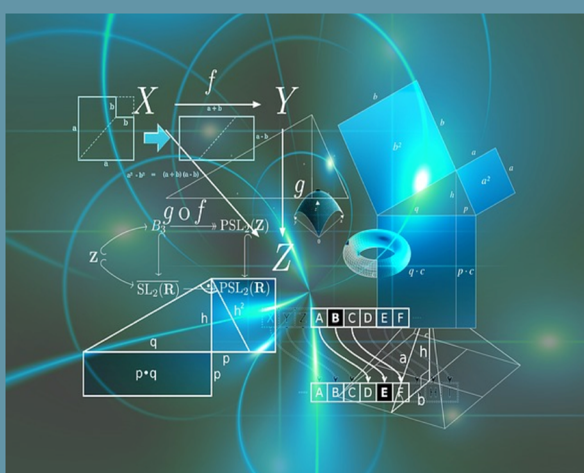
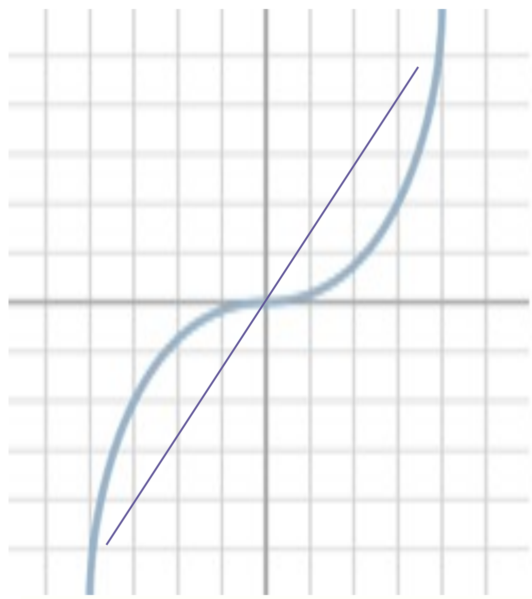
VARIABLES and CONSTANTS

a2i accessing algebra through inquiry

Final Report
Summer 2017





Rockman et al
Research & Evaluation

Acknowledgments

This study would not have been possible without the support of the many people who took part in a2i. We would especially like to thank the teachers and administrators who provided invaluable feedback—in surveys, interviews, and focus groups—and welcomed us into their schools, classrooms, and team meetings. They not only helped us understand a2i from a practitioner's perspective, but also made sure that consent forms and surveys were distributed to over 4,000 New York City high school students. Our thanks, too, to all those students who shared their views about math.

We would also like to thank Fatih Unlu for his expert guidance in the design of the study, and Irene Montonti, our U.S. Department of Education program officer, for her ongoing support. We, like the project team, also benefitted from the wisdom of Magdalene Lampert and David Foster.

As a2i evolved, we spent a good deal of time with the a2i team—at New Visions, in schools, at PD sessions and debriefings, and on subways—learning about all the moving pieces in the a2i project. We want to thank all the Instructional Specialists, along with Erik Laby and the data team, for answering our many questions and data requests. A special thank you goes to Russell West and Danny Voloch, whose thoughtful, incisive comments always improved our work.

Alison Allen

Adam Moylan

Saul Rockman

Kay Sloan

This study was supported by the U.S. Department of Education's Office of Innovation and Improvement, Grant PR Award #U411B110116. The opinions expressed here are those of the authors and do not represent the views of the U.S. Department of Education.

Rockman et al is an independent research, evaluation, and consulting firm that specializes in examining critical issues in formal and informal education. Rockman works with education providers (serving learners of all ages), government agencies, and organizations with wide-ranging educational projects and wide-ranging community, school, or consumer audiences. Rockman's headquarters are in San Francisco, with offices in Bloomington, IN, Columbus, OH, and Pittsburgh, PA. For further information and examples of projects and reports, see www.rockman.com.

TABLE OF CONTENTS

TABLES AND FIGURES

TABLES

FIGURES

a2i accessing algebra through inquiry

Final Report

VARIABLES and CONSTANTS

OVERVIEW

This report shares findings from Rockman et al's comprehensive evaluation of Accessing Algebra Through Inquiry (a2i), a five-year high school mathematics project launched in 2012 by New Visions for Public Schools, in partnership with the New York City Department of Education (NYC DOE), the Silicon Valley Mathematics Initiative (SVMI), and the Boston Plan for Excellence and its Boston Teacher Residency. During a2i's first year, 14 New York City public high schools took part in the project, which over the next four years expanded to 33 schools and over 100 math teachers and 4,000 students. Funded by a validation grant from the U.S. Department of Education's Invest in Innovation (i3) program, a2i proposed to close achievement gaps by helping students—many of them in some of the city's highest-need schools—understand math, become problem-solvers, and think like mathematicians.

Phil Daro, a member of a2i's advisory committee, explained this approach to teaching mathematics as "rethinking grain size," and focusing more on big ideas and processes—relationships, patterns, and the variety of ways mathematical expressions can be represented—and less on mastery of basic facts and discrete operations (Daro, 2014). The Common Core State Standards (CCSS) for Mathematics (National Governors Association Center for Best Practices (2010), which Daro helped craft, also endorses this approach, so the a2i model had the potential to not only strengthen math teaching and address the grant goals of improving academic outcomes for students, but also help teachers translate the new standards—adopted by New York in 2010—into practice and prepare students for the companion assessments.

On baseline surveys and in focus groups and informal conversations, a2i teachers repeatedly said that this was the way math should be taught. New and veteran teachers alike preferred depth to breadth, and those with more classroom experience were relieved to no longer have to teach "a million topics." Frustrated by the hazy year-to-year connections in math standards and curriculum, experienced teachers also welcomed new standards that vertically articulated goals across grades and charted learning progressions for students. All were pleased to be moving toward an approach that would "draw students in and reassure them that it's OK to make mistakes."

Teachers were also aware of challenges: Although they agreed with the approach, teachers were not used to teaching this way, nor were their students used to this way of learning. It would, they noted, be a while before students arrived with eight years of carefully sequenced math classes behind them, ready to tackle algebraic and geometric relationships. Teachers and students would need tools and training to make the transition.

For a2i's designers, the key was formative assessment, which became the central focus—and a constant—of the project. Focusing on student work and integrating assessment feedback into classroom practices would help teachers chart instructional pathways for students and personalize learning. According to Dylan Wiliam, whose Embedded Formative Assessment embodied a2i's vision, formative assessment was a way to figure out "where students are in their learning, where they're going, and how to get there" (Wiliam, 2011, p. 45).

Formative assessment was also one of the absolute priorities for the i3 grant, which stipulated that funded projects incorporate not formative assessment alone, but formative assessments "aligned with high-quality student content and academic achievement standards." Partners brought experience with this caliber of math content to the collaboration: SVMI had introduced teachers around the country to the Mathematics Assessment Resource Service (MARS) tasks and SVMI's own Mathematics Assessment Collaborative (MAC) summative performance tasks, which were intended to expose teachers to Common Core-like tasks and give students practice with problems they would eventually encounter on the Regents exams (Fisher, 2007; Foster, 2007). The Shell Centre's Formative Assessment Lessons (FALs), already in use in some New Visions Schools, provided examples of the rich, multi-step formative tasks called for in the a2i plans, again providing teachers and students with standards-aligned tasks like those planned for the new Partnership Assessment of Readiness for College and Careers (PARCC) exams, slated to begin roll-out in a2i's last school year (2016–17).

The a2i plans also included support mechanisms to help teachers create effective math learning environments. School inquiry teams, already mandated by the NYC DOE, would provide a forum for regular discussions of student work and data-driven instructional practices in algebra and geometry, and Instructional Specialists or coaches would meet regularly with teachers and teams to guide them in translating assessment feedback into instructional practices. Again, partners brought experience to the table: SVMI had developed and tested coaching strategies to support teachers in their work. New Visions had developed, assessed, and provided training on the use of data and formative assessment to inform standards-based instruction, and actively supported the inquiry work of school teams in its 70+ school network.

The broader support mechanism was a2i's theory of change. Drawing on the Chicago Consortium for School Reform model (Sebring, Allensworth, Bryk, Easton, & Luppescu, 2006), the a2i team built in structures to set a uniform pace for teachers, even as they were making instructional decisions based on formative assessments and their own students' work. Their premise was that:

- If inquiry groups within and across schools are looking at the same student tasks at roughly the same time to make instructional decisions,
- and teachers are teaching the same algebraic concepts at roughly the same time and embedding formative assessment strategies in instruction roughly every day,
- then student achievement in mathematics will measurably increase.

There were some assumptions and probabilities embedded in this logic, and some distance between the "if" and the "then"—which, as a2i moved forward, required a series of adjustments related to inquiry teams, pacing, and measures of student achievement. Even though teachers embraced a2i's overall approach, they also felt a little unmoored with a general framework instead of the customary math curriculum. According to Daniel Voloch, New Visions' Vice President for Curriculum and Instruction, handing teachers a curriculum in some ways "went against the idea that teachers should figure out what students need," but the teachers' needs took priority and, in Year 3, project staff began developing an a2i curriculum. Additional curricular resources were developed as *Instructional Activities* (IAs) that teachers could integrate into their daily lessons. Both the curriculum and the IAs were available to teachers on the project website in Year 4. The curriculum is now publicly accessible at curriculum.newvisions.org/math.

Data tools and systems evolved as well. Teachers had expressed concerns about whether the formative assessment lessons (FALs) and initial and final assessment measures were a good fit for urban students, the best gauge of their needs and skills, and a reliable data source for teachers. In Year 3, the internal formative assessment measure and systems shifted, and a2i teachers began uploading the results into a newly developed balanced assessment tool. This new balanced assessment tool was designed using Google Spreadsheets, to give teachers immediate visualizations of student progress once they entered in student performance data from a balanced assessment. This tool became the prototype for a more advanced data tool built to help schools understand student performance data from a Mock Regents exam.

The composition and focus of inquiry teams varied across NYC public schools, which required the a2i team to customize school coaching activities in Year 1. Nearly doubling the numbers of participating schools in Year 2—and designing support around the needs of less experienced teachers—required additional adjustments in coaching strategies and professional development. The level of staffing and effort required by ever-larger numbers of schools, and by the addition of Geometry and Algebra II courses, was also a barrier to scaling the effort based on personnel and funding. Given these constraints, and guided by Senior Advisor Magdalene Lampert, the a2i Instructional Specialists developed protocols and routines that would allow them to systematize their school-based coaching efforts based on teacher practices observed or discussed during school visits. Moving from 1:1 teacher support to support for math inquiry teams or departments made the coaching model more sustainable and scalable coaching model and carried an added benefit of building school capacity.

One of the biggest challenges for a2i partners and teachers—and the evaluation—was the uncertainty around the state's high-stakes assessments. Delays in the roll-out of the new state Common-Core aligned Regents Algebra I and Geometry exams created conflicts for teachers who had signed on to a new set of standards and an approach to teaching math but were held to an old set of accountability measures. Students in different cohorts were also held to different examination requirements, including one cohort that could take both old and new exams. The flux in exams that lasted throughout a2i's first two and a half years—and uncertainty that lingered in Years 3 and 4—led the a2i team to make repeated adjustments in the a2i model to guide and support teachers.

This report explores what the a2i team and teachers implemented over four years, describing what core components remained constant despite a mutable assessment landscape, what changed, and what others could replicate.

Organization of the Report

Following summaries of methods and findings, the report begins, in Section 1, with a more detailed description of how the a2i team and a2i teachers and schools defined and redefined the core components of the a2i model. The first section also includes the implementation fidelity results for Years 1–4, and a2i teachers' ratings of the project's durable elements. Section 2 examines how the a2i model affected student achievement, based on results of the formal Year 4 impact study, designed to examine how students in a2i Algebra I and Geometry teachers' classrooms performed on Common-Core aligned state assessments, compared to non-a2i students in matched schools.

Sections 3 and 4 examine more proximal changes, including changes in teachers' beliefs and practices and classroom learning environments. These sections also explore whether the length or level of teachers' involvement was a factor in implementation changes. Section 5 returns to student outcomes, reporting findings on a2i's impact on students' self-efficacy and interest in math. The final section discusses changes in collaboration and culture in a2i schools.

SUMMARY OF METHODS

The mixed methods evaluation included an implementation study in all four years of the project and an impact study in Years 3 and 4. Throughout the project, the evaluation team collaborated with a2i partners to develop instruments and determine ways to combine internal and external data to document project progress and to generate evidence that would allow practitioners and policymakers to understand what a2i practices or strategies worked, with whom, and in what contexts. This section summarizes the data collection activities, sources, samples, and analytic strategies. Discussions of findings in the body of the report include further details.

Impact Study

In the impact study, the evaluators examined student achievement in Algebra and Geometry, as measured primarily by the new New York State Regents examinations. A quasi-experimental study with a matched comparison design was used. The study involved 20 a2i schools and 20 matched comparison NYCDOE schools. The Algebra I outcomes during school years 3 and 4 and the Geometry outcomes in school year 4 were included in the impact study. This impact study tested two confirmatory research questions:

> R1. Does the a2i model have a greater effect than regular or current instruction on math achievement as measured by standardized state mathematics examinations in high school Algebra after at least two years of implementation? (confirmatory contrast)
>
> R2. Does the a2i model have a greater effect than regular or current instruction on math achievement as measured by standardized state mathematics examinations in high school Geometry after at least two years of implementation? (confirmatory contrast)

The comparison condition for the impact study consisted of Algebra I and Geometry courses taught by teachers in non-a2i schools. These schools were made up of similar student populations based on key demographics and 8th grade math achievement on the NYS examination. They were drawn from the 75+ schools in the New Visions affinity network. The teachers of comparison schools had some access to the same categories of resources; that is, resources about inquiry teams and formative assessment, the state-developed EngageNY resources (www.engageny.org), as well as some level of support from district-funded instructional coaches. However, the consistency and level of engagement with and professional development about those resources was not as extensive as in a2i schools in comparison schools. See Appendix B for a description of the matching procedures.

We assessed baseline equivalence between a2i and matched comparison groups for the two primary analytic samples of students (Common Core algebra and Common Core geometry), which only consisted of individuals with non-missing values for both the baseline measure and the outcome measure. To assess baseline equivalence, we used student scores in math and reading from the 8th grade state assessments. See Appendix C for further discussion of the baseline equivalence testing.

To explore potential improvements associated with the addition of Instructional Activities to the a2i curriculum and restructuring of the PD sessions in Year 4 (the 2nd year of the two-year impact study), we examined school-level algebra Regents exam pass rates and college readiness rates for 9th graders at all a2i and comparison schools for June 2015 and June 2016. These two indicators are key metrics used by New Visions schools to measure progress in algebra achievement. For the Regents Algebra I exam, the minimum score for passing is 65, while the minimum score to indicate college readiness is 70.

Implementation Study

Our approach to studying implementation was based on a combined theory-of-change (Connell & Kubisch, 1995) and outcomes-based model (Schalock, 2001). The use of both approaches allowed us to document a2i's evolution, providing feedback to program designers about changes in the implementation of program components and shifts in the hypothesized relationships laid out in the project's logic model (see Figure 2, p.16).

The questions guiding the implementation study included:

1. What are the critical, measurable features of the a2i program's professional development and instructional components?
2. How did a2i influence classroom practice? Did teachers' level of involvement matter?
3. Did a2i bring about changes in learning environments?
4. Did a2i affect students' attitudes about math, including their sense of efficacy and interest in math courses or careers?
5. Did a2i bring about changes in school collaboration and inquiry around formative assessment?

Data Sources and Samples

Fidelity of Implementation Matrix

In a2i's first year, evaluators worked with the project team to define and measure the project's critical features, charted in a fidelity matrix that was refined over the project's next two years. The fidelity of program implementation was assessed by attending to the five key components (or constructs) of the a2i model as shown in the logic model, which are: 1) professional development, 2) school support for collaborative inquiry, 3) training for administrators, 4) instructional practice changes, and 5) routine use of a student data management system to monitor students' progress. Table 2 (p. 21), briefly lists the key components along with indicators for benchmarking implementation. We defined two sets of measures of fidelity: one for the process variables (inquiry team work, school context, etc.), and one for teachers' classroom implementation of the pedagogy and content.

As part of the study of implementation and fidelity, we developed measures to monitor professional development and classroom instruction fidelity, facilitated by New Visions' attendance records, a web-based database system of student variables, and an accompanying database of teacher variables.

Observations, Interviews, and Focus Groups

Throughout the implementation study, evaluators conducted observations, interviews, and focus groups designed to gather more in-depth data on professional development, collaborative inquiry work, and their effects on teachers' practice. In the project's first two years, evaluators observed PD sessions, inquiry team meetings, coaching sessions, and classrooms. In subsequent years, evaluators continued to observe a sample of school team coaching sessions and professional development sessions, including the week-long August sessions, quarterly sessions, and a sample of unit-based monthly PDs. Additional feedback came from beginning and end-of-year focus groups with teachers, based on structured protocols.

In Year 2, a two-tiered process allowed the evaluators to take advantage of opportunities to observe classrooms and team and coaching sessions and get a broad view of implementation, but also conduct more in-depth studies of a representative sample of 5-6 schools that were visited multiple times. Evaluators also worked with the a2i Instructional Specialists to norm a new classroom observation tool that combined features of the Year 1 instrument with a New Visions instrument based on the Dylan Wiliam formative assessment practices. In Years 3 and 4, evaluators continued to observe a sample of all-day PD sessions, school team coaching sessions, and classrooms.

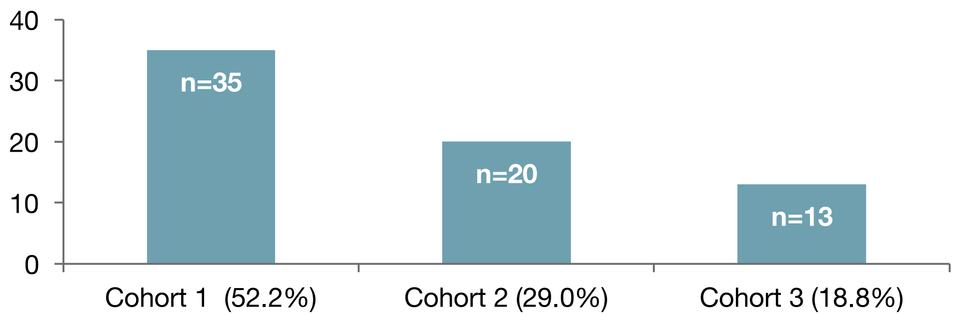
Throughout the project, REA conducted periodic formal interviews with the a2i project team, including the Instructional Specialists, using semi-structured protocols, along with informal interviews during project activities. Evaluators also interviewed administrators or math department chairs at a sample of comparison schools. Lastly, schools were invited to take a brief, online survey with questions about their shift to the Common Core standards and assessments, and the resources teachers consulted on professional development activities they took part in as part of that transition.

Teacher Surveys and Samples

Each year of the project, REA invited teachers to complete baseline and/or follow-up surveys, all administered online. Cohort 1 teachers completed the baseline survey at the beginning of the project, either in the spring or summer before the project's first school year (2012–13), and a follow-up survey at the end of the school year. Each subsequent cohort completed the baseline survey in August, prior to the school year, and a follow-up survey at the year's end. The baseline survey gathered information about teachers' experience, beliefs and practices, school conditions, collaborative activities, and expectations for the project.

Survey responses came from 71 teachers who completed the annual survey administered in May 2016, as a2i wrapped up the final year. The survey sample included respondents from all four project years and all three school cohorts, Cohort 1 (2012–13/Year 1), Cohort 2 (2013–14/Year 2), and Cohort 3 (2014–15/Year 3). Figure 1 shows the numbers of survey respondents, based on teachers' cohort or the year they began their a2i participation, and the percentages of the full sample comprised by each of the three cohorts.

FIGURE 1. NUMBER OF SURVEY RESPONDENTS BY YEAR TEACHER BEGAN a2I



Of the 71 Year 4 respondents, 86% had advanced degrees beyond a Bachelor's degree (Master's degree, 83%; Doctorate or Professional degree, 3%). Thirty-two (45%) teachers entered teaching through an alternative certification program, most through the NYC Teaching Fellows program (56%); other certification programs include New Visions – Hunter Residency programs and Teach for America (both, 9.4%). Although the Year 4 teachers taught multiple grades, they were concentrated in grades nine and ten. Similarly, they taught a wide range of subjects—from 8th grade math to Calculus and International Baccalaureate mathematics—but with larger percentages teaching Algebra I (60.5%) and Geometry (40.8%). In the final program year, more special education teachers participated in a2i compared to previous years. The majority (83%) of respondents identified as regular education math teachers; 17%, as special education teachers team-teaching in math classes. See Section 1 for additional background information on the Year 4 teacher survey sample.

The Year 4 survey sample skewed to teachers that had started in Years 3 and 4, so in our analyses we looked at the survey sample as a whole and created survey subgroups as needed to look for common trends among groups. For example, we divided teachers into *early* and *recent* cohort groups to explore the effect of extended participation; and into *a2i* and *a2i mix* to explore differences between the two levels of implementation.

Using Year 4 with the prior waves of survey data, we created matched groups of teachers who completed two surveys—an initial and a follow-up survey—to examine responses from the same teachers. We conducted mean comparisons on items using *t*-tests. For items that were dichotomous or categorical, we conducted chi-square analyses. Bonferroni corrections were used for analyses where there were multiple items.

Student Survey Scales and Samples

In Year 1, we began development of a survey designed to gather feedback from high school students about their attitudes toward math and a2i classroom learning environments. Questions focused on seven constructs: math self-efficacy, math interest/intrinsic value, math utility value, perseverance, student behaviors, teacher practices/behaviors, and classroom culture, validated through a series of factor analyses. See Appendix E for the survey scales and response options.

In our analyses of the student data—for Section 4, on changes in classroom learning environments, and in Section 5, on changes in student attitudes—we compared pre-post (Fall and Spring) responses from all students, and from matched groups of students who completed both surveys. In calculating significance, we conducted chi-square analyses for items that were dichotomous or categorical. Bonferroni corrections were used for analyses where there were multiple items.

In SY1, we distributed parent consent and student assent forms required by the NYCDOE to all participating a2i schools. Due to low return rates for the consents, we did not administer student surveys the first year. Consent and survey return rates were higher In SY2, when we received approximately 700 completed student surveys from consented students. In Years 3 and 4, the NYCDOE granted permission to use passive consents that parents signed only if they did not want their children to complete surveys.

That approval meant a smoother survey administration and far larger samples. The SY4 student survey sample, the basis of the findings reported here, included 10,599 students across 28 schools. (Only one school did not collect student surveys. See Appendix E, Table E1, p. 99, for the breakdown of the survey sample of schools.)

- Males and females were equally represented in the pre-post surveys (though there were fewer who included gender information at posttest).
- Approximately one-third of the students were enrolled in 9th or 10th grade at pretest. At posttest, the largest numbers of students were again enrolled in 9th grade (41.4%) or 10th grade (30.8%). Similar percentages of the students at pretest and posttest indicated that it was their first year at the school (36.0% and 58.6%, respectively), likely because of the larger numbers of freshmen or sophomores taking Algebra I and Geometry.

SUMMARY OF FINDINGS

Overall Findings

- Our multiyear studies of a2i portray meaningful impacts on students' engagement with mathematics and teachers' classroom practice. We did not, however, find improvements on the state's Common-Core aligned measures of academic performance in mathematics when a2i students are compared with similar students in similar schools.

- A supplemental, exploratory analysis of college readiness and algebra pass rates among ninth graders found that while pass rates were nearly identical between a2i and non-a2i schools in Year 1 of the 2-year impact study, the addition of Instructional Activities to the a2i practices the following year showed greater improvement in both pass rates and college readiness at a2i schools than at the comparison schools. These exploratory findings suggest further study of impacts associated with the final version of a2i is warranted.

- Both the fidelity ratings and teachers' reports of a2i's critical components reflect the changes in a2i's last two years. Changes in the a2i model meant that schools did not meet fidelity thresholds during those two years, but also suggest that the a2i model is adaptable, and that certain components remained at the core of the model, if gradually and variously redefined.

- Feedback from teachers indicated two implementation tracks based on length of involvement: Returning teachers from the early, first two cohorts comprised one track, and teachers new to a2i, many of them new to the profession, the other. Both took ownership of the project, but prioritized core components differently. The returning group attached more importance to the original curricular components such as the unit schematic and formative assessment lessons higher; the newer group, reassured by strategies they could easily incorporate into their toolkits and daily lessons, defined the project by curricular elements such as the Instructional Activities.

- Teachers' feedback also indicated two levels of implementation: those who considered their classrooms to be a2i classrooms, and those who chose to mix a2i with other instructional strategies and resources. The differences between these groups, and between the early and recent groups of teachers, point to trends rather than certainties, and are generally descriptive rather than statistically significant. However, in the composite, they do suggest that the a2i model has staying power and, even with variable implementation, can lead to desired changes.

- Learning environments shifted with a2i, as students engaged in more group work, talked more about math, and solved problems together. Although achieving this goal did not lead directly to student achievement goals, evidence appeared early and remained one of a2i's durable impacts.

- Our findings confirmed the importance of context—factors external to or beyond the control of the project, including uncertainty around the Common Core exams, changes in teacher evaluations, elimination of the inquiry mandate; and internal school factors, including collaborative structures and precedents and decisions about which students would sit for exams. a2i's implementation and its impact on inquiry practices around formative assessment were dependent on both.

Implementation and Fidelity

- Fidelity ratings for a2i's first two years show a marked difference between the first two years of the project, when schools met thresholds for fidelity of implementation, and the latter two years—following some significant project changes, including restructuring professional development and developing curriculum to meet the evolving needs of teachers. Changes in the a2i model meant that schools did not meet the originally defined fidelity thresholds during those two years, but also suggest that the a2i model is adaptable, and that certain components remained at the core of the model, if gradually and variously redefined.

- Teachers consistently assigned high ratings to a2i's curricular components, and their ratings for Professional Development and online resources formed an important triad around curriculum: among the PD sessions they rated highest were the monthly sessions devoted to curriculum. In their final (2015–16) survey assessments, the core feature that all teachers found most critical, most effective, and most instrumental in changing practice was the frequently visited a2i website, with its curricular resources.

- Inquiry-team work declined, due, in large part, to a shift in district policy lifting the mandate for school inquiry teams. Departmental work, however, remained important. Teachers continued to welcome the opportunity to discuss strategies, unit goals and designs, and applications of the IAs with colleagues in all-day professional development sessions.

- Teachers also rated a2i's core support elements differently, depending on how long they had participated. For example, professional development was more important for those newer to the project, as was individual support from Instructional Specialists (83.9%).

Impacts on Student Achievement

Confirmatory Analyses

- Findings showed that, both with and without controlling for covariates (including student-level prior math and reading achievement, ELL status, IEP status, and ethnicity), there was no statistically significant effect associated with a2i on students' New York State Regents Common Core Algebra scores.

- Like the findings for algebra, there was no statistically significant effect of a2i treatment on geometry performance despite controlling for key covariates (including student-level prior math and reading achievement, ELL status, IEP status, and ethnicity). Unadjusted differences in means between a2i and non-a2i were compared, and results again showed no statistically significant difference.

Exploratory Analyses in Algebra

- In comparing group averages, we found that while pass rates (Algebra Regents score of 65 or higher) were nearly identical Year 3 between a2i and non-a2i schools, there was a greater improvement in the pass rates at the a2i impact study schools than in comparison schools (9.96 percentage point increase vs. 5.83) in Year 4, after professional development was restructured and instructional activities (IAs) were introduced. In addition, the improvement in the pass rate was even greater when the a2i charter schools and the third wave of a2i schools were also considered (12.99 percentage points vs. 5.83). Likewise, college readiness rates (Algebra Regents score of 70 or higher) for both the impact study a2i group and the group consisting of all a2i schools showed a greater improvement in percentage points (17.82 and 20.27, respectively) than among the comparison group (14.25). Clearly, regardless of a2i status, there was improvement in these rates overall, but the greater gains for a2i after the introduction of a new professional development structure and additional curricular materials suggest that modification to the intervention may have led to improvements in algebra learning for a2i students.

Teachers' Beliefs and Practices

- Throughout the project, most teachers endorsed a2i's approach to teaching math. Differences in length or level of involvement with a2i generally did not affect beliefs, although some teachers newer to the project were uncertain about the interplay between basic operations and big ideas. This may have been less a matter of teaching experience than of experience with the new Common Core-aligned Regents exams and concerns about students' performance.

- a2i teachers taught for understanding rather than memorization. Students acknowledged this approach, and indicated that they were not afraid of making mistakes as long as they were learning. Students also indicated that teachers checked to make sure they understood what was being taught—and conveyed that it was okay not to understand something.

- Most a2i teachers reported at least modest changes in their approach to teaching math, peer collaboration, and use of data and formative assessment. Those with the project longer—who also had more time to adjust to the a2i approach and assessments introduced by standards—cited greater changes in students' understanding of basic concepts and processes.

