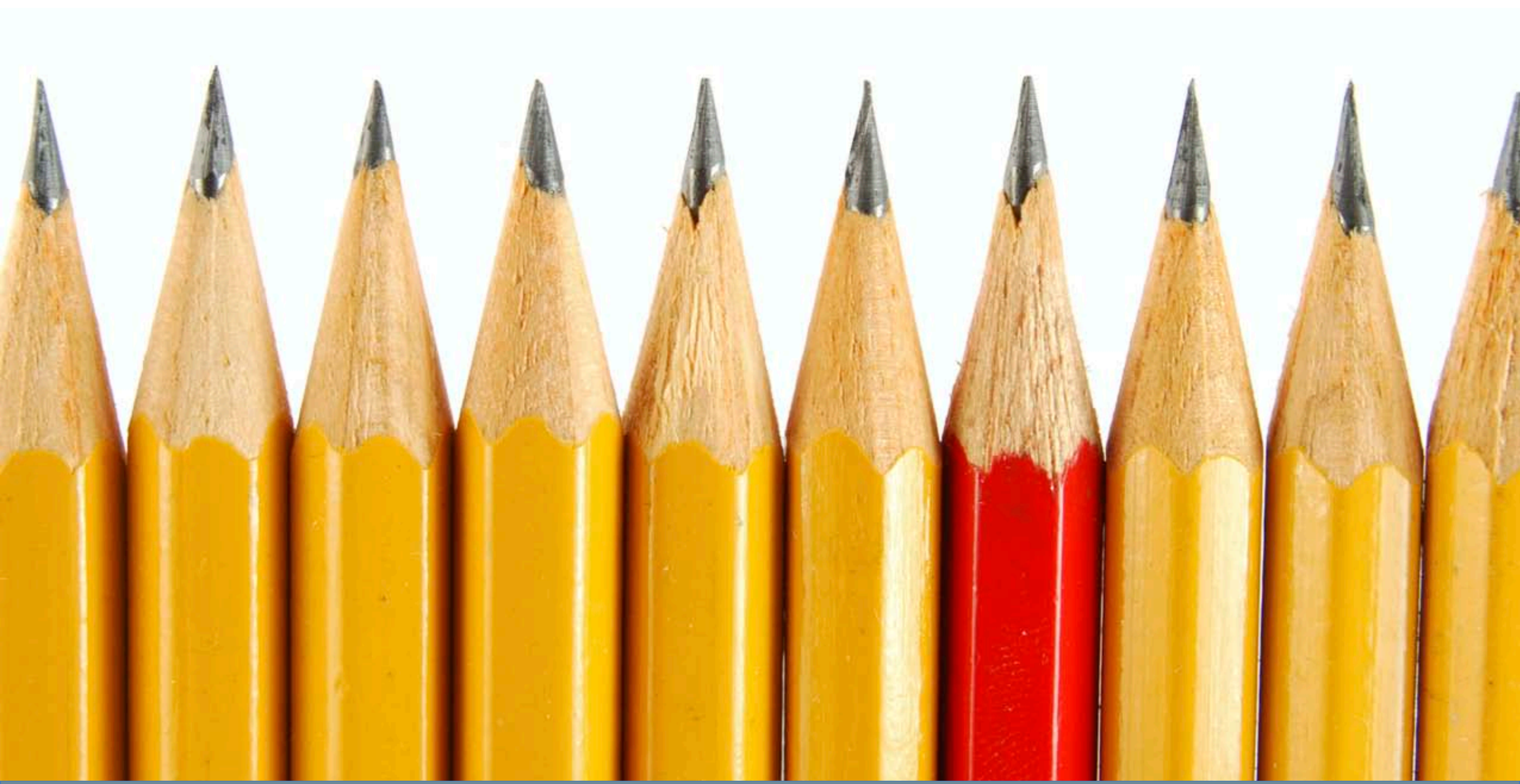
The Research Alliance for New York City Schools

Assessing the Early Impact of School of One: Evidence from Three School-Wide Pilots



Rachel Cole

James J. Kemple

Micha D. Segeritz

June 2012

NYU Steinhardt
Steinhardt School of Culture, Education, and Human Development

NEW YORK UNIVERSITY

The Research Alliance for New York City Schools

Assessing the Early Impact of School of One: Evidence from Three School-Wide Pilots

Rachel Cole
New York University

James J. Kemple
The Research alliance for New York City Schools

Micha D. Segeritz
The Research Alliance for New York City Schools

June 2012

ACKNOWLEDGEMENTS

The authors would like to acknowledge the many individuals who contributed to this report. Special thanks go to the Fund for Public Schools for supporting the study and to their staff, especially Caroline Gonzalez and Elizabeth (Sunny) Lawrence, for giving thoughtful feedback on earlier drafts of the report. We also want to thank the staff of the New York City Department of Education and the School of One, especially Joel Rose, for taking the initiative to seek a rigorous study of the program. We are also grateful to Anne-Marie Hoxie, Stacey Gillette, Wendy Lee, Chris Rush, Robert Swanwick, Tres Watson, David Weiner, Jonathan Werle, who provided useful comments and suggestions on the report and helped us better understand the theory of action behind School of One, and how it works in practice. The authors wish to thank our Research Alliance colleagues, who played vital roles preparing the data for analysis and providing constructive comments on multiple drafts of the report, especially Janet Brand and Thomas Gold. Finally, we would like to offer a special thanks to Chelsea Farley for her careful editing, helpful advice on presentation style, and thoughtful guidance with the Research Alliance's evolving publication guidelines.

CONTENTS

EXECUTIVE SUMMARY

Across the country, students are performing poorly in mathematics, particularly in middle schools. According to the U.S. Department of Education, achievement in the U.S. lags below that of other developed nations, with about two thirds of 8^{th} graders scoring below proficiency on standardized math tests. Efforts to boost achievement are complicated by the diversity of the student population and the wide range of prior math knowledge and skills that they bring to their classrooms. Teachers, principals, and curriculum developers often face extraordinary challenges in meeting this array of needs efficiently and effectively. Technological innovations combined with better tools for the systematic diagnosis of learning challenges are coalescing around the country to spur innovative approaches that individualize teaching and learning strategies.

School of One (SO1) is an individualized, technology-enhanced math instructional program that responds to the challenges of diverse learners. The program was piloted in three New York City middle schools in the 2010-11 school year. In this report, we evaluate the impact of SO1 in its first year of school-wide implementation by addressing the following questions:

- What is the impact of the initial whole-school version of SO1 on students' math achievement, as measured by the New York State math test?
- Do the effects of SO1 on math achievement differ across subgroups of students?
- Is exposure to more SO1 material, and/or mastery of SO1 skills, associated with improved math performance?

It is important to note that, given the early stage of SO1's development and implementation and the limited number of schools that have piloted the program, this evaluation cannot reach definitive conclusions about SO1's effectiveness. The findings presented in this report provide a preliminary assessment of SO1's initial impact on students' math achievement and offer insights into achievement trends that may assist with the program's ongoing development. Future studies of SO1 should combine the rigorous assessment of impacts with analyses of its implementation and of teachers' and students' experiences with the program.

About SO1

SO1's theory of action is based on the premise that students cannot learn grade-level content when they are missing precursor skills from earlier grades. Similarly, more advanced students should be able to move on to higher-level skills when they are ready. SO1 seeks to meet each student wherever he or she is on the continuum of math knowledge and skills, while acknowledging that it may take several years to see the results of this strategy. The SO1 instructional approach begins with an in-depth diagnostic assessment of each student's math skills. Results from this assessment are used to create an individualized learning plan that specifies the skills on which the student should work. Students are then grouped to receive instruction in large or small clusters, or to do independent work. At the end of each class period, students take a short assessment of the skill that was the focus of their lesson. The results of this assessment are used to develop a new learning plan for the next lesson. Both teachers and SO1 staff monitor students' progress and adapt the learning plans to meet their evolving needs on a daily basis.

SO1 represents some important adjustments from "business as usual" for both students and teachers. From a student's perspective, SO1 begins when she walks into the classroom and looks to a large screen to find out where she will be working that day. She then checks into the SO1 web portal to learn what skill she will be working on during the session. For teachers, SO1's set progression of skills allows them to predict generally what lessons they will teach, days in advance. Every school-day afternoon, however, they learn which particular students will receive the lesson and how those students have recently performed on that and other related skills.

The 2010-2011 school year marked the first attempt to implement SO1 as the school-wide math curriculum (following pilot tests of the program held during the summer and as an afterschool option). Three schools were chosen to pilot the school-wide program, after applying and demonstrating that they could support the technical infrastructure required. These schools are diverse, situated in three different boroughs of New York City, and serve populations of varying ethnic composition and socioeconomic status. All three schools implemented SO1 for their students in grade six. Two of the schools also implemented SO1 with students in grades seven and eight.

SO1 staff reported that both teachers and students needed some time to adjust to the program structure and its new teaching and learning modalities. SO1 staff reported that they made a number of midcourse modifications to the program during this initial year as part of their effort to continuously improve its functionality and learn from implementation challenges. Many of these adjustments aimed to help teachers adapt to new roles and ensure that the program was aligned with expectations for student performance on state assessments.

