Evaluation Report on the School of One i3 Expansion

Jonah E. Rockoff
Columbia Business School

September, 2015*

This report presents quantitative and qualitative evidence on the expansion of the New York City Department of Education's School of One middle school mathematics program. The School of One program began pilot testing in the summer of 2009 and was first used as a full-time mathematics program by three NYC middle schools in the school year 2010-11.[1] An expansion of School of One into four additional middle schools, funded by a federal "i3" grant, occurred over the school years 2012-13 and 2013-14. As part of the grant requirements, an evaluation of program impacts was performed by the authors of this report, who are independent researchers. This evaluation is based on a randomized control trial (RCT) methodology, where the School of One "treatment" was assigned randomly among a set of candidate schools. School of One is therefore being compared to traditional mathematics instruction programs in the remaining "control" schools.

Overall, program impacts on student and teacher outcomes are imprecise because of the very small effective sample size. This was anticipated at the time of the grant application, but is nevertheless an important limitation of the RCT method for evaluating this developing program. Very large positive or negative program effects are rejected, but moderate effects or the null hypothesis of program effects being zero cannot be rejected. Survey evidence is also mixed. Students in later grades assigned to School of One expressed significantly worse views of their math curriculum in the first year of the expansion (although not in the second year), while teachers of School of One expressed significantly positive views.

* Address: Uris Hall 603, Columbia University, 3022 Broadway, NY NY 10027. Email: jr2331@gsb.columbia.edu.
This evaluation would not have been possible without the help of a large number of individuals. The author would like to thank Leo Bullaro, Anne-Marie Hoxie, Alana Laudone, Brent Morita, Michelle Joseph, Yvonne Wang, Tres Watson, and Jonathan Werle, from the NYC Department of Education; Wendy Lee, Jessica Licata, Jennifer Padilla-Howard, Joel Rose, Chris Rush, Christine Sargent, Dana Schmidek, Jennifer Stillman, and Mark Williams of School of One/New Classrooms; Leigh Linden, Ofer Malamud, Jesse Margolis, and Geoffrey Borman for advice on the empirical design; Ama Awotwi, Angela Deng, Carlos Marin, Eric Hardy, Nekpen Osuan, and Jimmy Salamon for outstanding research support; Hana Lahr who conducted summer interviews and wrote the related qualitative reports, and a special thanks to Miriam Fenton who worked tirelessly to manage the evaluation process.
[1] Early research on the School of One program is described in Appendix A.

Additional detailed qualitative evidence on the implementation of School of One is provided via reports based on interviews with administrators and teachers in the four School of One expansion schools, which were conducted in the summers of 2013 and 2014.[2] These interviews reveal a number of difficulties faced during the expansion years, but also the belief by some teachers and administrators that the program was effective.

The remainder of this report proceeds as follows. Section 1 provides background information on the School of One program and Section 2 discusses the design of the grant-funded expansion and evaluation and how this changed over time due to unforeseen circumstances such as non-compliance. Section 3 describes our data and the sample we use for analysis. Section 4 shows baseline comparisons of treatment and control groups and assesses sample attrition. Section 5 presents descriptive evidence from surveys on how School of One was viewed by students and teachers, and Section 6 contains treatment effect estimates on math tests and survey data. Section 7 provides a brief discussion and conclusion.

1. Background on the School of One Program[3]

School of One (So1) is a mathematics instruction program which differs in many respects from the traditional classroom experience. A technology platform provides instructional plans and materials dynamically, and students receive a mix of live, online, and collaborative instruction that is tailored to their academic needs, interests, and learning preferences. So1's theory of action is based on the premise that students learn at different speeds and in different ways, and that fully differentiated instruction is not possible under the traditional "one teacher" model. So1 seeks to meet each student wherever he or she is on the continuum of math knowledge and skills.

So1 begins the school year with a diagnostic assessment of each student's math skills, which is used to create an individualized learning plan ("playlist") that specifies the skills on which the student

[2] These reports are included as Appendices D and E.

[3] The text on the background of School of One borrows heavily from the i3 Study Design Plan for the Evaluation of the School of One Expansion. The analysis provided in this document, however, provides a more comprehensive quantitative and qualitative analysis of the School of One Expansion than the Impact and Implementation analyses required under the conditions of the i3 grant.

should work. Each class period, students receive instruction in a variety of methods of instruction ("modalities"), and, 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 these assessments are used to inform the students' learning plans for future class periods. Both teachers and So1 staff monitor students' progress and can adapt the learning plans to meet their evolving needs on a daily basis.

So1 represents some important adjustments from traditional math instruction for both students and teachers. The most basic visual difference is in classroom design. A group of roughly 90 students, who would typically be split into multiple traditional classrooms, instead learn in one large room with multiple teachers. Students' walk into the classroom and look to a large screen to find out where they will be working and what skills they will be working on during the session. For teachers, So1's dynamic algorithm means that the sequence of lessons they will teach, and which particular students will receive the lessons, cannot be known at the start of the school year; new information on students' learning plans are provided to teachers after school each day.

From a technical perspective, the So1 program has five key components. First is a "learning progression": the discrete set of skills students must master and research-based evidence on the relationships among those skills. Second, there is a baseline model which integrates all of the available data about each student, administers an additional diagnostic instrument, and, based on that information, creates a "playlist" for each student- a unique set of skills to focus on over a period of time. Third, So1 sources instructional content from publishers, software providers, and other educational groups across nine instructional modalities: life instruction, live reinforcement of prior lessons, live tutoring, small group collaboration, independent practice, virtual computerized instruction, virtual live instruction, virtual live tutoring and homework. Fourth, a "learning algorithm" captures and analyzes the data from each lesson and recommends to teachers a unique daily schedule for each student that they can adjust as necessary. Fifth, at the end of each day, students take a unique assessment- "playlist update"- to measure mastery of the skill they studied.[4]

[4] For students with disabilities, scheduling learning activities can be provided in a manner consistent with the Individualized Education Plan and in a way that enables special needs students to learn alongside their general

As mentioned above, So1 was developed by the NYC DOE and its curriculum was designed to work within the New York State mathematics standards. However, in 2011 the intellectual property for So1 moved to a new nonprofit organization—New Classrooms Innovation Partners ("New Classrooms"). While the So1 program continued to be within the purview of the DOE, its operation during the expansion was handled by New Classrooms, who continued to develop their instructional platform under the name "Teach to One." In addition to other changes made to improve the program, updates to the platform were made to meet changes in curriculum standards in New York (e.g., the adoption of the Common Core), as well as to meet standards in other districts and states where New Classrooms has expanded over time.

