used to Create Predictor Variables

Correlates of Interest Description	Common Core State Standards	Metadata Combinations Matching Correlates of Interest
Predictor: Number Sense		
A1a - Number sense of whole numbers	3.NBT.A.1 4.NBT.A.1* 4.NBT.A.2 4.NBT.A.3	1: Claim 1 & Standard 3.NBT.A.1 2: Claim 1 & Standard 4.NBT.A.1 & TM 5a 3: Claim 1 & Standard 4.NBT.A.2 4: Claim 1 & Standard 4.NBT.A.3
A1b - Number sense of fractions	3.NF.A* 4.NF.A* 5.NF.B.5	1: Claim 1 & Standard 3.NF.A.2 2: Claim 1 & Standard 3.NF.A.3c 3: Claim 1 & Standard 3.NF.A.3d 4: Claim 1 & Standard 4.NF.A.1 & TM 1a 5: Claim 1 & Standard 4.NF.A.1 & TM 1b 6: Claim 1 & Standard 4.NF.A.1 & TM 1d 7: Claim 1 & Standard 4.NF.A.1 & TM 2b 8: Claim 1 & Standard 4.NF.A.2 9: Claim 1 & Standard 5.NF.B.5a 10: Claim 1 & Standard 5.NF.B.5b
A1c - Number sense of decimals and understanding of the place value system	4.NF.C.6 4.NF.C.7 5.NBT.1* 5.NBT.2 5.NBT.3a 5 NBT 3b 5.NBT.4	1: Claim 1 & Standard 4.NF.C.6 2: Claim 1 & Standard 4.NF.C.7 3: Claim 1 & Standard 5.NBT.A.2 4: Claim 1 & Standard 5.NBT.A.3a 5: Claim 1 & Standard 5.NBT.A.3b 6: Claim 1 & Standard 5.NBT.A.4
Predictor: Problem Solving		
A2 - Elementary Modeling: one step problems that correspond to those in tables 1 and 2, pp. 88–89, CCSS		

Correlates of Interest Description	Common Core State Standards	Metadata Combinations Matching Correlates of Interest
A2a - One-step word problems involving only whole numbers	3.OA.A.3*@ 3.MD.A.1@ 3.MD.A.2@ 3.MD.B.3@ 3.MD.C.7b@ 4.OA.A.2*@ 4.MD.A.2*@ 4.MD.A.3@ 5.MD.C.5b@ {any 5.NBT that match the language of A2a}	1: Claim 1 & Standard 3.OA.A.3 2: Claim 1 & Standard 3.MD.A.1 & TM 2 3: Claim 1 & Standard 3.MD.A.2 4: Claim 1 & Standard 3.MD.B.3 & TM 2 5: Claim 1 & Standard 4.OA.A.2 & TM 1a 6: Claim 1 & Standard 4.OA.A.2 & TM 1b 7: Claim 1 & Standard 4.OA.A.2 & TM 1c 8: Claim 1 & Standard 4.MD.A.3 & TM 4a 9: Claim 1 & Standard 5.MD.C.5b & TM 2b 10: Claim 1 & Standard 5.MD.C.5b & TM 2e
A2b - One-step word problems involving fractions	4.NF.B.3d*@ 4.NF.B.4c*@ 4.MD.A.2*@ 4.MD.A.3@ 5.NF.B.3*@ 5.NF.B.4 5.NF.B.6*@ 5.NF.B.7c*@ 5.MD.B.2@ 6.NS.A.1*@	1: Claim 1 & Standard 4.NF.B.3 & TM 3a 2: Claim 1 & Standard 4.NF.B.3 & TM 3b 3: Claim 1 & Standard 4.NF.B.3 & TM 3c 4: Claim 1 & Standard 4.NF.B.3 & TM 3d 5: Claim 1 & Standard 4.NF.B.4 & TM 6a 6: Claim 1 & Standard 4.NF.B.4 & TM 6b 7: Claim 1 & Standard 5.NF.B.3 & TM 1b 8: Claim 1 & Standard 5.NF.B.3 & TM 2 9: Claim 1 & Standard 5.NF.B.4 & TM 4b 10: Claim 1 & Standard 5.NF.B.6 11: Claim 1 & Standard 5.NF.B.7 12: Claim 1 & Standard 5.MD.B.2 & TM2 13: Claim 1 & Standard 6.NS.A.1 &TM 2b 14: Claim 1 & Standard 6.NS.A.1 &TM 2c
Predictor: Reasoning & Communicating		
A3a - Multi-step or higher complexity word problems with no fractions or decimals	3.OA.D.8* 3.MD.A.1@ 3.MD.B.3@ 3.MD.C.7b@ 3.MD.C.7d@ 3.MD.D.8@ 4.OA.A.2*@ 4.OA.A.3* 4.MD.A.2*@ 4.MD.A.3@ 5.MD.C.5b@ 5.MD.C.5c@ 5.G.A.2@	Claim 2 & any standard listed to the left (omit 4.MD.A.3 and 5.G.A.2) Claim 4 & any standard listed to the left
A3b - Multi-step or higher complexity word problems with fractions or decimals	4.NF.B.3d*@ 4.NF.B.4c*@ 4.MD.A.2*@ 4.MD.A.3@ 5.NF.B.3*@ 5.NF.B.6*@ 5.NF.B.7c*@ 5.MD.1@ 5.MD.2@ 5.G.A.2@	1:Claim 2 & Standard 5.NF.B.3*@, 5.NF.B.6*@, 5.NF.B.7c*@ 2: Claim 2 & Standard 4.NF.B.3d*@, 4.NF.B.4c*@ 3: Claim 4 & Domain NF 4: Claim 4 & Standard 5.MD.B.2 5: Claim 4 & Standard 6.NS.A.1