Impact Findings

The evaluation uses a rigorous methodology, known as comparative interrupted time series (CITS) analysis, to isolate the unique effect of SO1. The method accounts for a wide range of potential external influences on student achievement, including ongoing conditions and initiatives in the participating schools and the potential effects of other system-level initiatives. The CITS design compares the achievement of SO1 students with that of previous cohorts of students in the same schools prior to the arrival of the program. It also draws comparisons with similar students in comparable New York City schools that did not implement SO1. Finally, the method controls for the influence of students' prior math achievement and demographic characteristics. Math achievement for all students was measured with the same New York State tests from 2007-2008 through 2010-2011. Key findings from the analyses include the following.

- SO1 produced a mix of positive, negative and neutral results across schools and grade levels.

Because all three SO1 schools served students in grade six, these results are the most robust. Table ES-1 shows that, on average across the three pilot schools, SO1 did not affect 6th grade students' math achievement, either positively or negatively. Overall, 6th graders in SO1 schools and comparison schools had virtually identical achievement and trends. The table also shows, however, that this overall neutral result is an artifact of positive and statistically significant impacts for School A, neutral results at School B, and negative and statistically significant impacts for School C. The difference in impacts across the three schools is statistically significant.

Table ES-1
First-Year Impacts of SO1, by School and Grade Level
(New York State Math Test, Scaled Scores)

Sample	SO1 Schools	Comparison	Estimated Difference	
6th grade				†
School A	682.0	672.6	9.5	***
School B	679.1	680.2	-1.1	
School C	652.8	660.8	-7.9	***
6th grade average	671.4	671.3	0.1	
7th grade				
School A	686.0	682.2	3.8	*
School B	679.9	684.1	-4.2	**
7th grade average	683.0	683.2	-0.2	
8th grade				
School A	686.2	692.8	-6.6	***
School B	685.0	686.6	-1.6	
8th grade average	685.6	689.7	-4.1	***

Source: Research Alliance analysis of New York State math test scores.
Notes: Statistical significance of estimated differences is indicated by: * = $p < .10$; ** = $p < .05$; *** = $p<.01$.
Statistical significance of variation in estimated differences across schools is indicated by † = $p< .01$.

Table ES-1 also shows that results for grades six, seven and eight are not consistent across schools. This raises questions about whether the variation in impacts are due to implementation challenges, the program's fit by grade level, or a variety of other school characteristics and contextual factors. For this reason, it is impossible to draw definitive conclusions about the overall effectiveness of the program or the conditions under which it might be more effective.

- Differences in SO1 impacts across subgroups of students do not follow a discernible pattern that would suggest SO1 is reliably more effective for some students and not for others.

Table ES-2 presents the impact of SO1 for subgroups of 6th grade students defined by gender, race and prior achievement levels, as well as the results for the special education students and English Language Learners (ELL) students who participated in SO1. (ELL students receiving bilingual instruction were not enrolled in SO1 nor were special education students who required instruction in self-contained classrooms).The table shows a mix of positive and negative differences, none of which are statistically significant. The lack of a discernible pattern of impacts across subgroups was similar for students in grades seven and eight. (See Appendix B in the full report for more information.)

Table ES-2
First-Year Impacts of SO1, by Student Subgroup
6th-Grade Students
(New York State Math Test, Scaled Scores)

Sample	SO1 Schools	Comparison	Estimated Difference
All 6th graders	671.4	671.3	0.1
Level on New York State math test in 5th grade			
Level 1	645.3	638.7	6.7
Level 2	652.2	655.3	-3.1
Level 3	678.5	676.9	1.6
Level 4	696.7	695.5	1.2
Race/ethnicity			
Asian	680.7	679.6	1.1
Black	656.0	656.1	-0.1
Hispanic	663.6	660.6	3.0
White	684.7	679.5	5.1
Gender			
Female	670.9	671.8	-0.9
Male	671.5	671.1	0.4
English as a Second Language	661.1	657.2	3.8
Mainstream special education	656.3	661.1	-4.8

Source: Research Alliance analysis of New York State math test scores.
Note: None of the differences are statistically significant.

The results in Table ES-2 suggest that it will be worthwhile to learn more about program mechanisms and strategies that could be helping students who entered SO1 with the lowest prior achievement levels (Level 1). Although not statistically significant, the positive difference for Level 1 students is noteworthy, because the SO1 theory of action suggests that these lowest performing students should have been the least likely to experience short-term gains on the New York State math test. SO1 program staff hypothesized that the test would be less sensitive to improvements among very low-performing students, since it focuses most heavily on grade-level-appropriate material. It will be important to follow these initially low-performing students over time to determine if the suggestive patterns of positive, yet statistically insignificant, gains translate into stronger long-term impacts.

Exploratory Analysis

In addition to assessing the first-year impacts of SO1, the study included an exploratory analysis of the relationship between students' exposure to and mastery of SO1 skills and their rate of improvement on the New York State math test. The analysis drew on internal data available through SO1, and provides insights into how the program's "meet students where they are" approach may be working.

- Student improvement on the New York State math test was associated with exposure to on-grade-level skills through SO1, even though the students may not have mastered these skills. This relationship was strongest for students who entered SO1 at the lowest levels of prior achievement.

Students who came to SO1 with low prior performance were exposed to approximately twice as many below-grade-level skills, compared to those who came with higher performance levels from prior grades. This is consistent with SO1's focus on filling gaps in students' understanding. However, these students mastered less than 15 percent of the skills to which they were exposed (as measured by SO1's daily assessments), compared to approximately 85 percent mastery for students who entered with higher prior performance. This finding may be counter to SO1's theory of action, which suggests that all students should achieve a high level of mastery if they are exposed to instructional material whose difficulty aligns with their current knowledge and skills.

When we looked within groups of similarly performing students, we found that those who were exposed to more on-grade-level skills experienced higher rates of growth on the New York State math test. This is consistent with the test's focus on grade-level-appropriate material. While there was a relationship between SO1 skill exposures and year-to-year growth on state test scores for all groups of students, it was particularly strong for the students who entered the program with lowest levels of prior performance. This suggests that a marginal increase in exposure to on-grade-level skills for students who start off at low performance levels may have a positive effect on their state test scores, even if they do not master a high proportion of these skills in SO1's daily assessments. This insight may need to be balanced against SO1's goal of "meeting students where they are" and ensuring that students master lower-level skills before moving on to more advanced material.

Implications and Next Steps

Evaluating this program involved a number of challenges that lead us to recommend caution in interpreting the findings—and to suggest several important refinements to SO1's learning agenda. Like any first-year pilot of an innovative and complex intervention, SO1 changed and evolved continuously during its initial implementation year. In fact, SO1 program staff hypothesized that schools might experience a variety of implementation and outcome "dips," in which instructional quality and student achievement might initially decline, as teachers adjusted to the new organization and delivery of the math curriculum. SO1 staff also hypothesized that students' math test scores might actually lag behind the scores for students in traditional classrooms, because of the program's focus on addressing gaps before moving on to grade-level content. While some assessment of SO1's implementation in these schools was conducted by another research group, the results of that work have not been integrated with the impact study. A systematic and coordinated process study would provide useful insights into of SO1's implementation and may help shed light on the mixed results we found across schools and grades.

A second caution arises from the fact that the outcome measure used in this study—the New York State math test—focuses mostly on grade-level material. Thus, it is possible that some students made progress on lower-level math skills that were not detected by the state test.

Finally, in general, educational innovation is exceedingly challenging: Program impact is often incremental, rather that abrupt and dramatic; the process of development and evidence building is iterative and dynamic, rather than linear and uni-directional; and it often takes years, rather than months, to establish program efficacy and a credible track record for expansion and scale. With this in mind, we offer the following suggestions for the ongoing development and study of SO1.

- Continue to measure the impact of SO1 on test scores as the program expands to other schools, and build in the capacity to follow students as they transition from their SO1 middle schools into and through high school. Establishing the program's impact for a wider range of schools and its effect on longer-term outcomes will be important to determine its efficacy.
- Track the progress of the lower-achieving students in light of the trends we found for this group—positive but not statistically significant impacts, combined with steeper improvements among those exposed to a higher proportion of on-grade-level skills. Despite the program's hypothesized pattern of effects, these students do not appear to have lagged behind their peers in traditional classrooms, as SO1 hypothesized. In future studies of SO1, it may be useful to assess students' learning progression in a more fine-grained and more frequent manner than is possible with the state assessments.
- Ensure that future research examines the implementation of the program as well as its impact. The current study points to a web of different effects across the three pilot schools and across grade levels. It would be useful to know whether some of the schools have been more effective in their implementation than others and whether these differences are associated with an evolving pattern of impacts.