- Based on teacher reports, a2i's greatest impact was related to the Common Core Standards for Math. Almost all—90% of the teachers completing the 2015–16 survey—cited a "modest" or "significant" change, in their understanding of the Common Core Standards, equal percentages (44.8%). A third also cited "significant" changes in their ability to prepare students for the Common Core-aligned Regents exams; another 70%, a "modest" change. Algebra I teachers who used a2i tended to have higher mean ratings of the amount of change in their own understanding of Common Core standards (*M* = 2.44 vs. *M* = 2.00, a2i vs. mix of a2i, respectively) and students' understanding, compared to those who reported using a mix of a2i and other strategies and resources (*M* = 2.13 vs. *M* = 1.70, a2i and mix of a2i, respectively).

Changes in Learning Environments

- Teachers' feedback indicates that learning environments changed, moving away from an emphasis on getting the right answer to understanding math—a change in which neither teachers' longevity nor fidelity seemed to be a factor. Based on the scales that included items related to both teacher and student behaviors, agreement and frequency levels rose from Fall to Spring.

- In the changed classrooms, students engaged in more group work, talked more about math, and solved problems together. Again, students' feedback mirrored teachers'.

Students' Interest and Efficacy

- The overall finding from the student surveys is that a2i students were positive about math: the majority were confident in their skills and interested in the subject—though a little less sure about its long-term utility. There were slight changes from Fall to Spring, but no clear differences or trends based on subject (i.e., Algebra, Geometry, and Algebra II) or whether teachers considered their classes to be a2i classes or a mix of a2i and other instructional strategies and approaches.

- At both survey administrations, almost all students in a2i classes indicated that they believe they can learn math (97% Fall and 96% Spring; *M* = 3.49, *M* = 3.45), and that they can get good grades (91%, 90%; *M* = 3.22, *M* = 3.22). Self-reported survey data also indicated that students thought it was important to do well in math class (97%, 94%; *M* = 3.47, *M* = 3.38). Fewer students, but still a majority, are confident they could do well in more advanced math classes (57%, 59%; *M* = 2.59, *M* = 2.66). Just under two-thirds of the students said their teacher made math interesting (61%, 62%; *M* = 3.73, *M* = 3.77) and enjoyable (62%, 62%; *M* = 3.76, *M* = 3.75).

- Other analyses showed that students were more confident overall that they could understand the basic concepts in their math classes than the harder ones, but the former edged down from Fall to Spring by a few hundredths of a percentage point (*M* = 3.21 to *M* = 3.19), and the latter edged up (*M* = 2.70 to *M* = 2.71) Students' ratings for interest in math and taking more math classes in high school were lower (*M* = 2.49 Fall; *M* = 2.55 Spring), as was their confidence in their ability to do well in more advanced math classes (M = 2.59 Fall; M = 2.66 Spring), but both increased, and the differences were statistically significant.

Changes in Schools

- The implementation fidelity ratings show that, at the school level, a2i schools met the thresholds set for inquiry team quality indicators all four years of the study. Our reviews of teachers' ratings of a2i's core components also show that teachers saw the inquiry-team work as critical, but not as effective or instrumental to their implementation as the PD, IS support, or curricular resources.

- Final year (2015–16) teacher survey results suggest relatively frequent engagement in collaborative inquiry activities, and school support for those activities. Over half of respondents, for example, agreed that teachers had common planning time with their departmental teams and that there was an emphasis on peer collaboration. Three-fourths of the respondents reported that they collaborated with colleagues to design math lessons or units "almost always" (34.4%) or "pretty often" (41.0%).

- Assessments of change indicate that, on average, 40% of the teachers reported engaging in inquiry or formative assessment practices more than they did prior to a2i. For the matched group of a2i teachers for whom we had baseline and follow-up survey responses, there was a statistically significant difference in the emphasis on peer collaboration between the two survey administrations (p = .003).

- For most statements in the school culture cluster of survey items, agreement levels were slightly higher among teachers who began a2i earlier, compared to the more recent group. The more recent group, however, agreed more strongly that there was a school-wide focus on formative assessment, and the difference between the two groups was statistically significant (p = .004). Other feedback suggests that, although inquiry team work declined in some schools, in others, departmental team work around the instructional activities and strategies, often embraced by newer teachers, grew stronger.

1 a2i's Core Elements
IMPLEMENTATION FIDELITY

❖ *What are the critical, measurable features of the a2i program's professional development and instructional components?*

One of the goals of i3 projects was to "codify" core elements, not only to confirm their importance but also to guide those adopting or replicating the effort. To identify the elements critical to the a2i model, we combined a theory-of-change and outcomes-based model. What this allowed us to do was provide formative feedback to project leaders about what was working, what was not, and why, and, as project leaders and teachers defined the core elements (see a2i logic model in Figure 2), examine the hypothesized relationships between them.

According to that model, teacher and student outcomes intersected at direct and indirect points: in the shorter term, changes in teaching strategies, including more group work and "math talk," would invigorate classroom learning environments. The less direct path led from teachers' engagement in core components—high-quality formative assessments, collective inquiry, coaching, ongoing professional development—to instructional changes that would, in the longer term, influence students' achievement and attitudes.

Mapping the critical components of a project and the relationships between them can keep a project on track and provide a blueprint for others. The i3 validation projects involve some particular challenges in that they require confirmatory evidence that the logic works, but also a scale and complexity that can stress the links. For a2i, testing of relationships became a kind of stress test with broader ramifications, to see if the links could withstand the variations introduced into the model as the project expanded, adapted to teacher and school needs, and pivoted based on uncertainties around high-stakes exams.

To begin to codify the elements of the model validated by a2i's implementation, this section describes a2i's core elements, adjustments project leaders made over the course of the project, final ratings for implementation fidelity, and teachers' ratings of elements critical to their implementation—and to changes in practice.

FIGURE 2. THE a2I LOGIC MODEL

Inputs/A2I Resources	**Key Components: Inputs & Activities**	**Outcomes / Impacts**		
		Short-term	Medium-term	Longer-term
• Diagnostic assessments designed by SVMI & NV • Formative Assessment Lessons (FALs) • Post-unit assessment • Instructional specialists (ISs); NV Leadership Development Facilitators (LDFs); SVMI facilitators • NV ongoing support • School conditions conducive to collaborative inquiry: Existing school inquiry teams • DataCation; Skedula, Standards based gradebook (SBGB) • CRM database • Expert advisory committee • External evaluation feedback • Federal and non-federal funds	• Professional development for teachers: ○ Weeklong intensive summer workshop & additional workshops ○ Quarterly cross-school inquiry & assessment analysis sessions ○ Monthly cross-school sessions • School support for collaborative inquiry: ○ Weekly math inquiry team meeting ○ Collective analysis of student work & identification of student learning needs in math inquiry teams • Training for admin.: Information sessions for principals and instructional leadership teams in inquiry framework • Instructional practice changes: ○ Common, year-long sequence of math units aligned with CCS; examination of same core math tasks at roughly same time, within & across sites; collaboratively developed re-engagement lessons for each unit (algebra and geometry courses each have 7 units) ○ Formative assessment strategies consistently embedded in instruction (each unit): Implementation of all FALs & all re-engagement lessons • Routine use of student data management system to monitor students' progress	• Increased confidence among teachers in using inquiry cycle • Cross-school course-specific teams & sharing • Increased use of student data management system to document/monitor work	• Increased teacher skill at cyclically designing effective learning environments based on formative assessment feedback (aligned with challenging content & CC standards) • Increased instructional effectiveness • Extended student work on problems or projects • Increased student skill & comfort in proposing/using multiple methods to solve problems • Translation of Common Core Standards into Practice • Increased school-wide focus on formative assessment • Initiative resources and findings publicly available through Dissemination & Knowledge Management Center (DKMC)	**Student Learning & Achievement** • Increased math learning & achievement: post-unit assessments, end-of-year SVMI summative performance test • Improved grades • Improved student achievement on standardized assessments (Regents) ○ 80% of Ss pass Algebra I Regents exam performance with a 65 or higher ○ 70% of Ss enrolled in Alg II pass NYS Alg II/Trig exam **Student Motivation & Persistence** • Increased confidence & interest in math • Enrollment in higher-level, higher challenge math courses ○ 75% of Ss in Alg I, Geom enroll in Alg II • Credit accumulation; progress toward graduation **School-wide Practice & Culture** • Inquiry & formative assessment part of math tchr effectiveness determination, all schools • Inquiry & formative assessment used in other content areas at 90% of schools • Readiness for new Common Core Assessments • Enhanced collective school efficacy • Utilization of DKMC in NYC, Boston, & beyond

The a2i Theory of Change

a2i's leaders and partners charted a path for participating teachers and schools that began with their introduction to a set of high-quality instructional resources, along with professional development, coaching, and data support structures designed to help them shift to a new approach to math instruction, aligned to the Common Core Standards for Mathematics, and defined by a focus on formative assessment and analysis of student work.

1. First, teachers used performance-based assessment tools focused on what students should understand and be able to do in their study of mathematics. A core feature of the a2i intervention introduced high-quality assessment tasks throughout each math unit.

2. Second, teachers were guided to engage in meaningful inquiry, using data to craft an instructional strategy that addresses gaps in comprehension or performance.

3. Third, teachers were supported to develop skills in implementing quality mathematics units, aligned to the Common Core Standards and designed to deepen student understanding and ultimately improve achievement, especially for underperforming students.

4. The fourth and final key goal for professional development encouraged teachers to engineer effective learning environments by using evidence about learning to inform subsequent instructional moves.

Evolution of a2i's Core Elements

As the project progressed, there were multiple changes, necessitated by the addition of more schools and teachers, teachers' experiences using a2i resources and assessment tools, and uncertainty around the new Common Core-aligned assessments. The changes, described in more detail below, included changes to the formative assessments, the coaching, the professional development, and the emphasis on inquiry teams.

Changes in Professional Development

Partly due to the continued participation of teachers from year to year, the a2i professional development sessions needed to introduce new topics from session to session. The a2i team had to balance the continuing teachers' needs for new (or deeper) content with the needs of teachers new to a2i. Professional development topics also changed over time because the a2i team developed new curriculum and resources. The structure of the PD sessions changed from cross-school, quarterly sessions and monthly, after-school workshops to full-day, unit-based workshops that met six to seven times during the school year.

Additions to professional development resources included a project handbook, which was provided to teachers starting in Year 3, and subject-area booklets with expanded unit plans structured around big ideas.

Coaching Changes

The a2i design, and plan for Years 1 and 2, called for coaching to be provided 1:1 for teachers, in addition to weekly support with each math inquiry team. However, as the number of a2i schools increased, the a2i coaching staff, even though it had doubled from three to six coaches, had to make changes and forego the unsustainable model of 1:1 coaching. Teachers could request coaching assistance when needed, but more effort was devoted to supporting the math inquiry team and department at each school.

Assessment Changes

There were other changes in the implementation of assessments, in part due to feedback from teachers about the fit of certain assessment tools with students' skills and needs. For example, after the first year, the Initial Performance Assessment Task (IPAT) was viewed more as an optional component of every unit. In the latter years, the summative assessment tools for each unit shifted from the Final Performance Assessment Task (FPAT) to the Balanced Assessment tool developed by the a2i team. A summative performance assessment, adapted from SVMI, was administered at the end of the year in all a2i algebra classes during Years 1 and 2 for measuring end-of-year performance, but discontinued and not replaced during the impact study years (Years 3 and 4). Instead, the elements of the end-of-year SVMI assessment were integrated into the Balanced Assessment tool administered at the end of each unit.

Beginning in Year 3, the a2i team also changed the teachers' process for using the data management system from uploading all the completed formative assessments (each scanned assessment document) to instead simply entering the balanced assessment results into a pre-populated spreadsheet. This change meant that teachers could immediately view summaries of student performance data instead of having to manually generate reports through the reporting system. It also reduced the error rate to virtually zero, which resulted in more reliable data.

Changes in Inquiry Teams

The original intention was to have a single math inquiry team at each school, which would serve as a model for the longer-term goal of having strong inquiry teams in each of the core content areas at a2i schools. As teachers came to increasingly speak a common language and school-wide support for inquiry grew, these inquiry teams would become an integral part of research-based instruction rather than a tangentially related collaborative activity. As identified in the original project proposal, the a2i teams would engage in an inquiry cycle that included the following key steps:

- Analysis of student work and data and identification of student needs prior to the implementation of the specific formative assessment lesson, or FAL

- Implementation of the FAL and an intensive examination of classroom discussion, assessments of student understandings evidenced by student discussion, work and self-identified need

- Identification of gaps in curriculum, pedagogical approach, and school support systems
- Modification of the units, selected re-engagement and re-teaching and the provision of additional opportunities for practice and analysis of key concepts or types of problems

The inquiry team goals had to be redefined for various reasons. When the project began, the NYC DOE mandated schools to have teacher inquiry teams, and the a2i program was developed in part as a response to the need for greater support in implementing inquiry teams of high quality and consistency. Yet, a couple of years into the project, NYC DOE no longer mandated inquiry teams, and the a2i team subsequently reduced its focus on inquiry teams. There were also operational challenges. Initially, the focus was on algebra in the inquiry teams, but when geometry implementation began, attention to each subject had to be divided in some fashion, since many teachers only taught one or the other subject during a given year. Cross-subject teams were not only an operational challenge due to a lack of common meeting times, but it was also more natural for teachers to focus on either algebra or geometry, or for there to be a separate inquiry team for each. As noted above, coaches' overall effort also had to be diminished for each school as the project scaled, and, by Years 3 and 4, the frequency of coach participation in inquiry teams intentionally decreased from the original goal of weekly meetings to a range of biweekly to monthly, depending on the choices of school.

Curricular Changes

Perhaps the most important shift in the a2i project was an increased focus, in the second half of Year 3, on a more fully developed curriculum. Many teachers had asked for curriculum—as opposed to a curricular framework—from the outset, to help them transition to the new math standards. They welcomed the common set and sequence of math units, and found many of the resources from SVMI's Mathematics Assessment Collaborative (MAC), the Shell Centre Formative Assessment Lessons (FALs), and Mathematics Assessment Resource Service (MARS) helpful, but still felt a need for more activities to implement each unit, and activities to incorporate into daily lessons.

In response to their needs, project leaders had fleshed out the structure somewhat in Year 2, adding more in Year 3, including the IAs, which focused on the day-to-day pedagogical practices that complemented or mirrored the practices embedded within the formative assessment activities.

The introduction of IAs into a2i practices was based on a confluence of teacher feedback, classroom observations, and unsatisfactory Regents performance, and viewed as a necessary and effective addition to the a2i intervention, but it was not wholly outside the original plan, or theory of change, which assumed that teachers would naturally start to use IA-like practices because of their experience with using the rich formative assessment activities of a2i. According to New Visions' Deputy Director of Instruction, Russell West, this shift to support for day-to-day instruction with more nimble activities and unit guides reflected an effort to "lighten the burden on teachers," so "instead of having to worry daily about curriculum, they could focus on the students."

By the project's end, the final set of curricular resources for every a2i unit included:

- An initial task so that teachers could get a sense of their students' knowledge coming into a unit
- A balanced assessment so that teachers could measure their students' progress relative to end-of-year state assessments
- A detailed description of the math to be taught, apportioned to week-long chunks called Big Ideas
- Formative assessment lessons so that teachers could respond to student needs before summatively assessing them
- General resources for implementing re-engagement lessons so that teachers could respond differently than during initial instruction
- Tasks (including lesson plans, printable posters, and slides) for instructional activities so that teachers would have resources available when implementing ideas from professional development sessions in their own classrooms.

See Table 1 for a yearly breakdown of developed curricular resources.

TABLE 1. NUMBER OF CURRICULAR RESOURCES CREATED BY a2I, BY YEAR

	Core Tasks Created	**Instructional Routine Tasks Created**	**Initial Tasks Curated**	**Balanced Assessment Tasks Created**
SY2012–13	0	0	7	7
SY2013–14	0	0	4	4
SY2014–15	0	23	14	14
SY2015–16	4	235	2	2

There were other factors that prompted curricular changes, including teachers' concerns that not all students were ready for a2i. This was in part a concern about the FALs, and the lack of an access point for students, and a belief that teachers would need to make up some ground for students whose first eight years of mathematics instruction did not always prepare them for Common Core practices. On teacher surveys, approximately 20% of teachers had indicated that their school provided pre-Algebra, Math lab, double-blocking in math as well as other resources. Schools also offered, repeater courses, numeracy classes, along with after-school tutoring and clubs or other extra-curricular activities to prep students for exams. In Year 3, a2i project leaders added to those offerings a companion curriculum, *Transition to Algebra* (Heinemann.com/transitiontoalgebra/), created by the Education Development Center to ready students for Common Core algebra. On the 2015–16 survey, just over one-third of teachers (36.4%) reported that their school used *Transition to Algebra* or another remediation program.

Fidelity Matrix Ratings

In collaboration with the a2i team, we developed a fidelity matrix based on the logic model—designed, as noted at the beginning of this section, to "codify" a2i's core elements. Fidelity is also a way to measure or confirm the “if” part of a conditional statement, the contingency upon which the “then”—the student achievement outcomes that are the subject of Section 2—depends.

Over the course of the project, the fidelity matrix remained a living document, and some elements in the matrix were revised and new thresholds and benchmarks set to reflect the project changes described above. Table 2 provides the fidelity status by component and year of implementation. By this metric, a2i schools met the benchmarks in Year 1 and 2, but did not attain thresholds for fidelity of implementation in the last two. While the a2i team believed that it was important for every school in the first two years to have common foundational experiences, in the latter two years, the a2i team developed a more differentiated approach to the support that schools received. Teachers appreciated this differentiated approach, as indicated in surveys and focus groups. Teachers who were newer to the project and to the classroom required more coaching support and practical instructional aids for daily lesson plans, like IAs, and saw them as a practical, accessible, concrete way to encourage students to look at mathematical structure. Teachers who began a2i earlier in the project’s life also welcomed the curricular support, and praised the IAs along with the curriculum units explored in monthly meetings, but relied less on project support, and tailored implementation more closely to their and their students’ needs.

In the final year of the project, a2i shifted to providing full-day professional development sessions that were unit-based and grounded in the curriculum that had been developed over the previous year. Schools in both cohorts met the fidelity criteria for substantial participation in Year 4, with 58 Algebra I teachers, 43 Geometry teachers, and 24 Algebra II teachers attending a total of 2,334 hours of professional development over the course of the school year (averaging nearly 19 hours of professional learning per teacher). In the final survey, most teachers ranked the professional development components among the most critical. Having conversations and collaborations with other a2i math teachers (87.5%) and attending all-day professional development sessions during the school year (82.4%) ranked as the most critical PD activities, and most effective in teachers’ implementation.

Due to the shift to full-day professional development sessions, only two cross-school quarterlies were offered in Year 4. All but one school (95%) participated in these cross-school sessions.

TABLE 2. COMPONENT AND INDICATOR STATUS FOR FIDELITY OF IMPLEMENTATION: YEARS 1–4

Key Components of a2i	Key Indicators for Each Key Component	Fidelity Status			
		Prior to Impact Study		Impact study	
		Y1	Y2	Y3	Y4
1. Professional Development	• Summer workshop (school level)	Met	Met	Unmet	Met
	• Quarterly cross-school assessment and analysis sessions (school level)	Met	Met	Met	Unmet
	• Monthly cross-school sessions (school level)	Met	Met	Met	Met
	Component-level status	Met	Met	Unmet	Unmet
2. School Support for Collaborative Inquiry	• Coaching for math department from Instructional Specialist (school level)	Met	Met	Met	Unmet
	• Weekly math inquiry team meeting (school level)	Met	Met	Unmet	Unmet
	• Inquiry team meeting quality indicators (school level)	Met	Met	Met	Met
	Component-level status	Met	Met	Unmet	Unmet
3. Training for Administrators	• Information sharing for principals or assistant principals in inquiry framework	Met	Met	Met	Met
	• Signed MOU	Met	Met	Met	Met
	Component-level status	Met	Met	Met	Met
4. Instructional Practice Changes	• Common set of math units	Met	Met	Met	Unmet
	• Common sequence of core math units	Met	Met	Met	Met
	• Implementation of all FALs	Met	Met	Unmet	Unmet
	• Enactment of all re-engagement experiences	Met	Met	Unmet	Unmet
	Component-level status	Met	Met	Unmet	Unmet
5. Routine Use of Student Data Management System to Monitor Students' Progress	• Routine use of student data management system (teacher level)	Met	Met	Unmet	Unmet
	Component-level status	Met	Met	Unmet	Unmet

Teachers' Perspectives on Core Components

As a2i moved forward, it was clear, as with any multi-year project, that project leaders would need to make some adjustments to accommodate teachers' needs and project expansion, even programmatic changes in core components like the adjustments described above. It was also clear that there was substantial variability in implementation: schools and teachers focused on certain components as they gauged their students' skills and needs and adapted to the new standards and assessment.

Ratings across Years

2013–15 Teacher Surveys

Starting in Year 2, we canvassed teachers about a2i's core components to see what they considered central to the project and their implementation. The Year 2 survey asked teachers to rate core components, on a scale that ranged from "not at all effective" to "a critical component." We then created four survey scales that, based on our factor analyses, formed closely related groups,[1] and continued surveying teachers through SY2015–16. This set of survey questions provided a way to track how teachers responded to project adjustments and a2i's evolution.

The Year 2 and the Year 3 results indicated that teachers thought most components were important, but prioritized them in different ways. Professional development was critical, but informal opportunities to share and collaborate with other teachers was as—if not more—important to teachers than the August sessions that served as a kick-off for the school year. The curricular elements were also critical, but the FALs and other assessment resources were not as important as the unit sequence. In Year 3, teachers ranked the re-designed a2i website as the project's most critical component.

2015–16 Teacher Survey

On the SY2015–16 survey, using the same survey scales, we asked teachers three sets of questions: 1) whether they considered individual elements in each scale or construct to be "critical components" of the a2i model; 2) how effective each element was in supporting implementation and instructional changes; and 3) what impact each had on their practice.

What we found was not so much a change in the core components as continuing shifts in priorities (see Table 3). Ratings for Instructional Specialists' support for school teams, for example, rose in importance, while ratings for inquiry team meetings declined. At the end of the project:

- The website was considered the most critical component, and the most effective in implementation and impact. By that point in the project, the website had become a key resource for curriculum (unit plans,

[1] To ensure internal consistency, we calculated Cronbach's alphas based on 2015–16 survey responses, which indicated acceptable levels of consistency. The alphas for 4 scales were: Professional Development (5 items, α = .667); Inquiry and Support (5 items, α = .843); Curricular Strategies and Assessment (6 items, α = .803); Online Resources (2 items, α = .518).

Instructional Activities) and project information, widely used not only by a2i teachers, but also, based on web analytics, by teachers beyond the project and the district: In the final year of the project, 18,900 teachers accessed 335,000 pages on the website.

- Most teachers ranked the professional development components among the most critical. Having conversations and collaborations with other a2i math teachers (87.5%) and attending all-day professional development sessions during the school year (82.4%) ranked as the most critical PD activities, and most effective in teachers' implementation.

- The unit design—a key part of the original a2i design—remained an effective element in implementation (M=2.23), but declined slightly as a critical component. The IAs were considered critical by 98% of the teachers, though not as effective in implementation as the unit sequence.

There were other indications that teachers thought certain components were still critical, but not altogether effective, or vice versa. Teachers found Instructional Specials' support for school leaders to be fairly effective, but not as critical as individual coaching or team support; they indicated that mock Regents support was critical, but not as effective as other PD activities.

These divisions and overall ratings may also reflect the fact that teachers appeared to be on two tracks: Those newer to the project and to the classroom required more coaching support and practical instructional aids for daily lesson plans, like the IAs, and saw them as a practical, accessible, concrete way to encourage students to look at mathematical structure. Teachers who began a2i earlier in the project's life also welcomed the curricular support, and praised the IAs along with the curriculum units explored in monthly meetings, but some relied less on project support and having strategies modeled or rehearsed, and tailored implementation more closely to their and their students' needs.

TABLE 3. TEACHERS' RATINGS OF a2I'S CORE COMPONENTS AND THEIR EFFECTIVENESS

		% Critical Component	*M*	Not/Somewhat effective	Generally effective	Highly effective
Inquiry & Support	Support or buy-in from school administrators	68.9%	2.05	15 (27.3%)	22 (40.0%)	18 (32.7%)
	School inquiry team meetings	72.5%	2.08	12 (24.0%)	22 (44.0%)	16 (32.0%)
	Individual support and coaching from Instructional Specialists	82.4%	2.19	10 (17.2%)	27 (46.6%)	21 (36.2%)
	Instructional Specialists' work with school teams	87.5%	2.20	10 (17.9%)	25 (32.9%)	21 (37.5%)
	Instructional Specialist support for teacher leaders	60.4%	2.12	9 (18.0%)	26 (52.0%)	15 (30.0%)
	Total Scale	72.5%	2.14			
Professional Development	August PD sessions	71.7%	2.12	11 (22.4%)	21 (42.9%)	17 (34.7%)
	Math department team meetings	75.5%	2.25	10 (16.9%)	24 (40.7%)	25 (42.4%)
	All-day a2i PD sessions during the school year	77.1%	2.27	11 (19.6%)	19 (33.9%)	26 (46.4%)
	Conversations, collaborations with other a2i teachers	71.7%	2.22	11 (19.0%)	23 (39.7%)	24 (41.4%)
	PD or support around Mock Regents prep	87.5%	2.04	15 (28.3%)	21 (39.6%)	17 (32.1%)
	Total Scale	85.1%	2.13			
Curricular Strategies & Assessment	unit design schematic (initial task, plan, FAL, re-engagement, final assessment)	68.8%	2.23	12 (21.1%)	20 (35.1%)	25 (43.9%)
	a2i curriculum	75.5%	2.16	10 (18.2%)	26 (47.3%)	19 (34.5%)
	Classroom Challenges [FALs])	73.5%	2.00	12 (21.1%)	33 (57.9%)	12 (21.1%)
	Re-engagement lessons	75.0%	1.96	16 (29.6%)	24 (44.4%)	14 (25.9%)
	Instructional Activities (e.g., Contemplate then Calculate)	98.0%	2.14	11 (19.0%)	28 (48.3%)	19 (32.8%)
	Balanced Assessment System	47.9%	2.02	17 (29.3%)	23 (39.7%)	18 (31.0%)
	Total Scale	60.0%	2.09			
Online Resources	a2i website – as a source for resources & information	68.9%	2.48	5 (8.9%)	19 (33.9%)	32 (57.1%)
	a2i website – as a place to share/interact	72.5%	1.80	22 (40.0%)	22 (40.0%)	11 (20.0%)
	Total Scale	82.4%	2.13			

Source: SY 2015–16 a2i Teacher Survey. Note. Teachers who marked "N/A" were not included in the table. Scores ranged from 1 (Not/Somewhat effective) to 3 (Highly effective).

Historical Teacher Survey Sample

Pre-post survey responses from matched sets of around 50 teachers from all cohorts (2012–13, 2013–14, 2014–15, and 2015–16) showed that they also assigned high ratings to collaborative sessions—departmental meetings, all-day PD sessions, conversations with colleagues—and considered them as some of the most critical components of the project, but ratings edged down slightly on their follow-up surveys. Their ratings for curricular resources—the a2i curriculum, the unit schematic, the Instructional Activities—also ticked down a few hundredths of a point, though teachers still considered them critical. The activities for which ratings rose among the matched group were Instructional Specialists' work with school teams, and the a2i website, for which the increase was statistically significant ($M = 3.08$ vs. $M = 3.51$; $p = .024$).

Core Components and Changes in Practice

On the 2015–16 survey, we also asked teachers how much each component had changed their practice. Ratings in general mirrored previous ratings, but also indicated differences in what teachers consider effective and what they credit with changing practice (see Table 4).

- The a2i website again received the highest ratings: 57.1% of the teachers said it had a "significant" impact on their practice; another 33.9%, a "modest" impact ($M = 2.48$).

- All-day PD sessions ranked highest on the PD scale ($M = 2.27$), followed closely by math department meetings ($M = 2.25$), and conversations with peers ($M = 2.22$); just under half of the teachers reported that these PD activities brought about a "significant" change in practice.

- For the school support scale, Instructional Specialists' work with school teams ($M = 2.20$) and individual support ($M = 2.19$) ranked highest, even though the latter was no longer a key part of the a2i model.

- The unit design schematic was considered most instrumental ($M = 2.23$) among instructional or curricular components.