Correlates of Interest Description	Common Core State Standards	Metadata Combinations Matching Correlates of Interest
	{any 5.NBT that match the language of A3b} 6.NS.A.1*@	
A4 - Reasoning & Communication Problems where students can clearly and precisely construct viable arguments to support their own reasoning and to critique the reasoning of others	3.OA.B 3.NF.A 3.NF.1 3.NF.2 3.NF.3 3.MD.A 3.MD.7 4.OA.3 4.NBT.A 4.NBT.5 4.NBT.6 4.NF.A 4.NF.1 4.NF.2 4.NF.3a 4.NF.3b 4.NF.3c 4.NF.4a 4.NF4b 4.NF.C 4.NF.7 5.NBT.2 5.NBT.7 5.NF.1 5.NF.2 5.NF.B 5.NF.3 5.NF.4 5.NF.7a 5.NF.7b 5.MD.5a 5.MD.5b	1: Claim 3 & any standard listed to the left.
Contrast Variable Least likely to predict later outcomes in Algebra	Task models in OA & NBT that are strictly procedural; geometry of shapes (non-formula based); MD items that only have	1: 3.G.A.1 2: 3.G.A.2 3: 4.G.A.1 4: 4.G.A.2 5: 4.G.A.3 6: 4.MD.C.6 7: 4.MD.C.7 8: 5.G.B.3 9: 5.G.B.4 10: 5.MD.C.3

Correlates of Interest Description	Common Core State Standards	Metadata Combinations Matching Correlates of Interest
	students produce a graph.	11: 5.MD.C.4

* A central standard or CCSS organizer for the correlate in question
@ A "bucket" containing tasks that match the correlate as well as tasks that don't match.
For codes with @ symbol attached, the principle for identifying tasks that match the correlate is to apply the language of the correlate as a filter. For example, the language of correlate A3a ("Multi-step or higher complexity word problems with no fractions or decimals") implies disregarding tasks in 5.G.2 if they are one-step problems or if they contain fractions or decimals. Note that the term *task* here refers not only to math problems, but also more generally to any student encounter with mathematics that generates the desired data.
HCII mapped standards codes in the data to the standards codes in Column A and Column B using both exact match and fuzzy match logic (to handle issues of data format—for example, codes of 3.MD.A vs. 3.MD.A.1). Limitations on available item metadata resulted in corresponding limitations on the coding process.

Table C2: CCSS and Metadata Used to Create Outcome Variables

Correlate of Interest	Common Core State Standards	Metadata Combinations Matching Correlates of Interest
B1 - Quantitative literacy (without use of variables) connecting magnitude sense to numbers and operations). Identify and express relationships among quantities. Represent quantities graphically. [Here we restricted the selection of quantitative literacy to that which is most applicable to the work of Algebra 1 and distant enough from the predictor variables identified in the "A" table.]	6.NS.C.5, 6.NS.C.7b, 6.NS.C.7c, 6.NS.C.8, 7.RP.A.3*, 7.NS.A.3*, 7.EE.B.3*, 7.G.A.1@, 7.G.B.6@, 8.EE.A.3, 8.F.B.5*, 8.G.B.7, 8.SP.A.1, 8.SP.A.2, HSN-Q.A.1, HSN-Q.A.3, {higher-intensity high school modeling tasks where present}, {"Column 6" high school modeling tasks where present}	1: 8.SP.A.1 2: 8.SP.A.2 3: 8.G.B.7 4: 8.F.B.5 5: 8.EE.A.3 6: 7.G.B.6 7: 7.G.A.1 8: 7.EE.B.3 9: 7.NS.A.3 10: 7.RP.A.3 11: Claim 3 & 6.NS.C.7b 12: Claim 3 & 6.NS.C.7c 13: Claim 3 & 6.NS.C.8 14: Claim 3 & 6.NS.C.6 15: Claim 3 & 6.NS.C.6b 16: Claim 3 & 6.NS.C.6c (This bucket includes both conceptual quantitative literacy and application-based problems; it doesn't require algebraic expressions or equations.)
B2 - Algebra as generalized arithmetic writing and reading expressions and equations, transforming expressions and equations into equivalent expressions or equations using the properties of arithmetic operations and equality. (linear, quadratic, exponential, or conceptual/general)	6.EE.A* 7.EE.A.1* 8.EE.C.7b HSA-SSE.A.2 HSA-SSE.B.3a* HSA-SSE.B.3b* HSA-SSE.B.3c* HSA-CED.A.4	1: Claim 1 & 6.EE.A.3 2: Claim 1 & 6.EE.A.4 3: Claim 1 & 7.EE.A.1 4: Claim 1 & 8.EE.C.7b (These are all non-application, or problems with no context.)
B3 - Algebra as functional thinking (linear, quadratic, exponential, or conceptual/general). Formulating, interpreting and using mathematical expressions, tables and graphs that refer to variable quantities and relationships between quantities.	6.EE.C.9*, 7.RP.A.2b, 7.RP.A.2c, 7.EE.B.4@, 8.F.B.5, HSA-REI.D.11, HSF-IF*@, HSF-BF*@, HSF-LE.A.1, HSF-LE.A.2, HSF-LE.A.3, HSF-LE.B.5	1: 6.EE.C.9 2: 7.RP.A.2b 3: 7.RP.A.2c 4: 8.F.B.4 5: 8.F.B.5 6: 7.RP.A.2a 7: 7.RP.A.2d