2. i3 Expansion and Evaluation

In the spring of 2010, the NYC DOE applied for a federal Investing in Innovation ("i3") development grant, which it was subsequently awarded in August of that year. The grant would fund an expansion of So1 into four middle schools where it would become the new school-wide math curriculum. An independent two-year evaluation of the program expansion was included as a requirement of the grant proposal and focuses on the following key questions:

1. Is the So1 mathematics curriculum associated with higher math achievement, as measured by high stakes and low stakes math tests, compared to the usual mathematics curriculum offered in New York City public schools?
2. Is the So1 mathematics curriculum associated with other improved student outcomes including academic behaviors and attitudes, compared to the usual mathematics curriculum offered in New York City public schools?

education peers. For students learning English as a second language, instructional content can be sourced and specifically scheduled for those students, including digital content that is translated into a number of different languages and collaborative content that requires students to interact with one another.

3. Is the So1 mathematics curriculum associated with improved teacher outcomes including attitudes towards teaching and technology, compared to the usual mathematics curriculum offered in New York City public schools?

The proposed evaluation was to use a cluster randomized controlled trial (RCT) comprising eight schools equally divided (by random assignment) into treatment and control groups. Randomization helps to reduce the risk that any potential confounding factors might lead to differences in outcomes between treatment and control groups that are not due to So1. In other words, since So1 was assigned randomly, any statistically significant differences between the groups ex post can be attributed to the impact of So1. Power tests conducted for the proposal indicated that this design would only have sufficient power to detect vary large (i.e., roughly 0.25 standard deviation) effects of the program. The reason why the evaluation is fairly low powered is that, despite a large number of students within each school, assignment was made at the school level and there are substantial correlated effects within a school. This effectively yields eight independent observations. Nevertheless, at the time, these effect sizes were thought to be within the potential of So1.

Originally, the evaluation was intended to proceed as follows. First, over the course of year one of the project, a sample of eight schools with the appropriate characteristics—e.g., technological infrastructure and buy-in from administrators and teachers—were to be identified as eligible. Second, in the eight schools identified for the sample, we were to collect administrative data (including end of year test scores) for all students and conduct a baseline survey of students and math teachers to measure the non-test outcomes described above. Third, we intended to assign four of these schools randomly to a treatment group and four to a control group. The program was then to be implemented in the four chosen schools, with evaluators collecting follow-up data (via surveys, testing, and administrative data requests) and performing statistical analyses.

So1 began working on recruitment and site selection in late 2010. An initial notice soliciting volunteer schools was released in late 2010 which detailed the requirements for volunteer middle schools in terms of technology, infrastructure, staff and students. Technology-wise, schools needed at least W3

wireless access and ATM or EVPL bandwidth pipe. Schools also needed the ability to adjust schedules and a willingness to consider space adjustments. Staff at participating schools must be willing to participate, be comfortable with instructional technology, and be available for professional development and feedback. Leaders must have a proven record of efficacy and be enthusiastic, responsive, and comfortable with the uncertainty that accompanies innovation. So1 personnel also endeavored to select schools with a strong need for the program, giving preference to schools serving greater percentages of students eligible for Title I, English as a Second Language (ESL), and special education services. Site visits began in January 2011 with repeat visits occurring in February. During this process, the field was narrowed down to the final 8 schools: R049, X022, K088, K211, K381, X368, R002, K014.

So1 expansion was to start in the school year 2011-12. However, as mentioned earlier, a number of issues led to the movement of the intellectual property for So1 from the NYC DOE to New Classrooms, and the decision was made in March 2011 to delay the expansion for one year. Randomization of the eight schools to the treatment (So1) or control (traditional math programs) conditions was conducted on October 20, 2011. A matched pair design, based on prior year math scores, was used to assign four schools to the treatment condition (R049, K088, R002, K381) and four to the control condition (X368, K014, X022, K211). Copies of randomization code and log files are included in Appendix B.

In addition to the delayed expansion, there are several substantial modifications to the originally proposed evaluation design. First, due to the delay in implementation as well as a long delay in finalizing the evaluation contract (which was completed in August 2012), baseline survey data were collected at the start of the school year 2012-13, rather than in the prior school year. Second, while building and renovating classrooms for So1 during the summer of 2012 it became clear that financially and structurally it would be too difficult to implement So1 in the entire school for two of the treatment schools (R049 and K088). These middle schools use an academy system, which means that students and teachers within each school are divided into three parallel systems, each with a particular theme but all with the same core curriculum and all under the direction of the school principal. Two academies within each middle school were chosen by the principal and So1 staff to be eligible for So1, and one of these academies in each

school was randomly selected (see code/log in Appendix B). As we now had a set of control students within each of these treatment schools, the paired control schools (X022 and K211) were dropped from the evaluation.[5] Although the study design is based on both between-school and within-school comparisons, for simplicity we sometimes refer to treatment and control "schools" to denote both school and academy level treatment/control units.