TABLE 4. WHICH a2I CORE ELEMENTS CHANGED PRACTICE (*N* = 49–58)

		M	Negative/ Little or no change	Modest change	Significant change
Professional Development	August PD sessions	2.12	22.4%	42.9%	34.7%
	Math department team meetings	2.25	16.9%	40.7%	42.4%
	All-day a2i PD sessions during the school year	2.27	19.6%	33.9%	46.4%
	Conversations and collaborations with other a2i teachers	2.22	19.0%	33.9%	46.4%
	PD or support around Mock Regents prep	2.04	28.3%	39.6%	32.1%
School Support	Support or buy-in from school administrators	2.05	27.3%	40.0%	32.7%
	School inquiry team meetings	2.08	24.0%	44.0%	32.0%
	Individual support and coaching from Instructional Specialists	2.19	17.2%	46.6%	36.2%
	Instructional Specialists' work with school teams	2.20	17.9%	32.9%	37.5%
	Instructional Specialist support for teacher leaders	2.12	18.0%	52.0%	30.0%
Instructional Strategies and Assessments	a2i unit design schematic	2.23	21.1%	35.1%	43.9%
	a2i curriculum	2.16	18.2%	47.3%	21.1%
	Classroom Challenges (FALs)	2.00	21.1%	57.9%	21.1%
	Re-engagement lessons	1.96	29.6%	44.4%	25.9%
	Instructional Activities	2.14	19.0%	48.3%	32.8%
	The Balanced Assessment System	2.02	29.3%	39.7%	31.0%
Online Resources	a2i website—as a source for resources, and information	2.48	8.9%	33.9%	57.1%
	a2i website – as a place to share/interact	1.80	40.0%	40.0%	20.0%

Source: SY2015–16 a2i Teacher Survey

2 Impact of the a2i Model

STUDENT ACHIEVEMENT

❖ *What impact did the a2i model have on Algebra I and Geometry instruction and student achievement on state exams?*

A constant throughout a2i—among project leaders and school principals—was the desire to support teachers and prepare students as they made the transition to the Common Core Math Standards and new exams. Project leaders assumed, as they initially aligned project activities, that, by the project's last years, teachers would be relatively familiar with the new exams.

Delays in the new exams thwarted those efforts. In Year 1, even though a2i's first cohort Algebra 1 teachers had begun folding the new standards in practice, their students were sitting for the old state Regents Integrated Algebra exam, aligned to New York State's 2005 math standards. There was some reassurance in the fact that, if students understood the big ideas, they would do fine on the old exams, but that was not necessarily the case, and, even as teachers were reassured, new metrics for teacher evaluations were introduced, which prompted more anxiety about the new standards and assessments.

Although the new Algebra I exam was available at the end of a2i's second year, uncertainties and doubts led to an option for schools: students could take either the new or the old exam, and some did both.

a2i Impact Study Design

The impact study used a school-level quasi-experimental design (QED) to estimate impacts of a2i on participating high school students in algebra and geometry compared to matched peers in matched schools (business as usual condition). The outcome measures were the New York State Regents Examination in Common Core Algebra, Regents Examination in Integrated Algebra, and the Regents Examination in Common Core Geometry. Algebra outcomes were investigated in school years 3 and 4 (2014–15 and 2015–16), after a minimum of two years of school involvement in a2i algebra. Since the a2i geometry intervention began in year 2 of the initiative, the study examined geometry performance in school year 4 (2015–16), after the schools participated in a2i geometry for two years.

This impact study was designed to answer two confirmatory research questions:

R1. Does the a2i model have a greater effect than regular or current instruction on math achievement as measured by standardized state mathematics examinations in high school Algebra after at least two years of implementation? (confirmatory contrast)

R2. Does the a2i model have a greater effect than regular or current instruction on math achievement as measured by standardized state mathematics examinations in high school Geometry after at least two years of implementation? (confirmatory contrast)

In the QED, the business-as-usual condition consisted of Algebra I and Geometry courses taught by high school teachers in non-a2i schools. These schools were selected to have similar student populations based on 8^{th} grade math achievement on the NYS examination and key demographics (including ethnicity, total enrollment, percent English learners, and percent students with IEPs). The schools were drawn from the schools in the New Visions network. Throughout the network schools, New Visions emphasized data-driven instruction and provided access to an online suite of applications for management of student data. The teachers of comparison schools had some access to resources about inquiry teams and formative assessment, as well as support from other New Visions Instructional Specialists; however, the consistency and level of engagement with those resources was expected to be far greater in a2i schools. The process of determining matched comparison schools and matched students is explained in the following sections.

Sample and Matching

The a2i schools in the impact and implementation studies were selected from the New York City public district and charter high schools that are associated with the New Visions network. In 2012, there were 71 high schools in the network, and the project objective was to introduce a2i to 30 of the schools by the end of the i3 grant. The implementation of a2i was rolled out to three cohorts of schools in Years 1–3. New Visions selected all treatment schools based on their readiness and willingness to implement the a2i program. The first cohort, involving 14 schools, began implementation in the 2012–13 school year (SY1). The second cohort, involving 12 schools, began in the 2013–14 school year (SY2). The third cohort of 5 schools began in 2014-15 (SY3). The non-charter schools in the first and second cohorts were included in the quasi-experimental impact study. The charter schools were not included because they would not have reasonable matched comparison counterparts since (a) New Visions required all their charter schools to implement a2i, and (b) given significant differences between charter and non-charter schools. The a2i schools in the third cohort were not included in the impact study because these schools did not complete the designated two years of program implementation prior to gathering of outcome data.

The pool of potential comparison schools was limited to those within the New Visions network, which operates differently than non-New Visions schools, to ensure willingness to participate in the evaluation and to increase cooperation with evaluation activities. The identification of a comparison group involved two phases of matching, one at the school level and a subsequent one at the student level.

School Level Matching Process

Comparison schools were selected through a 1:1 matching process that relied on a Mahalanobis distance (MD) metric. When calculating MDs for schools, we strived to include all baseline variables that were previously found to be highly correlated with Regents Algebra exam scores, one of the primary outcomes of interest. The variables used included scaled state test scores for 8th grade math and English language arts (ELA), percent Black or Hispanic, percent over age on entry to 9th grade, status of involvement (yes/no) in another local math initiative (UTR project), percent with disabilities, percent female, percent eligible for free lunch, and attendance rate. The process was conducted in year 1 for the first cohort and then again in year 2 for the second cohort.

Two schools that were a2i schools in Cohort 2 left the network prior to the start of year 3, the beginning of the 2-year impact study. Their matched comparison schools were dropped. As a result, another a2i school and matched comparison school were added to the impact study. This replacement a2i school was a Cohort 2 school that was originally going to be excluded from the impact study because it distributes the algebra course over two years rather than one. In summary, there were 20 a2i schools in the impact study (10 Cohort 1, 10 Cohort 2) and 20 comparison schools (10 Cohort 1, 10 Cohort 2). Note that one Cohort 1 a2i school dropped left the a2i initiative after the first year of the impact study, as did one cohort 2 a2i school. The matched comparison schools for these two treatment sites remained in the analytic samples for both years of the impact study. All algebra and geometry teachers from each Cohort 1 and 2 intervention schools that attempted to implement a2i during the impact study were included in the analytic sample for the year and subject of implementation. All teachers that were teaching algebra I or geometry in the matched comparison schools were included in the pool of possible comparison teachers/classrooms.

Student Level Matching Process

To be included in the impact analyses, which relied on a treatment-on-treated analysis, students had to possess pretest and posttest scores and had to be:

- in the ninth or tenth grade during Years 3 or 4 of the project (2014–15 and 2015–16),
- in one of the matched schools,
- enrolled in an Algebra 1 or Geometry course that year, and
- for the treatment group only, had to have the math course taught by a teacher that attempted to implement a2i that year.

Prior to the impact study, 1:1 matches were identified for each wave 1 or wave 2 a2i school that was not a charter school. In response to partial participation of math teachers at a2i schools in the lead up to the impact study, a second phase of matching was used to identify 1:1 matches for each student in an a2i classroom, which were the classrooms where the teacher attempted to implement the a2i intervention (as identified by match coach log records). The aim was to find 1:1 matches for students in a math class that was recorded to be an a2i class by the a2i Instructional Specialist. For Algebra I and Geometry, REA received records of teachers that were identified as having attempted to implement the a2i intervention.

The original planned analytic sample would have included all 9th and 10th graders with a June algebra exam score in either year 3 or year 4 or a June geometry exam score in year 4. However, because of our second phase of matching, we excluded "non a2i classes" from our primary impact contrasts. In addition, for the test of a2i effects in geometry, we included 11th graders, because although we had originally planned on only including grade 9 and 10 students, the number of grade 11 students with geometry outcome scores roughly outnumbered grade 9 students by 2:1; therefore, it seemed important to include those grade 11 students within our impact analyses. An initial concern with including grade 11 and 12 students was the number of students that could be exposed to multiple attempts at passing geometry. Students in 12th grade were deleted from the analytic sample.

We identified matched comparison students from the pool of students that had attempted the math course being investigated during the year of interest. First, we determined matches based on the covariates, and then we integrated the outcome test scores into the matched file. Next, we separately established the analytic sample for each test outcome, which meant that only students with a test outcome were included. The alternative strategy of beginning the student-level matching with only cases with a test outcome was not pursued because we did not feel confident in identifying the mechanisms for whether students completed the June Regents examination (i.e, outcome test).

After iterative attempts at propensity score matching, it was determined that using exact matching with the school pair variable, which was based on the a priori process of matching schools, was problematic to establishing balanced groups, and it was not included as a covariate. It was also determined to be beneficial to include comparison students from all comparison schools as possible matches rather than exclude students from schools that did not have a paired treatment school in the final sample for Year 4. Math and reading pretest scores were used as covariates, while exact matching was used for grade level, IEP status, EL status, sex, and ethnicity. See Appendix B for further details of student propensity score matching.

Dependent Measures

Algebra Achievement

The New York State Regents Examination in Algebra (Common Core) (www.nysedregents.org/algebraone) is a significantly different assessment than the Regents Examination in Integrated Algebra (www.nysedregents.org/IntegratedAlgebra) that it replaced. Both examinations were administered during the first year of the impact study; however, the outgoing Integrated Algebra exam was only completed by a small subset of algebra students that year, so our estimates of impact do not include those outcomes.

Geometry Achievement

The New York State Regents Examination in Geometry (Common Core) (www.nysedregents.org/geometrycc) is a new, significantly different assessment than the traditional Regents Examination in Geometry, which was phased out in project year 3. The impact study examined effects on geometry for project Year 4, which allowed for there to be at least two years of school-level implementation of a2i in geometry, as deemed necessary to attain significant intervention effects.

Impact Study Findings

This section begins with a brief description of the analytic samples and then summarizes the results of the confirmatory tests of impacts on student achievement. Details about both the analytic approach, baseline equivalence testing, and results of final model testing are in Appendix C.

Analytic Samples

After importing the outcome test scores to the algebra database of matched students, there were large differences between the treated and comparison group in the proportion of students with missing algebra outcome scores (see Table 5). There was a statistically significant difference between the treated and comparison group in the likelihood of missingness of the common core algebra examination ($X^2 = 58.49$, $p <$ .001); however, we did not attempt to impute missing data values.

TABLE 5. AVAILABILITY OF ALGEBRA OUTCOME DATA

Subsample		**Regents Common Core Algebra Exam**		
		n	Available	Missing
	Comparison	1361	913 (67%)	448 (33%)
	Treatment	1361	1089 (80%)	272 (20%)
Total		2722	2002 (74%)	720 (26%)

The resulting dataset available for student level matching in geometry had records for all 20 comparison schools. However, four of the 20 treatment schools (Bronx Theatre, CIMS, FDNY, and MACS) did not have the geometry outcome test scores. After student-level matching for geometry, outcome test scores were imported into the database. Geometry outcome test scores were available for 454 comparison students and 736 treated students in the matched file (see Table 6).

TABLE 6. AVAILABILITY OF GEOMETRY OUTCOME DATA

Subsample		Regents Common Core Geometry Exam		
		n	Available	Missing
	Comparison	1045	454 (43%)	591 (57%)
	Treatment	1045	736 (70%)	309 (30%)
Total		2090	1190 (57%)	900 (43%)

See Table 7 for a breakdown of key demographic and other background variables for both analytic samples.

TABLE 7. DEMOGRAPHIC DATA FOR COMMON CORE ALGEBRA AND GEOMETRY DATASETS

	Common Core Algebra		**Common Core Geometry**	
	a2i (*n*=2140)	Non-a2i (*n*=1731)	a2i (*n*=736)	Non-a2i (*n*=454)
Grade 9	39%	38%	20%	13%
Grade 10	61%	61%	52%	67%
Grade 11	1%	1%	28%	20%
Female	50%	51%	51%	50%
ELL	9%	9%	4%	2%
IEP	21%	21%	17%	13%
Native American	<1%	<1%	-	-
Asian	4%	5%	7%	9%
Hispanic/Latino	60%	59%	59%	59%
Black	33%	34%	29%	28%
White	2%	2%	4%	3%
Black or Hispanic	94%	93%	88%	87%

Confirmatory Contrasts

R1: Does the a2i model have a greater effect than regular or current instruction on math achievement as measured by standardized state mathematics examinations in high school Algebra after at least two years of implementation?

Common Core Algebra

Hierarchical linear modeling (HLM) was used to analyze the Regents Common Core Algebra dataset, with students at level 1, nested within teachers at level 2, and teachers nested within schools at level 3. There were 3871 students, 137 teachers, and 40 schools. All models had the Regents Common Core Algebra score (from tests completed at the end of the school year in June) at the student level as the dependent variable. The predictor and demographic variables tested in our models were 8^{th} grade state test scores in math and reading, student sex (1 = female, 0 = male), Black or Hispanic status (1 = Black or Hispanic, 0 = not Black or Hispanic), English Language Learner (ELL) status (1 = ELL, 0 = not ELL), and Individualized Education Plan (IEP) status (1 = has IEP, 0 = no IEP).

Model testing began with the unconditional model (intercept-only model). The unconditional model showed statistically significant variance in the Regents Common Core Algebra outcomes at level 2 ($X^2_{97} = 883.87$, $p < .001$) and at level 3 ($X^2_{39} = 73.42$, $p < .001$). While 70% of the variance in Common Core Algebra outcomes was explained at the student level, 22% was explained at the teacher level and 7% at the school level.

Subsequently, the random intercepts model was conducted with the grand mean centered 8^{th} grade state test in math as a level 1 predictor. This model confirmed the positive relationship between prior math achievement and Regents Common Core Algebra outcome ($b = 11.81$, $p < .001$). The effect size was computed to be .275; meaning, the 8^{th} grade math scores explained about 27% of the variance in Common Core Algebra outcome. Next, with the 8^{th} grade state test in reading also grand mean centered at the student level, results showed a positive relationship between prior reading achievement and algebra outcome (8^{th} grade math: $b = 9.76$, $p < .001$; 8^{th} grade reading: $b = 3.41$, $p < .001$).

To test the significance and direction of the relationship between the a2i intervention and Regents Common Core Algebra outcome, we conducted a means as outcomes model. This model only included a2i status (at the school level) as a predictor of Regents outcome. The results showed no statistically significant relationship between school a2i status and Regents outcome ($b = 1.98$, $p = .544$). When 8^{th} grade math performance was added back to the model, grand mean centered at level 1, there remained no significant a2i treatment effect. Iteratively, we tested models with the remaining possible predictor variables available at the three levels.

The final model for estimating a2i treatment effects on algebra, which was limited to variables that best fit the model, included covariates at all three levels. At level 1, the covariates were 8^{th} grade state test score in math, 8^{th} grade state test score in reading, Black or Hispanic status, ELL status, and IEP status. At level 2, the model included grade level and 8^{th} grade math test grand mean centered at teacher level. Level 3 predictor variables were a2i status and percentage ELL at school (grand mean centered). The results showed no statistically significant effect of the a2i intervention on student scores on the algebra outcome ($b = -0.55$, $p = .58$). Thus, despite controlling for prior achievement and key demographic variables, differences in means between the a2i and non-a2i groups were not statistically significant.

> ***R2: Does the a2i model have a greater effect than regular or current instruction on math achievement as measured by standardized state mathematics examinations in high school Geometry after at least two years of implementation?***

The same HLM approach was used to analyze the Common Core Geometry dataset as for the algebra datasets. There were 1190 students at level 1, 65 teachers at level 2, and 35 schools at level 3. The Regents

Geometry scores (completed at the end of the school year in June) at the student level were the dependent variable in all models. The predictor and demographic variables tested in our models were:

- Level one: 8th grade math test, 8th grade reading test, sex, Black or Hispanic status, ELL status, and IEP status.
- Level two: teacher total years at school, class mean 8th grade math test, class mean 8th grade reading test, class percent ELL, and class percent IEP.
- Level three: a2i status, school percent Black or Hispanic, school percent ELL, school percent IEP, school mean 8th grade math test, and school mean 8th grade reading test.

First, an unconditional model was conducted with no covariate variables at any level. The unconditional model showed statistically significant variance in the Regents Geometry outcomes at level 2 ($X^2_{30} = 158.91$, $p < .001$) and level 3 ($X^2_{34} = 79.27$, $p < .001$). The proportion of variance in the outcome at the student level was 65%, while at the teacher level it was 18% and at the school level it was 17%.

Next, the random intercepts model was run with the grand centered math pretest (8th grade state math test) entered as a level 1 predictor and both error terms were included in the level 2 model. The results supported the expected relationship between the math pretest and Regents Geometry outcome ($b = 11.68$, $p < .001$). The effect size was computed to be .364, showing 8th grade math pretest to explain 36% of variance in geometry outcome.

When grand centered reading pretest was also entered at level 1, results supported the smaller, but statistically significant relationship between reading pretest and Regents Geometry outcome ($b = 1.58$, $p < .05$). Grand centered math pretest at the teacher level had a significant relationship with the geometry outcome ($b = 19.45$, $p <.001$). When reading pretests was also grand centered at the teacher level, it showed a positive, but not a statistically significant relationship with geometry outcome (8th grade math: $b = 12.52$, $p < .01$; 8th grade reading: $b = 9.31$, $p = .06$).

The means as outcomes model tested the significance and direction of the relationship between a2i status at level 3 on Regents Geometry outcome. The results of this analysis did not support the hypothesis that a2i status predicts geometry outcome as the results showed no statistically significant relationship ($b = -4.21$, $p = .15$). The inclusion of prior math achievement (grand mean centered) at the student level as a predictor, in addition to the a2i status predictor variable at the school level, still showed no statistically significant effect for the a2i treatment (b = -1,27, $p = .21$). In subsequent models, we tested all remaining variables as predictors.
In the final HLM (shown immediately below), covariates were limited to those that best fit the model. At the student-level, there were three binary variables for Black or Hispanic, ELL, and IEP status; and the 8th grade state test scores for math and reading. There were no covariates at the teacher level. At the school level, only a2i treatment status (1 = a2i, 0 = not a2i) was included as a predictor variable.

The results of the final HLM showed no statistically significant relationship of a2i on geometry performance ($b = -1.14$, $p = .26$). Therefore, similar to the results of the impact estimates on algebra outcomes, evidence did not support that there was an effect of a2i treatment on students' end-of-year performance on the state test.

Summary of Confirmatory Tests

In examining a2i impacts in algebra, the a2i and comparison analytic samples were determined to be equivalent at baseline in terms of students' prior math and reading achievement, as assessed by state 8th grade math and ELA tests. This supported the integrity of the matching process in establishing comparable groups. Counter to expectations, the multilevel analysis showed that—both with and without controlling for covariates (including prior math and reading scores, ELL status, IEP status, and ethnicity)—there was no statistically significant difference in student performance on the end-of-course state test for Common Core Algebra.

For the estimation of impacts in geometry, standards for baseline equivalence between the treatment and control conditions were met. Student performance on the end-of-course state test was again examined with a three-level HLM. Similar to the findings for algebra, there was no significant effect associated with the a2i intervention on geometry performance, despite controlling for key covariates (including student-level prior math and reading achievement, ELL status, IEP status, and ethnicity).

Limitations

Several issues limited the internal and external validity of the impact study findings. Internal validity of the findings is limited by the opportunities that comparison-school teachers had access to some a2i materials and PD opportunities as well as access to resources, coaching, and PD from the NYC school district and other sources. As teachers were transitioning to the Common Core standards, resources were also made available on the state's EngageNY website. Data about these opportunities are limited to some general interviews with administrators and math department chairs or teachers at comparison schools, so the possible exposure of individual comparison teachers to a2i treatment or a comparable treatment is unknown.

Data quality issues also limited internal validity. For example, data on implementation of instructional activities were incomplete, and therefore maybe unreliable. Implementation of instructional activities by teachers was monitored, but implementation at the individual classroom level was largely unknown. Students without 8th grade test scores had to be left out of analyses. Similarly, many students were not included in analyses because they had not been administered the June outcome Regents examination as a culminating assessment of their learning in the course. Overall, the a2i group administered the exam at the end of the course to a significantly larger proportion of enrolled students.

External validity of the impact study findings was limited to generalizations about the high schools within the network of schools served by New Visions for Public Schools. It is unknown to what extent findings could be generalized to other high schools. In addition, due to the incomplete fidelity to the a2i intervention in the overall sample, the findings cannot be generalized to a setting where a2i is implemented with complete fidelity.

Exploratory Algebra Cohort Analysis

In the second year of the impact study (i.e., SY2015-16, the final year of the project), the a2i team introduced Instructional Activities (IAs) to the a2i curriculum and did some restructuring of the PD sessions. The IAs were the day-to-day pedagogical practices that complemented or mirrored practices embedded within the formative assessment activities. Originally, the theory of change was that teachers would naturally start to use IA-like practices because of their classroom use of the rich formative assessment activities of a2i. Based on a confluence of teacher feedback, classroom observations, and unsatisfactory achievement, the introduction of the IAs was viewed as a necessary and effective addition to the a2i intervention being implemented in classrooms.

To explore potential improvements at a2i schools associated with this change and the PD restructuring, we examined June 2015 and June 2016 algebra exam pass rates and college readiness rates for 9th graders at all a2i and comparison schools. These two indicators are key metrics used by New Visions schools to measure progress in algebra achievement. For the Regents Algebra I exam, the minimum score for passing is 65, while the minimum score to indicate college readiness is 70. Two schools that were initially in the impact study as a2i schools had left the initiative, so they were not included in the database.

In comparing group averages, we found that while pass rates were nearly identical between a2i and non-a2i schools in Year 1 of the impact study, there was greater improvement in the pass rates at a2i impact study schools than at comparison schools after the full introduction of IAs in Year 2 (9.96 percentage point increase vs. 5.83, see Table 8). In addition, the improvement in the pass rate was even greater when the a2i charter schools and the third wave of a2i schools were also considered (12.99 percentage points). For the algebra college readiness rates, both the impact study a2i group and the group consisting of all a2i schools showed a greater improvement in percentage points (17.82 and 20.27, respectively) than the comparison group (14.25). Clearly, regardless of a2i status, there was year-to-year improvement overall, but the greater gains for a2i after the introduction of IA and PD restructuring suggests the modification to the intervention may have led to improvements in algebra learning for a2i students. Further research of the final a2i intervention/model is needed, including outcome tracking over multiple years.

TABLE 8. PASS RATES AND COLLEGE READINESS RATES IN ALGEBRA

		Pass Algebra			**College Ready in Algebra**		
	Number of schools	Cohort 2018	Cohort 2019	Change	Cohort 2018	Cohort 2019	Change
Impact study a2i schools	18	38.13	48.09	+ 9.96	15.60	33.42	+ 17.82
All a2i schools	27	39.12	52.10	+ 12.99	15.31	35.58	+ 20.27
Comparison schools	20	40.07	45.90	+ 5.83	20.41	34.66	+ 14.25

FIGURE 3. ALGEBRA PASS RATES BY COHORT AND GROUP

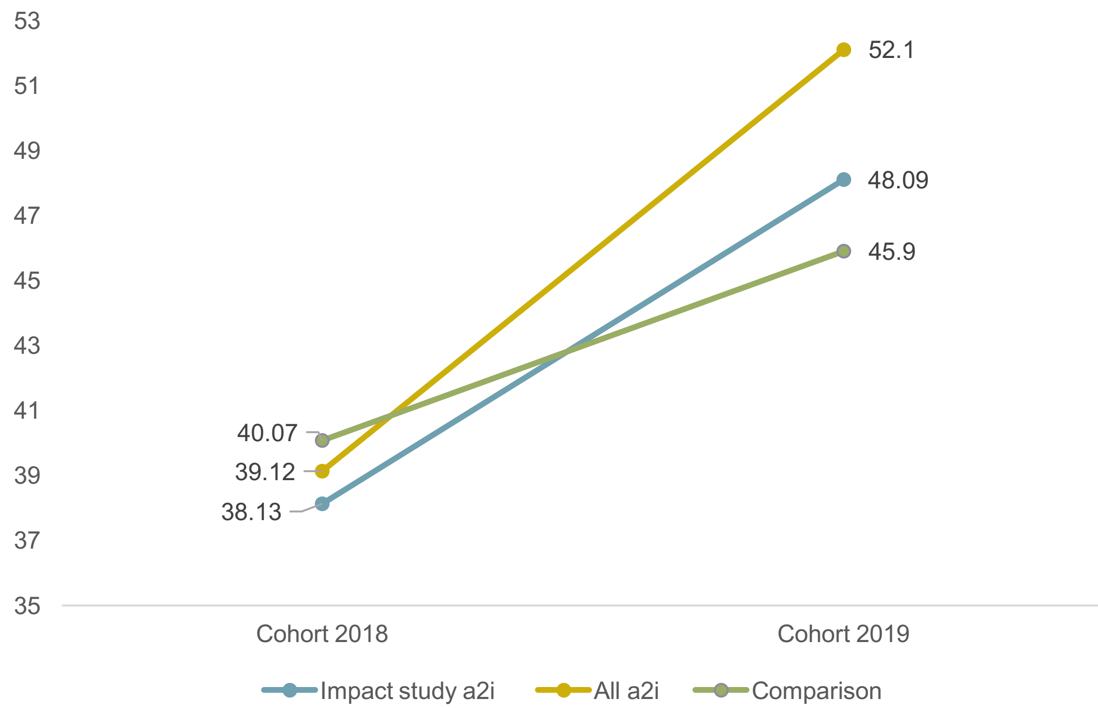
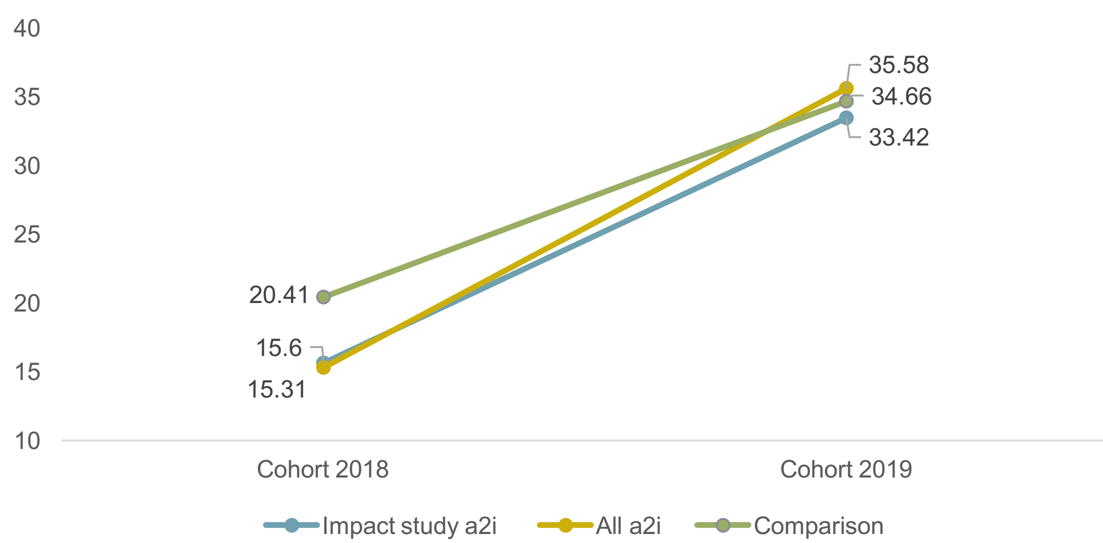


FIGURE 4. ALGEBRA COLLEGE READINESS RATES BY COHORT AND GROUP



3 Levels of a2i Involvement

CHANGES IN PRACTICE

❖ *How did a2i influence classroom practice? Did teachers' level of involvement matter?*

Although we do not know whether higher levels of fidelity to the core elements of a2i would have resulted in higher student achievement, we do know that teachers reported other changes in practice, which, for some teachers, exceeded their initial expectations.

Sections 3 and 4 of this report examine changes in beliefs and learning environments as well as practices—the medium-term outcomes charted in the a2i logic model. Both sections also explore the role played by teachers' level of involvement in a2i. The methods depart somewhat from the longitudinal sub-study outlined in the original a2i evaluation plan. The original strategy required stable achievement metrics and a large and stable corps of teachers to see whether extended involvement led to improved practice and year-to-year gains in students' math achievement and attitudes. As noted in previous sections, these conditions were not met. Consequently, we conducted a series of sub-analyses using the SY2015–16 teacher survey responses, triangulating findings where possible with the historical teacher survey sample, made up of a matched set of teachers from all cohorts who completed a baseline survey when they started the project and a 2015–16 follow-up survey; focus group findings; and students' survey reports.