Correlate of Interest	Common Core State Standards	Metadata Combinations Matching Correlates of Interest
B4 - Algebra as writing and solving constraint equations to solve problems in modeling scenarios (linear, quadratic, exponential, or conceptual/general)	6.EE.B.5, 6.EE.B.6*, 6.EE.B.7*, 7.EE.B.4*, 8.EE.C.7@, 8.EE.C.8c*, HSA-REI.C.6, HSA-CED.A.3@, HSA-REI.A.1, HSA-REI.B.3, HSA-REI.B.4*, HSA-REI.D.10, F-LE.A.4, {word problems matching the language of B4 not otherwise captured}	1: Claim 4 & 6.EE.B.6 3: Claim 4 & 6.EE.B.7 4: Claim 4 & 7.EE.B.4 5: Claim 4 & 7.EE.B.4a 6: Claim 4 & 8.EE.C.7 7: Claim 4 & 8.EE.C.7a 8: Claim 4 & 8.EE.C.7b 9: Claim 4 & 8.EE.C.8c (These are currently all application, or problems that have real-world context)
Contrast Variable Least Associated with Algebra 1	1: 8.G.A.1 2: 8.G.A.2 3: 8.SP.A.4 4: 7.G.A.3 5: 7.G.A.2 6: 7.SP.A.1 7: 7.SP.A.2 8: 7.SP.B.3 9: 7.SP.B.4 10: 7.SP.C.5 11: 7.SP.C.6 12: 7.SP.C.7 13: 7.SP.C.8	1: 8.G.A.1 2: 8.G.A.2 3: 8.SP.A.4 4: 7.G.A.3 5: 7.G.A.2 6: 7.SP.A.1 7: 7.SP.A.2 8: 7.SP.B.3 9: 7.SP.B.4 10: 7.SP.C.5 11: 7.SP.C.6 12: 7.SP.C.7 13: 7.SP.C.8

* A central standard or CCSS organizer for the correlate in question
@ A “bucket” containing tasks that match the correlate as well as tasks that don’t match.
For codes with @ symbol attached, the principle for identifying tasks that match the correlate is to apply the language of the correlate as a filter. For example, the language of correlate A3a (“Multi-step or higher complexity word problems with no fractions or decimals”) implies disregarding tasks in 5.G.2 if they are one-step problems or if they contain fractions or decimals. Note that the term *task* here refers not only to math problems, but also more generally to any student encounter with mathematics that generates the desired data.
HCII mapped standards codes in the data to the standards codes in Column A and Column B using both exact match and fuzzy match logic (to handle issues of data format—for example, codes of 3.MD.A vs. 3.MD.A.1). Limitations on available item metadata resulted in corresponding limitations on the coding process.

Appendix D: Descriptive Information for Predictor and Outcomes Items for Idaho

The descriptions of the prealgebra predictors (e.g., A1a) and algebra outcomes (e.g., B1) are in Table 16 of the main text. :Overall Prealgebra" is based on all of the A items, and "Overall Algebra" is based on all of the algebra outcome items (i.e., across B1 to B4, inclusive, from Table C2).

Table D1: Descriptive Information on the Number of Items Available Per Student for Idaho

	Min	Qt1	Median	Qt3	Max	Mean	*SD*
Prealgebra Predictors							
Overall Prealgebra	0	14	27	36	55	25.84	11.55
Geometry	0	3	5	6	11	4.82	2.01
A1a	0	0	1	2	5	1.21	1.04
A1b	0	3	4	6	11	4.46	2.14
A1c	0	2	2	3	5	2.32	0.57
A2a	0	0	0	2	10	1.16	1.86
A2b	0	1	1	2	9	1.64	1.28
A3a	0	1	4	5	11	3.30	2.50
A3b	0	2	3	3	8	2.79	1.06
A4	0	5	9	13	22	8.97	4.60
Algebra Outcomes							
Overall Algebra	0	4	17	23	39	14.69	9.51
Geometry	0	0	2	4	11	2.53	2.35
B1	0	1	9	12	25	7.19	5.37
B2	0	1	3	4	9	3.10	1.71
B3	0	1	4	6	15	4.15	3.09
B4	0	0	0	1	5	0.40	0.67

Appendix E: Descriptive Information for Predictor and Outcomes Clusters for California and Idaho

The descriptions of the prealgebra predictors (e.g., A1a) and algebra outcomes (e.g., B1) are in Table 16 of the main text. The outcome B is based on all of the algebra outcome items (i.e., across B1 to B4, inclusive, from Table C2).

For example, for cluster A1b (Number Sense: Fractions), there were 117 items with observed responses; 108 were dichotomously scored, and 9 were polytomously scored. Of the 850,057 students in the combined dataset, 632,374 had three or more responses. Across all 850,057 students, the number of responses ranged from 0 to 9, with an overall mean of 3.51 responses. After excluding students with responses to fewer than three items, the mean number of responses increased to 4.19. Hence, even though the number of identified items for the cluster is quite large, due to the SBAC design, each student received relatively few of these items. Despite the small number of observed responses per student for these students, the marginal reliability is 0.61.