Last, but not least, despite repeated attempts at telephone and in-person communication with the principal by the evaluation team and high ranking NYC DOE personnel, we were unable to obtain survey or low-stakes test data for one of the control schools (K014) due to non-compliance. Thus, we can include the schools R002 and K014 when examining impacts on high stakes, NY state tests, but when looking at survey or low-stakes test outcomes we drop them from the sample. The table below summarizes information on the original and actual schools and the corresponding randomization assignments.

Randomization Assignments

ORIGINAL PROPOSED COMPARISON		**ACTUAL COMPARISON**	
Treatment Group	*Control Group*	*Treatment Group*	*Control Group*
R049	X022	R049 academy: Academy of Medical Science and Technology	R049 academy: Academy of Journalism Science and Technology
K088	K211	K088 academy: The School for Media Arts Research and Technology	K088 academy: The School for Medical Health Careers
K381	X368	K381	X368
R002	K014	R002	K014 *(NY test only)*

Before moving to our presentation of findings, it is important also to note that Hurricane Sandy, which hit New York City and the surrounding areas at the end of October, 2012, caused serious damage to one of the treatment schools, R002, which is located near the eastern coast of Staten Island.[6] Students there had to be temporarily relocated and did not use So1 for several months. Subsequently, in the second year of program expansion, only roughly one-third of students in R002 used So1. We adjust for less-than-full

[5] In addition to the financial and logistic benefits, this change to a within-school comparison (rather than a between-school comparison) may reduce the amount of random error in the analysis and therefore improve the precision of results. It may also reduce the amount of systematic error in the analysis since students from the same schools may be more comparable that students in different schools.

[6] The schools R002 and R049 are located in Staten Island, K381, K014, and K088 are located in Brooklyn, and X368 is located in the Bronx.

take-up of So1 in our analysis below by using assignment to treatment as an instrument for actual use of So1. For analyses other than NY state test scores, R002 is dropped (because its pair, K014, was non-compliant), but it should be kept in mind that any direct effect of Hurricane Sandy and displacement will be included in the NY state test outcomes of R002 students. We therefore also present NY state test results with the subsample that drops R002 and its matched pair school.

3. Data and Analysis Sample

In the two years following expansion, we collected survey data on all students and math teachers to assess academic behaviors and attitudes, and for the treatment group, their experiences with the School of One program. Questions were chosen from previous surveys used by the NYC DOE as well as surveys from prior research studies on similar topics. Copies of survey instruments are included in Appendix C. Responses to individual survey questions are averaged to construct several indices, by design and supported by confirmatory factor analysis. For students, we form measures of Intrinsic Motivation, External Motivation, Amotivation (which we reverse score), Interest and Competence in Math, Self-Directed Learning, and Confidence/Competence in Using Technology. For teachers, we measure Self-Efficacy and Attitudes Towards Use of Technology in the Classroom. We also asked students and teachers for their opinions regarding the math curriculum they used at the end of each school year and, for students and teachers using So1, their opinions regarding specific aspects of the So1 program.

In addition, students in the fall and spring of each year took an online mathematics test developed by the Northwest Evaluation Association (NWEA). This is a nationally normed exam taken on a computer in an adaptive format, i.e. the questions (and their level of difficulty) change dynamically in response to students' prior answers, and the amount of time students spend on the test will vary with the amount of questions needed for NWEA to attain a sufficiently precise estimate of their knowledge.

Administrative records from the NYC DOE provide us with enrollment and demographic data as well as results from New York State math tests, which are high stakes for both students and schools. The New York State Mathematics Test is taken by students in grades 6-8 in late April of each year and is

administered on paper.[7] The NYS math test is not adaptive, i.e. all students are asked to answer the same questions in the same amount of time.

For the purpose of this analysis, both NYS and NWEA scores are normalized. A score of zero on the NYS test indicates the student has the average score for his/her grade level *in New York City*, and a score of +1 (or -1) indicates a score that is one standard deviation above (or below) this average. A score of zero on the NWEA exam indicates the average score for students in the same grade *nationwide* (based on NWEA national sampling), and a score of +1 (or -1) indicates a score that is one standard deviation above (or below) this average. This normalization is standard in the evaluation literature so that the magnitude of effects of different treatments can be compared across studies.

Finally, New Classrooms provided records on student participation in the So1 program. These are detailed records of the dates on which students worked on different aspects of the So1 learning progression, enabling us to distinguish any students who moved into or out of the program over time. While these records also provide information about the intensity of program usage (e.g. the number of lessons completed successfully, as measured by post-tests), variation in these measures is of course endogenous and we do not use them in our evaluation.

Our analysis sample consists of students who were enrolled in the relevant schools/academies in the fall of 2012, as well as first-time sixth graders who enrolled in the fall of 2013. For students who were in 8th grade in the school year 2012-13, we analyze outcomes for that year, but not in the school year 2013-14, when only a small fraction of students who repeat 8th grade are observed. Table 1 shows the number of students in our sample by randomization unit; the total number of students is roughly 5,000.

4. Baseline Comparisons

While schools and academies were selected randomly for the So1 program within pairs that were matched based on students' average prior math test scores, these matches were made using data available

[7] 2013 was the first year that the test was based on the Common Core Learning Standards (CCLS). Prior state test scores from spring 2012, which we use as a point of baseline comparison and as a control variable, were not based on CCLS. Nevertheless, the correlation in scores from 2012 and 2013 is 0.80 for the entire population of students in New York City and 0.75 in the analysis sample.

as of early 2011 and we have only a small number of units. Thus it may still be the case that student characteristics are significantly different, on average, between treatment and control groups. To assess these differences, we compare treatment and control groups along a number of important dimensions based on data from the prior school year (2011-12, or 2012-13 for new 6th graders in 2013-14). We also test using a weighted least squares regression (weighted by number of students) of school average characteristics on an indicator for assignment to the So1 treatment. These results are shown in Table 2. We see that treatment and control groups are fairly similar in terms of the percent of students in poverty (as measured by receipt of free/reduced price lunch), percent who received special education services, and the fraction of days they were absent, but there is a far larger fraction of English language learners (ELL) in the control group (22%) than in the treatment group (6%).[8] This is likely the reason why control group schools have a significantly lower fraction with prior English Language Arts (ELA) test scores, and why their prior ELA (and, to a lesser extent, math) test scores are worse, on average.