We first reviewed overall changes reported by teachers completing surveys, then divided respondents into *early* and *recent* groups, based on when they joined the project, which also told us whether they began before or after the changes instituted in Year 3. We also disaggregated data based on teachers' self-reports of whether they used the a2i model intact, or mixed a2i and other instructional strategies and resources. The survey options included a third, "not a2i" option, but very few teachers selected that option, so they were not included in the analysis.

There were cross overs in the groups—recent teachers who embraced the revised model but self-identified as "a2i," and early, veteran teachers who, over three or four years of a2i experience, customized their implementation by using parts of a2i but not the full model. These groups, while not proxies for implementation fidelity, do represent a level of ownership of a2i and a commitment to an adaptable set of core elements.

Overall Feedback: What Changed?

Feedback from the 71 teachers completing the SY2015–16 teacher survey showed that the a2i intervention, especially certain core components, had a positive impact for most teachers. Interestingly—given the uncertainty around the new exams and lack of assessment evidence of differential a2i impacts—what a2i teachers say changed most was their understanding of the Common Core Standards: about 90% of the teachers completing the Year 4 survey cited a "modest" or "significant" change in understanding (see Table 9). A similar percentage (86.5%) reported modest or significant improvement in their ability to prepare students for Common Core-aligned assessments. Other data provided additional evidence of proximal changes:

…teachers need examples of what the CCSM are expecting our students to do and how we can help them get there. A2i has helped us recognize deficits and figure out routines and activities to help support them.

- Even though baseline surveys showed agreement with a2i's approach, 96% of the respondents still reported a change in their approach to teaching math: 61.2% reported a "modest" change; 34.3%, a "significant" change (M = 2.30).

- Close to 90% of the teachers reported changes in peer collaborations (M = 2.28), and, slightly fewer (82%) reported a change in their use of data to inform instruction (M = 2.13) and evidence-based strategies (M = 2.01).

These teacher-reported changes corroborate the pathway charted in the a2i logic model (Figure 2): teachers' participation in professional development and inquiry work and their use of the a2i curricular resources led to desired changes in practice.

Yet, the teacher survey data also indicated where theorized links were more tenuous, or where objectives less directly related to teacher practice were not so easily attained. For instance, although 50–60% of the teachers reported "modest" changes in students' understanding of math concepts, confidence, and interest, fairly large portions (28.4% and 38.8%) reported no change (M = 1.84, M = 1.72, and M = 1.66, respectively). Over half of teachers (56%) reported little to no change in students' enrollment in higher-level math courses. According to Russell West, that goal was likely affected by the "unstable" structure of the new Common-Core-aligned exams. What did change, he noted, was how teachers viewed Algebra I: no longer a "sorting" course that dictated whether students would enroll in higher level courses, Algebra I instead became, over the course of the a2i project, the first of a sequence of math courses aligned from year to year, reinforcing many of the same mathematical concepts.

TABLE 9. CHANGES TEACHERS ATTRIBUTED TO a2I PARTICIPATION

	M	Negative/ little or no change	A modest change	A significant change
Your understanding of the Common Core Standards	2.34	10.4%	44.8%	44.8%
Your approach to teaching math	2.30	4.5%	61.2%	34.3%
Peer collaboration/ way inquiry or departmental teams work	2.28	10.4%	50.7%	38.8%
Your ability to prepare students for Common Core-aligned assessments	2.19	13.4%	53.7%	32.8%
Your use of data to inform instruction	2.13	17.9%	50.7%	31.3%
Your classroom use of evidence-based strategies	2.01	17.9%	62.7%	19.4%
School conditions that support inquiry or other teamwork	1.94	32.8%	40.3%	26.9%
Students' understanding of math concepts and practices	1.84	28.4%	59.7%	11.9%
Students' confidence in doing math	1.72	38.8%	50.7%	10.4%
Students' interest in math	1.66	38.8%	56.7%	4.5%
Students enrollment in higher level math courses	1.50	56.0%	37.9%	6.1%

Source: SY2015–16 Teacher Follow-up Survey. Means calculated on a 3-point Likert scale:1=*little or no change*, 2=*a modest change*, 3=*a significant change*. The survey used a 4-point scale, with a *negative change*, but very few respondents indicated that practice moved backward, and those responses were combined with option 1 in a 3-point scale.

Some of teachers' expectations about a2i outcomes did not parallel certain changes evidenced in survey data. Based on matched pre-post teacher survey data, the practices or activities that changed most—peer collaboration, and a better understanding of the CCMS—also changed more than teachers had anticipated. The activities that changed least—generally those related to students, including their interest in math and understanding of math concepts—changed less than they had anticipated.

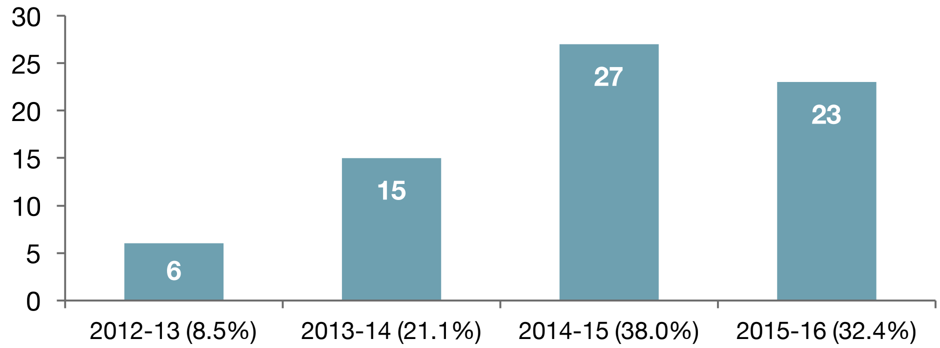
How Length of Involvement Affected Changes

In Section 1, we noted that teachers seemed to be on different tracks as the project progressed, based on whether they were new or returning a2i teachers. To further explore differences associated with length of involvement, we divided teachers into two groups:

- "Early" teachers were those who reported beginning the project in Year 1 or 2. Although the level of participation may have varied over the four-year project, these were the teachers who had been with the project the longest and had been introduced to the a2i model prior to the Year 3 adjustments. This was the smaller of the two groups, and included 21 teachers, which was about 30% of the sample.
- "Recent" teachers were those who reported beginning in Year 3 or 4, and thus came on board after changes were made to the curriculum, coaching, and inquiry team work. This larger group included 50 teachers, or 70% of the sample.

Figure 5 shows the numbers of teachers in the respondent group, by the year they began their a2i participation, and the percentages in each cohort.

FIGURE 5. NUMBER OF RESPONDENTS BY YEAR TEACHERS BEGAN a2I



Source: SY 2015–16 Teacher Follow-up Survey.

Early and Recent Teachers' Reports of Changes in Practice

Responses to survey questions about change in practice (see Table 10) showed that,

- For most items (11 out of 13; shaded in Table 10), teachers in the early group indicated greater change. Even in the case of the items where teachers saw fewer changes, the early group indicated more change than the recent group. (Items

- In two cases, the differences approached statistical significance .The early group reported greater changes in students' understanding of math concepts and practices (M = 3.05 vs. M = 2.74; p = 0.064)), and in students' enrollment in higher-level math courses (M = 2.70 vs. M = 2.35; p = 0.054). The latter may relate to West's observation about changes in perceptions if not recorded enrollment.

- In certain cases, though mean differences were not significant, the differences in percentages of respondents reporting a "significant" change appeared noteworthy. For example, percentages of early teachers reporting a "significant" change in use of data to inform instruction exceeded the recent group's reports by almost 20 percentage points. Though all teachers reported fewer student-related changes, percentages of teachers reporting "significant" changes were also notably higher among the early group.

TABLE 10. REPORTED LEVELS OF CHANGE, BY TEACHER GROUP

Item	Group	n	*M*	Negative/ Little or no change	A modest change	A significant change
Your approach to teaching math	Early	20	3.40	0 (0.0%)	12 (60.0%)	8 (40.0%)
	Recent	47	3.26	3 (6.4%)	29 (61.7%)	15 (32.9%)
Peer collaboration/way inquiry or departmental teams work	Early	20	3.45	1 (5.0%)	9 (45.0%)	10 (50.0%)
	Recent	47	3.21	6 (12.8%)	25 (53.2%)	16 (34.0%)
School conditions that support inquiry or other teamwork	Early	20	3.00	4 (20.0%)	12 (60.0%)	4 (20.0%)
	Recent	47	2.91	18 (38.3%)	15 (31.9%)	14 (29.8%)
Your use of data to inform instruction	Early	20	3.25	4 (20.0%)	7 (35.0%)	9 (45.0%)
	Recent	47	3.09	8 (17.0%)	27 (57.4%)	12 (25.5%)
Your classroom use of evidence-based strategies	Early	20	3.10	3 (15.0%)	12 (60.0%)	5 (25.0%)
	Recent	47	2.98	9 (19.1%)	30 (63.8%)	8 (17.0%)
Students' interest in math	Early	20	2.70	7 (35.0%)	12 (60.0%)	1 (5.0%)
	Recent	47	2.64	19 (40.4%)	26 (55.3%)	2 (4.3%)
Students' confidence in doing math	Early	20	2.80	6 (30.0%)	12 (60.0%)	2 (10.0%)
	Recent	47	2.64	20 (42.6%)	22 (46.8%)	5 (10.6%)
*Students' understanding of math concepts and practices	Early	20	3.05	3 (15.0%)	13 (65.0%)	4 (20.0%)
	Recent	47	2.74	16 (34.0%)	27 (57.4%)	4 (8.5%)
Student achievement in math as measured by the 2005-aligned Regents	Early	18	2.56	8 (44.5%)	9 (50.0%)	1 (5.6%)
	Recent	40	2.58	19 (47.5%)	19 (47.5%)	2 (5.0%)
Student achievement in math as measured CC-aligned Regents	Early	18	2.89	5 (27.8%)	9 (50.0%)	4 (22.2%)
	Recent	43	2.65	17 (39.5%)	23 (53.5%)	3 (7.0%)
*Students' enrollment in higher level math courses	Early	20	2.70	7 (35.0%)	11 (55.0%)	2 (10.0%)
	Recent	46	2.35	30 (65.2%)	14 (30.4%)	2 (4.3%)
Your understanding of the Common Core Standards	Early	20	3.30	1 (5.0%)	12 (60.0%)	7 (35.0%)
	Recent	47	3.36	6 (12.8%)	18 (38.3%)	23 (48.9%)
Your ability to prepare students for Common Core-aligned assessments	Early	20	3.30	0 (0.0%)	14 (70.0%)	6 (30.0%)
	Recent	47	3.15	9 (19.1%)	22 (46.8%)	16 (34.0%)

Source: SY 2015–16 Teacher Follow-up Survey.

* Mean comparisons on items were conducted using *t*-tests. For two items, differences in means approached statistical significance: Students' understanding of math concepts, $p = 0.064$; Students' enrollment in higher level math course, $p = 0.54$).

Changes in Use of Resources

As noted previously, teachers assigned high ratings to the a2i website for its resources and information. This held true for both the early and recent groups. Yet, although both groups reported frequent use of the a2i website, the recent group reportedly used it more often, and they also used EngageNY and other online resources more often. By contrast, the early group reportedly used Delta Math and textbooks more often. It is important to point out that the a2i project website was less rich and developed during the first two years of the project. The more robust website available in Years 3 and 4 showed high usage. Between August 2014 and August 2016, 28,742 users visited the a2i website and accessed 215,543 total pages.

Views on Core Components

In Section 1, discussion of teachers' perceptions of core components indicated that teachers thought most elements were important, but prioritized them differently. When we looked at how early and recent groups prioritized core components, some interesting differences emerged. The shading in Tables 11–13 indicates which group—early or recent teachers—perceived the component to be more important.

Professional Development

The a2i PD was perceived as more important by teachers newer to the project. The biggest differences were for the August PD sessions and PD around Regents prep, with a 30 percentage-point difference between recent and early teachers (see Table 11).

TABLE 11. CRITICAL COMPONENTS — PROFESSIONAL DEVELOPMENT

	n	All	Early	Recent
August PD sessions	45	68.9%	50.0%	81.5%
Math department team meetings	51	72.5%	66.7%	75.8%
All-day a2i PD sessions during the school year	51	82.4%	72.2%	87.9%
Conversations & collaborations with a2i teachers	48	87.5%	83.3%	90.0%
PD or support around Mock Regents Prep	48	60.4%	38.9%	73.3%

Source: SY 2015–16 Teacher Follow-up Survey.

Inquiry and Support

Instructional Specialists' support changed over the course of a2i. In Years 1 and 2, they focused on supporting individual teachers and teams. Once Cohort 3 schools were added, they continued to support inquiry teams and team leaders, but not individual teachers. Feedback indicated that teachers newer to a2i continued to want individual support from Instructional Specialists (83.9%). The early teachers who worked with Instructional Specialists in both settings found school team inquiry work more critical than individual support (77.8% vs. 61.1%), and those who began the project earlier also put more emphasis on support from administrators (88.9% vs. 63.6%). (See Table 12. Noted differences between the groups are shaded in the table.)

TABLE 12. CRITICAL COMPONENTS—INQUIRY AND SUPPORT

	n	**All**	**Early**	**Recent**
Support/buy-in, administrators	51	72.5%	88.9%	63.6%
School inquiry team meetings	46	71.7%	72.2%	71.4%
Individual coaching from Instructional Specialists	49	75.5%	61.1%	83.9%
Instructional Specialists' work with school teams	48	77.1%	77.8%	76.7%
Instructional Specialists' support for teacher leaders	46	71.7%	66.7%	75.0%

Source: SY 2015–16 Teacher Follow-up Survey.

Curricular Strategies and Assessment

All teachers agreed that the a2i unit structure and a2i curriculum were critical components of the a2i project (87.5% and 85.1%, respectively). Three-fourths (75.5%) considered IAs, added in Year 4 to help meet teachers' day-to-day lesson needs, as a critical piece of a2i.

Differences between the early and recent group were not statistically significant, but there were some notable differences (shaded in Table 13 below). For the early group, components that were key parts of the original design—the unit design schematic, the re-engagement lessons, and the classroom challenges (FALs)—continued to play a more critical role.

TABLE 13. CRITICAL COMPONENTS CURRICULAR STRATEGIES AND ASSESSMENT

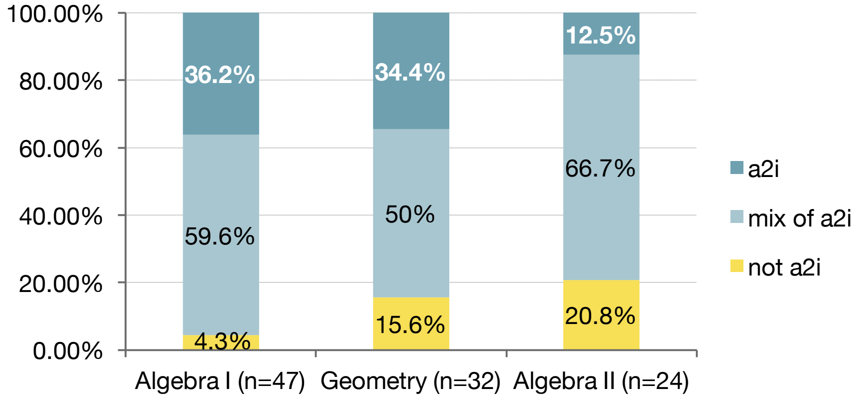
	n	**All**	**Early**	**Recent**
a2i unit design schematic (IPAT, FAL, Re-engagement, FPAT)	48	87.5%	94.4%	83.3%
a2i curriculum	47	85.1%	82.4%	86.7%
Re-engagement lessons	48	68.8%	77.8%	63.3%
Instructional Activities (IAs)	49	75.5%	77.8%	74.2%
Classroom Challenges (FALs)	49	73.5%	82.4%	68.8%
Balanced Assessment System	48	75.0%	76.5%	74.2%

Source: SY 2015–16 Teacher Follow-up Survey.

How *Level* of Involvement Affected Changes

We also compared the Year 4 survey respondents based on whether they identified their classrooms as "a2i" or "mix of a2i." Most teachers considered their classes a mix of a2i and other resources and strategies. Just over a third of the Algebra I (36.2%) and Geometry (34.4%) teachers said their classes were a2i classes, along with smaller numbers of Algebra II teachers (12.5%). (See Figure 6.)

FIGURE 6. DEGREE OF a2i IMPLEMENTATION



Source: SY 2015–16 Teacher Follow-up Survey

Differences in Reported Changes

Sub-analyses also revealed some descriptive differences or patterns in Algebra I and Geometry teachers' self-reports of their level of a2i involvement.

- Algebra I teachers who considered their classrooms a2i classrooms were more positive about the amount of improvement in students' understanding of math concepts (M = 2.13 vs. M = 1.70, a2i and mix of a2i, respectively). Algebra I teachers who used a mix of a2i reported a "negative" or "little to no" change (31.2%), compared to only 9.1% of a2i teachers.

- For teachers' own understanding of the Common Core standards, more a2i teachers reported a significant change (54.5%), compared to 25.0% of teachers using a mix of a2i (M = 2.44 vs. M = 2.00).

- Geometry teachers who considered their classrooms a2i classrooms tended to report a higher amount of change in their understanding of the Common Core standards than those teaching a mix (M = 2.64 vs. M = 2.19, respectively). Teachers who used a2i (72.7%) tended to state that a significant change had occurred due to the project, whereas for teachers using a mix of a2i, only 21.2% stated that there had been a significant change because of the project.

Differences by Years of Experience

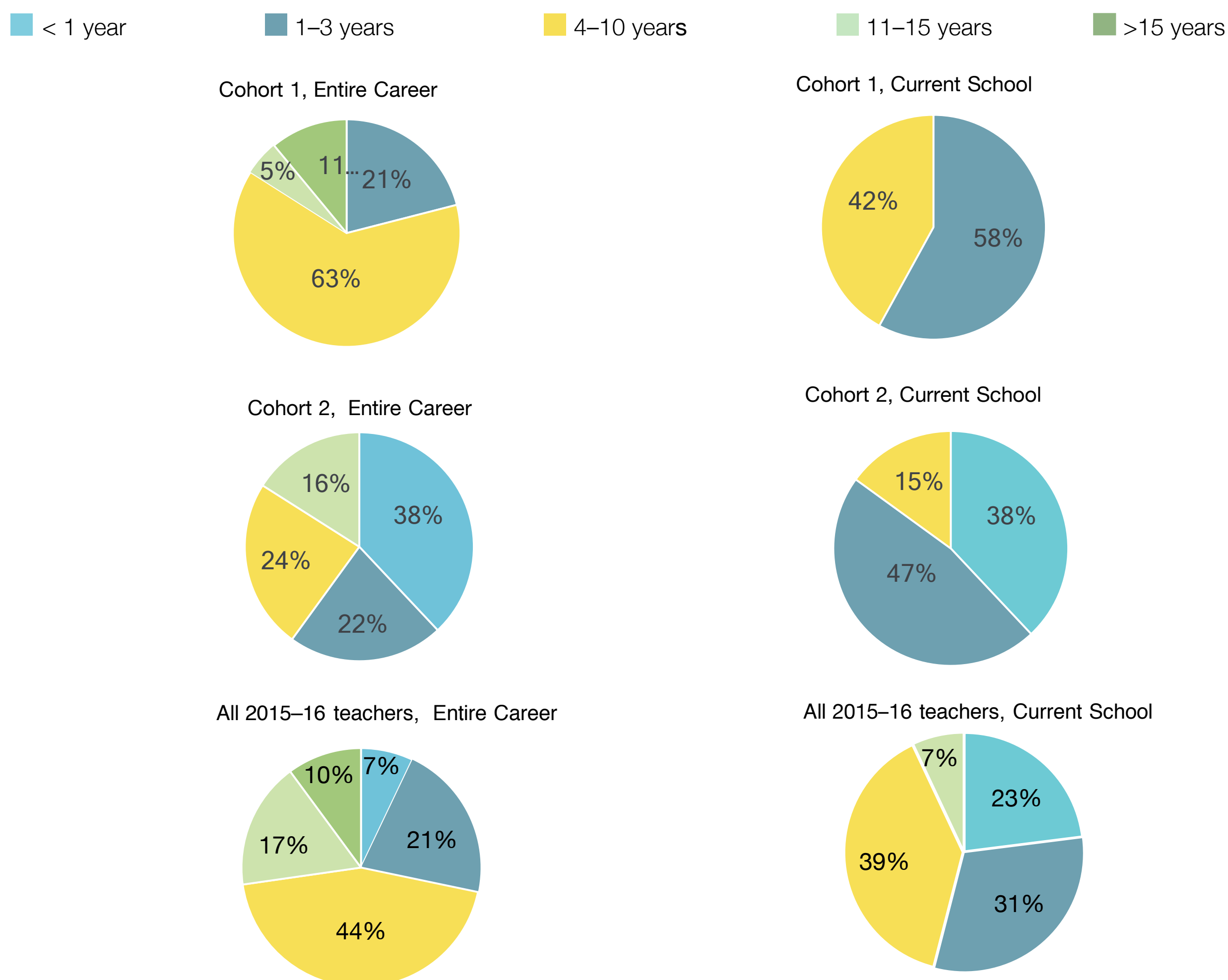
Knowing that many of the teachers in Year 2 were new not just to the project but also to teaching (and that in New York City teacher mobility and attrition rates are high), we looked at teaching experience across cohorts, and at whether level of experience seemed to play a role in project implementation (see Figure 7).

- The Year 1 teachers were relatively experienced: over their entire careers, most (63%) had taught from four to 10 years; another 16 % had taught 11 years or more. Although a fifth had one to three years' experience—and over half (58%) had the same number of years at their current school—no teachers were new to their schools or the profession.

- In Year 2, as more schools were added to the project, that changed. It was the first year of teaching—overall and at their current school—for over a third of the new a2i teachers (38%). Fairly large percentages had between 4–10 (24%) and 11–15 (16%) years of experience, but only 15% had more than four years at their current schools.

- Year 4 survey respondents included all cohorts, and more balance in experience. Still, over their entire careers, over a fourth of the a2i teachers (28%) had three years of teaching experience or less; that figure almost doubled (54%) for the percentage of teachers teaching three or less years at their current school. For close to a fourth of the teachers, Year 4 was their first year of teaching. Both breakdowns showed that the largest portion of teachers had between four and 10 years in the profession.

FIGURE 7. A2I TEACHER'S EXPERIENCE, OVER ENTIRE CAREER AND IN CURRENT SCHOOL



Source: SY2013–14, SY2014–15, SY2015–16 Surveys

Teaching Experience and Level of Implementation

When we further disaggregated data by teaching experience, we found that experience, and likely comfort in the classroom, did seem to be a factor in how teachers used a2i: those with fewer years at their current school tended to mix a2i and other resources. There also seemed to be a sweet spot, a level of experience where fidelity was higher: those with 11–15 years' experience were more likely to consider their classrooms a2i classrooms rather than a mix.

- Over 60% of Algebra I teachers indicated that they used a mix of a2i across, regardless of years of experience, except for those who had taught 4–10 years over their entire careers. We did not see the same trend in those who had been at their current school for 4–10 years.

- We did see, that, for both Algebra I and Geometry, three-fourths and two-thirds of the teachers, respectively, who had been at a given school for 11–15 years indicated that they used a2i (75% and 66.7%, respectively). Teachers for the other ranges of years of experience tended to indicate that they used a mix of a2i in their classrooms.

TABLE 14. TEACHERS' LEVEL OF IMPLEMENTATION, BY YEARS TEACHING (CURRENT SCHOOL)

		This is my first year (n=22)		**1–3 years (n=34)**		**4–10 years (n=37)**		**11–15 years (n=10)**		**More than 15 years (n=22)**	
		Freq.	%	Freq.	%	Freq.	%	Freq.	%	Freq.	%
Algebra I	a2i	2	16.7%	6	40.0%	6	37.5%	3	75.0%	2	16.7%
	Mix of a2i	8	66.7%	9	60.0%	10	62.5%	1	25.0%	8	66.7%
	Not a2i	2	16.7%	0	0.0%	0	0.0%	0	0.0%	2	16.7%
Geometry	a2i	3	33.3%	3	27.3%	3	33.3%	2	66.7%	3	33.3%
	Mix of a2i	5	55.6%	5	45.5%	6	66.7%	0	0.0%	5	55.6%
	Not a2i	1	11.1%	3	27.3%	0	0.0%	1	33.3%	1	11.1%
Algebra II	a2i	0	0.0%	0	0.0%	2	16.7%	1	33.3%	0	0.0%
	Mix of a2i	1	100.0%	5	62.5%	9	75.0%	1	33.3%	1	100.0%
	Not a2i	0	0.0%	3	37.5%	1	8.3%	1	33.3%	0	0.0%

Source: SY 2015–16 Teacher Follow-up Survey

4 Changes in Classrooms

LEARNING ENVIRONMENTS

❖ *Did a2i bring about changes in classroom learning environments?*

One of the first differences that principals and teachers observed in a2i classrooms was that students were talking, to each other, and about math. That change, another of the medium-term outcomes identified in the logic model, appeared early—during the project's first year—and remained visible. It was a change that became a constant, an outcome that deepened over time as students continued to work in groups, tackle multi-step problems, and rely on one another as instructional resources.

What also held steady was teachers' endorsement of the a2i approach: from the first cohort on, teacher said they preferred depth over breadth—an endorsement as well of the Common Core Standards for Math. Although it required some changes from conventional and comfortable classroom practice, teachers wanted their students to talk about math and come up with different ways to solve problems, not just memorize formulas. As the a2i team hoped, the classroom emphasis was on process and on understanding math.

Feedback from teachers also indicated that, for some teachers, this did not always mean privileging big ideas over basic operations. This section shares feedback about the interplay between the two, continuing the discussion of differences or trends based on the extent of teachers' involvement and adding students' reports about the learning environment in the math classrooms.

Beliefs about Math Practices

Overall Trends

Even with initial consensus about a2i's approach to math instruction, levels of agreement among teachers to most survey belief statements ticked up, among respondents to annual surveys and the matched group of teachers who completed both an initial and a follow-up survey (see Table 15).

- On baseline surveys completed by Cohort 1 and 2 teachers, 91% and 97% respectively said it was important to let students puzzle things out for themselves. On the Year 4 survey, agreement was unanimous, with 71.2% strongly agreeing.

- While 62% of the Cohort 1 teachers agreed that they regularly had their students work through real-life math problems of interest to them, that percentage rose to 93% for Cohort 2. In the Year 4 survey, the percentage was 98%.

TABLE 15. LEVELS OF AGREEMENT ABOUT MATH PRACTICES

	n	*M*	**Strongly/ Somewhat Disagree**	**Moderately Agree**	**Strongly Agree**
I regularly have my students work through applied math problems.	59	3.37	1 (1.7%)	35 (59.3%)	23 (39.0%)
It's important to let students puzzle things out for themselves.	59	3.71	0 (0.0%)	17 (28.8%)	42 (71.2%)
I like my students to master basic mathematical operations before they tackle complex problems.	59	3.10	12 (20.4%)	24 (40.7%)	23 (39.0%)
If students understand big ideas in math, they should do well on standardized tests.	59	2.97	16 (27.12%)	27 (45.8%)	16 (27.1%)
When students are working on math problems, the emphasis should be on getting the correct answer rather than on the process followed.	59	2.02	20 (33.9%)	25 (42.4%)	14 (23.8%)

Source: SY 2015–16 Teacher Follow-up Survey. "Strongly disagree" and "Somewhat disagree" response options were combined as very few (if any) respondents marked "Strongly disagree" on the items. The item "When students are working on math problems, I think the emphasis should be on getting the correct answer rather than on the process followed" was scored the opposite of the other items, with the "Moderately agree" and "Strongly agree" options combined.

Basic Operations vs. Big Ideas

Teachers' responses to items about the interplay between basic operations and big ideas—a dichotomy that in some ways distinguished a2i and the Common Core approach to math instruction from more traditional or conventional approaches—revealed some uncertainty about the balance.

- The Year 4 respondents agreed far more strongly than earlier cohorts that students should master basic operations before tackling complex problems.

- More teachers in the latter group also indicated that, when working through math problems, the emphasis should be on getting the correct answer.

- A larger percentage of both Algebra I and Geometry teachers in the Year 4 respondent group who identified their class as being a mix of a2i reported that they do like students to master basic operations before tackling complex problems, compared to those who identified their class as an a2i class (61.5% vs. 20% strongly agreed, for mix of a2i vs. a2i respectively). (See Table 16. Note: Although there was a notable observed difference, it was not statistically significant due to the small sample and the Bonferroni correction.)

TABLE 16. DESCRIPTIVE COMPARISONS BETWEEN a2I VERSUS MIX OF a2I CLASSES, 1

I like my students to master basic mathematical operations before they tackle complex problems.

		n	Strongly/Somewhat Disagree	Moderately Agree	Strongly Agree
Algebra					
	a2i	15	4 (26.7%)	8 (53.3%)	3 (20.0%)
	Mix of a2i	25	4 (16.0%)	5 (20.0%)	16 (64.0%)
Geometry					
	a2i	10	3 (30.0%)	5 (50.0%)	2 (20.0%)
	Mix of a2i	13	1 (7.7%)	4 (30.8%)	8 (61.5%)

- A larger percentage of teachers who identified their class as being an a2i class disagreed with the statement that "If students understand the big ideas in math, they should do well on standardized tests" compared to those who taught a mix of a2i (53.3% vs. 16.0%, see Table 17).