- Provide SO1 with formative feedback on implementation challenges through systematic, observations and interviews with teachers and SO1 program staff. Such field research should focus on the challenges teachers face as they adapt to the program and how they are supported with professional development opportunities and collaboration. It will be useful to document how teachers are trained to use this innovative model, and to identify supports that help teachers address issues that emerge throughout the school year. Toward this end, future researchers may want to observe SO1's professional development activities and conduct focus groups with teachers to gain their perspective on the challenges of implementing the program.

Just as SO1 challenges its teachers and students to continually assess their progress and make adjustments in response to those assessments, the program's developers are committed to a learning process that allows them to refine and improve the model. SO1 continues to evolve, and its developers are seeking opportunities to expand its use in selected New York City middle schools. The program was recently awarded a coveted "development grant" from the U.S. Department of Education's Investing in Innovation (I3) Fund, which will support improvements to the program and further research on its impact and implementation. The grant provides a unique opportunity to execute some of the recommendations presented above.

I: INTRODUCTION

School of One (SO1) is a technology-rich, individualized math instructional program designed for middle school students. The program served as the primary math instructional program in three New York City middle schools during the 2010-11 school year. Despite the program's early stage of development, SO1 leadership requested an evaluation of its impact on student test scores during this pilot year. This report presents the results. It begins with a brief discussion of the context for SO1 and describes the theory of action behind the program. It also provides basic information about how SO1 was rolled out in the 2010-11 school year. The report then examines the impact of SO1 on New York State math test scores, first for all participating students and then for various subgroups defined by background characteristics and prior achievement levels. The findings include exploratory analyses of the relationship between SO1 skill mastery and test score growth. The report concludes with recommendations for SO1's ongoing development and for efforts to build evidence about the program's implementation and impact.

Context for SO1

U.S. students are not performing strongly in math. In recent years, only 39 percent of 4th graders, 34 percent of 8th graders, and 26 percent of 12th graders performed at or above proficiency on the National Assessment of Educational Progress (National Center for Education Statistics (NCES), 2009; NCES, 2010). Gains on successive NAEP tests have been meager, and racial achievement gaps persist. Math achievement in the U.S. lags below that of other developed nations. The 2009 average score of U.S. students on the Program for International Student Assessment (PISA) was lower than the average score for Organisation for Economic Co-operation and Development (OECD) countries, despite small gains since the 2006 assessment (OECD, 2010).

In the development of math skills over the course of primary and secondary education, the middle years are a key time when some students' performance begins to decline (Lee and Fish, 2010). As a result, middle school classes may include students with a variety of levels of skills and knowledge. To address this challenge, educators and researchers are seeking out instructional methods that allow teachers to meet students' individual needs (Davis, 2011). Students need instruction in different skills, and have diverse interests and learning styles. If a teacher responds to students' diverse skill levels by aiming her instruction at the middle, the material is likely to be too difficult for weaker students and a boring review for stronger students (Ackerman, 1987), leading to inefficient instruction. Technology-based programs offer promise for developing highly individualized lessons. Findings about the impact of these programs on student outcomes are mixed, but have tended to be positive, particularly in mathematics (Barrow, Markman, and Rouse, 2009; Bannerjee et al, 2007). Such instructional techniques have been found to be especially effective in working with at-risk students (Hamby, 1989).

SO1 is a new, individualized, technology-rich math program that offers a high level of customization for each student based on her current skill level. SO1's approach has generated a great deal of interest across the country. The program is the recipient of a three-year, five-million-dollar Investing in Innovation (I3) development grant from the federal Department of Education, and it was named one of the top 50 inventions of 2009 by Time magazine (Kluger, 2009).

SO1 was implemented as the school-wide math instructional program in three New York City middle schools in 2010-2011. This report refers to these as Schools A, B, and C. Before school-wide implementation of SO1, the program went through three previous pilot versions. In summer 2009, SO1 provided a four-week summer school program for 80 rising 7th graders in School A. In spring 2010, it was adapted into a seven-week, after-school program for 240 6th graders in all three middle schools. Later that spring, SO1 became the school-day math instructional program for 200 6th graders for six weeks at School B.

This report examines the effectiveness of SO1 in improving math test scores in its first year of school-wide implementation in the three initial pilot schools. This is the first independent evaluation of this new, expanding program. Rather than provide definitive evidence of SO1 efficacy, the study offers a preliminary assessment of SO1's impact in the hope of contributing to the program's development and of informing ongoing and future research on the initiative.

II: Program Description and Theory of Action[1]

SO1 offers "mass customization" of student learning in response to the diverse levels of math proficiency that students bring into the classroom. The program is premised on the notion that students will have difficulty learning grade-level content when they are missing precursor skills from earlier grades. Therefore, SO1 meets each student where he or she is. According to program staff, the underlying theory of action for SO1 is that individualization of both the pace of skill exposure and learning modalities will allow student learning gaps to be diagnosed and addressed more quickly and efficiently that traditional whole-group instruction. In addition, a focus on skill mastery, rather than curricular scope and sequence, should ensure that student build precursor skills before moving on to grade-level material. As a result, students move on to grade-level material when they are ready, rather than when the curriculum or textbook says they should. Depending on how much below-grade-level material a student needs, it may take several years for their progress to be evident on tests that measure achievement based on grade-level content.

SO1's developers began developing the program by identifying—at a granular level—the skills and competencies that make up the New York State Math Standards for middle school students. They considered material from 4th- through 9th-grade-level standards. They worked to

[1] This section draws largely on information posted on SO1's website, http://schoolofone.org/, and from our research team's participation in a guided tour.

sequence these skills and competencies logically, so that the program could create a clear path for each student between their current abilities and their goals.

With the skill map in place, the SO1 team developed a scheduling algorithm to determine the lesson-by-lesson progression that teachers and students should follow as they fill gaps and move through instructional material. They developed diagnostic assessments to determine each student's placement and progress through the skill sequence. The algorithm specifies an optimal configuration of students, teachers, teaching technology, and space so that each student receives instruction in the skill he needs and in the teaching and learning modality best suited to his development. SO1's teaching and learning modality options are divided into four "learning zones:"

- Live Learning Zone (LLZ) – students receive instruction from a teacher or a student-teacher. LLZ includes both small and large groups, of approximately 10 and 25, respectively.
- Collaborative Learning Zone (CLZ) – small groups of students work together on shared tasks.
- Virtual Live Instruction (VLI) and Virtual Live Reinforcement (VLR) – students work with online tutors to learn new skills or review ones they have already mastered.
- Individual Learning Zone (ILZ) – students work individually on assignments, both online and on paper.

After completing a lesson, each student then takes a short online assessment on the material on which he worked that day and receives immediate feedback on his performance. The algorithm uses this information to determine if he is ready to move on to the next skill for the following school day, or if he needs further work in the current skill. SO1 staff and school administrators can override the algorithm if they prefer a different configuration of students and teachers. Teachers can also override the algorithm's placement if they believe a child has been placed incorrectly.

From a student's perspective, SO1 begins when she walks into the classroom and looks to a large screen to find out where she will be working that day and in which instructional modality. She then logs on to the SO1 "portal," a website that can be accessed from any internet-ready device, to both confirm her schedule and see what skill she will be working on during the session. Through the portal, she can view a variety of materials: math textbook pages that explain the skill she is working on and upcoming skills, additional problems with which she can challenge herself, and sometimes other kinds of instructional tools, such as videos of tutors explaining content or games designed to build a particular skill. Electronic copies of her daily homework assignment are also available on the portal, although all students receive hard copies of their homework assignment from their homeroom instructor. If a student has an internet-ready computer at home, she can access all this content there as well.

From a teacher's perspective, SO1's dynamic nature changes the process of lesson planning significantly. By logging in, the teacher can see the algorithm's predictions about what skills he will teach over the next few days, linked to lesson plans from the teacher's guides of multiple math text books. This advanced notice allows the teacher to become familiar with the provided lesson plans and modify them as he sees fit. By about 5 p.m. on each school day, the teacher is emailed his precise teaching assignments for the next day, including the skill to focus on, the students he will teach, and information on whether each student was exposed to this skill before or not. Given this information he can customize his lesson plan for these students. Frequently teachers have the opportunity to teach the same lesson plan multiple times in a short period, as a given skill is assigned for different groups of students. Teachers can use data on how students performed to help reflect on their practice. This repetition allows teachers to hone their lesson plans and implementation over time.

Teachers' calculations of grades also take a different form in SO1. Each day, teachers use the portal to grade the students in their assigned groups for class work, homework and participation. These data are combined with the daily short assessment students take at the end of each class. With these records, a homeroom teacher can easily use all this accumulated data for grading, even though he may only teach his homeroom students a few times a week. Different schools and teachers have different policies about how to weight class work, homework, participation, and performance, but once these decisions are made, the portal makes it easy to translate the weighted average of these indicators into grades.

As with many innovative and complex initiatives, SO1's developers anticipated that the program's initial implementation might result in "dips" in student performance and program effectiveness, at least in the short term.[2] First, the program developers suggested that SO1 might produce a "gap dip" in which student progress on grade-level material would decline relative to a traditional classroom, because SO1 begins by filling the gaps in lower level precursor skills. Depending on how far below grade level the student is performing, this "gap dip" may last a few months or a few years until student catch up with the grade-level material that their peers in traditional classrooms have been exposed to all along. This gap dip may be particularly acute for students who are new to SO1.