The precision of our comparisons between treatment and control schools' outcomes may be improved by the inclusion of control variables. At the baseline stage, it is worthwhile asking whether the differences in test scores are likely driven by differences in the demographic population served by these schools—and particularly the sharp difference in students learning English—or whether there are remaining differences in achievement levels that cannot be easily explained. Table 2 therefore includes a comparison of "adjusted" test scores, which have been conditioned on students' prior year ELL, special education, and free lunch status, as well as their prior year absence rate.[9] We can see that these basic observables account for over half of the test score differences between treatment and control groups, and that these differences are no longer statistically significant.

[8] Correspondence with School of One personnel indicates that there may be an undercounting of ELL students in the K088 So1 academy. Our data indicate very few ELL students in that sub-population, while their data (based on another indicator) indicates roughly 50% ELL students. If we use their classification then the overall treatment-control contrast for ELL students in Table 2 is no longer significant. However, it is currently unclear whether the So1 ELL indicator is more accurate than the indicator we receive for all of New York City, and we therefore do not rely on their coding elsewhere in the analysis.

[9] Adjusted ELA and Math scores are the residuals from a regression of prior scores on prior demographics that includes randomization block by treatment assignment fixed effects. The fixed effects, which are included in the residual, ensure that any "effects" of demographics which we are removing are not driven by differences in the quality of education across schools.

Because much of our analysis excludes R002 and K014, we present baseline comparisons for this subsample as well (Table 3). While the control group still has a significantly higher percentage of ELL students, differences in prior test scores are smaller and can be completely explained by pre-existing differences in student demographics; i.e., the treatment and control groups' adjusted prior test score averages are almost identical. For the NWEA baseline test we see that there are more missing observations for the control group, which reflects the fact that New Classrooms had less success getting teachers in the control classrooms to administer the exam to their students. Average scores on the baseline NWEA were also lower (by roughly 0.22 standard deviations) in the control group, although the majority of this gap can be explained by differences in students' pre-existing demographics; the difference in adjusted NWEA scores is only 0.08 standard deviations. Baseline surveys indicate small and statistically insignificant differences between treatment and control students' self-reported attitudes and behaviors.

In addition to comparisons at baseline, we are interested in whether information on outcomes later in time may be missing in such a way as to influence our results. While this is not generally testable, we first assess whether the fraction of students leaving their original school or rates of missing outcome data are different between treatment and control groups (Table 4). We can see that rates of attrition from the original school and rates of missing data for NY state test scores are quite low, around 10%, and are almost exactly the same for treatment and control. The latter likely reflects the fact that this test is required and used for high stakes decisions, as well as the fact that we can follow students and measure their test scores even if they move to another school within New York City. Rates of missing data for surveys are higher (around 25%), not surprising given 10% of the sample has moved and take-up was imperfect, but also very similar for both groups. However, for the NWEA test, rates of missing data are larger and the rate is significantly higher for the control group (55%) than the treatment group (33%). This reflects the fact that New Classrooms had some difficulty getting control school teachers to administer the NWEA exam in the spring of 2014. Thus, our results for the NWEA exam should be

interpreted with caution, given that they may not be representative of the full set of students in our analysis sample.

5. Descriptive Evidence on So1 Implementation from Student and Teacher Surveys

Before presenting treatment effect estimates, we first show descriptive evidence on students' and teachers' opinions regarding their math curriculum in each of the two expansion years. We do not have any baseline measures but we asked similar questions of all students and teachers in spring 2013 and spring 2014. The responses we received are quite informative for thinking about the implementation of School of One, as students and teachers are the main stakeholders and have a large amount of information from which to make their opinions.

Table 5 shows the average responses of students enrolled in So1 and not enrolled in So1 in both years, restricting the sample by dropping R002 since K014 was not surveyed.[10] Survey responses have been re-scaled from zero to one, where zero (one) indicates the least (most) agreement with phrases such as "I like the way math was taught this year." We can see that in spring 2013, students in So1 had generally lower opinions of their math curriculum on a host of items like, "I think I learned a lot in math this year" or "I like the way that math was taught this year." However, these differences shrink considerably in the spring 2014 survey, although average responses are still slightly lower for So1 students. Notably, the one question in which So1 students' responses were not noticeably worse in 2013 (and were slightly better in 2014) relative to non-So1 students was the item on the pace of math class being too easy, just right, or too hard. This may indicate some success of the So1 program in targeting lessons to the level of each individual student.

Anecdotal evidence from interviews with So1 teachers and principals suggested that 6th graders took to the So1 model more readily than the 8th graders, most of whom had already been using a more traditional model of math instruction at these middle schools for two years. Support for this notion can be

[10] Since this is merely a descriptive exercise, we show results with the sample split by actual receipt of So1; later we will use a more rigorous approach to estimating these differences using treatment assignment as an instrument for receipt of So1.

seen when we break out students' responses by grade level. In 2013, opinions are just slightly lower for 6th grade students in So1, but there are large gaps for 7th and (especially) 8th graders (Appendix Table 1a). In 2014, the differences in average opinions between students in So1 and those in traditional math programs are higher for 8th graders than younger students, but to a much smaller extent (Appendix Table 1b). We also check whether students' opinions on their math curriculum differed by their special education status (Appendix Table 2) but find hardly any differences along this dimension.