TABLE 17. DESCRIPTIVE COMPARISONS BETWEEN a2i VERSUS MIX OF a2I CLASSES, 2

If students understand the big ideas in math, they should do well on standardized tests.

		n	Strongly/Somewhat Disagree	Moderately Agree	Strongly Agree
Algebra					
	a2i	10	3 (30.0%)	6 (60.0%)	1 (10.0%)
	Mix of a2i	13	1 (7.7%)	7 (53.8%)	5 (38.5%)
Geometry					
	a2i	15	8 (53.3%)	4 (26.7%)	3 (20.0%)
	Mix of a2i	25	4 (16.0%)	12 (48.0%)	9 (36.0%)

SY2015–16 Teacher Follow-up Survey. These are descriptive differences only, and not statistically different.

Among the matched group, the trend was in the other direction: the teachers completing a baseline and follow-up survey agreed more strongly at the end of a2i that students need not master basic mathematical operations before they tackle complex problems, and that, if they understood the big ideas, they should do better on standardized tests. The baseline to follow-up differences were statistically significant (p = .010 and p = .012, respectively).

Changes in Learning Environments

Teachers' Reports

The Year 4 teacher survey results provided additional confirmation of initial observations that classrooms were changing—regardless of whether teachers considered their classes a2i classes or a mix of a2i and other strategies: We found no differences between the a2i and mix groups on the learning environment survey items. Even though, as Deputy Director of Instruction West noted, some teachers may have become "disenchanted" with the a2i approach because of some anxiety about students' performance on the new exams, they still subscribed to the project's emphasis on supportive learning environments where students are encouraged to believe that they can solve complex problems, excel at math, and develop a "growth mindset" (Dweck, 2008).

When asked how often these activities occurred in their math classes, over half of the teachers in the overall 2015–2016 survey respondent group reported that students work together in groups "almost always" (54.2%, see Table 18). Half also said that students "almost always" explain their problem-solving strategies (50.8%) and work on multi-step problems (54.2%).

TABLE 18. TEACHERS' RATINGS ON FREQUENCY OF MATH PRACTICES (*n*=59)

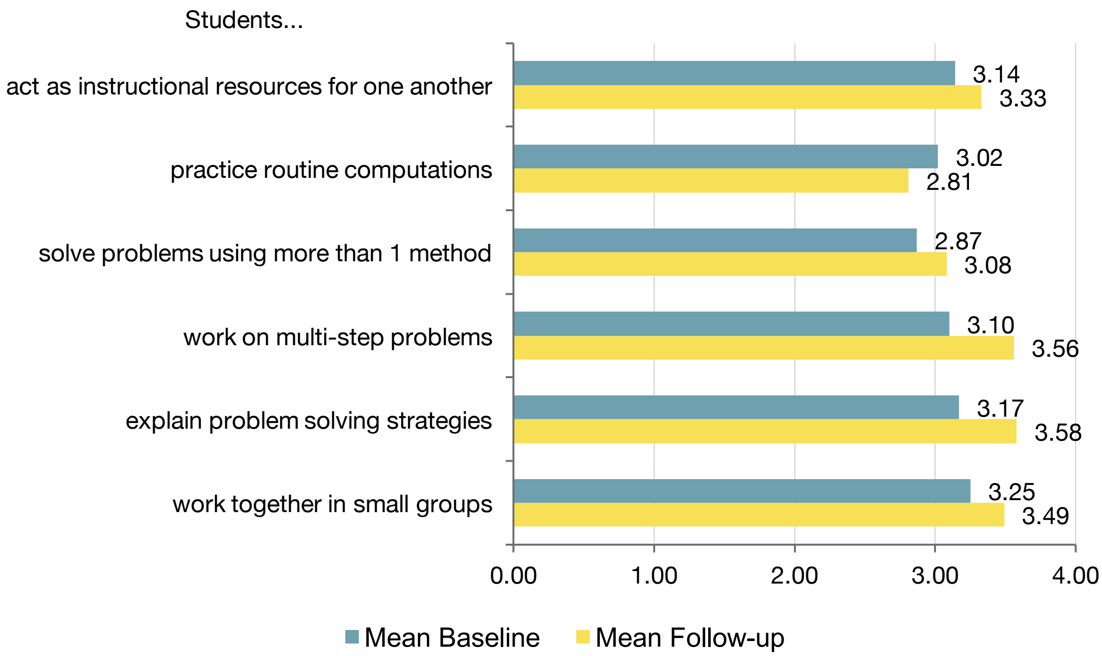
	M	**Never/ Rarely/ Once in a while**	**Pretty often**	**Almost always**
Students work on multi-step problems.	3.51	3.4%	42.4%	54.2%
Students explain their problem solving strategies.	3.47	3.4%	45.8%	50.8%
Students work in small groups.	3.41	11.9%	33.9%	54.2%
Students act as instructional resources for one another.	3.22	13.6%	50.8%	35.6%
Students solve the same problems using more than one method.	3.00	20.3%	59.3%	20.3%
Students practice routine computations.	2.81	35.6%	39.0%	25.4%

Source: SY2015–16 Teacher Follow-up Survey. Means were calculated on a 4-point scale; the *Never/Rarely* and *Once in a while* response options were combined as very few (if any) respondents marked *Never/Rarely* on the items.

Matched Groups

Teachers in the matched group also reported increases from baseline to follow-up—in group work, work on multi-step problems, use of multiple methods, and more reliance on classmates as resources (see Figure 8). In two cases, the increases were statistically significant: explaining how they solved problems (M = 3.10 vs. M = 3.56; p < .001) and using multi-step problems (M = 3.17 vs. M = 3.58; p < .001)

FIGURE 8. CHANGES IN STUDENTS' CLASSROOM ACTIVITIES



Beliefs and Length of Involvement

Although there were no differences—descriptive or otherwise—based on whether teachers considered their classes to be a2i classes, or a mix, there were some slight differences based on teachers' experience with a2i. For all items related to a2i-endorsed practices, the early participants indicated more frequent use (see Table 19). In one case—how often students explain their problem-solving strategies—the difference was statistically significant. For the single item that might not characterize an a2i classroom—"students practice routine computations"—the recent group reported higher frequency.

TABLE 19. DIFFERENCES IN STUDENTS' CLASSROOM ACTIVITIES BASED ON TEACHERS' INVOLVEMENT

	Teacher Group	*n*	*M*	Never/ Rarely/ Once in a while	Pretty often	Almost always
Students work together in small groups.	Early	19	3.58	1 (5.3%)	6 (31.6%)	12 (63.2%)
	Recent	40	3.33	6 (15.0%)	14 (35.0%)	20 (50.0%)
*Students explain their problem solving strategies.	Early	19	3.74	0 (0.0%)	5 (26.3%)	14 (73.7%)
	Recent	40	3.35	2 (5.0%)	22 (55.0%)	16 (40.0%)
Students work on multi-step problems.	Early	19	3.63	0 (0.0%)	7 (36.8%)	12 (63.2%)
	Recent	40	3.45	2 (5.0%)	18 (45.0%)	20 (50.0%)
Students solve the same problems using more than one method.	Early	19	3.05	2 (10.5%)	14 (73.7%)	3 (15.8%)
	Recent	40	2.98	10 (25.0%)	21 (52.5%)	9 (22.5%)
Students practice routine computations.	Early	19	2.63	10 (52.7%)	5 (26.3%)	4 (21.1%)
	Recent	40	2.90	11 (27.5%)	18 (45.0%)	11 (27.5%)
Students act as instructional resources for one another.	Early	19	3.37	1 (5.3%)	10 (52.6%)	8 (42.1%)
	Recent	40	3.15	7 (17.5%)	20 (50.0%)	13 (32.5%)

Source: SY2015–16 Teacher Follow-up Survey.

* *The difference in means for this item were not significant using the Bonferroni correction (which would have had to exceed .008). However, it was significant at the traditional .05 level (t[57] = 2.559, p = .013).*

Students' Reports on Classroom Environments

Students' reports of classroom behaviors and activities largely aligned with teachers' reports:

- Working in small groups: 88.1% of teachers indicated that students worked in small groups "pretty often" to "almost always" in their math classes. Students' responses were consistent with teachers in that 79.7% stated that they "pretty often" to "often" worked in small groups to think through math problems.

- Explaining problem solving: Almost all teachers (96.6%) indicated that students pretty often to almost always explained their problem-solving strategies in math class. Students were consistent with teachers in that 91.5% indicated that they pretty often to often explained how they solved problems in math class, 84.0% were encouraged to discuss different solutions to problems, but 57.8% said they were asked to explain their work out loud or in writing only once in a while. When reporting on teacher practices, most of students stated that it was mostly true that teachers wanted them to explain more about their answers (81.6%) and why they thought the way that they did (80.5%).

- Solving problems: Over three-fourths of the teachers (79.6%) said they often gave students problems that they could solve using more than one method. Students' responses were consistent: 85.4% indicated that teachers often gave them problems that could be solved in a lot of different ways.

- Serving as resources for one another: 86.4% of teachers indicated they pretty often to almost always students act as instructional resources for one another. Students' responses were consistent with teachers. Most students (81.7%) also said that they often work with each other to find their mistakes. As reported above, students often worked in small groups to think through math problems.

On average, both targeted student behaviors and teacher practices increased ($p < .001$) in Year 4 from the Fall to the Spring (see Table 20). For example, students said they worked in small groups to think through math problems ($M = 3.01$ Fall to $M = 3.21$ Spring), worked through problems that they didn't know how to solve at first ($M = 2.97$ to $M = 3.03$), explained their work out loud or in writing ($M = 2.69$ to $M = 2.80$), and asked questions or take part in class discussions ($M = 2.68$ to $M = 2.78$). Students also indicated teachers often asked them how they solved problems ($M = 3.45$ to $M = 3.51$), figure out where they made mistakes in math problems ($M = 3.19$ to $M= 3.28$), and gave math problems that can be solved in multiple ways ($M = 3.18$ to $M = 3.26$). See Appendix E for further details.

TABLE 20. PRE/POST MEANS FOR CLASSROOM ENVIRONMENT OUTCOMES

		Pre (Fall)		Post (Spring)			
Construct	*n*	*M*	SD	*M*	SD	*t* (*df*)	*p*
Student Behaviors	1836	2.87	.55	2.97	.60	-7.316 (1835)	<.001
Teacher Practices	1762	3.20	.53	3.28	.56	-5.763 (1761)	<.001
Classroom Culture [Students' Understanding]	1793	3.36	.48	3.35	.55	.494 (1792)	.621
Classroom Culture [Learning in General]*	1806	3.90	.76	3.90	.82	-.013 (1805)	.989
Feedback Practices*	1714	4.12	.63	4.13	.71	-905 *17113)	.365

Source: SY2015–16 Student Surveys. *Using a Bonferroni correction for the number of analyses conducted, the p-value needed for statistical significance was .01.*
*All scales were measured using a 4-point scale, except for Feedback Practices and Classroom Culture [Learning in General] which used a 5-point scale.

As reported in Section 3, teachers had tended to place emphasis on students' understanding math rather than rote memorization. Students reported that teachers checked to make sure they understood what was being taught. Students also agree that, in their a2i math classes, it's okay not to understand something: mistakes were okay as long as they were learning.

5 Changes in Students' Attitudes

EFFICACY AND INTEREST

❖ *Did a2i affect students' attitudes about math, including their sense of efficacy and interest in math courses or careers?*

Although the impact of changes in teachers' practice did not translate to higher performance on statewide exams, responses to the SY2015–16 student surveys, completed by close to 2,000 a2i students in the Fall and Spring of the school year, seemed to reflect changes in learning environments. As reported in Section 3, teachers emphasized their students' understanding the work as opposed to simply memorizing it. Students also agreed that, in their a2i math classes, it's okay not to understand something: mistakes were okay as long as they were learning.

This section looks at students' attitudes and engagement in their math classes, based on students' self-reports in the Fall and Spring surveys. Most averages are on four-point frequency or agreement scale, with the majority at a 3.0 or higher. Items in bold (or highlighted in gray) indicate there was statistical significance for the construct or individual item.

Attitudes toward Math

The overall finding from the student surveys is that a2i students were positive about math: the majority were confident in their skills and interested in the subject—though a little less sure about its long-term utility. There were some, albeit slight, changes from Fall to Spring, some minor differences between classes that teachers considered to be a2i classes and those they described as a mix, and some minor differences by subject (i.e., Algebra, Geometry, and Algebra II).

At both survey administrations, almost all students in a2i classes (97% Fall, 96% Spring) indicated that they believe they can learn math, and that they can get good grades (91%, 90%). Self-reported survey data also indicated that students thought it was important to do well in math class (97%, 94%). Fewer students, but still a majority, are confident they could do well in more advanced math classes (57%, 59%). Just under two-thirds of the students said their teacher made math interesting (61%, 62%) and enjoyable (62%). See Figure 9 for a summary of Fall and Spring percentages, by item.

Other analyses showed that students were more confident overall that they could understand the basic concepts in their math classes than the harder ones, but the former edged down from Fall to Spring by a few hundredths of a percentage point (M = 3.21 to M = 3.19), and the latter edged up (M = 2.70 to M = 2.71) Students' ratings for interest in math and taking more math classes in high school were lower (M = 2.49 Fall; M = 2.55 Spring), as was their confidence in their ability to do well in more advanced math classes (M = 2.59 Fall; M = 2.66 Spring), but both increased, and the differences were statistically significant (see Appendix E, p. 100–104 for individual construct breakdowns).

The overall scale ratings show increases from Fall to Spring in intrinsic value items, and decreases in utility value; differences for both were statistically significant. (See Appendix E, for items in each scale.)

TABLE 21. PRE/POST MEANS FOR STUDENT MATH ATTITUDINAL OUTCOMES

		Pre (Fall)		**Post (Spring)**			
Construct	n	*M*	SD	*M*	SD	t (df)	*p*
Math Self-Efficacy	1852	3.05	.51	3.07	.56	-2.314 (1851)	.021
Math Interest/ Intrinsic Value	1859	2.49	.66	2.53	.72	-3.313 (1858)	.001
Math Value/ Utility Value	1759	3.15	.52	3.07	.57	6.864 (1758)	<.001
Perseverance	1810	3.09	.57	3.06	.66	2.255 (1809)	.024
Enjoyment*	1807	3.74	1.00	3.76	1.07	-.507 (1806)	.612

Source: SY2015–16 Student Surveys. *Using a Bonferroni correction for the number of analyses conducted, the p-value needed for statistical significance was .01.*
*All scales were measured using a 4-point scale, except for Enjoyment which used a 5-point scale.

Favorite Subjects

Students ranked math as their favorite subjects, in both the Fall (18.4%) and Spring (18.3%), with physical education (16.9% in the Fall and 17.8% in the Spring) as a close second (see Table 22).

TABLE 22. STUDENTS' FAVORITE SUBJECTS – MATCHED (n = 1,930)

	Fall		**Spring**	
Subject	Frequency	Percent & Rank	Frequency	Percent & Rank
Math	362	18.4% (1)	373	18.3% (1)
Physical Education	332	16.9% (2)	362	17.8% (2)
History/Social Studies	283	14.4% (4)	308	15.1% (3)
English	275	14.0% (5)	297	14.6% (4)
Science	289	14.7% (3)	277	13.6% (5)
Art	108	5.5% (7)	122	6.0% (7)
Music	63	3.2% (10)	47	2.3% (9)
Foreign Language	65	3.3% (9)	62	3.0% (10)
Technology	117	5.9% (6)	124	6.1% (6)
Other	76	3.9% (8)	64	3.1% (8)
Total	1,970	100%	2,036	100%

Source: SY2015–16 Student Surveys. The frequencies add up to over the sample size of 1,930 because some students chose more than one favorite subject.

FIGURE 9: SUMMARY OF STUDENTS' ATTITUDES ABOUT MATH AND MATH CLASSES

How do students feel about math?

➢ ***Most students think math is interesting and valuable—and they think they can learn it.***

SELF-EFFICACY

- 97% (Fall) & 96% (Spring) of students said they can learn math.
- 95%/94% said they can understand the basic concepts in their math classes.
- 65%/69% said they can understand the harder ones.*
- 57%/59% said math was an easy class.*

ENJOYMENT

- 61%/62% like the way they learn in math class.
- 62% (Fall & Spring) said teachers made math interesting.

PERSEVERANCE

- 88%/82% said they kept working until they finished, even if assignments were hard.*
- 81%/80% felt successful when they worked hard on something, even if they didn't get the right answer.

INTRINSIC VALUE

- 70% (Fall & Spring) of the students found math interesting.
- 50%/54% would like to take more math classes in high school.*

UTILITY

- 97%/94% said it was important to do well in math.*
- 86%/82% said math was important to their daily lives.*
- 87%/84% thought math will help them get a good job.*
- 47%/48% thought they could have a job that requires math.

What do students say about their math classrooms?

➢ ***Students are positive about math classes: they work in groups, explain their work—and they think their teachers respect their ideas.***

STUDENT BEHAVIORS

- 75%/83% of the students often worked in small groups to think through problems.*
- 72%/76% said they worked through problems they didn't understand at first.*
- 55%/59% of students ask questions and take part in class discussions.

INSTRUCTION

- 91%/93% said teachers often asked them to explain how they solved a problem.*
- 84%/87% of teachers give problems that can be solved in multiple ways.*

FEEDBACK

- 84%/80% said teachers respected their ideas & suggestions.
- 69% said the comments they get help them understand how to improve.

CLASSROOM CULTURE

- 95%/92% said it's okay in this class to say you don't understand something.
- 94%/92% agree it's okay to make mistakes as long as they are learning.

*Statistically significant results, differences exceed value needed using Bonferroni correction and can be interpreted as differences.

Source: SY2015–16 Student Survey

The survey included 45 questions, plus background items; these 23 bulleted items are a snapshot of students' responses.

Comparison of Teacher and Student Perceptions

We matched teacher and student survey items based on questions relating to students' attitudes towards math, problem solving, classroom environment, and teachers' behaviors in the classroom. While questions between surveys did not align perfectly, trends did emerge between how teachers and students view students' attitudes in math and what both groups say happens in their math classrooms.

Attitudes toward Math

In Year 4, most teachers indicated a modest change in their students' interest in, confidence in, and understanding of math concepts and practices. Student responses about their own interest in, confidence in, and understanding of math started off positive in the Fall and showed little to no differences by Spring.

- Interest: 61.2% of the teachers said their students' interest in math had a modest change because of the a2i program. While students' interest ratings were not as high as confidence ratings (*M* = 2.84 Fall and *M* = 2.80 Spring), 67.9% still agreed that math was interesting to them.

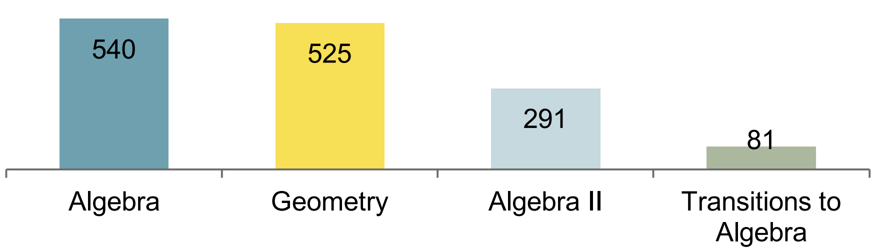
- Confidence: 61.1% of the teachers reported that a2i brought about a modest change in students' confidence in doing math. Although we cannot wholly attribute confidence levels to a2i, 95.8% of the students agreed that they can learn math, 88.3% were confident they could get good grades in math classes, and 54.1% said that math was an easy subject.

- Understanding: 71.6% of teachers reported that a modest to significant change in their students' understanding of math concepts and practices resulted from a2i. As noted above, students tended to believe that they could understand the basic concepts taught in math class (92.1%), but had less confidence in their ability to understand the harder concepts (64.9%).

❖ *Were there differences in students' interest and confidence by a2i course?*

The student surveys asked what classes students were taking (i.e., Algebra I, Geometry, Algebra II, Transitions to Algebra). Responses indicated that, in the survey sample, most students were enrolled in Algebra and Geometry during Year 4, with fewer student respondents in Algebra II and Transitions to Algebra (see Figure 10).

FIGURE 10. NUMBER OF STUDENTS ENROLLED IN EACH COURSE SUBJECT



Note. Students enrolled in Transitions to Algebra were not included in further analyses.

There were no statistically significant differences between Algebra I, Geometry and Algebra II students' attitudes toward math and their reports about the classroom learning environment (see Table 23).

- Students' came into their math classes with a high level of self-efficacy, which rose even higher from the Fall to Spring in each math subject, Algebra 1 (M = 3.05 to M = 3.08), Geometry (M = 3.09 to M = 3.13) and Algebra II (M = 3.18 to M = 3.21).

- In the Fall, students reported strong value and utility ratings, which, by Spring, dropped slightly in each math subject area: Algebra 1 (M = 3.15 to M = 3.06), Geometry (M = 3.19 to M = 3.14) and Algebra II (M = 3.21 to M = 3.13).

- Levels of perseverance in math were strong in the Fall for all groups. By Spring, Algebra I students' perseverance dropped slightly (M = 3.08 to M = 3.03); in Geometry and Algebra II, there were slight gains (M = 3.16 to M = 3.17 and M = 3.12 to M = 3.13).

TABLE 23. PRE/POST MEANS FOR ATTITUDES TOWARDS MATH AND CLASSROOM ENVIRONMENT, BY COURSE SUBJECT

	Algebra 1 (*N*=540)			**Geometry (*N*=525)**			**Algebra II/Trig (*N*=291)**		
	n	Pre	Post	*n*	Pre	Post	*n*	Pre	Post
Math Self-Efficacy	521	3.05	3.08	497	3.09	3.13	281	3.18	3.21
Math Interest/Intrinsic Value	516	2.46	2.50	501	2.56	2.62	282	2.63	2.69
Math Value/Utility Value	491	3.15	3.06	493	3.19	3.14	270	3.21	3.13
Perseverance	505	3.08	3.03	503	3.16	3.17	282	3.12	3.13
Enjoyment*	499	3.72	3.67	506	3.92	3.99	285	3.75	3.87
Student Behaviors	503	2.85	2.96	506	2.91	3.04	284	3.01	3.10
Teacher Practices/Behaviors	483	3.23	3.29	490	3.23	3.34	273	3.31	3.40
Classroom Culture [Students' Understanding]	492	3.37	3.35	501	3.40	3.45	281	3.45	3.45
Classroom Culture [Learning in General]*	499	3.92	3.90	506	4.00	4.03	282	3.94	4.04
Feedback Practices*	473	4.14	4.11	481	4.19	4.26	276	4.20	4.25

*All scales were measured using a 4-point scale, except for Enjoyment, Feedback Practices and Classroom Culture [Learning in General] which used a 5-point scale.

We also explored differences based on whether teachers considered their classes to be a2i classes, a mix of a2i classes, or not a2i classes at all. Again, differences were slight, with no discernible trends. Tables 24–26 show the breakdowns by subject and a2i classification.

TABLE 24. PRE/POST MEANS FOR ATTITUDES TOWARDS MATH, ALGEBRA I

	a2i Only			Mix of A2i			Both a2i and Mix of a2i		
	n	Pre	Post	*n*	Pre	Post	*n*	Pre	Post
Math Self-Efficacy	292	2.99	3.04	127	3.08	3.14	102	3.20	3.09
Math Interest/Intrinsic Value	289	2.37	2.43	123	2.53	2.64	104	2.65	2.54
Math Value/Utility Value	285	3.07	2.99	117	3.24	3.14	89	3.28	3.16
Perseverance	292	3.01	3.00	120	3.08	3.04	93	3.31	3.10
Enjoyment*	288	3.56	3.62	121	3.99	3.77	90	3.91	3.70
Student Behaviors	283	2.84	2.97	124	2.75	2.91	96	3.01	2.98
Teacher Practices/ Behaviors	279	3.21	3.29	118	3.17	3.30	86	3.37	3.29
Classroom Culture [Students' Understanding]	283	3.37	3.34	117	3.33	3.41	92	3.40	3.32
Classroom Culture [Learning in General]*	288	3.81	3.89	120	3.99	3.96	91	4.14	3.87
Feedback Practices*	277	4.10	4.08	115	4.14	4.18	81	4.26	4.08

*All scales were measured using a 4-point scale, except for Enjoyment, Feedback Practices and Classroom Culture [Learning in General] which used a 5-point scale.

TABLE 25. PRE/POST MEANS FOR ATTITUDES TOWARDS MATH, GEOMETRY

	a2i Only			Mix of A2i		
	n	Pre	Post	*n*	Pre	Post
Math Self-Efficacy	257	3.02	3.07	225	3.16	3.18
Math Interest/Intrinsic Value	256	2.50	2.53	229	2.63	2.71
Math Value/Utility Value	253	3.23	3.13	225	3.15	3.16
Perseverance	260	3.16	3.11	227	3.15	3.22
Enjoyment*	260	4.03	3.98	229	3.76	3.97
Student Behaviors (in Math Class)	261	2.91	3.00	230	2.92	3.08
Teacher Practices/ Behaviors (in Math Class)	251	3.28	3.37	222	3.17	3.30
Classroom Culture [Students' Understanding]	256	3.39	3.42	229	3.39	3.46
Classroom Culture [Learning in General]*	259	4.08	3.99	230	3.90	4.05
Feedback Practices*	245	4.26	4.29	219	4.12	4.21

*All scales were measured using a 4-point scale, except for Enjoyment, Feedback Practices and Classroom Culture [Learning in General] which used a 5-point scale.

TABLE 26. PRE/POST MEANS FOR ATTITUDES TOWARDS MATH, ALGEBRA II

	a2i Only			Mix of A2i		
	n	Pre	Post	*n*	Pre	Post
Math Self-Efficacy	22	3.07	3.06	241	3.18	3.22
Math Interest/Intrinsic Value	23	2.66	2.76	241	2.62	2.69
Math Value/Utility Value	23	3.16	3.19	230	3.22	3.13
Perseverance	23	3.20	3.26	241	3.12	3.13
Enjoyment	24	3.50	4.06	242	3.80	3.88
Student Behaviors (in Math Class)	24	3.03	3.13	242	3.02	3.12
Teacher Practices/ Behaviors (in Math Class)	23	3.33	3.40	232	3.32	3.42
Classroom Culture [Students' Understanding]*	24	3.42	3.39	238	3.46	3.46
Classroom Culture [Learning in General]*	24	3.96	3.96	239	3.96	4.06
Feedback Practices	24	4.14	4.18	236	4.21	4.26

*All scales were measured using a 4-point scale, except for Enjoyment, Feedback Practices and Classroom Culture [Learning in General] which used a 5-point scale.

6 Changes in Schools
COLLABORATION & CULTURE

❖ *Did a2i bring about changes in school collaboration and inquiry around formative assessment?*

A constant in our conversations with a2i teachers, principals, and the project team was that schools are different. Variations were apparent in our earliest conversations with teachers about school inquiry teams, and in our final interviews with Instructional Specialists sharing final insights about their work in building capacity in school teams and supporting teacher leaders. These observations corroborate a finding that runs through this report, one important but not surprising to others who might replicate the a2i model: change takes time, and context matters.

The variations in schools and our assessments of changes did not always form clear categories: changes in formative assessment practices school-wide—a medium-term goal in the logic model—took place in schools with different histories and structures around formative assessment: in schools where inquiry teams and school structures supporting them were already strong; in schools where inquiry team work was limited, especially after it was no longer mandated by the city, but math department teams were strong and provided a model for school-wide work; and in schools where, prior to a2i, neither inquiry nor math teams had gelled, but a2i strategies and support from Instructional Specialists provided the focus and structure needed to engage in effective inquiry around student work.

The factor that did seem to move school teams forward was flexibility on the part of the Instructional Specialists, and support designed to help teachers make better decisions and become teacher leaders. This final section explores teachers' perceptions of support and changes in school collaboration, as they evolved over four years of a2i implementation. It summarizes overall survey findings and again reviews differences based on teachers' level of involvement, and also includes discussions of four schools not so much representative of school types as of the variation in a2i's 31 schools.

Collaboration and Support around Inquiry

The implementation fidelity ratings show that, at the school level, a2i schools met the thresholds set for inquiry team quality indicators in all four years of the study. Our reviews of teachers' ratings of a2i's core components also show that teachers saw the inquiry-team work as critical, but not as effective or instrumental part of their implementation as the PD, IS support, or curricular resources.