Table E1: Cluster Information for Predictor Variables for California

	Prealgebra Predictors								
	A1a	A1b	A1c	A2a	A2b	A3a	A3b	A4	Geometry
Number of Items	74	117	113	3	86	161	113	379	142
Dichotomous	74	108	113	3	82	133	63	287	141
Polytomous	0	9	0	0	4	28	50	92	1
N Students (w/3+ Resp)	27378	632374	246416	0	158922	473373	604023	849875	714222
Min Num Responses	0	0	0	0	0	0	0	0	0
Max Num Responses	4	9	4	2	8	15	11	20	11
Mean Num Responses	0.90	3.51	2.26	0.02	1.38	3.94	3.54	14.68	4.28
Mean Num Responses (3+)	3.01	4.19	3.00	NA	3.35	6.43	4.25	14.69	4.77
Reliability	0.42	0.61	0.30	NA	0.46	0.64	0.55	0.74	0.48

Table E2: Cluster Information for Outcome Variables for California

	Algebra Outcomes					
	B	**B1**	**B2**	**B3**	**B4**	**Geometry**
Number of Items	521	283	118	114	6	74
Dichotomous	498	262	118	112	6	73
Polytomous	23	21	0	2	0	1
N Students (w/3+ Resp)	849775	849733	727779	617096	0	431431
Min Num Responses	0	0	0	0	0	0
Max Num Responses	29	22	5	10	2	5
Mean Num Responses	17.15	10.84	2.85	3.41	0.04	2.62
Mean Num Responses (3+)	17.15	10.85	3.00	4.09	NA	3.24
Reliability	0.86	0.80	0.48	0.33	NA	0.25

Table E3: Cluster Information for Predictor Variables for Idaho

	Prealgebra Predictors								
	A1a	**A1b**	**A1c**	**A2a**	**A2b**	**A3a**	**A3b**	**A4**	**Geometry**
Number of Items	80	137	113	25	88	157	106	269	157
Dichotomous	80	120	113	25	83	140	63	215	157
Polytomous	0	17	0	0	5	17	43	54	0
Number of Students	20236	41829	42375	1115	31616	29583	42354	42473	20236
Min Num Responses	0	0	0	0	0	0	0	0	0
Max Num Responses	4	8	4	5	7	9	7	17	10
Mean Num Responses	0.84	3.52	2.27	0.03	1.43	2.32	2.78	6.44	4.03
Reliability	0.35	0.62	0.29	0.29	0.26	0.41	0.43	0.59	0.40

Table E4: Cluster Information for Outcome Variables for Idaho

	Algebra Outcomes					
	B	**B1**	**B2**	**B3**	**B4**	**Geometry**
Number of Items	785	355	169	264	18	136
Dichotomous	766	343	169	259	16	126
Polytomous	19	12	0	5	2	10
Number of Students	42474	42472	42375	42468	18494	42473
Min Num Responses	3	0	0	0	0	1
Max Num Responses	39	25	9	15	5	11
Mean Num Responses	21.31	10.90	4.18	5.89	0.57	3.99
Reliability	0.69	0.57	0.40	0.36	0.21	0.26

Appendix F: Correlations Among Predictors and Outcomes for California

The descriptions of the prealgebra predictors (e.g., A1a) and algebra outcomes (e.g., B1) are in Table 16 of the main text. Math is the overall fifth-grade mathematics competence (i.e., theta) score, and ELA is the overall score for English Language Arts. The variable B is based on all of the algebra outcome items (i.e., across B1 to B4, inclusive, from Table C2).

Table F1: Pairwise Correlations based on Non-Imputed Scores for California

	Math	ELA	A1a	A1b	A1c	A2b	A3a	A3b	A4	AG	B	B1	B2	B3	B5
Math	1.00	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ELA	0.81	1.00	-	-	-	-	-	-	-	-	-	-	-	-	-
A1a	0.44	0.39	1.00	-	-	-	-	-	-	-	-	-	-	-	-
A1b	0.65	0.51	0.28	1.00	-	-	-	-	-	-	-	-	-	-	-
A1c	0.26	0.18	0.15	0.07	1.00	-	-	-	-	-	-	-	-	-	-
A2b	0.40	0.32	0.19	0.16	0.10	1.00	-	-	-	-	-	-	-	-	-
A3a	0.65	0.59	0.38	0.41	0.16	0.29	1.00	-	-	-	-	-	-	-	-
A3b	0.72	0.62	0.33	0.43	0.12	0.26	0.49	1.00	-	-	-	-	-	-	-
A4	0.85	0.74	0.43	0.53	0.20	0.34	0.62	0.62	1.00	-	-	-	-	-	-
AG	0.34	0.30	0.17	0.10	0.10	0.11	0.24	0.18	0.27	1.00	-	-	-	-	-
B	0.80	0.72	0.40	0.52	0.21	0.33	0.56	0.61	0.73	0.28	1.00	-	-	-	-
B1	0.76	0.67	0.38	0.50	0.19	0.32	0.53	0.58	0.70	0.26	0.96	1.00	-	-	-
B2	0.42	0.39	0.22	0.25	0.12	0.17	0.29	0.31	0.38	0.17	0.58	0.42	1.00	-	-
B3	0.41	0.39	0.19	0.26	0.10	0.16	0.28	0.31	0.37	0.15	0.53	0.39	0.25	1.00	-
B5	0.40	0.38	0.19	0.26	0.09	0.18	0.28	0.31	0.37	0.14	0.39	0.37	0.20	0.21	1.00