We exclude R002 to compare treatment and control in Table 5, but we are also interested in the survey responses of these students and teachers because we use R002 in our examination of NY State test scores. In Appendix Tables 3a and 3b, we show results for So1 and non-So1 students by school/academy. We can see clearly there that, with the exception of their opinions on pacing, students in R002 who were using So1 tend to be more heavily negative on their math curriculum, especially in spring 2013. While we have no survey data from K014, given the distribution of responses across the various treatment and control schools/academies, our belief is that responses for K014 students would not be equally low to those of R002 if we had been able to survey them.

In contrast, an examination of teacher surveys (Table 6) shows that So1 teachers (excluding R002) expressed more positive feelings about their math curriculum than control teachers, in both spring 2013 and 2014. This contrast between teachers' and students' views is interesting and may have something to do with the selection process, i.e. schools eligible for So1 had teachers who wanted the program, while students arrived to their school to find a very new and different math curriculum waiting for them. However, when we look by school (Appendix Tables 2a and 2b), we can see that teachers of So1 within R002 had far worse evaluations of their math curriculum than teachers using a traditional setup, and this difference shows up to a greater extent in 2013. While this may speak more to the specific circumstances of R002 than anything general about the So1 program, it is worth noting nonetheless.

In addition to the questions we asked of both treatment and control groups, we asked additional questions of students and teachers using So1. These answers may tell us something about which aspects of the program seemed to work well or not, but they are of course only suggestive. Since we have no

control group with which to compare, we simply show average responses by school. A few patterns in student surveys are consistent with what we have seen earlier (see Table 7). First, it is clear that student satisfaction with So1 grew from 2013 to 2014, with the exception of R002, where student opinions were quite low and remained low in both years. Second, students consistently had the most positive reaction to the statement "The School of One program told me how well I was doing in math," supporting the view that personalization/targeting was a distinguishing feature of the program. Third, students consistently felt that they learned more in School of One when working "directly with a teacher," followed by "working on a computer," and least when "working with other students." This ordering is similar in both years, but the gaps in opinion across these three types of lessons shrank considerably from 2013 to 2014.

For teachers of So1 outside of R002, evaluations are fairly high and similar in magnitude across the two years (Tables 8a/b).[11] In contrast, So1 teachers in R002 gave very low ratings on a number of dimensions, particularly those items dealing with So1 logistics (e.g. "Student noise levels are typically appropriate during instructional time"), and ratings were especially low in 2013. Teachers also indicate a belief that students learned most when working directly with teachers, although these patterns are not as stark as in student responses.

This descriptive evidence suggests a few takeaways.[12] Most clearly, students and teachers in R002 had a poor experience. Although this may have been driven by the hardships caused by Hurricane Sandy, the feelings expressed continued into the second year when the program was scaled down. In the remaining So1 expansion schools, survey responses regarding the math curriculum paint a picture which is decidedly mixed. Teachers were quite positive on the program but students (especially in the first year of the expansion) were noticeably less so. While these opinions are helping in providing context for our impact estimates, it is important to realize that these are only self-reports.

[11] In spring 2013, teachers in K381 were inadvertently administered the survey meant for control teachers which omitted questions specifically regarding So1. Thus, our statement regarding changes across years should be taken with some caution.

[12] We return to some of these survey responses below when examining treatment effects on student and teacher attitudes. Although we have no baseline measures, we can still test whether differences are statistically significant under the assumption of the RCT creating a balance between treatment and control groups.

6. Treatment Effect Estimates on Math Tests and Survey Measures

At the outset, it is important to note that selection into So1 treatment does not necessarily mean that a student received the So1 program. For example, we know that in R002 the rate of usage of So1 was substantially lower than in other treatment schools. This analysis is therefore based on a "treatment on the treated" framework, comparing outcomes of *all* students assigned to treatment with *all* students assigned to control, and then adjusted for the fact that usage of So1 was incomplete. Specifically, we first construct variables for the number of years a student used So1 (0, 1, or 2), where usage means presence in the New Classrooms So1 records before the end of October of each school year, and a student's "potential years of School of One"; the number of years the student would have used So1 if there had been perfect take-up of the assignment. This second variable, which is based solely on assignment to treatment in a particular year and grade will be used as an instrument for years of So1 in a two-stage least squares regression specification.[13]

We first turn to the New York State Math Test results. As noted above, we deal with the interdependence of outcomes among students within the same school or academy using an aggregated regression of school/academy means on treatment indicator (and other controls). To address the fact that take-up in R002 was much lower, we allow our first-stage regression coefficient on "potential years of So1" to differ for R002 students. Not surprisingly, the first stage coefficients are quite strong (Column 1 of Table 9). For each year of "potential" So1, students' actual years of So1 rise by 0.93, except in R002, where the coefficient is only 0.53 but still highly significant.

In results that do not control for prior test scores or demographics, years in So1 are associated with a rise of just under 0.10 standard deviations, but these estimates are not statistically significant. It is important to recall that treatment students already had higher math test scores at baseline. Adding controls causes the estimate to flip sign and go quite close to zero (-0.02 standard deviations in the full

[13] We use this approach to incorporate the fact that two cohorts—8th graders in 2012-13 and 6th graders in 2013-14—could only be exposed to So1 for one year, while the other two cohorts in our sample could be exposed for up to two years. However, results that simply look at any exposure to School of One (instrumented with assignment to treatment) are qualitatively quite similar.

sample, and less than -.01 in the sample that drops R002 and K014).[14] In other words, once we account for preexisting observable differences between the treatment and control groups, an additional year of So1 is expected to have almost no effect on NYS math test performance. However, the standard error on this estimate is roughly 0.10 standard deviations, so that we cannot rule out modest positive or negative effects of the program.