The Year 4 survey results also point to this modest level of success and nascent progress. For example, a third of the respondents (34.4%) reported that they collaborated with colleagues to design math lessons or units "almost always"; another 41.0% reported doing so "pretty often." About two in five (42.6%) of the teacher respondents indicated that they only "once in a while" examined student work as part of their inquiry or departmental team, and close to the same number (39.3%) said it happened "pretty often." Over half the respondents (58.7%) strongly agreed teachers have common planning time with their departmental teams, 50.8% indicated that there was an emphasis on peer collaboration, and 42.9% of teachers drew on the expertise of their colleagues. Assessments of change indicate that, on average, 40% of the teachers were engaging in inquiry or formative assessment practices more than they did prior to a2i (see Table 27).

TABLE 27. TEACHERS' REPORTS OF CHANGES IN INQUIRY WORK

	n	**More**	**Less**	**About the Same**
Collaborate with colleagues to design math lessons or units	57	25 (43.9%)	6 (10.5%)	26 (45.6%)
Examine student work as a part of inquiry team or departmental meetings	58	25 (43.1%)	10 (17.2%)	23 (39.7%)
Use formative assessments to gauge students' math skills and plan instruction	58	22 (37.9%)	5 (8.6%)	31 (53.4%)
Revise instruction based on formative assessments	57	24 (42.1%)	4 (7.0%)	29 (50.9%)
Discuss next steps in instruction based on assessments with your inquiry team	58	18 (31.0%)	7 (12.1%)	33 (56.9%)

Survey responses from the matched groups of teachers—those completing both a baseline and follow-up survey—mirror these findings. Ratings for how frequently they engaged in inquiry work ticked up over the course of the a2i participation, though ratings for how often teachers used formative assessment were generally higher than ratings for how often they did so as a team.

…as a team, we have made leaps and bounds compared to previous years when we were barely working together. This is something we will work towards next year.

Where we did see some significant differences in the matched group's ratings was in items related to school-wide practices.

TABLE 28. MATCHED GROUP: CHANGES IN SCHOOL WIDE INQUIRY ACTIVITIES

To what extent do you agree that the following statements apply to your school?	*N*	Mean Baseline	*SD*	Mean Follow-up	*SD*	Change	Sig.
Teachers draw on the expertise of colleagues (department chair, other teachers, coaches)	48	3.21	0.85	3.46	0.58	0.25	Ns
Teachers get ongoing guidance from a2i Instructional Specialists	47	2.96	1.06	3.34	0.84	0.38	p = .043
There is an emphasis for peer collaboration (inquiry teams, PLCs, lesson sharing)	27	3.00	0.96	3.63	0.57	0.63	p = .003
Teachers have common planning time with their departments	47	3.00	1.10	3.26	0.92	0.26	Ns
Teachers can translate Common Core Standards into practice	38	3.24	0.68	3.50	0.56	0.26	p = .048
There is a school wide focus on formative assessment	38	3.34	0.63	3.47	0.73	0.13	Ns

Effects of Extended Involvement

Results based on whether teachers described their classrooms as a2i or a mix, or whether teachers were in the earlier or more recent cohorts, were mixed. Descriptive differences showed that Geometry teachers who used a mix of a2i reported that they revised instruction more often compared to previous years than a2i teachers ($M = 2.54$ vs. $M = 1.70$, $F(1, 22) = 5.443$, $p = .030$).

For most items, agreement levels with statements related to school culture were slightly higher among teachers who began a2i earlier, compared to the more recent group (see Table 29). There was, however, a statistically significant difference between the two groups in response to the question about a school-wide focus on formative assessment: Recent cohorts scored higher than early cohorts teachers, $t_{61} = -2.968$, $p = .004$. This result exceeded the Bonferroni correction p-value limit of .0125.

TABLE 29. TEACHERS' RATINGS FOR SCHOOL COLLABORATION

Item	**Teacher Group**	**n**	**Mean**	**Strongly/Somewhat Disagree**	**Moderately Agree**	**Strongly Agree**
Teachers draw on the expertise of colleagues (dept. chair, other teachers, etc.)	Early	20	3.45	0 (0.0%)	11 (55.0%)	9 (45.0%)
	Recent	43	3.28	4 (9.4%)	21 (48.8%)	18 (41.9%)
Teachers get ongoing guidance from a2i Instructional Specialists	Early	20	3.25	2 (10.0%)	10 (50.0%)	8 (40.0%)
	Recent	43	3.09	10 (23.3%)	16 (37.2%)	17 (39.5%)
There is an emphasis for peer collaboration (inquiry teams, PLCs, lesson sharing)	Early	20	3.55	0 (0.0%)	9 (45.0%)	11 (55.0%)
	Recent	43	3.37	4 (9.3%)	18 (41.9%)	21 (48.8%)
There is a school-wide focus on formative assessment	Early	20	3.10	4 (20.0%)	9 (45.0%)	7 (35.0%)
	Recent	43	3.60	0 (0.0%)	17 (39.5%)	26 (60.5%)

Cross-School Sharing

One of the original goals of a2i was sharing across schools. In Year 2, the project team organized cross-school visitations to encourage this level of collaboration. Although teachers were positive about their experiences, which included observing a2i teachers and debriefing afterwards, the logistics were complicated, and the activity was not continued.

We each teach a different subject since we are a small school…. We have improved in our ability to share what we are doing in our class, but I think we have to improve on how we share what students are learning—we need to analyze student work together more.

In a2i's final year, the project team facilitated a different kind of cross-school sharing on large campuses where individual schools were small, with too few math teachers to form a functioning team or support collaboration across math subjects. The responses from both administrators and teachers was positive, and feedback suggests that the cross-school work will continue.

CASE STUDIES OF FOUR a2I SCHOOLS

To examine a2i implementation and impact within particular contexts, we conducted a more in-depth analysis of the survey and performance data in four a2i schools, along with qualitative data from observations and interviews. Various factors went into the selection of the four schools, including the year the school started a2i, the number of participating teachers and survey respondents, visits to the schools by REA staff, and interviews with principals, teachers, and Instructional Specialists. Two of the schools, A and B, were part of a2i's first cohort, starting in 2012–13; Schools C and D joined the project in Year 2.

These cases provide some range in size, focus, demographics, and performance. They include some of the largest schools in the New Visions network, and some of the smallest. The school foci range from international studies to Career and Technical Education (CTE). There are schools with very stable faculties, including math teachers who have been at the school for five years or more, as well as schools with teachers new to the profession and teaching math. The schools provided some range in performance levels, with higher and lower performing schools among the four sites. Departmental or inquiry teams were relatively strong in two of the schools prior to a2i, while a2i participation helped strengthen the teams and structures in the other two schools.

At the same time, there were aspects of implementation at each school that allowed us to explore a2i components or implementation features. School A, for example, embraced the school team inquiry model and the Instructional Activities, and saw an opportunity, with the support of the IS, to develop leadership skills in new teachers. School B had a strong math team prior to a2i and teachers confident in their skills and willing to incorporate a2i strategies. Departmental and other support structures were also strong at School C, and seemed to provide the security to more fully institute a2i-supported changes and teacher-developed instructional activities. Prior to a2i, School C teachers had engaged in an in-depth, year-long instructional strategy that was school-wide except among the math teachers. However, with support from the a2i IS, math became part of that school-wide effort.

The cases are not meant to portray a fully articulated picture of a2i implementation, neither individually nor as a whole. The teacher sample is small, and the student sample, though sufficiently large, skews toward School D. What the studies are intended to provide, through school profiles and feedback from teachers and students, is a backdrop against which to explore the school impact of the a2i components or implementation features just described—on school structures, instructional strategies, and performance goals.

Teacher Survey Sample

A total of 20 teachers—four from School A, four from school B, five from School C, and seven from School D—completed the Year 4 survey, which was a primary data source for the case studies. The sample included teachers in all three subjects (Algebra I, Geometry, and Algebra II). Some teachers focused on implementing a2i, others combined a2i with other strategies and resources, and others chose not to use a2i (see Table 30).

TABLE 30. CASE STUDY TEACHERS' REPORTS OF LEVELS OF a2I USE

		a2i	Mix of a2i	Not a2i	TOTALS
Algebra I (n=10)	School A	2	2		4
	School B	1	1		2
	School C		2		2
	School D	2			2
Geometry (n=9)	School A	2			2
	School B		1	1	2
	School C		2		2
	School D	3			3
Algebra II (n=7)	School A	1			1
	School B		1	1	2
	School C		2		2
	School D		3		3
	TOTALS	11	14	2	27

Totals by subject exceed the numbers of survey respondents by school because some teachers taught more than one subject.

Student Survey Sample

Student survey data from the four schools also informed the case studies. The sample included close to 1500 students who took a Fall and/or Spring survey in Year 4. Table 31 shows the numbers of students per school, and the percentage of the full case study sample and the overall student survey sample they represented.

TABLE 31. STUDENT SURVEY SAMPLE BY SCHOOL

	Fall Respondents	% Fall Case Study Sample	% Full Fall Sample	Spring Respondents	% Spring Case Study Sample	% Full Spring Sample
School A	193	13.12%	5.39%	124	8.93%	3.70%
School B	100	6.80%	2.79%	95	6.84%	2.83%
School C	322	21.89%	8.99%	371	26.71%	11.06%
School D	856	58.19%	23.91%	799	57.52%	23.82%
TOTALS	1471	100%	41.08%	1389	100%	41.41%

School A Profile

School A is a technology-based school located in the South Bronx that enrolls students in grades 6–12. There are approximately 90 staff members and 684 students enrolled at School A. Almost a quarter (73%) of the teachers have more than three years of teaching experience, and the principal, who was a co-founder of the school, over 12 years of administrative experience. Of the four case study schools, School A has the highest number of special education students (31%) and the second highest number of English Language Learners (11%). The student attendance rate is 81%, with 59% of students reported as chronically absent (city averages are 89% and 30%). Most students identify as Hispanic (69.2%) or Black (27.9%).

School-wide, teachers focus on both the content knowledge and skills required for students to pass the Regents exams, using project-based learning and technology to guide instruction. Raising achievement levels has been a challenge for the school, and the principal has encouraged innovative efforts—including a2i—to help students succeed. As one of the 14 schools that made up a2i's first cohort, there was some initial uneasiness among teachers, especially about the lack of an a2i curriculum, but the teachers as well as the principal were eager to adopt new strategies to support the shift to the new Common Core standards.

Two of the six School A a2i teachers joined the project in Year 1; one transferred later from a different Year 1 a2i school; the other three joined the project in Years 2 and 3. Four of the six math teachers completed the Year 4 survey.

I followed all the structures for the big ideas, instructional routines and tried to maintain true faith to the spirit of a2i.

Teachers at School A reported that their math classes were a mixture of a2i and other approaches. All teachers frequently used the a2i website, one teacher used Delta math and other online resources, and 75% of teachers "sometimes" used ideas, lessons, and activities from other colleagues. One teacher explained that, "Additionally, we implement FALs and Balanced Assessments."

Visits to classrooms at School A matched teachers' reports that students almost always worked together in small groups (100%), acted as instructional resources for one another (100%), explained their problem-solving strategies (75%), and worked on multi-step problems "pretty often" (75%) to "almost always" (25%). Most teachers (75%) said it was important to let students puzzle things out for themselves.

In the classroom, we use many of the instructional routines. We focus on making sense of mathematical strategies and voicing our thinking.

The same was true of the inquiry team meetings observed in a2i's fourth year. Math teachers followed suggested inquiry team procedures and protocols. During inquiry team meetings, teachers rehearsed with colleagues, debriefed about Instructional Activity use in the classroom, and analyzed teachers' moves and students' work.

School B Profile

School B, also located in the Bronx, focuses on international studies, global awareness, and world languages. Most of the faculty have been teaching at the school for three or more years; all six a2i math teachers had been at the school for four or more years. Joining a2i in Year 1, the six math teachers remained the same throughout the project.

With a focus on international studies, students at the school can participate in cultural exchanges and are exposed to over 35 languages spoken at the school by students and staff. The school has a high number of English Language Learners (25%), the highest of the four focus schools, due to the international focus, and the lowest number of special education students (10%). Most students at the school identify as Hispanic (63.0%), 21.5% as Black, 9.3% Asian, and 5.6% White. The school prides itself on student and parental involvement, in part credited for a 94% attendance rate with only 19% of students having chronic absences.

I used the formative assessment lessons and balanced assessments, and the Connecting Representations protocol. Occasionally other resources as well. I did not use the FPATS as much this year, not in the same way as previous years.

Of the six math teachers at School B, four completed the final survey at the end of the program. School B was the only case study school where a math teacher identified as teaching a2i Algebra and taught other upper level math courses that were not considered a2i courses.

With a seasoned group of math teachers, the school did not meet the implementation fidelity requirements in either 2014–15 or 2015–16, but teachers remained committed to project features, especially core curricular components. Using a pick-and-choose approach, they implemented the parts of a2i that worked with their style of teaching and their students' learning needs. As one teacher noted, implementation changed some from year to year, as they combined math instructional resources. Of the teachers who took the 2015–16 survey, 67% said they frequently used the a2i website as a resource; a third also used Delta Math.

All School B teachers reported that their students almost always work together in small groups and explain their problem-solving strategies. All teachers agreed that it is important for students to puzzle things out for themselves, 67% strongly agree that students at their school regularly work through applied math problems, and 67% agreed that they like their students to master basic mathematical operations before they tackle complex problems. These responses matched what was observed by REA during classroom visits to School B.

School C Profile

School C is a small Career and Technical Education (CTE) high school located in Manhattan, with a focus on business and entrepreneurship. Teachers bring a "real-world" focus to their instruction, creating an atmosphere where students can build networks with professionals in the sports management industry and can earn industry certificates in Microsoft Office, Business Management, and Entrepreneurship upon graduating.

There are approximately 432 students in grades 9–12 at School C. A majority of students identify as Hispanic (62.5%); another 31.9%, as Black; and 2.1% Asian and White. Almost two-thirds of the students (61.6%) are male. (See SQS.) Close to a fourth (23%) of the student population qualifies for special services; 7% are English Language Learners. School C has a student attendance rate of 84%, but the number of students who are chronically absent—48%—is well above the city average of 30% (SQS).

A classroom activity that sets School C math classes apart from other a2i math classes is the use of the Socratic seminar. The school prides itself on this work, giving students the opportunity to explore a common theme and create class discussions and debates. In the years prior to and during a2i's first three years, students in most content areas participated in Socratic seminars, but math seemed less well-suited to the process of examining an issue, presenting evidence, and explaining a conclusion. The a2i approach, however, seemed to fit the Socratic goals. During SY4, math teachers worked with the Instructional Specialist, who tailored support for a2i work to this effort, incorporating Socratic seminar activities into math classes for the first time.

Joining a2i as part of Cohort 2, a total of four math teachers and one special education teacher at School C took the final survey. Teachers reported teaching a mix of the a2i curriculum and other strategies, one of which was the New Visions-endorsed *Transitions to Algebra* curriculum, which was recommended to schools in Year 3 to introduce students identified as not fully ready for Algebra I to algebraic thinking and problem-solving. School-wide, all freshman enrolled in Algebra I received both *Transitions to Algebra* and Introduction to Algebra in the same term. As part of the collaboration and planning, a2i teachers work to have the two courses complement each other.

[Transitions to Algebra] seems to be a stand alone due to its lesson plans. But through common planning we align topics with the Algebra course as much as possible.

In School C math classrooms, students were, according to 75% of the a2i teachers, almost always working together in small groups and working on multi-step problems. According to all the teachers, students also explained their problem-solving strategies and solved the same problems using more than one method "pretty often" to "almost always." All teachers at School C responding to the Year 4 survey agreed they regularly had students work through applied problems. All also agreed that when students are working on math problems, the emphasis should be on getting the correct answer rather than on the process followed.

School D Profile

School D is the largest of the four case study schools with a student population of approximately 1,277 students in grades 9–12. Students receive a "rigorous academic" curriculum with a liberal arts focus, using the most advanced technologies.

Located in Brooklyn, School D has the smallest percentage of English Language Learners (6%) of the four case study schools, and has 20% of students qualifying for special education services. Students had a high attendance rate (94%), with 20% of students reportedly having chronic absences. There were more males than females at the school (56.6% males and 43.4% females). A little over half the students (52.2%) identify as Hispanic, 19.5% as Asian, 17.2% White, and 8.1% Black.

School D joined a2i in Year 2 (2013–2014), and teachers showed a strong commitment throughout the project. That commitment was apparent in their survey responses: seven of the 12 math teachers completing the survey reported following the a2i model in Algebra I or Geometry.

School D also provided the structure and support needed for strong teamwork, and, as a math department, teachers worked in content groups. As one teacher explained, the

[We}…look at the overall course and find connections between different units. We now use longer units with more overlap instead of smaller content-packed units.

common planning time and the fact that multiple teachers were teaching the same course allowed them to make connections between units.

Our department has broken down the math practice standards into sub-skills and use them in our inquiry and lesson study cycles…all math teachers have a common planning period during the school day where we meet with our math content teams.

The numbers of teachers, the structures, and the emphasis on collaboration, along with support from the a2i IS, also helped teachers explore connections across courses.

School D teachers all agreed that they regularly have students work through applied math problems (85.7% agreed, 14.3% strongly agreed) and that it is important to let students' puzzle things out for themselves (28.6% agreed, 71.4% strongly agreed). Like a2i colleagues in other case study sites, they also agreed that it is important to have students master basic mathematical operations before they tackle complex problems (71.5% agreed). At School D, all teachers reported that students worked together in small groups almost always. According to most teachers (71.4%), students "almost always" explained their problem-solving strategies. Fewer, but over half of teachers (57.1%), "almost always" had students work on multi-step problems.

Implementation Fidelity in the Case Study Sites

The program fidelity ratings for the case-study sites, like the overall ratings, showed more fidelity in Year 3 than in Year 4. Findings for Schools B and C differed somewhat from the other sites, with more unmet components starting in Year 3.

Contextual knowledge about the schools also shed some light on the ratings. School A, which showed a strong commitment to the PD and the inquiry-team work provided or encouraged by a2i, had the highest fidelity ratings overall (see Table 32). The area where School C did meet the fidelity thresholds both years was in the use of the common sequence of core math units, an element of a2i that appealed to teachers participating in Year 1, and a component that remained a core part of their implementation. School C, by contrast, met few of the curricular or instructional fidelity benchmarks either year, and thus did not use the common set of units, FALs, or re-engagement activities, but the school did incorporate Transition to Algebra, which seemed a better fit for its students. Fidelity to the inquiry team quality indicators did not fall off in Year 4 in School D, which, of the four schools, had and sustained more collaborative teamwork around student work and lesson planning around it.

TABLE 32. COMPONENT AND INDICATOR STATUS FOR FIDELITY OF IMPLEMENTATION, YEARS 1–4

Key Components of a2i	Key Indicators for Each Key Component	School A		School B		School C		School D	
		Y3	Y4	Y3	Y4	Y3	Y4	Y3	Y4
1. Professional Development	• Summer workshop (school level)	M	M	U	U	M	M	M	M
	• Quarterly cross-school assessment and analysis sessions (school level)	M	U	U	U	M	U	M	U
	• Monthly cross-school sessions (school level)	M	U	U	U	M	M	M	M
	Component-level status	M	U	U	U	M	M	M	U
2. School Support for Collaborative Inquiry	• Coaching for math department from IS (school level)	M	M	M	U	U	U	M	U
	• Weekly math inquiry team meeting (school level)	M	M	U	U	U	U	U	U
	• Inquiry team meeting quality indicators (school level)	M	M	M	U	M	M	M	M
	Component-level status	M	M	U	U	U	U	M	U
3. Training for Administrators	• Information sharing for P or AP in inquiry framework	M	M	M	M	M	M	M	M
	• Signed MOU	M	M	M	M	M	M	M	M
	Component-level status	M	M	M	M	M	M	M	M
4. Instructional Practice Changes	• Common set of math units	M	M	M	U	U	U	M	U
	• Common sequence of core math units	M	M	M	M	M	M	M	M
	• Implementation of all FALs	M	U	M	U	U	U	U	U
	• Enactment of all re-engagement experiences	M	U	M	U	U	U	M	U
	Component-level status	M	U	M	U	U	U	U	U
5. Routine Use, Student Data Mgmt. Monitor Progress	• Routine use, student data mgmt. system (teacher level)	M	U	M	U	U	U	M	U
	Component-level status	M	U	M	U	U	U	M	U
INDIVIDUAL ITEM MET		13	8	9	3	7	6	11	6
COMPONENT LEVEL MET		5	2	3	1	2	2	4	1

Case Study Teachers' Ratings of a2i's Critical Components

Case study teachers' ratings of a2i's critical components generally mirrored the overall ratings, but, again, emphases by school emerged. Table 31 shows the ratings of critical components by school.

Professional Development

Survey responses indicated that the math teachers at the case study schools considered their department meetings as one of the most critical of the PD components. In addition:

- All math teachers at Schools A, C, and D agreed the all-day PD sessions during the school year were a critical component in helping implement a2i, while only one of the three math teachers at School B indicated this was a critical component. (Results were statistically significant).
- At School B, where teachers and administrators remained committed to a2i, they also tended to be selective about a2i strategies and PD sessions.

Inquiry and Support

Across the four schools, teachers agreed that support and buy-in from school administrators was a critical a2i component (School A: 75.0%, School B: 66.7%, School C: 66.7%, and School D: 85.7%) along with Instructional Specialists work with school teams (Schools B, C, and D: 66.7%, School A: 50.0%).

- Inquiry team meetings were rated highest at School A, where the IS helped teachers strengthen teamwork, and at School D, where structures were in place prior to a2i and supported further inquiry work.
- All teachers at School C, 66.7% at School B, and 50.0% at School D said having individual support and coaching from Instructional Specialists was critical to a2i.
- Teachers at School A all agreed that IS support for teacher leaders was a critical component to a2i, while 83.3% of teachers at School D and 66.7% at School C agreed.

Curricular Strategies and Assessments

Ratings from teachers at the case study schools were mixed when it came to the a2i curriculum.

- Most teachers found the a2i unit structure, composed of the Initial Task (IPAT), learning plan, Classroom Challenges (FAL), re-engagement lessons, and Final Assessment (FPAT) to be a critical component of a2i, but School B teachers, who were introduced to the unit structure in a2i's first year and selective about which elements they incorporated into practice, assigned the highest ratings to that component. Fewer teachers found the a2i curriculum, including the IAs, to be a core component of the a2i program.
- While all teachers at School B agreed that the FALs were a critical component for a2i, only 50.0% of teachers at School A, 33.3% at School C, and 28.6% at School D found the FALs to be critical.
- All teachers at Schools A and D, and 66.7% of the teachers at School C found the a2i curriculum, introduced later in the project, to be critical to the a2i program, compared to 33.3% at School B.
- All teachers at School A and 85.7% of teachers at School D (who had created IAs for Geometry) said the IAs were critical. Those figures ticked down some at Schools B (66.7%) and C (50.0%).
- In general, School C teachers rated curricular components a little lower than teachers at the other three sites. There, *Transition to Algebra* became an important part of their approach to supporting struggling math students.

Online Resources

Ratings for the a2i website as a source for resources and information were unanimously high: All teachers at the four case study schools agreed the website resources were a critical component. Three teachers (2 at School C and 1 at School D) indicated that the a2i website, as a place to share and interact with other math teachers, was also a critical feature.

TABLE 33. CRITICAL COMPONENTS IN HELPING IMPLEMENT a2i DURING THE 2015–16 SCHOOL YEAR BY SCHOOL

		School A	School B	School C	School D
Professional Development	August PD sessions	3 (75.0%)	0 (0.0%)	3(100.0%)	3 (50.0%)
	Math department team meetings	4 (100.0%)	2 (66.7%)	3(100.0%)	7 (100.0%)
	All-day a2i PD sessions during SY	4 (100.0%)	1 (33.3%)	3 (100.0%)	7 (100.0%)
	Collaborations with other a2i teachers	3 (75.0%)	2 (66.7%)	3 (100.0%)	7 (100.0%)
	PD or support fro Mock Regents prep	1(25.0%)	0 (0.0%)	1 (50.0%)	2 (33.3%)
Inquiry & Support	Support/ buy-in, administrators	3 (75.0%)	2 (66.7%)	2 (66.7%)	6 (85.7%)
	School inquiry team meetings	2 (50.0%)	2 (66.7%)	1 (33.3%)	4 (80.0%)
	Individual support & coaching, IS	1 (25.0%)	2 (66.7%)	3 (100.0%)	3 (50.0%)
	IS work with school teams	2 (50.0%)	2 (66.7%)	2 (66.7%)	4 (66.7%)
	IS support for teacher leaders	3 (100.0%)	1 (33.3%)	2 (66.7%)	5 (83.3%)
Curricular Strategies and Assessment	a2i unit design schematic	3 (75.0%)	3 (100.0%)	2 (66.7%)	5 (83.3%)
	a2i curriculum	3 (100.0%)	1 (33.3%)	2 (66.7%)	7 (100.0%)
	Re-engagement lessons	3 (75.0%)	2 (66.7%)	1 (33.3%)	2 (40.0%)
	Instructional Activities	4 (100.0%)	2(66.7%)	1 (50.0%)	6 (85.7%)
	Classroom Challenges ([FALs])	2 (50.0%)	3 (100.0%)	1 (33.3%)	2 (28.6%)
	The Balanced Assessment System	3 (75.0%)	2 (66.7%)	0 (0.0%)	5 (71.4%)
Online Resources	a2i website, resources & information	4 (100.0%)	3 (100.0%)	3 (100.0%)	7 (100.0%)
	a2i website, place to share/interact	0 (0.0%)	0 (0.0%)	2 (66.7%)	1 (16.7%)

Note. The table only includes the *n* and % for those indicating it was a Critical Component of the program.

Most teachers at the case study schools reported using the a2i website as their main source for math resources and materials. One teacher at School C indicated using the website occasionally, but also relying on resources such as Delta Math and other online resources. Teachers seemed to veer away from using textbooks, which were traditionally used prior to the Common Core and a2i. They also indicated less use of the EngageNY web resources, which were available to schools across New York as the state transitioned from the Regents exams to the Common Core exams. School A teachers were the most frequent users of the website, but also sought out other resources, which fit their reports of teaching math as a mixture of a2i and other approaches. a2i teachers new to the profession appeared eager to stock their toolkits with multiple resources (see Appendix F).

Engagement in Activities by School

As seen in Table 34, engagement in a2i activities not only varied by school, but also by teacher. The majority of teachers at the four schools said they use formative assessments to gauge students' math skills and plan instruction and use formative assessments to revise instruction "pretty often" to "almost always." Where responses somewhat shift is in the number of teachers indicating how often they engage with fellow teachers to examine student work and design next steps.

- The majority of teachers at School A (75% pretty often, 25% almost always), whose inquiry teamwork has evolved over the course of a2i, said that they use inquiry meetings to discuss next steps in instruction based on the assessments they give students. Teachers also used formative assessments (50% pretty often, 50% almost always) to gauge students' math skills and plan instruction.

- Teachers at School B, where math department procedures and instructional styles seemed to be set prior to a2i, engaged in a2i activities related to inquiry and departmental teams less frequently than teachers at the other case study schools. 100% of teachers at School B said "only once and a while" to they examine student work; 66.7% also said that discussing next steps in instruction only occurs once in a while as part of their inquiry team or departmental team meetings.

- At School C, teachers were divided with how often they discuss next steps in instruction with their inquiry teams: 33.3% said once in a while, 33.3% indicated pretty often, and the remaining 33.3% said almost always.

- Teachers at School D, with structures encouraging teamwork in place prior to a2i and evolved throughout the project, reported collaborating with colleagues to design math lessons or units more frequently than teachers at other case study schools: 71.4% said they collaborate pretty often; 14.3% said almost always.

TABLE 34. CASE STUDY TEACHERS' ENGAGEMENT IN COLLABORATIVE ACTIVITIES

Item	**School**	***n***	***Mean***	**Never, Rarely/ Once in a while**[a]	**Pretty often**	**Almost always**
Collaborate with colleagues to design math lessons or units	School A	4	3.25	1 (25.0%)	1 (25.0%)	2 (50.0%)
	School B	3	2.33	2 (66.7%)	1 (33.3%)	
	School C	4	3.25	1 (25.0%)	1 (25.0%)	2 (50.0%)
	School D	7	3.57	1 (14.3%)	1 (14.3%)	5 (71.4%)
Examine student work as a part of inquiry team or departmental meetings	School A	4	2.25	1[a] (25.0%)	1 (25.0%)	2 (50.0%)
	School B	3	2.00	3 (100.0%)		
	School C	4	3.00	1 (25.0%)	2 (50.0%)	1 (25.0%)
	School D	7	3.00	1 (14.3%)	5 (71.4%)	1 (14.3%)
Use formative assessments to gauge students' math skills and plan instruction	School A	4	3.50		2 (50.0%)	2 (50.0%)
	School B	3	3.67		1 (33.3%)	2 (66.7%)
	School C	4	2.75	1[b] (25.0%)	2 (50.0%)	1 (25.0%)
	School D	7	3.14		6 (85.7%)	1 (14.3%)
Revise instruction based on formative assessments	School A	4	3.25	1 (25.0%)	1 (25.0%)	2 (50.0%)
	School B	3	3.00	1 (33.3%)	1 (33.3%)	1 (33.3%)
	School C	3	3.33		2 (66.7%)	1 (33.3%)
	School D	7	3.57		3 (42.9%)	4 (57.1%)
Discuss next steps in instruction based on assessments with your inquiry team	School A	4	3.25		3 (75.0%)	1 (25.0%)
	School B	3	2.33	2 (66.7%)	1 (33.3%)	
	School C	3	3.00	1 (33.3%)	1 (33.3%)	1 (33.3%)
	School D	7	2.86	2 (28.6%)	4 (57.1%)	1 (14.3%)

[a] All values in this column are for the response option "Once in a while" unless otherwise noted. [b] Never/Rarely.
Mean scores at three or above are shaded.
Source: 2015-16 Teacher Survey

Impact on Teachers, Schools, and Students

Changes in Practice

On the 2015-16 survey, most teachers from the four schools reported a modest to significant change in their teaching behaviors since starting the a2i project.