Table F2: Pairwise Correlations based on Imputed Scores for California

	Math	ELA	A1a	A1b	A1c	A2b	A3a	A3b	A4	AG	B	B1	B2	B3	B5
Math	1.00	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ELA	0.81	1.00	-	-	-	-	-	-	-	-	-	-	-	-	-
A1a	0.42	0.37	1.00	-	-	-	-	-	-	-	-	-	-	-	-
A1b	0.66	0.51	0.26	1.00	-	-	-	-	-	-	-	-	-	-	-
A1c	0.26	0.18	0.12	0.07	1.00	-	-	-	-	-	-	-	-	-	-
A2b	0.39	0.30	0.15	0.15	0.10	1.00	-	-	-	-	-	-	-	-	-
A3a	0.64	0.58	0.33	0.41	0.15	0.26	1.00	-	-	-	-	-	-	-	-
A3b	0.72	0.62	0.31	0.43	0.12	0.25	0.49	1.00	-	-	-	-	-	-	-
A4	0.85	0.74	0.41	0.53	0.20	0.33	0.62	0.62	1.00	-	-	-	-	-	-
AG	0.34	0.30	0.14	0.10	0.10	0.11	0.23	0.18	0.27	1.00	-	-	-	-	-
B	0.80	0.72	0.38	0.52	0.21	0.32	0.56	0.61	0.73	0.28	1.00	-	-	-	-
B1	0.76	0.67	0.36	0.51	0.19	0.30	0.53	0.58	0.70	0.26	0.96	1.00	-	-	-
B2	0.42	0.39	0.20	0.25	0.12	0.16	0.29	0.31	0.38	0.17	0.58	0.42	1.00	-	-
B3	0.41	0.39	0.19	0.26	0.10	0.16	0.29	0.31	0.38	0.15	0.53	0.40	0.25	1	-
B5	0.40	0.38	0.19	0.27	0.09	0.16	0.28	0.31	0.37	0.14	0.39	0.37	0.20	0.21	1

Appendix G: Additional Regression Models IRT Residuals for Idaho

The following tables provide the detailed results associated with Table 18 in the main text.

Table G1: Standardized Regression Estimates of the Relations Between Early Prealgebra Predictors and Later Overall Algebra in Idaho

Predictor	Estimate	*SD*	*t*-value	*p*
Intercept	-0.017	0.001	-15.13	< 2e-16
A1a: Numbers	-0.002	0.002	-1.02	0.306
A1b: Fractions	0.009	0.001	5.93	0.000
A1c: Decimals and Place Value	0.003	0.001	1.94	0.052
A2a: PS: Whole Numbers	0.098	0.003	36.77	< 2e-16
A2b: PS: Fractions	-0.004	0.001	-3.43	0.001
A3a: CPS: Whole Numbers	0.003	0.002	2.06	0.039
A3b: CPS: Rational Numbers	0.002	0.001	1.21	0.227
A4: Mathematical Reasoning & Communication	-0.001	0.002	-0.60	0.550
ELA	0.007	0.002	4.29	0.000
Geometry and Measurement	0.011	0.001	13.71	< 2e-16

Table G2: Standardized Regression Estimates of the Relations Between Early Prealgebra Predictors and Later Geometry and Statistics in Idaho

Predictor	Estimate	*SE*	*t*-value	*p*
Intercept	-0.014	0.003	-1.05	0.294
A1a - Whole Numbers	-0.003	0.004	1.71	0.088
A1b - Fractions	0.013	0.004	2.03	0.042
A1c - Decimals and Place Value	-0.011	0.004	5.48	0.000
A2a - PS: Whole Numbers	0.081	0.008	-9.43	< 2e-16
A2b - PS: Fractions	-0.006	0.004	-2.02	0.043
A3a - CPS: Whole Numbers	-0.002	0.005	-1.98	0.047
A3b - CPS: Rational Numbers	0.002	0.004	-0.97	0.331
A4 - Mathematical Reasoning & Communicating	-0.002	0.006	2.18	0.029
ELA	0.011	0.004	-0.54	0.590
Geometry and Measurement	0.010	0.002	2.77	0.006

Table G3: Standardized Regression Estimates of the Relations Between Early Prealgebra Predictors and Later English Language Arts in Idaho

Predictor	Estimate	*SE*	*t*-value	*p*
Intercept	0.685	0.006	121.34	< 2e-16
A1a - Whole Numbers	-0.055	0.009	-5.95	0.000
A1b - Fractions	-0.050	0.008	-5.94	0.000
A1c - Decimals and Place Value	0.015	0.008	1.95	0.052
A2a - PS: Whole Numbers	-0.078	0.016	-5.06	0.000
A2b - PS: Fractions	-0.005	0.008	-0.62	0.535
A3a - CPS: Whole Numbers	0.050	0.010	5.25	0.000
A3b - CPA: Rational Numbers	-0.012	0.009	-1.41	0.157
A4 - Mathematical Reasoning & Communicating	0.097	0.013	7.62	0.000
Geometry and Measurement	0.044	0.009	4.86	0.000

Appendix H: Additional Regression Models IRT Residuals for California

The following tables supplement the results shown in Table 23 of the main text; the latter includes control of ELA and early overall mathematics competencies and the following includes only control of ELA (Table H1) or overall mathematics competencies (Table H2).