Turning to the NWEA exam scores, it is important to recall that we exclude two schools (R002 and K014) and that we are missing outcome measures for a significant fraction of the sample. Again, the first stage is quite strong (Table 10, Column 1), and results without controls show a positive but imprecisely estimated coefficient on years of So1 of 0.15 standard deviations, but the inclusion of controls for students' baseline NWEA test performance, baseline NY test performance, and demographics reduces this estimate to 0.06 (standard error of 0.06). This is suggestive of a potential modest positive effect on math achievement, but it is far too imprecise to draw any strong conclusions.[15]

We now turn to self-reported measures on attitudes /behaviors for students and teachers, taking the same two-stage least squares approach as above, and presenting results with and without controls for baseline survey response and demographics. For measures of student motivation and self-directed learning, the coefficients are small, statistically insignificant, and insensitive to the controls (Table 11). We find modest, negative, and marginally significant effects on students' self-reported interest and confidence in math, and for their knowledge and confidence in the use of computer technology. Given the personalization of So1 and its focus on instructional technology, a negative finding is unexpected. However, it is worth pointing out that, as these are self-reports, the impact of exposure to the So1

[14] To account for these student-level controls, we first regress outcome test scores on the control variables in a student-level regression with randomization unit – treatment fixed effects. We then take the residuals from this regression (including the fixed effects), average these at the randomization unit – treatment level, and run our aggregated regression with 8 (or 6) observations.

[15] In addition to this imprecision and issues of sample selection, another important caveat to an interpretation of positive effects is that the NWEA is a computer-based exam, and it is possible that an effect of the So1 program might come through increasing students' familiarity with computer-based testing. Without further information it is impossible to know the extent to which this explanation has merit, but it is worth mentioning given the frequent use of computer-based assessments in So1.

program may be to make students more aware of their own deficiencies in math and technological knowledge, resulting in lower self-reported confidence.[16]

Although we have discussed the descriptive evidence on students' end-of-year opinions on their math curriculum, we did not perform any tests to discern whether these were statistically significant. We do this separately for spring 2013 and spring 2014 surveys, instrumenting for receipt of So1 using the randomized assignment. Since we do not have a baseline response to use as a control variable, these comparisons are made under the assumption that, without the expansion of So1, average opinions of the math curriculum would be statistically indistinguishable in treatment and control groups.[17] These tests generally confirm the patterns discussed above; in spring 2013 we see students assigned to So1 reporting lower opinions of their math curriculum on a number of dimensions—notably excluding the pace of instruction—and these negative effects become small and statistically insignificant in spring 2014.

Treatment effect estimates of a year teaching So1 on teachers' self-efficacy and feelings towards the use of technology in the classroom are shown in Table 13. Point estimates for self-efficacy are negative but very imprecisely estimated. The estimate for use of technology is positive (0.17 standard deviations) and marginally significant (p-value = 0.13) without baseline controls, but controlling for baseline responses reduces the coefficient considerably and we cannot reject zero at standard significance levels (p-value = 0.21). To simplify our analysis of teachers' opinions on curriculum, we take the average response to these questions and normalize answers to have mean zero and standard deviation one using the control group distribution. We find large positive effects of teaching So1 in 2013 (0.84 standard deviations) which are statistically significant (p-value = 0.05); these effects get slightly smaller (0.74 standard deviations) but are still marginally significant (p-value = 0.11) in 2014.

[16] For an example of such spurious effects, see West et al. 2014 (http://cepr.harvard.edu/files/cepr/files/cepr-promise-paradox.pdf), who find that attending a highly successful charter school improves outcomes but makes students express lower levels of self-control, grit, and growth mindset due to changes in their peer reference group. Similarly, Linden et al., 2014 (http://www.leighlinden.com/Higher_Achievement.pdf) find lower academic attitudes but better achievement outcomes for a group of high achieving students selected into an intense after-school enrichment program, also likely to changes in their reference peer group.

[17] Again, because we compare treatment and control groups, both K014 and R002 are excluded.

7. Discussion and Conclusion

This evaluation provides an initial indication of the implementation and effects of the School of One program in the two years of the program expansion into four middle schools. The evaluation is based on a RCT design in which So1 was randomly assigned to particular schools. While RCT is considered by many to be the "gold standard" for program evaluation, in the case of school-wide expansion of a small developing program, a major drawback of this approach is a lack of precision. While this was anticipated at the stage of the evaluation proposal, given the difficulties in the implementation of program expansion, it may have been more illuminating had the evaluative design been more focused around issues of implementation. While our descriptive survey evidence and qualitative interviews serve to try to fill this gap in knowledge, this was not the main focus of the evaluation effort.

The broad takeaway from these findings is the So1 program had neither very large positive or large negative effects relative to the math instruction that students would have otherwise received in the program's first two years. The estimated effects for the New York State Mathematics Examination are very close to zero, while estimates for the NWEA mathematics exam are small and positive, but statistically insignificant. We do not see significant or consistent shifts in student or teacher attitudes and self-reported behaviors.

To provide some context for these impact estimates, we find students in older grades who were assigned to So1 provided worse opinions on their math curriculum in the expansion's first year, while So1 teachers (outside of R002) provided more positive views, relative to the control group. While we cannot compare student and teacher surveys in R002 to those from a control school, due non-compliance of K014, opinions of the So1 curriculum were quite low. However, this school faced extreme circumstances due to displacement for several months after Hurricane Sandy landed in the fall of 2012, and it is hard to know what this school's experience might tell us about the quality of So1 implementation generally.

The quantitative evidence from this evaluation provides a somewhat mixed picture on the impact of So1. Broadly speaking, the imprecision of the estimates and the lack of consistency across outcomes

means it is still difficult to know whether the program was an improvement over the traditional math curricula offered in control schools/academies. Whether these results are judged as a success or a failure depends on one's prior belief on the likely short-run effectiveness of So1 expansion, particularly given the logistical hurdles in implementing a radically new math instruction program. The continued monitoring of existing So1 programs, which may now have overcome these hurdles, may ultimately determine whether the program lives up to its aspirations of radically transforming and improving math instruction for all students.