Secondly, SO1 developers anticipated a "teacher change management dip." This could occur as teachers adapt to the individualized pacing of material and to the program's multiple teaching and learning modalities. Teachers also need to adapt to the daily interaction with the SO1 algorithm which plays a central role in the assignment of students and teachers to lessons and modalities. This dip too could be acute for teachers who are new to the program.

Finally, SO1 developers anticipated a "systems stability dip." This was particularly a concern during the first year of school-wide implementation as the program underwent a series of technical

[2] Based on conversations with SO1 staff.

midcourse corrections and modifications and as it attempted to adapt to feedback from teachers, administrators and students.

The 2010-2011 SO1 Pilot[3]

In September 2009, SO1 opened an application process for all New York City middle schools interested in piloting the program as its primary math curriculum. SO1 required that interested schools meet the following requirements:

- Strong leadership/principal buy-in.
- Sufficient laptops for students use in school (at least one laptop for every student in the SO1 space at the same time).
- Available space to redesign classrooms.

Table 1, on the next page, presents the background characteristics of students in the three middle schools selected for the SO1 pilot phase. The table shows that Schools A, B, and C serve somewhat different populations of students, with School A serving a large proportion of Asian students and many ELLs, School B serving a mixed population, and School C serving a large proportion of Black and Hispanic students. The 2010 performance of the three schools was highest in School A, then B, then C; by contrast, the average yearly change in math scores was highest in School C, then B, then A.

After consultation with the schools selected for the program, SO1 replaced the traditional math instruction and curriculum for 6th grade students in all three schools and for 7th and 8th grade students in Schools A and B. Several groups of students were not included in SO1 because of their special learning needs. This included students receiving bilingual instruction, because SO1 was available in English only. It also included special education students whose individualized education plans (IEPs) required that they be in classrooms with small student-to-teacher ratios.

SO1 and school staff also reported several modifications to the assignment of students to the program during the initial pilot year. For example, at one school, several students were identified as being so far behind grade level that even 4th-grade skills were too advanced for them. These students were pulled out of SO1, in some cases temporarily and in other cases for the duration of the pilot year, to receive intensive academic support. At another school, a group of 8th graders participated in SO1 for only three periods per week (rather than the scheduled eight periods), as a supplement to the traditional math instruction they received for the other five periods a week. At this same school, there were a number of bilingual and special education students who were included in SO1 because the principal decided it would be the best program for them. Finally, some students in each of the three schools did not receive full exposure to SO1 because they were chronically absent or because they transferred to the other schools part way through the school year. It was not possible for the research team to identify these students accurately or to determine the precise timing of their movement in and out of SO1. However, we conducted a

3 This section draws largely from conversations with SO1 staff.

variety of analyses to determine the sensitivity of the impact findings to the inclusion or exclusion of students with characteristics associated with the placement and replacement decisions.

Table 1
Characteristics of Students Served in SO1 Schools
(2010-2011 School Year)

Characteristic	School A	School B	School C
English Language Learner (ELL) status (%)			
Bilingual	26	6	10
English as a Second Language	16	9	18
Non-ELL	58	85	72
Special education status (%)			
Small class	3	5	7
Mainstream	14	8	11
General education	83	87	82
Race/ethnicity (%)			
Black	6	14	34
Hispanic	12	24	64
Asian	82	34	1
White	1	28	0
Gender (%)			
Female	41	43	44
Male	59	57	56
Grade (%)			
6th Grade	20	29	32
7th Grade	34	30	33
8th Grade	46	40	35
Over age for grade[a] (%)	35	19	42
Free/reduced lunch (%)	88	79	91
2010 attendance rate (%)	97	95	91
Average scaled score on 2010 NYS math test	681	678	650
Average yearly trend in math scores[b], 2006-2010	4	5	11
Total enrollment	708	815	760

Source: Research Alliance analysis of student characteristics.
[a] Over age for grade designates students who were aged 13 or older on December 31 of their 6th grade year, 14 or older on December 31 of their 7th grade year, or 15 or older on December 31 of their 7th grade year.
[b] The average yearly trend in math scores gives the typical yearly change in the school's average test scores in the five years before SO1 was implemented.

SO1 Program Features and Their Implications

There are a number of features of the SO1 pilot program in 2010-2011 that lead the research team to recommend caution in interpreting the results at this stage of an evaluation. First, SO1 is a technically complex intervention that evolved over the course of the year. It requires the major actors in schools—teachers and students—to approach the work of teaching and learning in a different way than they have in the past. The process of acclimating to the structure of SO1 for teachers and students took considerable flexibility and adaptation. In addition, the program was undergoing changes during the year. As a result, the findings may not indicative of a more mature version of the program nor of its longer-term impact on student outcomes.

Second, SO1 is designed to meet students' individual needs, even if that means covering large amounts of material below their current grade level. Program operators shared with our research team that this design element was sometimes a source of concern for teachers. Some teachers worried that, for example, a 6^{th}-grade student would face a state test with 6^{th}-grade material, even if she was working on 4^{th}-grade skills in SO1. Focusing on below-grade-level material to the detriment of on-grade-level material could leave the student underprepared for the state test, which carries high stakes. For students, these tests can affect promotion decisions; for teachers, their students' scores can influence whether or not they receive tenure; and for schools, scores play a large role in accountability processes and can even prompt closure or reorganization. SO1 staff report that, as a result of these concerns, the proportion of on-grade material grew somewhat over the course of the year, for different groups of students, and across the three schools.[4]

A further caution stems from the primary outcome variable used for this study—student scores on the New York State math test. The design of the New York State math test may have important implications for interpreting the results we find in this analysis. The New York State math test is designed to differentiate students into four categories: significantly below standards (Level 1), approaching standards (Level 2), meeting standards (Level 3), and exceeding standards (Level 4). However, state accountability focuses largely on the distinction between Levels 2 and 3, so the test itself is designed to be most accurate in measuring this distinction (New York State Education Department, 1999). For a given test, the level of difficulty of the majority of the questions aims at differentiating between students who meet the state standards and those who are approaching the state standards. There are many fewer questions whose level of difficulty aims to differentiate students in Level 1 and 2, or students in Level 3 and 4, because the designers of the test determined that the accuracy of these distinctions is less important than accurately determining which students are meeting standards and which are not. Further, little can be determined from the New York State test about distinctions of performance within a performance level (NCEE, 2011). While it is theoretically possible for a student to learn an extraordinary amount within a school year and still have the same performance level, the New York State math test cannot report on this growth with accuracy.

These characteristics of the New York State math test matter for SO1 because the program attempts to individualize instruction for each student. Students that began the 2010-11 school year performing far below level and missing a large number of precursor skills may have mastered these skills over the course of the school year but not proceeded significantly into their grade level's material. Because the New York State math test primarily measures grade-level material

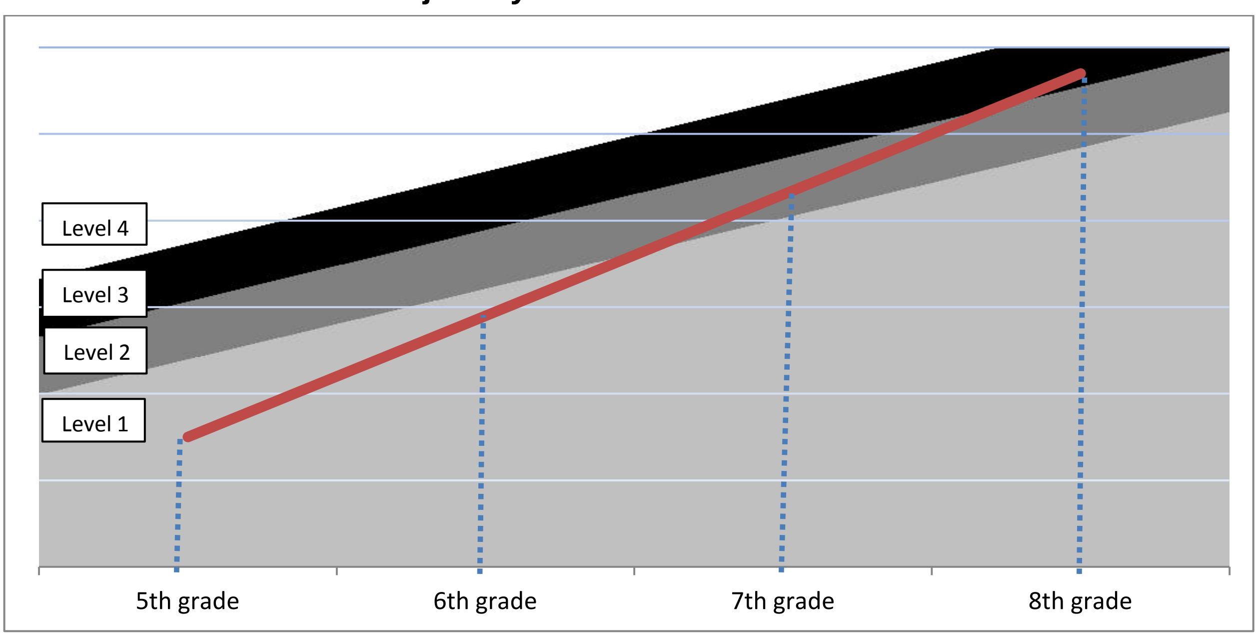
[4] Email with SO1 staff, August 10, 2011.

and not the material from the grade below, a student could make significant growth without it being apparent on the New York State math test. Such students would likely score within Levels 1 and 2, where performance is measured with less accuracy. Similarly, students who began the 2010-11 school year performing above grade level may have mastered large amounts of material beyond their grade level, which would not be reflected in their New York State math test performance, or would be reflected with limited accuracy.