Table H1: Model 2 Standardized Regression Coefficients for IRT Residuals for California Controlling Overall ELA Scores

Outcome Variables	*Overall Algebra Beta*	*Overall Algebra SE*	*Quantitative Literacy Beta*	*Quantitative Literacy SE*	*Generalized Arithmetic Beta*	*Generalized Arithmetic SE*	*Functional Thinking Beta*	*Functional Thinking SE*	*Geometry & Statistics Beta*	*Geometry & Statistics SE*
Predictor Variables										
A1a - Whole Numbers	-0.0005	0.0016	0.0012	0.0017	-0.0005	0.0016	-0.0027	0.0013	-0.0009	0.0019
A1b - Fractions	0.0012	0.0010	0.0004	0.0010	0.0015	0.0011	-0.0001	0.0011	-0.0015	0.0011
A1c - Decimals	-0.0010	0.0011	-0.0003	0.0011	-0.0014	0.0010	-0.0013	0.0011	0.0008	0.0010
A2b - Basic Problem-Solving: Fractions	0.0002	0.0012	0.0004	0.0012	-0.0013	0.0014	0.0020	0.0013	-0.0009	0.0014
A3a - Complex Problem Solving: Whole Numbers	0.0005	0.0015	-0.0006	0.0014	0.0000	0.0014	0.0024	0.0014	-0.0013	0.0012
A3b - Complex Problem Solving: Fractions	0.0004	0.0013	-0.0006	0.0012	0.0021	0.0012	0.0001	0.0011	-0.0001	0.0010
A4 - Mathematical Reasoning and Communication	0.0005	0.0011	0.0007	0.0011	0.0003	0.0011	-0.0009	0.0011	-0.0004	0.0011
Geometry & Measurement	0.0001	0.0014	-0.0014	0.0014	0.0016	0.0011	0.0010	0.0013	0.0004	0.0011
ELA	-0.0012	0.0010	-0.0002	0.0010	-0.0007	0.0010	-0.0016	0.0011	0.0014	0.0011

Table H2: Model 3 Standardized Regression Coefficients for IRT Residuals for California Controlling Overall Mathematics Scores

Outcome Variables	*Overall Algebra Beta*	*Overall Algebra SE*	*Quantitative Literacy Beta*	*Quantitative Literacy SE*	*Generalized Arithmetic Beta*	*Generalized Arithmetic SE*	*Functional Thinking Beta*	*Functional Thinking SE*	*Geometry & Statistics Beta*	*Geometry & Statistics SE*
Predictor Variables										
A1a - Whole Numbers	-0.0005	0.0016	0.0012	0.0017	-0.0005	0.0016	-0.0027	0.0013	-0.0009	0.0019
A1b - Fractions	0.0012	0.0010	0.0004	0.0010	0.0015	0.0011	-0.0001	0.0011	-0.0015	0.0011
A1c - Decimals	-0.0010	0.0011	-0.0003	0.0011	-0.0014	0.0010	-0.0013	0.0011	0.0008	0.0010
A2b - Basic Problem-Solving: Fractions	0.0002	0.0012	0.0004	0.0012	-0.0013	0.0014	0.0019	0.0013	-0.0009	0.0014
A3a - Complex Problem Solving: Whole Numbers	-0.0004	0.0015	-0.0010	0.0014	-0.0007	0.0014	0.0023	0.0014	-0.0010	0.0012
A3b - Complex Problem Solving: Fractions	0.0004	0.0013	-0.0006	0.0012	0.0021	0.0012	0.0001	0.0011	-0.0001	0.0010
A4 - Mathematical Reasoning and Communication	0.0005	0.0011	0.0007	0.0011	0.0003	0.0011	-0.0009	0.0011	-0.0004	0.0011
Geometry & Measurement	0.0001	0.0014	-0.0014	0.0014	0.0016	0.0011	0.0010	0.0013	0.0004	0.0011
Math	0.0003	0.0010	0.0005	0.0011	0.0005	0.0010	-0.0013	0.0011	0.0008	0.0011

Appendix I: Variance Explained for IRT Residual Models for California

The table shows the overall explained variance for the IRT residual models.

Table I1: Overall Variance Explained (R^2) for IRT Residuals for California

Outcome Variables	*Overall Algebra*	*Quantitative Literacy*	*Generalized Arithmetic*	*Functional Thinking*	*Geometry & Statistics*
Model 1	0.000006	0.000005	0.000015	0.000020	0.000008
Model 2	0.000007	0.000005	0.000015	0.000020	0.000008
Model 3	0.000006	0.000005	0.000015	0.000020	0.000008
Model 4	0.000009	0.000005	0.000016	0.000020	0.000009

Appendix J: Additional Regression Models IRT Subscores for California

The following tables are for Model 2 (Table J1) and Model 3 (Table J2) that supplement the results for Model 1 (Table 25) and Model 4 (Table 26) in the main text.