Table 1: Student Counts in Analysis Data

Randomization Group	School Code	Number of Students	
		Treatment	Control
1	R002	1267	
	K014		718
2	K381	558	
	X368		642
3	R049	394	463
4	K088	452	576
Total		2,671	2,399

Note: Sample includes all students enrolled in schools involved in randomization during school year 2012-13 as well as first time 6th graders newly enrolled in these schools in school year 2013-14. Randomization group denotes schools that were paired for random selection to the School of One (So1) program treatment or a school where one of two eligible academies was randomly selected for the So1 program.

Table 2: Baseline Comparisons, Full Sample

Characteristic	Treatment	Control	Difference (T-C)	P-value
English Language Learner	5.9%	21.7%	-15.8%	**0.00**
Free Lunch	83.6%	88.7%	-5.2%	0.39
Special Education	21.6%	26.9%	-5.4%	0.30
% Days Absent	5.6%	6.1%	-0.5%	0.61
Has Prior NY ELA Score	93.0%	88.6%	4.5%	**0.02**
Prior NY ELA Score	0.04	-0.29	0.33	**0.03**
Adjusted Prior NY ELA Score	0.07	-0.08	0.15	0.16
Has Prior NY Math Score	93.6%	91.0%	2.5%	**0.09**
Prior NY Math Score	-0.03	-0.26	0.23	0.15
Adjusted Prior NY Math Score	0.05	-0.06	0.11	0.35

Note: Treatment (So1) and Control columns show the mean value for each student chacteristic for students in schools/academies based on random selection for the So1 program. So1 treatment assignment does not necessarily mean the student received the So1 program. P-value on the difference between treatment and control is based on a regression of data collapsed to the school/academy level, weighted by the number of students. ELA stands for English language arts. Adjusted ELA and Math scores are the residuals from a regression of prior scores on prior demographics that includes randomization block by treatment assignment fixed effects. The fixed effects are included in the residual. Bolded cells denote P-values less than 0.10.

Table 3: Baseline Comparisons, NWEA/Survey Subsample

Characteristic	Treatment	Control	Difference (T-C)	P-value
Prior Year Characteristics				
English Lang. Learner	4.0%	23.4%	-19.3%	**0.00**
Free Lunch	92.0%	87.7%	4.2%	0.19
Special Education	21.5%	27.3%	-5.8%	0.48
% Days Absent	5.3%	5.8%	-0.5%	0.75
Baseline Test Scores				
Has Prior NY ELA Score	92.9%	88.5%	4.4%	0.12
Prior NY ELA Score	-0.07	-0.26	0.19	0.25
Adjusted Prior NY ELA Score	-0.04	-0.05	0.01	0.95
Has Prior NY Math Score	93.2%	90.8%	2.5%	0.26
Prior NY Math Score	-0.11	-0.25	0.14	0.54
Adjusted Prior NY Math Score	-0.04	-0.05	0.00	0.99
Has Baseline NWEA Score	85%	64%	21%	**0.08**
Baseline NWEA Score	-0.34	-0.56	0.22	0.24
Adjusted Baseline NWEA Score	0.02	-0.06	0.08	0.50
Baseline Survey Measures				
Intrinsic Motivation	0.11	0.10	0.01	0.85
External Motivation	-0.03	-0.09	0.06	0.79
Amotivation	0.03	-0.03	0.06	0.46
Math Interest/Competence	0.10	0.10	0.00	0.69
Self-Directed Learner	0.09	-0.01	0.10	0.27
Technology Confidence/Knowledge	0.04	-0.05	0.09	0.53

Note: Sample is limited to randomization blocks 2 through 4, where NWEA tests and surveys were administered in both treatment and control schools/academies. Treatment (So1) and Control columns show the mean value for each student chacteristic for students in schools/academies based on selection for the So1 program. P-value on the difference between treatment and control is based on a regression of data collapsed to the school/academy level, weighted by the number of students. ELA stands for English language arts. Adjusted ELA and Math scores are the residuals from a regression of prior scores on prior demographics that includes randomization block by treatment assignement fixed effects. The fixed effects are included in the residual. Bolded cells denote P-values less than 0.10. So1 treatment selection does not necessarily mean the student received the So1 program.

Table 4: Missing Outcomes by Student Treatment Assignment

		Missing Outcome Measure		
	Left RCT School	NY Test	NWEA Test	Survey
	(1)	(2)	(3)	(4)
Assigned to School of One	-0.02	0.01	-0.22	0.02
	(0.02)	(0.02)	(0.10)+	(0.03)
Constant (Control Mean)	0.10	0.10	0.55	0.25
	(0.02)**	(0.01)**	(0.06)**	(0.02)**
Number of Students	5,070	5,070	3,085	3,085
Number of Schools/Academies	8	8	6	6

Note: Each column presents a coefficient estimate (and standard error, in parentheses) for assignment to School of One from a separate regression using data collapsed to the treatment unit level (i.e. school or academy). Dependent variable in column 1 is an indicator for leaving the assigned RCT school prior to the period at which treatment is measured; dependent variables in columns 2-4 are indicators for missing outcome measures. So1 treatment assignment does not necessarily mean the student received the So1 program.