With low-performing students, improved performance would be expected to become clear over time: If SO1 helps low-performing students, each additional year of SO1 will bring them closer to performing on grade level, at which point grade-level tests will more accurately reflect their progress. Figure 1 shows how a student could theoretically make substantial progress without it being evident on the New York State math test, at least in the early years.

In this image, the shaded areas represent the distinctions the New York State math test can make between students performing at Levels 1, 2, 3 or 4. The red line represents the theoretical progress of a SO1 student who began the school year with a low skill level. Supposing the student was first exposed to SO1 in 6th grade, her 6th grade test does not reflect the substantial progress she made because she still scores a Level 1. Only by following her through 7th and 8th grade do we see that she catches up with the skill level expected for her grade, achieving a Level 3 in 8th grade. While our analysis cannot confirm whether the above image represents a typical student trajectory under SO1, we present the image to illustrate the limitations of the New York State math test for measuring progress, particularly in the program's first year of implementation.

Figure 1
Theoretical Score Trajectory for a Below-Grade-Level Student in SO1



Source: Research Alliance theoretical illustration of the implications of the New York State math test design for interpreting SO1 impact estimates.

III: Research Design

This study was designed to answer three core questions about SO1's pilot year:

1. What is the impact of the initial whole-school version of SO1 on students' math achievement, as measured by performance on the New York State math test?
2. To what extent does the impact of SO1 differ by prior mathematics achievement and across other subgroups?
3. What is the relationship between math achievement and SO1 skill exposure and mastery?

To address the first two questions, the study used what is known as comparative interrupted time series design, a method used widely in education research and evaluation to assess the impact of school-wide programs and systemic policies on student outcomes.[5][6] The SO1 study was able to use a particularly rigorous version of the comparative interrupted time series design.[7] The design first compares achievement levels for SO1 students during the implementation year with the achievement trends of students from the same schools during the previous five years. This first comparison documents deviations during the SO1 implementation year from historical trends in achievement for those schools. The design incorporates the same comparison for New York City middle schools with similar characteristics over the same time period. This second comparison documents deviations from historical trends during the 2010-2011 school year that may be due to broader influences from city, state, or federal initiatives. The difference between the two comparisons provides a credible estimate of the impact of SO1 during the 2010-2011 school year over and above ongoing trends in the SO1 schools and over and above other external influences on student achievement. The design further controls for differences among students in their prior achievement and demographic characteristics.

Operationally, the design relies first on school average test scores from 2006 to 2010 to construct a five-year baseline trend and predict the expected school average test score for 2011, the first year of SO1's implementation. Here, the analysis relies on past scores from each school as the best indication of what future scores would be. Next, we identified a group of six comparison schools for each SO1 school and constructed a baseline trend for the comparison group. While the three SO1 schools are very different from one another, a contrast of the characteristics of each school with those of its comparison schools shows that they are well-matched (see descriptive statistics tables in Appendix A). By observing how the comparison

---

[5] This statistical methodology has been used widely in education research and evaluation (see Bloom, 1999 and Shadish, Cook, & Campbell, 2002). As in this paper, comparative interrupted time series analyses have been applied primarily to study broad systemic policies and interventions, such as the federal No Child Left Behind Act of 2002 (see Dee & Jacob, 2008 and Wong, Cook, & Steiner, 2009); accountability systems (see Jacob, 2005); and comprehensive school reforms, such as Accelerated Schools (see Bloom, 2001) and Talent Development High Schools (see Kemple, Herlihy, & Smith, 2005).

[6] See Chapter 5, page 15, for a discussion of the data and methods used to answer the study's third research questions.

[7] See Appendix B for a detailed discussion of the research design.

schools' 2011 scores deviate from the scores predicted by their baseline trend, we were able to account for any citywide factors that may have impacted all schools, such as changes in DOE policies or in the New York State math test itself. Finally, we calculated a "difference in differences:" 1) how much the actual performance of SO1 schools differed from their predicted trends and 2) how much the performance of the comparison schools difference from their predicted trends. The difference between these two differences is our estimate of the impact of SO1 on test scores.

The central strength of this methodology is that it accounts for many factors that may have produced changes in math achievement in the SO1 schools besides the implementation of the school-wide SO1 program in the 2010-11 school year. The goal of accounting for these factors is to construct the best estimate of math achievement levels that were likely to have occurred in the SO1 schools in the absence of the program.

The primary findings for this study are based on New York State math assessment scores for 6th grade students from each SO1 school and each comparison school. Because all three SO1 schools served 6th grade students and because none of these students had participated in the summer or afterschool SO1 pilot initiatives before the 2010-2011 school year , the results for this group provide the most robust and valid indications of SO1's early impacts on math achievement. The analysis also focused on 7th and 8th graders at Schools A and B and their comparison schools. Because of the smaller sample sizes, results for these students will be less reliable.

In keeping with SO1's target population, the study sample does not include students receiving bilingual and special education services at the SO1 schools and comparison schools. In a separate analysis, we estimate the impact of SO1 on the achievement of those special education students that received SO1 instruction—that is, "mainstreamed" special education students who are not mandated to be in small classrooms. These students are not included in our overall estimates. [8]

The analysis includes several groups of students who may have received limited exposure to SO1. As noted above SO1 and school staff opted move several groups of students in and out of the program during the course of the pilot year (see page 5 for details). The research team was not able to identify these students individually and thus, could not extract them from the analysis. In addition, it was not possible to identify counterparts for these students in the previous cohorts from the SO1 schools or from the comparison schools in order to retain a balanced and unbiased analysis. However, the research team did conduct a series of tests to determine the sensitivity of the findings to excluding groups of students with related characteristics from both the SO1 and comparison schools. These sensitivity tests found no systematically different pattern of effects.

[8] We excluded "mainstreamed" students with special education needs from our overall estimate of the impact of SO1 for two reasons. First, these students received additional accommodations and services that may confound our assessment of SO1. Second, their performance levels were very different from those of the general education population, making comparisons difficult.

In addition to examining SO1's effects on the full sample of students in the selected schools, we also estimated its impact on different subgroups of students separately. The analytic strategy for these analyses was the same as the strategy described above, except that we focused on discrete subgroups of students defined by prior performance levels and other student characteristics. Two important factors make these estimates less reliable than the overall impact estimates. First, subgroups of students were unevenly distributed across SO1 schools and their matched comparison counterparts. This uneven mix may skew the estimates of SO1 impacts, depending on whether the distribution of students with certain characteristics is associated with the relative effectiveness of SO1. This may mean that any conclusions about the SO1's effectiveness for certain subgroups of students are confounded by its effectiveness under related operating conditions. Second, the number of students in each subgroup is smaller than the overall sample of students in each school. The smaller samples of student subgroups will make the impact estimates less reliable than the overall impact estimates. In light of these limitations, we present the subgroup findings as suggestive rather than definitive and to motivate further research.

IV: First-Year Impact Findings

Table 2 presents the core findings from the early analysis of SO1 impacts on student math achievement as measured by the New York State math test in the 2010-2011 school year. The table indicates that, on average across the three schools in which it was piloted, the program did not produce a systematic difference between the SO1 schools and the comparison. In other words, the average math scale score for 6^{th} graders at SO1 schools was virtually the same as that of the estimated comparison. These estimated differences account for changes in student achievement over time, changes in the demographic composition of the student body of schools over time, and difference in achievement and demographics between the SO1 schools and the comparison schools.

The most reliable estimate of SO1's impact is the estimate that is averaged across the three schools. However, looking at school-specific results adds further nuance to the findings even though the individual estimates may be less reliable that the overall average. Despite the lower statistical power, Table 2 indicates that the estimated impact of SO1 varied substantively across the three schools. The estimated impact for School A was positive and statistically significant (effect size: 0.28). The estimated impact for School B was nearly zero and not statistically significant. The estimated impact for School C was negative and statistically significant (effect size: -0.23). In the literature on education impacts (Hill et al., 2007), both the effects at School A and School C are considered relatively "large" effects. Further, the difference in 6^{th} grade estimated impacts across the three schools was statistically significant, suggesting that the differences are not likely to be due to chance. Nonetheless, with only one year of program operation and no consistent pattern of results in other grades, one should be extremely cautious about drawing inferences about the potential sources of this variation.