Table J1: Model 2 Standardized Regression Coefficients for IRT Subscores for California Controlling Overall ELA Scores

Outcome Variables	*Overall Algebra Beta*	*Overall Algebra SE*	*Quantitative Literacy Beta*	*Quantitative Literacy SE*	*Generalized Arithmetic Beta*	*Generalized Arithmetic SE*	*Functional Thinking Beta*	*Functional Thinking SE*	*Geometry & Statistics Beta*	*Geometry & Statistics SE*
Predictor Variables										
A1a - Whole Numbers	**0.050**	0.001	**0.048**	0.001	**0.030**	0.002	**0.016**	0.002	**0.018**	0.002
A1b - Fractions	**0.127**	0.001	**0.131**	0.001	**0.035**	0.002	**0.048**	0.002	**0.061**	0.002
A1c - Decimals	**0.050**	0.001	**0.046**	0.001	**0.032**	0.001	**0.021**	0.001	**0.015**	0.001
A2b - Basic Problem-Solving: Fractions	**0.059**	0.001	**0.060**	0.001	**0.024**	0.002	**0.019**	0.001	**0.034**	0.002
A3a - Complex Problem Solving: Whole Numbers	**0.062**	0.001	**0.059**	0.001	**0.026**	0.001	**0.020**	0.002	**0.025**	0.002
A3b - Complex Problem Solving: Fractions	**0.146**	0.001	**0.149**	0.001	**0.056**	0.002	**0.061**	0.002	**0.066**	0.002
A4 - Mathematical Reasoning and Communication	**0.283**	0.002	**0.281**	0.002	**0.134**	0.003	**0.124**	0.003	**0.114**	0.003
AG - Geometry & Measurement	**0.056**	0.001	**0.049**	0.001	**0.050**	0.001	**0.032**	0.001	**0.025**	0.001
ELA	**0.252**	0.002	**0.213**	0.002	**0.181**	0.003	**0.204**	0.002	**0.182**	0.002

Table J2: Model 3 Standardized Regression Coefficients for IRT Subscores for California Controlling Overall Mathematics Scores

Outcome Variables	*Overall Algebra Beta*	*Overall Algebra SE*	*Quantitative Literacy Beta*	*Quantitative Literacy SE*	*Generalized Arithmetic Beta*	*Generalized Arithmetic SE*	*Functional Thinking Beta*	*Functional Thinking SE*	*Geometry & Statistics Beta*	*Geometry & Statistics SE*
Predictor Variables										
A1a - Whole Numbers	**0.040**	0.001	**0.038**	0.001	**0.025**	0.002	**0.012**	0.002	**0.016**	0.002
A1b - Fractions	**0.020**	0.001	**0.030**	0.001	**-0.029**	0.002	**-0.006**	0.002	**0.021**	0.002
A1c - Decimals	**0.010**	0.001	**0.009**	0.001	**0.008**	0.001	-0.001	0.001	-0.001	0.001
A2b - Basic Problem-Solving: Fractions	**0.015**	0.001	**0.018**	0.001	-0.003	0.002	**-0.004**	0.001	**0.017**	0.002
A3a - Complex Problem Solving: Whole Numbers	**0.049**	0.001	**0.045**	0.002	**0.020**	0.002	**0.020**	0.002	**0.028**	0.002
A3b - Complex Problem Solving: Fractions	**0.064**	0.002	**0.069**	0.002	**0.009**	0.002	**0.026**	0.002	**0.043**	0.002
A4 - Mathematical Reasoning & Communication	**0.160**	0.002	**0.159**	0.002	**0.066**	0.004	**0.077**	0.003	**0.086**	0.002
Geometry and Measurement	**0.016**	0.001	**0.010**	0.001	**0.027**	0.001	**0.014**	0.001	**0.013**	0.001
Math	**0.544**	0.004	**0.502**	0.005	**0.340**	0.004	**0.313**	0.004	**0.245**	0.004

Appendix K: CCSS Mastered by All Students in the MATHia Analyses Samples

The following includes CCSS that were mastered by all students who were exposed to them and thus not included in the Mathia analyses.

Table K1: Standards Mastered by All Students in the 2017-2018 MATHia Analyses Sample

Common Core State Standards Workspaces Mastered by All Students 2017-2018	Number of Students
6.EE.A.2	3
6.EE.A.2.b~~6.EE.A.3~~6.EE.A.4	3
6.EE.A.2.c	3
6.EE.A.3	3
6.EE.B.5	3
6.EE.B.6~~6.EE.B.7	3
6.EE.B.6~~6.EE.C.9	3
6.EE.B.7	3
6.EE.B.8	3
6.EE.C.9	2
6.G.A.1	3
6.G.A.1~~7.G.B.6	3
6.G.A.2	3
6.G.A.3~~6.NS.C.8	2
6.NS.A.1	3
6.NS.B.3	3
6.NS.B.4	3
6.NS.C.6	6
6.NS.C.6.a	6
6.NS.C.6.c	2
6.NS.C.7.c	6
6.RP.A.1	3
6.RP.A.3.a	3
6.RP.A.3.a~~7.RP.A.2	3