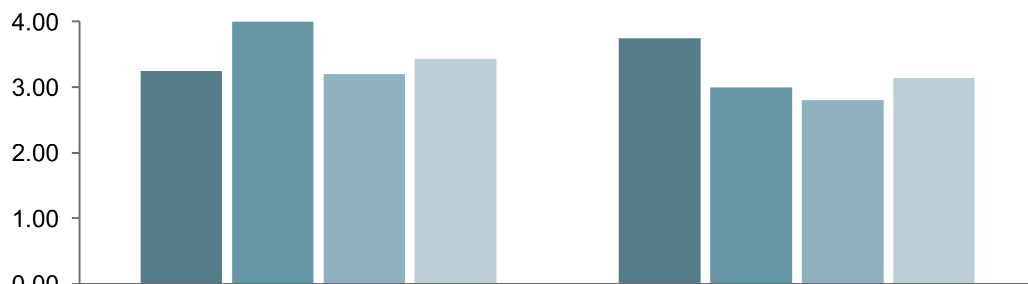
- Teachers at School A reported the highest levels of change in their use of data to inform instruction, possibly attributable to the focus on formative assessment and analysis of student work in inquiry teams.

- School B reported higher levels of change in their approach to teaching math, with all teachers rating the changes as "significant". As with other findings, this may reflect an early investment in a2i strategies and the unit structure. Other teachers, at Schools A, C, and D said a2i had a "modest" to "significant" effect on their approach to teaching math.

- Means for use of formative assessment were highest for School A, where 75.0% said a2i helped them use data to inform their instruction. Means were also high for Schools B and D; however, one teacher at each school stated a2i did not help them use data to inform instruction.

- School D reported higher levels of change in ways their school supports and implements inquiry work.

FIGURE 11. CASE STUDY TEACHERS' REPORTS OF CHANGES RESULTING FROM a2I

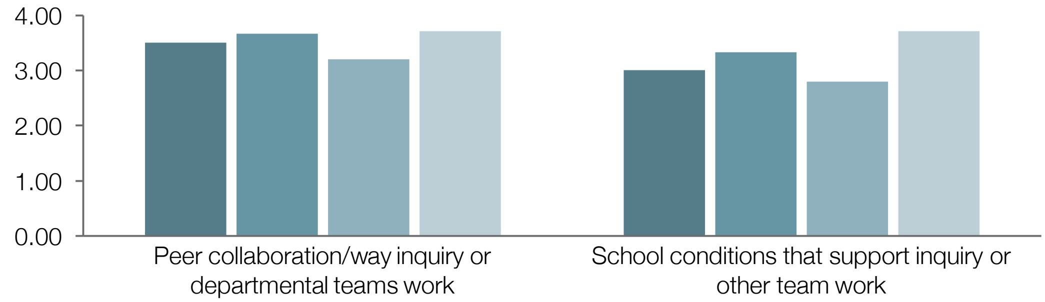


	Your approach to teaching math	Your use of data to inform instruction
School A	3.25	3.75
School B	4.00	3.00
School C	3.20	2.80
School D	3.43	3.14

Changes in School Support

Case study teachers also reported modest to significant changes in school conditions or support, with fewer differences between schools. Means were highest for peer collaboration or the ways departmental teams work, somewhat lower for school conditions that support inquiry. With a strong base to build from, teachers at School D reported the greatest degree of change in both. (See Figure 12.)

FIGURE 12. DEPARTMENTAL CHANGES TO SUPPORT INQUIRY BY SCHOOL



	Peer collaboration/way inquiry or departmental teams work	School conditions that support inquiry or other team work
School A	3.50	3.00
School B	3.67	3.33
School C	3.20	2.80
School D	3.71	3.71

Changes in Students' Pass Rates and College Readiness

One-year pass rates for Algebra ninth graders at case study sites A–C increased from the first to second cohorts (School D waits until 10th grade to teach Algebra). Rates also increased in comparison schools. In the School B pairing, a2i rates were slightly higher. The year-to-year improvement was greatest for School C (16.8%), but not as great as the improvements in the comparison school. There were also marked gains in the algebra college readiness rates in Schools A–C, in all cases exceeding the improvements in the comparison sites (see Table 35).

TABLE 35. PASS RATES AND COLLEGE READINESS RATES IN ALGEBRA

	Pass Algebra			**College Ready in Algebra**		
	Cohort 1 (2018)	Cohort 2 (2019)	Change	Cohort 1 (2018)	Cohort 2 (2019)	Change
School A	34.2%	47.4%	13.2%	7.0%	34.5%	27.5%
Comparison	8.5%	6.8%	-1.7%	2.1%	6.8%	4.7%
School B	48.5%	63.1%	14.6%	24.3%	58.6%	34.3%
Comparison	50.4%	62.5%	12.1%	22.3%	40.0%	17.7%
School C	37.7%	54.5%	16.8%	10.4%	29.1%	18.7%
Comparison	34.9%	61.8%	26.9%	19.8%	37.3%	17.5%

Learning Environment and Attitudinal Changes

In three of the four case study schools, student feedback indicated positive learning environments in their math classes, and modest changes from Fall to Spring in most schools, in areas such as group work and feedback practices. The most consistent changes were in School C, although for some of the items, the Fall or pre-means were lower; the highest levels, pre- and post-, were in Schools B and D. (See Table 36.)

TABLE 36. STUDENTS' FEEDBACK ON CLASSROOM ENVIRONMENT, CASE STUDIES

	School A			School B			School C			School D		
	n	Pre	Post	*n*	Pre	Post	*n*	Pre	Post	*n*	Pre	Post
Student Behaviors (Math Class)	73	2.94	2.96	50	3.08	3.03	229	2.72	2.93	662	2.94	3.06
Teacher Practices/Behaviors	75	3.28	3.26	46	3.37	3.41	224	3.08	3.23	646	3.29	3.38
Classroom Culture [A]	76	3.38	3.29	49	3.38	3.50	225	3.31	3.35	654	3.43	3.44
Classroom Culture [B]	77	3.84	3.77	49	4.26	4.15	227	3.75	3.79	659	4.04	4.08
Feedback Practices	72	4.12	3.97	41	4.38	4.44	216	3.96	4.02	641	4.24	4.29

Students' had a strong level of agreement at all schools that they could understand the basic math concepts taught in the class. Means from Fall to Spring dropped slightly at Schools A and B, and increased slightly at Schools C and D. In the Spring, students at School D had the highest level of agreement that they understood the harder concepts in the class.

Students at School C also reported the most consistently positive changes in their interest and confidence in math. Again, the starting means were lower, but, except for the utility of math, means increased from Fall to Spring (see Table 37).

TABLE 37. STUDENTS' ATTITUDES TOWARDS MATH, CASE STUDY SCHOOLS

	School A			School B			School C			School D		
	n	Pre	Post	N	Pre	Post	n	Pre	Post	n	Pre	Post
Math Self-Efficacy	72	3.14	3.07	52	3.26	3.02	230	3.03	3.08	660	3.06	3.14
Math Interest/Intrinsic Value	76	2.49	2.45	55	2.77	2.56	231	2.45	2.60	655	2.50	2.58
Math Value/Utility Value	74	3.19	3.01	49	3.39	3.22	214	3.15	3.10	646	3.19	3.10
Perseverance	76	3.23	3.05	50	3.30	3.19	227	3.05	3.09	666	3.14	3.12
Enjoyment	77	3.60	3.55	50	4.31	4.20	228	3.61	3.66	664	3.93	3.97

Teachers' ratings of their students' interest in math mirrored students' agreement in math on the student surveys. Ratings from teachers were lower than those of students, except at School C, where teacher ratings were higher than students.

References

Allison, P.D. (2001). *Missing Data*. Sage University Papers Series on Quantitative Applications in the Social Sciences, 07-136. Thousand Oaks, CA: Sage.

Bates, D., Maechler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software, 67(1),* 1–48.

Bowers, J., Fredrickson, M. and Hansen, B. (2016). *RItools: Randomization Inference Tools*. R package version 0.1-13.

Connell, J. P., & Kubich, A.C. (1995). Applying a theory of change approach to the evaluation of comprehensive community initiatives: Progress, prospects, and problems. In K. Fulbright-Anderson, A.C, Kubisch, & J. P. Connell (Eds., *New approaches to evaluating community initiatives: Theory, measurement, and analysis* (Vol. 2), pp. 15–44. Washington, DC: Aspen Institute.

Daro, P. (2014). *CCSS Mathematics* [PowerPoint slides]. Retrieved from: www.svmimac.org/images/PILMtg021214.PhilDaro.ppt

Dweck, C. S. (2008). *Mindset: The new psychology of success*. New York: Ballantine Books, 2008.

Fisher, L. (2007). *Mathematics Assessment Collaborative Final Report*. Palo Alto, California: The Noyce Foundation, https//www.noycefdn.org/math/

Foster, D. (2007) *When assessment guides instruction: Silicon Valley's mathematics assessment collaborative. Assessing Mathematical Proficiency*. Berkeley, California: Mathematical Sciences Research Institute Publications, 53.

Hansen, B. B. (2004). Full matching in an observational study of coaching for the SAT. *Journal of the American Statistical Association, 99*, 609–618.

Hansen, B. B., & Klopfer, S. O. (2006). Optimal full matching and related designs via network flows. *Journal of Computational and Graphical Statistics, 15,* 609–627.

Hansen, B. B., & Bowers, J. (2008). Covariate balance in simple, stratified and clustered comparative studies. *Statistical Science, 23(2),* 219–236.

Ho, D. E., Imai, K., King, G., & Stuart, E. A. (2011). MatchIt: Nonparametric preprocessing for parametric causal inference. *Journal of Statistical Software, 42(8*), 1–28.

Ho, D., Imai, K., King, G., & Stuart, E. (2007), Matching as nonparametric preprocessing for reducing model dependence in parametric causal inference. *Political Analysis, 15,* 199–236.

Iacus, S. M., King, G., & Porro, G. (2009). CEM: Software for coarsened exact matching. *Journal of Statistical Software, 30*, 1–27.

National Governors Association Center for Best Practices, Council of Chief State School Officers (2010). *Common Core State Standards for Mathematics*. National Governors Association Center for Best Practices, Council of Chief State School Officers: Washington D.C.

R Core Team (2015). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. URL http://www.R-project.org/.

Sebring, P. B., Allensworth, E., Bryk, A. S., Easton, J. Q., & Luppescu, S. (2006). *The essential supports for school improvement.* Chicago, IL: Consortium on Chicago School Research at the University of Chicago.

Thoemmes, F. (2012). *Propensity score matching in SPSS*. arXiv:1201.6385 [stat.AP]

Thoemmes, F., & Liao, W. (2013, May). *Propensity score matching (with multilevel data) using SPSS and R*. Modern Modeling Methods Conference.

Wiliam, Dylan (2011). *Embedded formative assessment—practical strategies and tools for K–12 teachers*. Bloomington, IN: Solution Tree.

APPENDIX A: FIDELITY FINDINGS, BY YEAR

For further details about the fidelity findings during each year of the project, fidelity of implementation is reported separately by year in the following tables. Findings in the following tables includes all the a2i schools in the impact study analyses.

TABLE A1. FINDINGS FROM IMPLEMENTATION YEAR 1 (SY2012-13)

Intervention Components	Number of indicators representing each component	Sample Size at the Sample Level (# of schools, districts, etc.)	Representativeness of sample: Measured on All (A), Some (S), or None (N) of the units representing the intervention group in the impact analyses[b]	Component Level Threshold for Fidelity of Implementation for the Unit that is the Basis for the Sample-Level	Evaluator's Criteria for "Implemented with Fidelity" at Sample Level	Component Level Fidelity Score for the Entire Sample	Implemented with Fidelity? (Yes, No, N/A)
Professional Development	3	10 schools	A	9	9	Average school score = 9	Yes
School Support for Collaborative Inquiry	3	10	A	7	7	Average school score = 7	Yes
Training for Administrators	2	10	A	2	2	Average school score = 2	Yes
Instructional Practice Changes	4	10	A	4	4	Average school score = 4	Yes
Routine Use of Student Data Management System to Monitor Students' Progress	1	10	A	2	2	Average school score = 2	Yes

TABLE A2. FINDINGS FROM IMPLEMENTATION YEAR 2 (SY2013-14)

Intervention Components	Number of indicators representing each component	Sample Size at the Sample Level (# of schools, districts, etc.)	Representativeness of sample: Measured on All (A), Some (S), or None (N) of the units representing the intervention group in the impact analyses[b]	Component Level Threshold for Fidelity of Implementation for the Unit that is the Basis for the Sample-Level	Evaluator's Criteria for "Implemented with Fidelity" at Sample Level	Component Level Fidelity Score for the Entire Sample	Implemented with Fidelity? (Yes, No, N/A)
Professional Development	3	21	A	9	9	Average school score = 9	Yes
School Support for Collaborative Inquiry	3	21	A	7	7	Average school score = 7	Yes
Training for Administrators	2	21	A	2	2	Average school score = 2	Yes
Instructional Practice Changes	4	21	A	4	4	Average school score = 4	Yes
Routine Use of Student Data Management System to Monitor Students' Progress	1	21	A	2	2	Average school score = 2	Yes

TABLE A3. FINDINGS FROM IMPLEMENTATION YEAR 3 (SY2014-15)

Intervention Components	Number of indicators representing each component	Sample Size at Sample Level (# of schools, districts, etc.)	Representativeness of sample: Measured on All (A), Some (S), or None (N) of the units representing the intervention group in the impact analyses[b]	Component Level Threshold for Fidelity of Implementation for the Unit that is the Basis for the Sample-Level	Evaluator's Criteria for "Implemented with Fidelity" at Sample Level	Component Level Fidelity Score for the Entire Sample	Implemented with Fidelity? (Yes, No, N/A)
Professional Development	3	20	A	9	9	8	No
School Support for Collaborative Inquiry	3	20	A	7	7	6	No
Training for Administrators	2	20	A	2	2	3	Yes
Instructional Practice Changes	4	20	A	10*	10*	9	No
Routine Use of Student Data Management System to Monitor Students' Progress	1	20	A	3	3	2	No

** Beginning in year 3, the thresholds changed for the "instructional practice changes" and "routine use of student data" components to create a more nuanced scoring system (changed dichotomous (1,0) scoring to 1-4 scale).*

TABLE A4. FINDINGS FROM IMPLEMENTATION YEAR 4 (SY2015-16)

Intervention Components	Number of indicators representing each component	Sample Size at the Sample Level (# of schools, districts, etc.)	Representativeness of sample: Measured on All (A), Some (S), or None (N) of the units representing the intervention group in the impact analyses[b]	Component Level Threshold for Fidelity of Implementation for the Unit that is the Basis for the Sample-Level	Evaluator's Criteria for "Implemented with Fidelity" at Sample Level	Component Level Fidelity Score for the Entire Sample	Implemented with Fidelity? (Yes, No, N/A)
Professional Development	3	20	A	9	9	6	No
School Support for Collaborative Inquiry	3	20	A	7	7	5	No
Training for Administrators	2	20	A	2	2	3	Yes
Instructional Practice Changes	4	20	A	10*	10*	4	No
Routine Use of Student Data Management System to Monitor Students' Progress	1	20	A	3	3	2	No

******* *Beginning in year 3, the thresholds changed for the "instructional practice changes" and "routine use of student data" components to create a more nuanced scoring system (changed dichotomous (1,0) scoring to 1-4 scale).*

APPENDIX B: STUDENT PROPENSITY SCORE MATCHING

To examine the use of nearest neighbor matching, we used the SPSS custom dialogue for propensity score matching, the "psmatching" program, which was developed by Thoemmes and Liao (Thoemmes & Liao, 2013). The psmatching program uses the following R packages to conduct the analyses: the "MatchIt" package (Ho, Imai, King, & Stuart, 2007), "RItools" package (Bowers, Fredersickson, & Hansen, 2010), and "cem" package (Iacus, King, & Porro, 2009).

A potential drawback of using the nearest neighbor method of matching is that it can result in unmatched treated cases. However, given that both optimal matching and nearest neighbor matching performed relatively similar with each year of algebra data (year 3 and year 4), the final choice was to use nearest neighbor matching for the final determination of 1:1 student matching based on propensity scores computed separately for each dataset. Although optimal matching identifies matches for all treatment cases, unlike the nearest neighbor approach, which leaves out a set of unmatched treatment cases, it was deemed more important to include the specified parameters for exact matching that is allowed under nearest neighbor matching. Those parameters are ignored for optimal matching.

Using the algebra data from the first year of the impact study as an example, with a sample of 1,536 treated cases and 2,658 control cases, the use of nearest neighbor matching yielded 1,442 treatment/control pairs of students for year 3 algebra estimates of impact. There were 94 treated cases that were unmatched and thus not retained in the dataset. The overall balance test (Hansen & Bowers, 2010) had a X^2 (df = 2) of 1.747, p = .417, indicating the null hypothesis of imbalance could be rejected. The measure of relative multivariate imbalance (Iacus, King, & Porro, 2010) was L1 = .237 after matching, while before matching it was .239; thus, matching did little to improve the balance between covariates. No standardized mean differences between covariates were greater than .25.

For the geometry dataset, after using exact matching on the key demographic variables (grade level, IEP, ELL, sex, and ethnicity) and nearest neighbor matching for math and reading pretests, the overall relative multivariate imbalance improved slightly; although, imbalance was negligible before matching. One-to-one matches were determined for 2,090 students (1,045 treated, 1,045 comparison), which left 264 treated and 1,223 comparison students unmatched. The overall balance test had a X^2 (df = 2) of 1.99, p = .40, which indicated the null hypothesis of imbalance could be rejected. Relative multivariate imbalance (L1) slightly improved from .193 before matching to .166 after matching. Although the absolute standardized difference in means increased slightly for each covariate, the differences were still very small (<.06).

APPENDIX C: ANALYTIC APPROACH, BASELINE EQUIVALENCE TESTING, AND FINAL HLM RESULTS

Analytic Approach

For **confirmatory achievement contrasts**, the student measures were modeled with a school level, a classroom level, and a student level, as follows:

Level 1 (student)

At Level 1, student-level test scores are modeled as a function of student-level baseline covariates:

(1) $Y_{ijk} = \beta_{0jk} + \Sigma\beta_{njk} X_{nijk} + e_{ijk}$

where:

Y_{ijk} = outcome for student i in classroom j in school k (e.g., Algebra I test score)

X_{xijk}= the n^{th} baseline student characteristic (n = 1,2,…,N) included in the model as a covariate. Student-level covariates include the pre-test variable (i.e., 8th grade Math score of student i) as well as other characteristics listed in Table 2 above.

e_{ijk} = student-level error term assumed to be distributed with mean of zero and variance of σ_e^2. This term is also assumed to be independent of the other error terms (classroom and school) included in the model.

Level 2 (teacher/classroom)

At the second level of the model, the intercept from the first-level, β_{0jk}, is modeled as a function of teacher and classroom characteristics:

(2) $\beta_{0jk} = \gamma_{00k} + \Sigma\gamma_{0mk} Z_{mjk} + \eta_{0jk}$

(3) $\beta_{njk} = \gamma_{n0k}$ for n=1,2,…,N

where:

Z_{mjk} = the m^{th} baseline teacher or classroom characteristic (m = 1,2,…,M) included in the model as a covariate (e.g., teacher experience). This also includes the classroom-level average of the pre-test variable selected for this outcome.

η_{0jk}= classroom-level error term assumed to be distributed with mean of zero and variance of σ_η^2. This term is also assumed to be independent of the other error terms (student and school) included in the model.

Level 3 (school)

At the third level, the intercept from the second-level, γ_{00k}, is modeled as a function of the treatment indicator, which indicates a2i schools, and school-level baseline characteristics.

(4) $\gamma_{00k} = \pi_{000} + \pi_{001}T_k + \Sigma\, \pi_{00(s+1)}\, W_{sk} + \Sigma\, \pi_{00(S+p+1)}\, PD_{sk} + \omega_{00k}$

(5) $\gamma_{m0k} = \gamma_{m00}$ for m=1,2,…,M

where:

T_k= treatment indicator which equals 1 if school k is an a2i school and 0 otherwise.

W_{sk} = s^{th} school-level baseline school characteristic included in the model as a covariate (e.g., percentage of students eligible for free lunch). This also includes the school-level average of the pre-test variable selected for this outcome as well as an indicator for observations from SY4.

> PD_{sk} = indicator variable for the p^{th} a2i-matched comparison pair where p = 1,2,…19 (i.e., we will include 19 pairs. Note that we do not have to control for the cohort variable because it is accounted for by these pair indicators.
>
> ω_{00k} = school-level error term assumed to be distributed with mean of zero and variance of σ_{ω}^{2}. This term is also assumed to be independent of the other error terms (student and classroom) included in the model.

In this system of equations, the coefficient on the treatment indicator, π_{001}, captures the average effect of the a2i initiative on this student-level outcome.

Baseline Equivalence Testing

Baseline equivalence between a2i and matched comparison groups was assessed for each primary analytic sample of students (Common Core Algebra and Common Core Geometry), which only consisted of individuals with non-missing values for both the baseline measure and the outcome measure. The student scores in math and reading from the eighth-grade state assessments were used for assessing baseline equivalence. Unconditional means and standard deviations at the student-level were used. The pooled within-group standard deviation of the baseline variable was calculated as indicated below (as described in Appendix B of the WWC Procedures and Standards Handbook, Version 2.1):

$$S_{pooled} = \sqrt{\frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{(n_1 + n_2 - 2)}}$$

where n_1 and n_2 are the student sample sizes, and S_1 and S_2 are the student-level standard deviations for the intervention group and the comparison group, respectively. Intervention effects were estimated by dividing the differences in adjusted means between the a2i and comparison groups by the unadjusted pooled within-group standard deviation (Hedge's g):

$$g = \frac{\gamma}{\sqrt{\frac{(n_1-1)S_1^2 + (n_2-1)S_2^2}{(n_1+n_2-2)}}}$$

Evidence of baseline equivalence was deemed met if the standardized mean difference was less than .25. For impact estimates in algebra, the timing of baseline measurement was spring 2013 for student Cohort 1, grade 10; spring 2014 for Cohort 1, grade 9 and Cohort 2, grade 10; and spring 2015 for Cohort 2, grade 9. For impact estimates in geometry, the timing of baseline measurement was Spring 2014 for grade 10 and spring 2015 for grade 9.

Due to matching occurring prior to inclusion of dependent variables, the final numbers for each analytic sample varied. Although there were equal numbers of students in each condition after the matching procedures, there were uneven numbers of students in each condition with the outcome test scores. The discrepancies were particularly large for the Common Core Algebra and geometry samples, with the a2i condition having roughly 20% more students in the Common Core Algebra sample and 40% more in the geometry sample.

The effect sizes for each analytic sample (see Table C1) indicated that the pretest differences were less than 0.25 of a standard deviation. However, several effects were between 0.05 and 0.25, which argues for statistical adjustment of these differences in the impact modeling. The analytic approach we took included math and reading test scores as covariates in all our analyses to increase precision of impact estimates.

TABLE C1. BASELINE EQUIVALENCE TESTING FOR PRIOR ACHIEVEMENT BY ANALYTIC SAMPLE

		a2i			Non-a2i			***Effect size (Hedge's g)***
Analytic Sample	State 8th Grade Test	*N*	*M*	*SD*	*N*	*M*	*SD*	
Algebra	Math	2140	2.2611	.53555	1731	2.2916	.55768	.05591
	Reading	2140	2.3395	.54524	1731	2.3635	.56504	.04331
Geometry	Math	736	2.5110	.64569	454	2.6125	.69851	.15233
	Reading	736	2.5335	.62642	454	2.5916	.64576	.11555

NOTE: Proficiency ratings range: 1.00–4.50. Baseline equivalence tests were performed for each analytic sample. These samples only included cases that met each of the following criteria: student in a school in the impact study, credits were attempted in the appropriate course during the school year being examined, and the appropriate outcome test score was available from the June test administration at the end of the school year.

Results of Final HLM Testing

The final HLM specified for Algebra was as follows:

Level-1 Model

$$R.RGTMAR_{iik} = \pi_{0ik} + \pi_{1ik}*(BLACKHIS_{iik}) + \pi_{2jk}*(ELL_{ijk}) + \pi_{3jk}*(R.GRD9_{ijk}) + \pi_{4jk}*(IEP_{ijk}) + \pi_{5jk}*(MA8PR_{ijk}) + \pi_{6jk}*(RD8PR_{ijk}) + e_{ijk}$$

Level-2 Model

$$\pi_{0ik} = \beta_{00k} + \beta_{01k}*(MA8PR\_ME_{jk}) + r_{0jk}$$
$$\pi_{1ik} = \beta_{10k}$$
$$\pi_{2ik} = \beta_{20k}$$
$$\pi_{3ik} = \beta_{30k}$$
$$\pi_{4ik} = \beta_{40k}$$
$$\pi_{5ik} = \beta_{50k}$$
$$\pi_{6jk} = \beta_{60k}$$

Level-3 Model

$$\beta_{00k} = \gamma_{000} + \gamma_{001}(A2I_k) + \gamma_{002}(ELL\_MEAN_k) + u_{00k}$$
$$\beta_{01k} = \gamma_{010}$$
$$\beta_{10k} = \gamma_{100}$$
$$\beta_{20k} = \gamma_{200}$$
$$\beta_{30k} = \gamma_{300}$$
$$\beta_{40k} = \gamma_{400}$$
$$\beta_{50k} = \gamma_{500}$$
$$\beta_{60k} = \gamma_{600}$$

MA8PR RD8PR have been centered around the grand mean.

MA8PR_ME has been centered around the grand mean.

ELL_MEAN has been centered around the grand mean.

Mixed Model

$$R.RGTMAR_{iik} = \gamma_{000} + \gamma_{001}*A2I_k + \gamma_{002}*ELL\_MEAN_k + \gamma_{010}*MA8PR\_ME_{jk}$$
$$+ \gamma_{100}*BLACKHIS_{iik} + \gamma_{200}*ELL_{iik} + \gamma_{300}*R.GRD9_{ijk} + \gamma_{400}*IEP_{ijk}$$
$$+ \gamma_{500}*MA8PR_{iik} + \gamma_{600}*RD8PR_{ijk}$$
$$+ r_{0jk} + u_{00k} + e_{ijk}$$

TABLE C2. FINAL ESTIMATION OF FIXED EFFECTS

		Standard		**Approx.**	
Fixed Effect	**Coefficient**	error	***t*-ratio**	*d.f.*	***p*-value**
For INTRCPT1, π_0					
For INTRCPT2, β_{00}					
INTRCPT3, γ_{000}	67.808449	0.909807	74.531	37	<0.001
A2I, γ_{001}	-0.578282	1.012716	-0.571	37	0.571
ELL_MEAN, γ_{002}	14.413882	5.622633	2.564	37	0.015
For MA8PR_ME, β_{01}					
INTRCPT3, γ_{010}	3.577541	1.376719	2.599	96	0.011
For BLACKHIS slope, π_1					
For INTRCPT2, β_{10}					
INTRCPT3, γ_{100}	-1.630934	0.587127	-2.778	3688	0.006
For ELL slope, π_2					
For INTRCPT2, β_{20}					
INTRCPT3, γ_{200}	-1.787267	0.551402	-3.241	3688	0.001
For R.GRD9 slope, π_3					
For INTRCPT2, β_{30}					
INTRCPT3, γ_{300}	-3.746352	0.312264	-11.997	3688	<0.001
For IEP slope, π_4					
For INTRCPT2, β_{40}					
INTRCPT3, γ_{400}	-3.225727	0.382579	-8.432	3688	<0.001
For MA8PR slope, π_5					
For INTRCPT2, β_{50}					
INTRCPT3, γ_{500}	8.450907	0.350328	24.123	3688	<0.001
For RD8PR slope, π_6					
For INTRCPT2, β_{60}					
INTRCPT3, γ_{600}	2.843932	0.334671	8.498	3688	<0.001

TABLE C3. FINAL ESTIMATION OF VARIANCE COMPONENTS

Random Effect	Standard Deviation	Variance Component	*d.f.*	χ^2	*p*-value
Level 1 and level 2					
INTRCPT1, r_0	3.68424	13.57361	96	527.23164	<0.001
Level 1, *e*	8.45368	71.46464			
Level 3					
INTRCPT1/INTRCPT2, u_{00}	1.90677	3.63577	37	65.3575	0.003

The final HLM specified for Geometry was as follows:

Level-1 Model

$R.RGTMAR_{iik} = \pi_{0ik} + \pi_{1ik}*(BLACKHIS_{ijk}) + \pi_{2jk}*(ELL_{ijk}) + \pi_{3jk}*(IEP_{ijk}) + \pi_{4jk}*(MA8PR_{ijk})$
$+ \pi_{5jk}*(RD8PR_{ijk}) + e_{ijk}$

Level-2 Model

$\pi_{0ik} = \beta_{00k} + r_{0jk}$
$\pi_{1ik} = \beta_{10k}$
$\pi_{2ik} = \beta_{20k}$
$\pi_{3ik} = \beta_{30k}$
$\pi_{4ik} = \beta_{40k}$
$\pi_{5jk} = \beta_{50k}$

Level-3 Model

$\beta_{00k} = \gamma_{000} + \gamma_{001}(A2I\_MEAN_k) + u_{00k}$
$\beta_{10k} = \gamma_{100}$
$\beta_{20k} = \gamma_{200}$
$\beta_{30k} = \gamma_{300}$
$\beta_{40k} = \gamma_{400}$
$\beta_{50k} = \gamma_{500}$

MA8PR RD8PR have been centered around the grand mean.