Table 2
First-Year Impacts of SO1, by School and Grade Level
(New York State Math Test, Scaled Scores)

Sample	SO1 Schools	Comparison	Estimated Difference	
6th grade				†
School A	682.0	672.6	9.5	***
School B	679.1	680.2	-1.1	
School C	652.8	660.8	-7.9	***
6th grade average	671.4	671.3	0.1	
7th grade				
School A	686.0	682.2	3.8	*
School B	679.9	684.1	-4.2	**
7th grade average	683.0	683.2	-0.2	
8th grade				
School A	686.2	692.8	-6.6	***
School B	685.0	686.6	-1.6	
8th grade average	685.6	689.7	-4.1	***

Source: Research Alliance analysis of New York State math test scores.
Notes: Statistical significance of estimated differences is indicated by: * = p < .10; ** = p < .05; *** = p<.01. Statistical significant of variation in estimated differences across schools is indicated by † = p< .01. Estimates are regression adjusted to control for differences between SO1 and comparison schools due to individual student characteristics (including race/ethnicity, gender, English Language Learner, special education, free lunch and holdover status, age, and prior test scores and attendance) and school-level trends in math achievement from 2006 to 2010.

In Table 2, the school-level impact estimates for 7th and 8th grade do not appear to be consistent with the 6th grade results. The table presents a mix of positive and negative estimates across grade levels and within schools, with no clearly discernible pattern. For example, while 6th and 7th grade pooled estimates were nearly zero, the pooled estimated impact for 8th grade was both negative and statistically significant. The estimated impacts for 6th and 7th grade in School A grade were positive, but the estimate for 8th grade was negative. The estimates for School B were all negative, but only the 7th grade estimate was statistically significant. While there is a chance that differences in the way SO1 functions in each of these different school and grade settings led to these different results, it is also possible that other factors caused the variations in effects.

Finally, we examined early SO1 impacts for a variety of subgroups in each grade and school. These subgroups were defined by prior New York State math test performance level, by special types of instruction (ESL and mainstream special education instruction[9]), by race/ethnicity, and by sex. Table 3 shows the results for subgroups in the 6th grade. None of the

[9] We focus on students receiving ESL and mainstream special education instruction, because these students were eligible for SO1. ELL students receiving bilingual instruction were not eligible; special education students whose IEPs mandated a small classroom were not eligible.

estimated differences in Table 3 were statistically significant, suggesting that the SO1 was not more effective for some students compared to others.

The positive results for Level 1 students may merit further study and additional follow-up data. These students entered 6th grade with the lowest levels of prior math achievement. Although the estimate is not statistically significant, the positive result runs counter to the SO1 theory of action. As noted earlier, SO1 staff suggested that these students were likely to experience an achievement dip because of the program's focus on "meeting students where they are," even through this may mean low performing students are not exposed to the grade level material that constitutes much of the New York State math test. It will be important to determine whether the focus on below-grade-level skills helped these students make more progress than students in traditional classes or if there is some other mechanism at work. A preliminary investigation of this is presented in the next section of the report.

Table 3
First-Year Impacts of SO1 on
6th Grade Students, by Subgroup
(New York State Math Test, Scaled Scores)

Sample	SO1 Schools	Comparison	Estimated Difference
All 6th graders	671.4	671.3	0.1
Level on New York State math test in 5th grade			
Level 1	645.3	638.7	6.7
Level 2	652.2	655.3	-3.1
Level 3	678.5	676.9	1.6
Level 4	696.7	695.5	1.2
Race/ethnicity			
Asian	680.7	679.6	1.1
Black	656.0	656.1	-0.1
Hispanic	663.6	660.6	3.0
White	684.7	679.5	5.1
Gender			
Female	670.9	671.8	-0.9
Male	671.5	671.1	0.4
ESL	661.1	657.2	3.8
Mainstream special education	656.3	661.1	-4.8

Source: Research Alliance analysis of New York State math test scores.
Note: None of the differences are statistically significant. Estimates are regression adjusted to control for differences between SO1 and comparison schools due to individual student characteristics (including race/ethnicity, gender, English Language Learner, special education, Free Lunch and holdover status, age, and prior test scores and attendance) and school-level trends in math achievement from 2006 to 2010.

Finally, Appendix B presents results for each student subgroup by grade and school, and tables include standard errors for estimates. These results are unstable and should be interpreted with a high degree of uncertainty. Appendix C presents results from our tests of the sensitivity of the sample specification; we find no systematically different pattern of effects.

V: Exploratory Analysis

This section describes the exploratory analysis, used to answer our third research question: Is exposure to more SO1 material, and/or mastery of SO1 skills, associated with improved math performance? First, the data are detailed and the methodology used is explained. Then we report on the relationship between exposure to SO1 instruction, mastery of SO1 skills, and prior student performance. Next, we discuss the association between SO1 skill mastery and test score growth. Finally, we describe the association between exposures to SO1 instruction and test score growth.

As a technology-rich program SO1 has a wealth of internal data. Every day, students are assigned to work on a particular skill at the level appropriate for them. Records of which skills each student worked on give us a sense of the proportion of time the student spent on material that was below-grade-level, on-grade-level, or above-grade level.

For each skill, SO1 staff estimated how many periods of instruction and practice were typically for a student to master the skill, an amount SO1 refers to as the skill's "par." Students receive "par points" when they master a skill. For example, student receive three the "par points" for mastering a skill that is estimated to require three lessons to master. They can earn these three "par points" regardless of whether they demonstrate mastery after being exposed to the material twice or they require five lessons. Therefore there may be a disjuncture between the number of *exposures* a student had to SO1 skills and the amount of *mastery* they demonstrate.

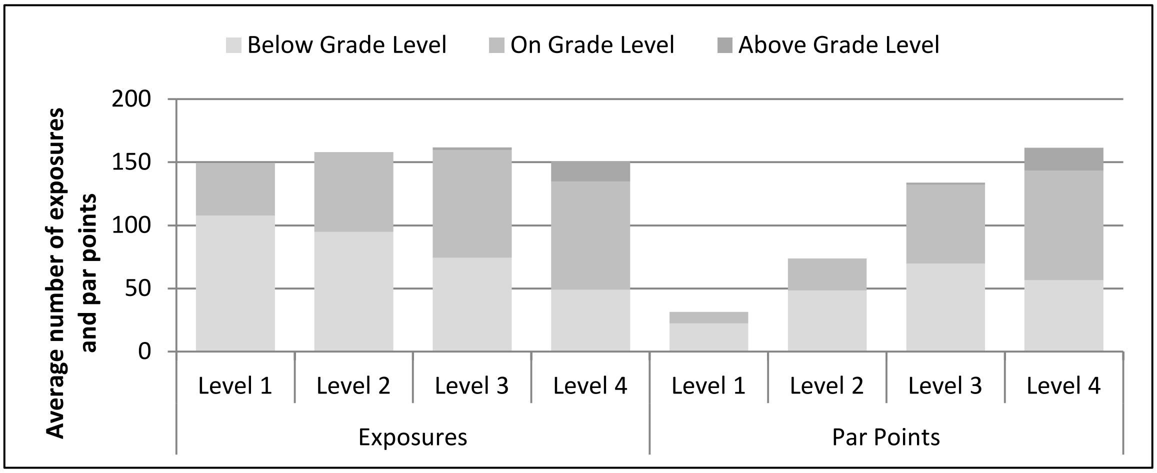
These data cannot be used for rigorous evaluation purposes, since they are not available in the comparison schools. Nevertheless, the exploratory analyses presented in this section may provide useful information for SO1.

In our exploratory analysis, we compare the average levels of exposure to and mastery of SO1 skills for groups of students based on prior performance. Then we use regression to look at the association of exposure to and mastery of SO1 skills with test score growth, in order to understand these relationships while controlling for other key factors.

Rates of Exposure and Par-Points by Prior Performance

We calculated the average number of exposures to skills at various levels and the average number of par points earned by prior performance to produce the following figure.

Figure 2
Average Number of Exposures and Par Points by 2010 New York State Math Performance Level



Source: Research Alliance analysis of SO1 internal data and NYS math test scores.