Common Core State Standards Workspaces Mastered by All Students 2017-2018	Number of Students
6.RP.A.3.c	3
6.RP.A.3.d	3
6.SP.A.3~~6.SP.B.5.c	1
6.SP.B.4~~6.SP.B.5	1
6.SP.B.5.c	1
8.F.A.1	141
8.G.C.9~~HSG.GMD.A.3	87
HSA.APR.A.1~~HSF.BF.A.1.b	5
HSA.APR.B.2	4
HSA.APR.B.3	4
HSA.SSE.A.1	6
HSA.SSE.A.1.a~~HSA.SSE.B.4~~HSF.BF.A.2	4
HSA.SSE.A.1.b~~HSF.BF.B.3	6
HSA.SSE.A.1~~HSA.SSE.A.2	6
HSA.SSE.A.2	4
HSA.SSE.B.4	4
HSF.IF.C.7.b	4
HSF.IF.C.8.b	1023
HSF.LE.A.4	4
HSF.TF.A.1~~HSF.TF.A.2	4
HSF.TF.B.5	4
HSF.TF.C.8	4
HSG.C.A.1~~HSG.C.A.2	12
HSG.C.A.2	12
HSG.C.A.3	11
HSG.C.B.5	11
HSG.CO.A.3	31
HSG.CO.A.4	31
HSG.CO.B.7~~HSG.CO.B.8	18

Common Core State Standards Workspaces Mastered by All Students 2017-2018	Number of Students
HSG.GMD.B.4	16
HSG.GPE.A.1	11
HSG.SRT.A.2	11
HSG.SRT.B.4	11
HSG.SRT.B.5	11
HSG.SRT.C.7	12
HSN.CN.A.1	6
HSN.CN.A.2	6
HSN.CN.C.7	5
HSS.CP.A.2	6
HSS.CP.A.3	6
HSS.CP.A.4	6
HSS.CP.A.5	6
HSS.CP.B.6~~HSS.CP.B.7	6

Table K2: Standards Mastered by All Students in the 2018-19 MATHia Analyses Sample

Common Core State Standards Workspaces Mastered by All Students 2018-2019	Number of Students
6.EE.A.2.b~~6.EE.A.3~~6.EE.A.4	1
6.EE.A.2.c	1
6.EE.A.3~~6.EE.A.4~~7.EE.A.1	1
6.EE.B.5	1
6.EE.B.6~~6.EE.B.7	1
6.EE.B.6~~6.EE.C.9	1
6.EE.B.7	1
6.EE.B.8	1
6.EE.C.9	1
6.G.A.1	2
6.G.A.1~~7.G.B.6	2
6.G.A.3~~6.NS.C.8	1
6.NS.A.1	1
6.NS.B.3	1
6.NS.B.4	2
6.NS.C.6	1
6.NS.C.6.a	1
6.NS.C.6.c	1
6.NS.C.7.c	1
6.RP.A.3.a	1
6.RP.A.3.a~~7.RP.A.2	1
6.RP.A.3.b	3
6.RP.A.3.c	1
6.RP.A.3.d	1
6.SP.A.3~~6.SP.B.5.c	1
6.SP.B.4~~6.SP.B.5	1
6.SP.B.4~~6.SP.B.5.a~~6.SP.B.5.b	1
6.SP.B.5.c	1

Common Core State Standards Workspaces Mastered by All Students 2018–2019	Number of Students
7.EE.A.1	1
7.EE.B.4	1
7.EE.B.4.a	44
7.EE.B.4.b	1
7.G.A.1	1
7.G.A.3~~HSG.GMD.B.4	19
7.G.B.4	3
7.NS.A.1	1
7.NS.A.2.a~~7.NS.A.2.b~~7.NS.A.2.c	1
7.NS.A.2.d	1
7.NS.A.3	1
7.RP.A.2.a	2
7.RP.A.2.a~~7.RP.A.2.b	2
7.RP.A.2.b	44
7.RP.A.2.b~~7.RP.A.2.c	2
7.RP.A.2.c	2
7.RP.A.3	2
8.EE.A.1	4
8.EE.A.2~~8.G.B.7	3
8.EE.A.2~~8.NS.A.1~~8.NS.A.2	3
8.EE.A.3	3
8.EE.A.4	3
8.EE.B.5~~8.F.A.3	42
8.EE.C.7.b	43
8.EE.C.8.a~~8.EE.C.8.b~~8.EE.C.8.c	42
8.EE.C.8.b	42
8.F.A.1	40
8.G.A.3	3
8.G.A.5	3

Common Core State Standards Workspaces Mastered by All Students 2018–2019	Number of Students
8.G.B.6	3
8.G.C.9	2
8.NS.A.1~~8.NS.A.2	3
8.SP.A.1~~8.SP.A.2	40
8.SP.A.2~~8.SP.A.3	40
HSA.APR.A.1~~HSF.BF.A.1.b	2
HSA.APR.B.2	1
HSA.APR.B.3	1
HSA.CED.A.3	16
HSA.REI.B.4.a	692
HSA.REI.D.11	626
HSA.SSE.A.1	4
HSA.SSE.A.1.b~~HSF.BF.B.3	4
HSA.SSE.A.1~~HSA.SSE.A.2	4
HSA.SSE.A.2	1
HSF.BF.A.1.c	3
HSF.IF.C.7.b	8
HSG.C.A.3	44
HSG.C.B.5	13
HSG.CO.B.7~~HSG.CO.B.8	45
HSG.CO.C.11	12
HSG.GMD.B.4	21
HSG.SRT.A.2	17
HSG.SRT.C.6	11
HSG.SRT.C.7	11
HSN.CN.A.1	4
HSN.CN.C.7	2