Mixed Model

$R.RGTMAR_{iik} = \gamma_{000} + \gamma_{001}*A2I\_MEAN_k + \gamma_{100}*BLACKHIS_{ijk} + \gamma_{200}*ELL_{ijk}$
$+ \gamma_{300}*IEP_{iik} + \gamma_{400}*MA8PR_{ijk} + \gamma_{500}*RD8PR_{ijk}$
$+ r_{0jk} + u_{00k} + e_{ijk}$

TABLE C4. FINAL ESTIMATION OF FIXED EFFECTS: GEOMETRY

Fixed Effect	**Coefficient**	**Standard error**	***t*-ratio**	**Approx.*d.f.***	***p*-value**
For INTRCPT1, π_0					
For INTRCPT2, β_{00}					
INTRCPT3, γ_{000}	60.304103	1.614993	37.340	33	<0.001
A2I_MEAN, γ_{001}	-2.212025	1.943626	-1.138	33	0.263
For BLACKHIS slope, π_1					
For INTRCPT2, β_{10}					
INTRCPT3, γ_{100}	-2.126071	0.950697	-2.236	1085	0.026
For ELL slope, π_2					
For INTRCPT2, β_{20}					
INTRCPT3, γ_{200}	-4.516110	1.646456	-2.743	1085	0.006
For IEP slope, π_3					
For INTRCPT2, β_{30}					
INTRCPT3, γ_{300}	-3.156456	0.828606	-3.809	1085	<0.001
For MA8PR slope, π_4					
For INTRCPT2, β_{40}					
INTRCPT3, γ_{400}	10.939150	0.629765	17.370	1085	<0.001
For RD8PR slope, π_5					
For INTRCPT2, β_{50}					
INTRCPT3, γ_{500}	1.115033	0.638375	1.747	1085	0.081

TABLE C5. FINAL ESTIMATION OF LEVEL-1 AND LEVEL -2 VARIANCE COMPONENTS: GEOMETRY

Random Effect	**SD**	**Variance Component**	***d.f.***	**χ^2**	***p*-value**
Level 1 and 2					
INTRCPT1,r_0	2.26994	5.15261	30	51.11715	0.010
level-1, *e*	9.45546	89.40575			
Level 3					
INTRCPT1/INTRCPT2,u_{00}	4.82061	23.23827	33	149.35109	<0.001

TABLE C6. UNADJUSTED OUTCOME MEANS FOR ALGEBRA BY SCHOOL

a2i		**non-a2i**	
School	*M*	School	*M*
School 1	73.39	Pair 1	75.18
School 2	57.49	Pair 2	64.87
School 4	64.31	Pair 4	75.27
School 6	62.93	Pair 6	56.07
School 7	67.63	Pair 7	61.80
School 9	57.46	Pair 9	68.55
School 10	58.87	Pair 10	63.18
School 11	62.22	Pair 11	71.60
School 12	67.22	Pair 12	63.90
School 14	66.42	Pair 14	70.30
School 15	59.00	Pair 15	62.09
School 16	61.04	Pair 16	55.21
School 17	65.99	Pair 17	67.98
School 18	61.60	Pair 18	66.79
School 19	58.88	Pair 19	58.95
School 20	58.79	Pair 20	61.99
School 22	67.86	Pair 22	62.32
School 23	63.00	Pair 23	57.72
School 24	59.59	Pair 24	70.24
School 32	69.13	Pair 32	57.02

TABLE C7. UNADJUSTTED OUTCOME MEANS FOR GEOMETRY BY SCHOOL

a2i		**non-a2i**	
School	*M*	School	*M*
School 1	67.10	Pair 1	77.71
School 2	50.69	Pair 2	60.20
School 4	59.00	Pair 4	NA
School 6	55.15	Pair 6	52.19
School 7	56.86	Pair 7	54.38
School 9	54.28	Pair 9	63.60
School 10	40.50	Pair 10	56.97
School 11	NA	Pair 11	62.35
School 12	59.74	Pair 12	53.88
School 14	65.85	Pair 14	67.39
School 15	44.55	Pair 15	60.67
School 16	NA	Pair 16	43.02
School 17	62.07	Pair 17	64.00
School 18	54.52	Pair 18	61.00
School 19	56.12	Pair 19	53.27
School 20	42.71	Pair 20	50.43
School 22	52.50	Pair 22	56.74
School 23	44.36	Pair 23	58.67
School 24	NA	Pair 24	73.80
School 32	NA	Pair 32	46.89

APPENDIX D: TEACHER SURVEY DATA

TABLE D1. TEACHER SURVEY RESPONDENTS, BY SCHOOL AND YEAR

	School	2012-13	2013-14	2014-15	2015-16	FULL SAMPLE
COHORT 1	1				1	1
	2		1		2	3
	3		1			1
	4	2	1	1	1	5
	5		1			1
	6		1	1		2
	7					
	8	1		2	1	4
	9					
	10			1	1	2
	11			1		1
	12	2	2			4
	13		1		2	3
COHORT 2	14				2	2
	15		1	2	2	5
	16					
	17		1	1		2
	18		2	1		3
	19		1	3	1	5
	20	1			3	4
	21			1	1	2
	22		1	4	1	6
	23		1	1	1	3
	24					
	25			1		1
	26					
COHORT 3	27			1		1
	28			2	1	3
	29			3		3
	30					
	31*				2	2
	Total	**6**	**15**	**26**	**22**	**69**

* School 31 joined a2i in SY 2015-16

TABLE D2. TEACHER SURVEY REPONDENTS BY GRANDES AND SUBJECTS TAUGHT

		Number	**Percentage**
Grade Level	8th	2	2.6%
	9th	43	56.6%
	10th	46	60.5%
	11th	36	47.4%
	12th	26	34.2%
Subjects	Algebra I	46	60.5%
	Geometry	31	40.8%
	Algebra II	22	28.9%
	Calculus	2	2.6%
	Other:	14	18.4%

Percentages exceed 100 because teachers may have taught more than one grade level and subject.

TABLE D3. FREQUENCY OF REGULAR AND SPECIAL EDUCATION TEACHERS TEACHING a2i, MIX OF a2i, AND NOT a2i

		Regular Education		**Special Education**	
		n	%	n	%
Algebra I	a2i	11	31.4%	5	45.5%
	Mix of a2i	22	62.9%	6	54.5%
	Not a2i	2	5.7%	0	0.0%
	Total	35	100.0%	11	100.0%
Geometry	a2i	11	39.3%	0	0.0%
	Mix of a2i	13	46.4%	3	75.0%
	Not a2i	4	14.3%	1	25.0%
	Total	28	100.0%	4	100.0%
Algebra II	a2i	3	14.3%	0	0.0%
	Mix of a2i	13	61.9%	2	100.0%
	Not a2i	5	23.8%	0	0.0%
	Total	21	100.0%	2	100.0%

TABLE D4. USE OF THE FOLLOWING RESOURCES AND SUPPLEMENTARY MATERIALS

Item	**Teacher Group**	***n***	***M***	**Never or rarely**	**Once in a while**	**Sometimes**	**Frequently or regularly**
a2i website	Early	20	3.35	1 (5.0%)	2 (10.0%)	6 (30.0%)	11 (55.0%)
	Recent	47	3.45	0 (0.0%)	8 (17.0%)	10 (21.3%)	29 (61.7%)
Delta Math	Early	20	1.90	12 (60.0%)	1 (5.0%)	4 (20.0%)	3 (15.0%)
	Recent	44	1.70	29 (65.9%)	5 (11.4%)	4 (9.1%)	6 (13.6%)
Textbook(s)	Early	20	2.10	6 (30.0%)	7 (35.0%)	6 (30.0%)	1 (5.0%)
	Recent	47	1.91	25 (53.2%)	9 (19.1%)	5 (10.6%)	8 (17.0%)
EngageNY	Early	20	1.95	9 (45.0%)	4 (20.0%)	6 (30.0%)	1 (5.0%)
	Recent	47	2.06	14 (29.8%)	20 (42.6%)	9 (19.1%)	4 (8.5%)
Online resources (other than a2i website)	Early	20	3.20	1 (5.0%)	3 (15.0%)	7 (35.0%)	9 (45.0%)
	Recent	47	3.51	0 (0.0%)	4 (8.5%)	15 (31.9%)	28 (59.6%)
Ideas, lessons, or activities from colleagues	Early	20	2.85	2 (10.0%)	3 (15.0%)	11 (55.0%)	4 (20.0%)
	Recent	47	3.19	0 (0.0%)	9 (19.1%)	20 (42.6%)	18 (38.3%)
Other online resources (describe):	Early	8	3.00	1 (12.5%)	1 (12.5%)	3 (37.5%)	3 (37.5%)
	Recent	22	3.23	3 (13.6%)	1 (4.5%)	6 (27.3%)	12 (54.5%)

APPENDIX E: STUDENT SURVEY DATA

TABLE E1. STUDENT RESPONDENTS BY SCHOOL

A total of 702 a2i Integrated Algebra and Geometry students from 19 of the 21(90%) a2i non-charter

	School	Fall Respondents	% Of Sample	Spring Respondents	% Of Sample
COHORT 1	4	193	5.39%	124	3.70%
	5			168	5.01%
	6	43	1.20%	197	5.87%
	7	34	0.95%		
	8			100	2.98%
	10			43	1.28%
	12	100	2.79%	95	2.83%
	13	253	7.07%	5	0.15%
COHORT 2	14	146	4.08%	65	1.94%
	15	322	8.99%	371	11.06%
	17	121	3.38%	124	3.70%
	18	76	2.12%	156	4.65%
	19	188	5.25%	184	5.48%
	20	243	6.79%	181	5.39%
	21	224	6.26%		
	22	856	23.91%	799	23.82%
	23	114	3.18%	127	3.79%
	24	98	2.74%		
	25	167	4.66%	187	5.57%
COHORT 3	27	77	2.15%		
	28	105	2.93%	81	2.41%
	30	220	6.15%	241	7.18%
	31*			107	3.19%
	Total	**3,580**		**3,355**	

*School 31 joined a2i in SY 2015-16

TABLE E2. STUDENT DEMOGRAPHIC DATA AT FALL AND SPRING

		Fall		Spring	
		n	%	*n*	%
Gender	Male	5496	51.9%	1659	51.0%
	Female	4809	45.4%	1595	49.0%
Current Grade Level	8	168	1.8%	1	.1%
	9	3085	33.4%	1373	41.4%
	10	2803	30.3%	1023	30.8%
	11	2048	22.1%	707	21.3%
	12	1145	12.4%	214	6.4%
First Year at School	Yes	1263	36.0%	1740	58.6%
	No	2245	64.0%	1228	41.4%

TABLE E3. CLASSROOM ENVIRONMENT: STUDENT BEHAVIORS, MATCHED SAMPLE (*N*=1,836)

	Fall			Spring			
	% Pretty Often/ Often	*M*	*SD*	% Pretty Often/ Often	*M*	*SD*	**Mean Change**
Work in small groups to think through math problems.	75.1%	3.01	.82	82.7%	3.21	.77	.20 (<.001)
Work through problems that you don't know how to solve at first.	72.4%	2.97	.81	75.5%	3.03	.81	.06 (.003)
Think about mistakes you've made in the past when working on a new problem or assignment.	72.2%	2.99	.85	74.0%	3.03	.86	.04
Explain your work, out loud or in writing.	56.0%	2.69	.85	61.0%	2.80	.89	.11 (<.001)
Ask questions or take part in class discussions.	54.5%	2.68	.80	59.4%	2.78	.86	.10 (<.001)
MEAN - Classroom Environment – Student Behaviors	--	2.87	.55	--	2.97	.60	.10 (<.001)

4-point scale: 1 = Never, 2 = Once in a while, 3 = Pretty often, 4 = Often

To be a statistically significant difference, obtained *p*-values had to exceed .008 (.05/6) using the Bonferroni correction.

TABLE E4. CLASSROOM ENVIRONMENT: TEACHER PRACTICE, MATCHED (*Nn*= 1,762)

	Fall			Spring			
	% Pretty Often/Often	*M*	*SD*	% Pretty Often/Often	*M*	*SD*	**Mean Change**
Asks us to explain how we solved a problem.	90.9%	3.45	.67	92.8%	3.51	.65	.06 (.002)
Asks us to figure out where we made a mistake in solving a math problem.	83.2%	3.19	.75	87.0%	3.28	.74	.09 (<.001)
Makes sure the work we do really makes us think.	89.1%	3.35	.70	89.7%	3.38	.71	.03
Encourages us to discuss different solutions to a problem.	84.8%	3.22	.73	85.8%	3.26	.74	.04
Asks us to work with each other to find our mistakes in a math problem.	80.1%	3.15	.81	84.0%	3.24	.78	.09 (<.001)
Gives us math problems that can be solved in a lot of different ways.	83.5%	3.18	.71	87.2%	3.26	.71	.08 (<.001)
Gives us math problems with real situations or uses.	68.5%	2.88	.81	76.6%	3.06	.79	.18 (<.001)
MEAN - Classroom Environment - Teacher Behaviors	--	3.20	.53	--	3.28	.56	.08 (<.001)

4-point scale: 1 = Never, 2 = Once in a while, 3 = Pretty often, 4 = Often

To be a statistically significant difference, obtained p-values had to exceed .006 (.05/8) using the Bonferroni correction.

TABLE E5. CLASSROOM CULTURE: STUDENTS' UNDERSTANDING AND LEARNING IN GENERAL, MATCHED

		Fall			Spring			
		% SA/Yes/ A/Mostly Yes	*M*	*SD*	% SA/Yes/ A/Mostly Yes	*M*	*SD*	**Mean Change**
Students' Understanding [A] (*n* =1,793)	My teacher wants us to understand the work, not just memorize it.	95.7%	3.44	.61	94.5%	3.44	.65	.00
	It's okay in this class to say you don't understand something.	95.1%	3.37	.60	92.2%	3.34	.69	-.03
	In this class, my teacher thinks mistakes are okay as long as we are learning.	93.5%	3.26	.62	92.1%	3.27	.65	.01
	MEAN - Classroom Culture [A]	--	3.36	.48	--	3.35	.55	-.01
Learning in General [B] (*n*=1,806)	My teacher doesn't let people give up when the work gets hard.	78.1%	4.05	.82	76.7%	4.04	.88	-.01
	I like the way we learn in this class.	62.8%	3.75	.92	61.8%	3.76	.99	.01
	MEAN – Classroom Culture [B]	--	3.90	.76	--	3.90	.82	.00

Students' Understanding: 4-point scale: 1 = Strongly disagree, 2 = Disagree, 3 = Agree, 4 = Strongly agree
Learning in General:5-point scale: 1 = No/Never/Totally Untrue, 2 = Mostly Not/Mostly Untrue, 3 =Maybe/Sometimes/Somewhat, 4 = Mostly Yes/Mostly true, 5 = Yes/Always true/Totally true.
To be a statistically significant difference, obtained p-values had to exceed .01 (.05/4) for students' understanding and .02 (.05/3) for learning in general using the Bonferroni correction.

TABLE E6. PERSERVRANCE AND MATH VALUE, MATCHED

		Fall			Spring			
		% SA/A	*M*	SD	% SA/A	*M*	SD	**Mean Change**
Perseverance (*n* = 1,810)	Even when math assignments are hard, I keep working until I finish.	87.5%	3.15	.64	82.2%	3.06	.74	-.09 (<.001)
	I feel successful in this class when I work hard on something, even if I don't get the right answer.	81.4%	3.04	.73	79.9%	3.05	.78	.01
	MEAN – Perseverance	--	3.09	.57	--	3.06	.66	-.03 (.024)
Math Value (*n* = 1,759)	I'll use what I learn in this math class later in life.	81.0%	3.02	.74	75.3%	2.92	.80	-.10 (<.001)
	I'll use what I learn in this math class in other classes.	75.5%	2.89	.74	74.2%	2.90	.78	.01
	It's important to know about math in my daily life.	85.6%	3.16	.74	81.5%	3.06	.77	-.10 (<.001)
	Learning math will help me get a good job.	87.3%	3.19	.70	83.7%	3.07	.74	-.12 (<.001)
	It's important to do well in math.	97.3%	3.49	.58	94.3%	3.38	.64	-.11 (<.001)
	MEAN – Math Value	--	3.15	.52	--	3.07	.57	-.08 (<.001)

4-point scale: 1 = Strongly disagree, 2 = Disagree, 3 = Agree, 4 = Strongly agree
To be a statistically significant difference, obtained *p*-values had to exceed .02 (.05/3) for perseverance and .008 (.05/6) for value using the Bonferroni correction.

TABLE E7. SELF-EFFICACY, MATCHED (*N*=1,852)

	Fall			Spring			Mean Change
	% SA/A	*M*	*SD*	% SA/A	*M*	*SD*	
I can learn math.	97.4%	3.49	.57	95.8%	3.45	.62	-.04 (.021)
I can understand the basic concepts taught in this class.	95.0%	3.23	.56	93.9%	3.25	.59	.02
I can get good grades in math.	91.3%	3.22	.64	90.1%	3.22	.67	.00
I can understand the harder concepts taught in this class.	65.1%	2.70	.74	69.0%	2.79	.78	.09 (<.001)
Math is an easy subject/class for me.	56.9%	2.60	.82	58.8%	2.67	.85	.07 (<.001)
MEAN – Self-efficacy	--	3.05	.51	--	3.08	.56	.03 (.021)

4-point scale: 1 = Strongly disagree, 2 = Disagree, 3 = Agree, 4 = Strongly agree

***p* < .01

To be a statistically significant difference, obtained p-values had to exceed .008 (.05/6) using the Bonferroni correction.

TABLE E8. INTEREST, MATCHED (N=1,859)

	Fall			Spring			Mean Change
	% SA/A	*M*	*SD*	% SA/A	*M*	*SD*	
Math is interesting to me.	70.2%	2.84	.82	70.1%	2.84	.87	.00
I think I can do well in more advanced math classes.	56.5%	2.59	.88	59.3%	2.66	.91	.07 (<.001)
I would like to take more math classes in high school.	50.0%	2.49	.87	54.1%	2.55	.90	.06 (.006)
I think I could have a job or career that requires math.	47.0%	2.42	.89	47.9%	2.44	.93	.02
In my spare time, I like to solve puzzles or do other math activities.	29.2%	2.09	.86	32.1%	2.15	.90	.06 (.004)
MEAN – Interest	--	2.49	.66	--	2.53	.72	.04 (.001)

4-point scale: 1 = Strongly disagree, 2 = Disagree, 3 = Agree, 4 = Strongly agree

***p* < .01

To be a statistically significant difference, obtained p-values had to exceed .008 (.05/6) using the Bonferroni correction.

TABLE E9. FEEDBACK PRACTICES, MATCHED (N= 1,714)

	Fall			Spring			
	% Always/Mostly true	*M*	SD	% Always/Mostly true	*M*	SD	**Mean Change**
My teacher doesn't let people give up when the work gets hard.	78.5%	4.06	.82	76.9%	4.05	.88	-.01
Students speak up and share their ideas about class work.	73.3%	4.00	.84	76.2%	4.07	.86	.07 (.002)
The comments I get on my work in this class help me understand how to improve.	69.1%	3.88	.90	69.3%	3.90	.94	.02
My teacher asks students to explain more about the answers they give.	83.5%	4.18	.78	83.0%	4.22	.81	.04
My teacher asks questions to make sure we are following along.	84.1%	4.20	.78	82.1%	4.21	.84	.01
My teacher wants me to explain my answers-why I think what I think.	83.1%	4.20	.79	82.0%	4.23	.82	.03
My teacher respects my ideas and suggestions.	83.6%	4.21	.79	79.8%	4.18	.88	-.03
My teacher checks to make sure we understand what s/he is teaching us.	84.3%	4.22	.80	81.5%	4.21	.89	-.01
MEAN – Feedback Practices	--	4.12	.63	--	4.13	.71	.01

5-point scale: 1 = No/Never/Totally Untrue, 2 = Mostly Not/Mostly Untrue, 3 =Maybe/Sometimes/Somewhat, 4 = Mostly Yes/Mostly true, 5 = Yes/Always true/Totally true.
To be a statistically significant difference, obtained p-values had to exceed .006 (.05/9) using the Bonferroni correction.

TABLE E10. ENJOYMENT, MATCHED (*N*=1,807)

	Fall			Spring			
	% Always/Mostly true	*M*	SD	% Always/Mostly true	*M*	SD	**Mean Change**
My teacher makes learning enjoyable	61.7%	3.76	1.03	61.8%	3.75	1.10	-.01
My teacher makes lessons interesting	60.7%	3.73	1.04	61.8%	3.77	1.10	.04
MEAN – Enjoyment	--	3.74	1.00	--	3.76	1.07	.02

5-point scale: 1 = No/Never/Totally Untrue, 2 = Mostly Not/Mostly Untrue, 3 =Maybe/Sometimes/Somewhat, 4 = Mostly Yes/Mostly true, 5 = Yes/Always true/Totally true.
To be a statistically significant difference, obtained p-values had to exceed .02 (.05/3) using the Bonferroni correction.

TABLE E11. ENJOYMENT, MATCHED (N=1,807)

	Fall			Spring			Mean Change
	% Always/Mostly true	*M*	SD	% Always/Mostly true	*M*	SD	
After this class, I'm curious about what we're going to do next time.	65.3%	2.70	.77	62.1%	2.65	.80	-.05 (.02)
This math class is different from other math classes I've taken.	66.9%	2.84	.78	73.5%	2.95	.78	.11 (<.001)

4-point scale: 1 = Strongly disagree, 2 = Disagree, 3 = Agree, 4 = Strongly agree

To be a statistically significant difference, obtained *p*-values had to exceed .025 (.05/2) using the Bonferroni correction.

TABLE E12. NUMBER OF STUDENTS AT EACH SCHOOL, ALGEBRA I

School	a2i	Mix of a2i	Both a2i and mix of a2i
4			48
23			29
28			15
12			20
15		87	
20		42	
19	108		
6	25		
22	147		
18	19		

TABLE E13. OVERALL MEANS AT FALL AND SPRING FOR ATTITUDES TOWARDS MATH, ALGEBRA I

	a2i Only			Mix of A2i			Both a2i and Mix of a2i		
	n	Fall	Spring	*n*	Fall	Spring	*n*	Fall	Spring
Math Self-Efficacy	292	2.99	3.04	127	3.08	3.14	102	3.20	3.09
Math Interest/Intrinsic Value	289	2.37	2.43	123	2.53	2.64	104	2.65	2.54
Math Value/Utility Value	285	3.07	2.99	117	3.24	3.14	89	3.28	3.16
Perseverance	292	3.01	3.00	120	3.08	3.04	93	3.31	3.10
Enjoyment	288	3.56	3.62	121	3.99	3.77	90	3.91	3.70
Student Behaviors	283	2.84	2.97	124	2.75	2.91	96	3.01	2.98
Teacher Practices/ Behaviors (in Math Class)	279	3.21	3.29	118	3.17	3.30	86	3.37	3.29
Classroom Culture [A]	283	3.37	3.34	117	3.33	3.41	92	3.40	3.32
Classroom Culture [B]	288	3.81	3.89	120	3.99	3.96	91	4.14	3.87
Feedback Practices	277	4.10	4.08	115	4.14	4.18	81	4.26	4.08

TABLE E14. SCHOOLS, GEOMETRY

School	a2i	a2i/Mix of a2i
4	12	
28	9	
22	248	
15		77
23		17
20		32
12		57
14		62
18		14

TABLE E15. OVERALL MEANS, FALL AND SPRING FOR ATTITUDES TOWARDS MATH, GEOMETRY

	a2i Only			Mix of A2i		
	n	Fall	Spring	*n*	Fall	Spring
Math Self-Efficacy	257	3.02	3.07	225	3.16	3.18
Math Interest/Intrinsic Value	256	2.50	2.53	229	2.63	2.71
Math Value/Utility Value	253	3.23	3.13	225	3.15	3.16
Perseverance	260	3.16	3.11	227	3.15	3.22
Enjoyment	260	4.03	3.98	229	3.76	3.97
Student Behaviors (in Math Class)	261	2.91	3.00	230	2.92	3.08
Teacher Practices/ Behaviors (in Math Class)	251	3.28	3.37	222	3.17	3.30
Classroom Culture [A]	256	3.39	3.42	229	3.39	3.46
Classroom Culture [B]	259	4.08	3.99	230	3.90	4.05
Feedback Practices	245	4.26	4.29	219	4.12	4.21

TABLE E16. SCHOOLS, ALGEBRA II

School Name	a2i	Mix of a2i
4	10	
28	14	
19		3
15		19
23		8
20		25
22		193

TABLE E17. OVERALL MEANS AT FALL AND SPRING FOR ATTITUDES TOWARDS MATH, ALGEBRA II

	a2i Only			**Mix of A2i**		
	n	Pre	Post	n	Pre	Post
Math Self-Efficacy	22	3.07	3.06	241	3.18	3.22
Math Interest/Intrinsic Value	23	2.66	2.76	241	2.62	2.69
Math Value/Utility Value	23	3.16	3.19	230	3.22	3.13
Perseverance	23	3.20	3.26	241	3.12	3.13
Enjoyment	24	3.50	4.06	242	3.80	3.88
Student Behaviors (in Math Class)	24	3.03	3.13	242	3.02	3.12
Teacher Practices/ Behaviors (in Math Class)	23	3.33	3.40	232	3.32	3.42
Classroom Culture [A]	24	3.42	3.39	238	3.46	3.46
Classroom Culture [B]	24	3.96	3.96	239	3.96	4.06
Feedback Practices	24	4.14	4.18	236	4.21	4.26

APPENDIX F: CASE STUDIES

TABLE F1. RESOURCE USE, CASE-STUDY SITES

Resource	Year	*n*	*M*	Never or Rarely	Once in a while	Sometimes	Frequently
a2i website	School A	4	4.00				4 (100%)
	School B	3	3.67			1 (33.3%)	2 (66.7%)
	School C	5	3.40		1(20.0%)	1 (20.0%)	3 (60.0%)
	School D	7	3.86			1 (14.3%)	6 (85.7%)
Delta Math	School A	4	1.75	3 (75.0%)			1 (25.0%)
	School B	3	2.00	2 (66.7%)			1(33.3%)
	School C	5	2.00	4 (80.0%)		1 (20.0%)	
	School D	7	2.43	3 (42.9%)		2 (28.6%)	2 (28.6%)
Textbooks	School A	4	2.50		2 (50.0%)	2 (50.0%)	
	School B	3	1.33	2(66.7%)	1(33.3%)		
	School C	5	1.40	4(80.0%)		1(20.0%)	
	School D	7	1.57	5(71.4%)	1(14.3%)		
EngageNY	School A	4	1.00	4(100%)			
	School B	3	2.00	1(33.3%)	1(33.3%)	1(33.3%)	
	School C	5	2.00	1(20.0%)	3(60.0%)	1(20.0%)	
	School D	7	1.43	4(57.1%)	3(42.9%)		
Online resources from sites other than the a2i website	School A	4	2.75		2(50.0%)	1(25.0%)	1(25.0%)
	School B	3	2.67	1(33.3%)	1(33.3%)	1(33.3%)	
	School C	5	3.80			1(20.0%)	4(80.0%)
	School D	7	3.14		2(28.6%)	2(28.6%)	3(42.9%)
Ideas, lessons, or activities from colleagues	School A	4	2.75	1(25.0%)		2(50.0%)	1(25.0%)
	School B	3	2.33	1(33.3%)	1(33.3%)	1(33.3%)	
	School C	5	2.80		2(40.0%)	2(40.0%)	1(20.0%)
	School D	7	3.57		1(14.3%)	1(14.3%)	5(71.4%)