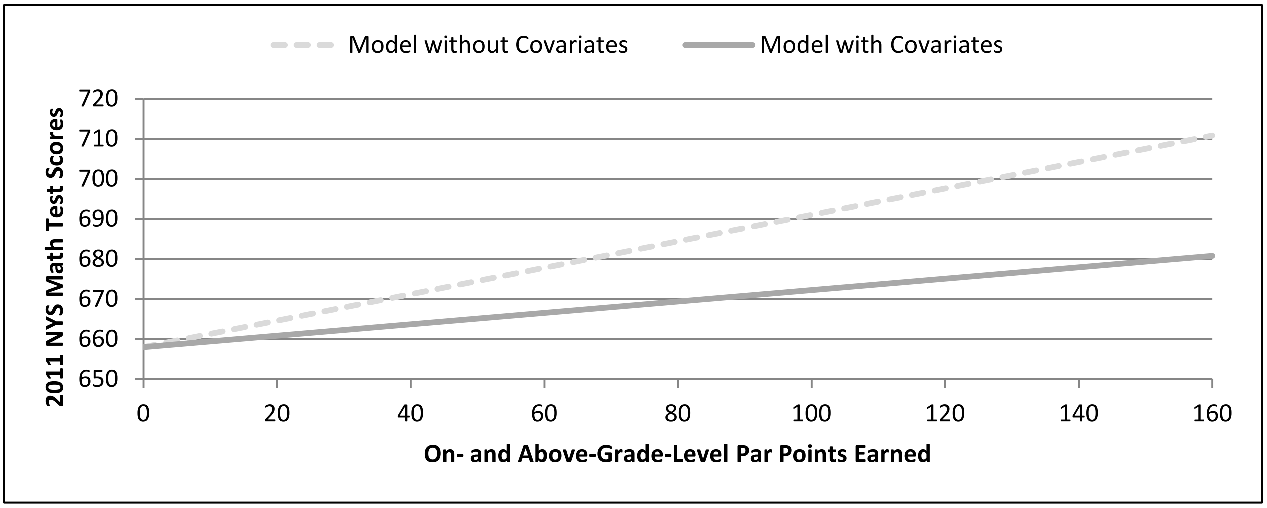
Students gained exposures for every day they are in attendance. Not surprisingly, students who performed better on the 2010 New York State math test tended to receive more on- or above-grade-level lessons than those who performed more poorly. The total number of exposures, however, varies only slightly by performance level.

By contrast, the average total of par points earned—which indicates the level of skill mastery—show substantive differences by prior performance levels. This figure shows that not only did lower-performing students receive more instruction on below-grade-level skills, they also earned fewer par points at these low levels. The SO1 theory of change relies on the idea that, given instruction appropriate to their level, low-performing students will make rapid progress and be able to catch up to grade level. The above figure can give us no indication of the rate of skill mastery that would allow such students to catch up to grade level, or whether or not SO1 students achieved that rate in the first year of implementation. However, the figure does make it clear that students with low prior test scores tended to master skills at a slower rate than students with high prior test scores, despite the fact that the former were exposed to a greater proportion of below-grade-level, presumably easier, skills. This finding suggests that students who came to SO1 with lower levels of achievement continued to struggle despite exposure to instruction that was presumably at the appropriate level for them.

The Relationship Between SO1 Skill Mastery and Test Score Growth

One way to look at the impact of SO1 is to consider the relationship between the intensity of students' SO1 experiences and their test scores. For example, we could say that students who earned more par points experienced SO1 more intensely. We can examine the relationship between the number of par points students earned and the growth in their test scores by performing a regression analysis. If we look at the relationships between the number of on- and above-grade-level par points students earn and their test scores, we find what appears to be a strong, positive relationship. However, the number of par points students earn are closely related to students' prior math performance and a number of other background characteristics that are completely independent of SO1. Below in Figure 3, we model the relationship between par points and 2011 New York State math test scores in two ways: first without taking into account other factors that influence performance, and then using statistical methods to attempt to account for a set of key background characteristics.[10]

Figure 3
Two Models of the Association Between Par Points and Test Scores



Source: Research Alliance analysis of SO1 internal data. The solid line is regression adjusted to control for individual student characteristics (including school race/ethnicity, gender, English Language Learner, special education, free lunch and holdover status, age, and prior test scores and attendance).

When we include only these basic covariates, the strength of the relationship between par points and 2011 test score drops by more than half. This model does not include important factors such as student's motivation or propensity to show effort in their work. If such difficult-to-measure variables were included in our model, it is likely that the relationship between par points and test scores would become even weaker. For these reasons, it is unlikely that the relationship between par points and test scores provides useful information about the impact of SO1.

The Relationship Between Exposure to On-Grade-Level Skills and Growth

SO1 works with each student at her own level, even if that means teaching material below grade level. This focus raises questions, because even students who are performing far below grade level must take the New York State math test intended for their grade. In the long run it is possible that the strategy of "meeting the students where they are," may be more beneficial for

[10] The characteristics we include are: prior test performance and attendance, grade level, school, race/ethnicity, gender, free lunch status, assignment to ESL instruction, and whether the student is over age for grade or was retained in the last year.

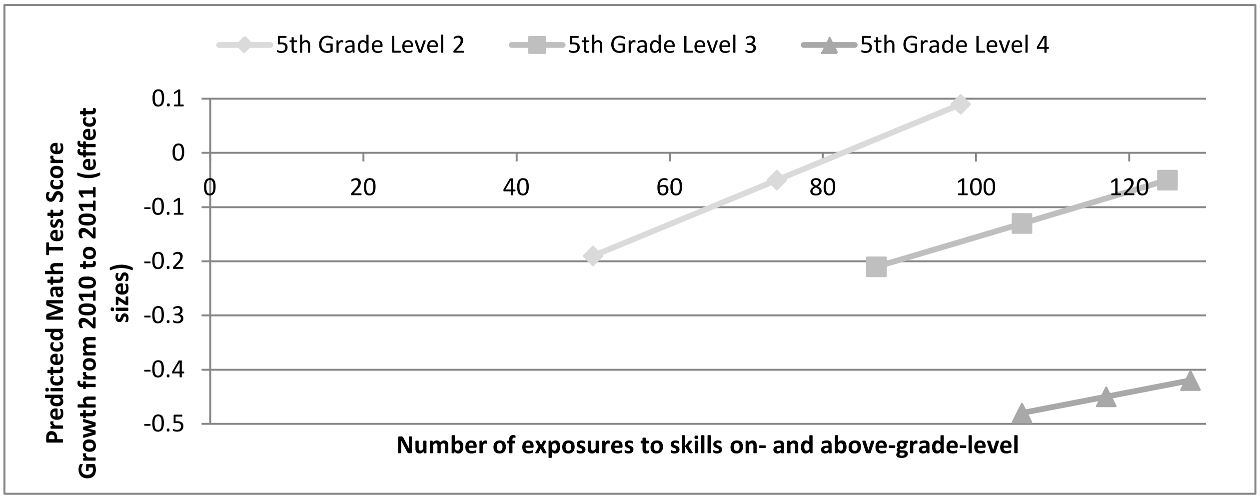
students performing below grade level than it would be to focus on grade level material. The question remains, however: what is the short-term impact of marginally increasing the amount of on-grade-level material? Without drastically altering SO1's theory of change, would it benefit students to be exposed to slightly more on-grade-level skills than they are currently exposed to?

As above, in this analysis we rely on internal SO1 data, so the variables we look at are not available for students in the comparison schools. Therefore, what follows is not a causal analysis that can definitively answer the questions posed, but a descriptive analysis that can illuminate patterns in the data. We use regression analysis to model the relationship between students' test score growth from 2010 to 2011 and the number of exposures to on- and above-grade-level skills. By doing this we can see if, on average, students who were exposed to more on-grade-level skills have higher test score growth. Because we already know that students with lower prior math performance are exposed to fewer on-grade-level skills, we looked at how the relationship plays out within groups of students who scored the same performance level on the 2010 New York State math test. Within each performance level and grade, we identified the 25th, 50th, and 75th percentile levels (Q1, Q2, and Q3) for the number of on- and above-grade-level skill exposures students received. We use our regression model to compute the predicted test score growth for students with Q1, Q2, and Q3 on-grade-level skill exposures. These predictions, shown in Figure 4 on the next page, suggest that marginal increases in on- and above-grade-level skill exposures were associated with higher or lower test score growth.

Each line shows the differences between growth of students with Q1 and Q3 levels of exposures to on- and above-grade-level skills. We do not chart the relationship for students with Level 1 because there were so few of these students in the 6th grade that we could not accurately estimate this relationship. For students that scored a Level 2 (lightest line), 3 (medium line) or 4 (darkest line) on the 2010 New York State math test, those who were exposed to a lower level of on- and above-grade-level skills had significantly lower growth than those exposed to a higher level of on- and above-grade-level skills. For Level 2 students, for example, changing a student's number of on- or above-grade-level exposures from around 50 to around 75 corresponds with a change in predicted growth from -0.20 to -0.05 effect sizes.[11] Changing from 75 to 100 on- or above-grade-level exposures corresponds with a change in predicted growth from -0.05 to 0.10 effects sizes.

[11] Effect sizes are commonly used to compare scores across years while taking into account slight differences in the test from year to year.

Figure 4
Associations Between Exposures to On- and Above-Grade-Level Skills and Test Score Growth, By Prior Year Performance Level, 6th Grade[12]



Source: Research Alliance analysis of SO1 internal data and NYS math test scores. Predictions are regression adjusted to control for individual student characteristics (including school race/ethnicity, gender, English Language Learner, special education, free lunch and holdover status, age, and prior test scores and attendance).

While there was an association between exposures to on- and above-grade-level skills and test score growth, this relationship varied depending on students' prior performance. The relationship was stronger in Level 2 than in Level 3, and it was stronger in Level 3 than in Level 4. That is, for students who began the year with lower skill levels, on average each additional exposure to an on- or above-grade-level skill corresponded to higher growth than an additional exposure for a higher-performing student. The figure above shows 6th graders only, but charts for all grades are available in Appendix D and all show similar patterns. This finding suggests that the test scores of students who enter SO1 performing below level may benefit from marginal increases in exposures to on- and above-grade-level skills more than for students who start the year performing on or above grade level. Of course, this analysis cannot address the long-term impacts of such marginal increases, or whether even small increases might undermine SO1's theory of change.

[12] It may be striking to the reader that, in this figure, test score growth declines as prior-year performance increases. While this idea may seem counterintuitive, it is commonly observed for two reasons: ceiling effects (limits on test score growth for high performers because they simply can't earn any more points than they have before) and regression to the mean (the idea that if a student scores extremely high or low on a prior year test, they will tend to be closer to the average on a future test).

VI: Conclusions and Next Steps

Because this study estimated the impact of SO1 in the first three schools to use the program and its initial year of school-wide implementation, the findings should not be interpreted as a definitive indication of the SO1's impact on student achievement. Rather, the findings are presented as initial feedback for SO1 in an effort to guide their ongoing development of the program model and to contribute to future studies of the program. The design, execution, and findings for the current study offer several lessons and recommendations for future research on SO1. Among these are the following.

Future efforts to expand the deployment of SO1 should assess its impact on student achievement. This should include the capacity to follow students throughout their middle school careers and assess impacts on their transitions into and through high school. Establishing the programs impact for a wider range of schools and its effect on longer term outcomes will be critical to establishing its efficacy.

It may be particularly important to track the progress of the lower-achieving students in light of the trends we find for this group—positive but not statistically significant impacts, combined with steeper improvements among those exposed to higher proportion of on-grade-level skills. These students do not appear to have experienced the initial dip in achievement that might have been expected, given that that most of the skills they worked on were below grade level. To explore this dynamic further, future studies of SO1 should consider assessing students' learning progression in a more fine-grained and more frequent manner than is possible with the state assessments.

Future research on SO1 should include systematic assessments of program implementation as well impact on student achievement. The current study points to a web of different estimated impacts across the three pilot schools and across grade levels. It would be useful to know whether some of the schools have been more effective in their implementation than others and whether these differences are associated with an evolving pattern of estimated impacts. Additional research should try to understand the challenges of implementing this program and to identify ways to strengthen the program's effectiveness. Addressing these questions might involve interviews with teachers and students and observations of the SO1 teaching and learning activities.

Studies of SO1 deployment should focus the challenges teachers face as they adapt to the program and how they are supported with professional development and collaboration. It will be useful to document how teachers are trained to engage with this innovative instructional model, and to identify supports that help teachers address issues that emerge throughout the school year. Toward this end, it may be appropriate for future researchers to observe SO1's professional development activities and to conduct focus groups with teachers to gain their perspective on the challenges of implementing the program.

Finally, it should be noted that while SO1 demands rigorous assessment, monitoring, and expectations in the classroom, its commitment to these values is reflected in its approach to program development and evidence-building. Just as SO1 challenges its teachers and students to continually assess their progress and make adjustments in response to those assessments, the program's developers are committed to a learning process that allows them to refine and improve the model. SO1 continues to evolve, and its developers are seeking opportunities to expand its use in selected New York City middle schools. The program was recently awarded a coveted "development grant" from the U.S. Department of Education's Investing in Innovation (I3) Fund. The award will enable SO1 to improve the program and conduct further research on its impact and implementation. The grant provides a unique opportunity to execute some of the recommendations presented above.

References

Ackerman, P. L. (1987). Individual differences in skill learning: An integration of psychometric and information processing perspectives. *Psychological Bulletin*, 102(1) 3–27.

Banerjee, A., Duflo, E., & Linden, L. (2007). “Computer-assisted learning project with Pratham in India.” *The Quarterly Journal of Economics*, August 2007.

Barrow, L., Markman, L., & Rouse, C.E. (2009). Technology’s edge: The educational benefits of computer-aided instruction. *American Economic Journal: Economic Policy,* 2009, 1(1), 52–74.

Bloom, H. S. (1995). Minimum detectable effects: A simple way to report the statistical power of experimental designs. *Evaluation Review, 19*(5), 547-556.

Bloom, H. S. (1999). *Estimation program impacts on student achievement using “short” interrupted time series*. New York, NY: MDRC. Retrieved September 2, 2010, from http://www.mdrc.org/publications/82/full.pdf.

Bloom, H. S. (2001). *Measuring the impacts of whole-school reforms: Methodological lessons from an evaluation of accelerated schools*. New York, NY: MDRC. Retrieved September 2, 2010, from http://www.mdrc.org/publications/76/full.pdf.

Bloom, H. S. (2003). Using “short” interrupted time-series analysis to measure the impacts of whole-school reforms. *Evaluation Review, 27*(1), 3-49.

Davis, Michelle R. (2011, March 17). Moving beyond one-size-fits-all.” *Education Week, 30* (25), 10. Hamby, J. V. (1989). How to get an “A” on your dropout prevention report card. *Educational Leadership, 46*(5), 21-28.

Hill, C. J., Bloom, H. S., Black, A. R., & Lipsey, M. W. (2007). *Empirical benchmarks for interpreting effect sizes in research*. New York, NY: MDRC.

Dee, T., & Jacob, B. (2008). *The impact of No Child Left Behind on student achievement.* (NBER Working Paper 15531). Cambridge, MA: National Bureau of Economic Research. Retrieved September 2, 2010 from http://www.nber.org/papers/w15531.pdf.

Fullan, M. (2001). *Leading in a Culture of Change*. San Francisco: Jossey-Bass.

Jacob, B. A. (2005). Accountability, incentives and behavior: The impact of high-stakes testing in the Chicago Public Schools. *Journal of Public Economics, 89*(5-6), 761-796.

Kemple, J. J., Herlihy, C. M., & Smith, T. J. (2005). *Making progress toward graduation: Evidence from the Talent Development High School Model*. New York, NY: MDRC. Retrieved September 2, 2010 from http://www.mdrc.org/publications/408/full.pdf.

The School of One. (2009, November 12). *Time.* Retrieved April 25, 2012, from: http://www.time.com/time/specials/packages/article/0,28804,1934027_1934003_1933977,00.html

Lee, Jaekyung & Fish, Reva M. (2010). International and interstate gaps in value-added math achievement: Multilevel instrumental variable analysis of age effect and grade effect. *American Journal of Education*, *117* (1), 109-137.

National Center for Education Statistics. (2011). *The nation's report card: Mathematics 2009 Trial Urban District Assessment*. Alexandria, VA: NCES. Retrieved September 2, 2010, from http://nces.ed.gov/pubsearch/pubsinfo.asp?pubid=2010452rev.

National Center for Education Statistics. (2010). *The nation's report card: Grade 12 reading and mathematics 2009 national and pilot state results.* Alexandria, VA: NCES. Retrieved September 2, 2010, from http://nces.ed.gov/pubsearch/pubsinfo.asp?pubid=2011455.

Nationan Center on Education and the Economy. (2011). *Variability in pretest-posttest correlation coefficients by student achievement level*. Washington, DC: NCEE. Retrieved September 2, 2010, from http://ies.ed.gov/ncee/pubs/20114033/pdf/20114033.pdf.

New York State Education Department. (1999). Standard setting and equating on the new generation of New York State assessments. Albany, NY: NYSED. Retrieved August 30, 2011 from http://www.p12.New York Stateed.gov/apda/assesspubs/pubsarch/sse-newgen.pdf.

Organisation for Economic Cooperation and Development. (2010). *PISA 2009 at a glance.* Paris: OECD Publishing. Retrieved August 30, 2011 from http://dx.doi.org/10.1787/9789264095298-en.

Shadish, W. R., Cook, T. D., & Campbell, D. T. (2002). *Experimental and quasi-experimental designs for generalized causal inference*. Boston: Houghton-Mifflin.

Wong, M., Cook, T. D., & Steiner, P. M. (2009). *No Child Left Behind: An interim evaluation of its effects on learning using two interrupted time series each with its own non-equivalent comparison series.* (IPR Working Paper WP-09-11). Evanston, IL: Institute for Policy Research, Northwestern University. Retrieved September 2, 2010 from http://www.northwestern.edu/ipr/publications/papers/2009/wp0911.pdf.

The Research Alliance for New York City Schools

285 Mercer Street, 3rd Floor | New York, New York 10003-9502
212 992 7697 | 212 995 4910 fax
research.alliance@nyu.edu | www.steinhardt.nyu.edu/research_alliance

The Research Alliance for New York City Schools conducts rigorous studies on topics that matter to the city's public schools. We strive to advance equity and excellence in education by providing non-partisan evidence about policies and practices that promote students' development and academic success.