Table 5: End of Year Survey Questions on Math Class, Excluding R002 and K014

	Spring 2013		Spring 2014	
Survey Question:	Control	Treatment	Control	Treatment
I think I learned a lot in math this year	0.79	0.63	0.77	0.72
(0: Least Agreement; 1: Most Agreement)	(0.27)	(0.34)	(0.24)	(0.25)
	[996]	[793]	[951]	[715]
I liked the way that math was taught this year	0.72	0.50	0.70	0.61
(0: Least Agreement; 1: Most Agreement)	(0.31)	(0.38)	(0.28)	(0.30)
	[995]	[797]	[948]	[715]
I understood what the math teacher(s) wanted me to do in class this year	0.74	0.62	0.73	0.69
(0: Least Agreement; 1: Most Agreement)	(0.28)	(0.32)	(0.23)	(0.23)
	[1,002]	[793]	[948]	[717]
I thought that all my needs were met in math class this year	0.65	0.55	0.66	0.62
(0: Least Agreement; 1: Most Agreement)	(0.30)	(0.32)	(0.25)	(0.24)
	[994]	[782]	[945]	[714]
Math class this year was…	0.81	0.79	0.80	0.82
(0: way too easy/hard; 1: just right)	(0.29)	(0.31)	(0.30)	(0.28)
	[986]	[769]	[897]	[685]

Note: Sample excludes students in K014 (who did not take surveys) and R002 (its matched pair), as well as students who switched out of their original schools. Responses are divided based on actual receipt of School of One. Response scales to questions about learning in math class differed across years (4 point scale in 2013 from "Not at al true" to "Very True" and 5 point scale in 2014 from "Strongly Disagree" to "Strongly Agree") and therefore responses have been rescaled from 0 to 1. For the pace of math class, students in both years provided a response on a 5 point scale ranging from "Way too Easy" to "Way too Hard" with "Just Right" as the middle of the scale. Answers here are normalized so that "Just Right" is 1 and either "Way too" response is equal to 0. Standard deviations are shown in paretheses and number of student responses are shown in brackets.

Table 6: End of Year Teacher Survey on Math Curriculum, Excluding R002 and K014

	Spring 2013		Spring 2014	
Survey Question:	Control	Treatment	Control	Treatment
I liked the math curriculum that I used this year	0.67	0.83	0.61	0.77
	(0.18)	(0.19)	(0.22)	(0.07)
	[22]	[13]	[23]	[12]
The students' learned a lot in math this year	0.76	0.81	0.71	0.77
	(0.16)	(0.18)	(0.16)	(0.17)
	[22]	[13]	[23]	[12]
Most of the students' needs were met in math class this year	0.66	0.75	0.64	0.71
	(0.16)	(0.18)	(0.24)	(0.14)
	[22]	[13]	[23]	[12]
Math teaching materials are effective in helping students…				
...improve in math	0.69	0.83	0.65	0.71
	(0.13)	(0.19)	(0.16)	(0.14)
	[22]	[13]	[23]	[12]
...on the statewide test	0.75	0.73	0.63	0.69
	(0.15)	(0.16)	(0.22)	(0.16)
	[22]	[13]	[23]	[12]
Students were aware of their learning goals			0.78	0.79
			(0.11)	(0.18)
			[23]	[12]
Students who started <u>below</u> grade level			0.58	0.77
were well served by the math curriculum			(0.22)	(0.17)
			[20]	[12]
Students who started <u>above</u> grade level			0.68	0.77
were well served by the math curriculum			(0.18)	(0.20)
			[22]	[12]

Note: Sample excludes teachers in K014 (who did not take surveys) and R002 (its matched pair) as well as teachers who entered the schools after the school year 2012-2013. Responses are divided based on actual work in the School of One program. Response scales have been rescaled from 0 (least agreement) to 1 (most agreement). Standard deviations are shown in paretheses and number of teacher responses are shown in brackets.

Table 7: End of Year Survey Questions on So1 Program, by School

Survey Question:	Spring 2013				Spring 2014			
	K088	K381	R002	R049	K088	K381	R002	R049
I liked using So1 to learn math	0.56	0.50	0.36	0.46	0.70	0.73	0.36	0.54
	(0.41)	(0.40)	(0.40)	(0.38)	(0.32)	(0.31)	(0.39)	(0.35)
	[239]	[328]	[557]	[203]	[257]	[279]	[197]	[167]
School of One is better than the math classes I have had in the past	0.52	0.45	0.30	0.40	0.68	0.66	0.30	0.47
	(0.43)	(0.40)	(0.38)	(0.40)	(0.33)	(0.35)	(0.36)	(0.36)
	[238]	[329]	[557]	[205]	[257]	[278]	[198]	[167]
I thought that the playlist was useful	0.54	0.49	0.35	0.48	0.61	0.66	0.35	0.56
	(0.34)	(0.33)	(0.34)	(0.34)	(0.29)	(0.27)	(0.30)	(0.29)
	[239]	[328]	[554]	[206]	[256]	[281]	[195]	[167]
The School of One program told me how well I was doing in math	0.67	0.58	0.45	0.59	0.75	0.74	0.46	0.62
	(0.32)	(0.33)	(0.36)	(0.34)	(0.26)	(0.26)	(0.34)	(0.30)
	[241]	[330]	[558]	[207]	[257]	[281]	[197]	[167]
During So1 I usually learned a lot when working...								
... with other students	0.52	0.43	0.29	0.44	0.59	0.58	0.35	0.52
	(0.32)	(0.33)	(0.30)	(0.34)	(0.28)	(0.27)	(0.30)	(0.27)
	[242]	[330]	[555]	[203]	[257]	[281]	[197]	[167]
... directly with a teacher	0.70	0.58	0.52	0.66	0.73	0.64	0.51	0.66
	(0.29)	(0.33)	(0.34)	(0.33)	(0.23)	(0.28)	(0.30)	(0.26)
	[242]	[329]	[555]	[203]	[257]	[280]	[196]	[167]
... on a computer	0.54	0.56	0.41	0.62	0.64	0.71	0.35	0.62
	(0.36)	(0.35)	(0.37)	(0.35)	(0.28)	(0.28)	(0.36)	(0.30)
	[241]	[328]	[555]	[201]	[256]	[281]	[197]	[166]

Note: Sample excludes students who switched out of their original schools or from their original assignment to School of One. Response scales to questions about School of One differed across years (4 point scale in 2013 from "Not at al true" to "Very True" and 5 point scale in 2014 from "Strongly Disagree" to "Strongly Agree") and therefore responses have been rescaled from 0 (least agreement) to 1 (most agreement). Standard deviations are shown in paretheses and number of student responses are shown in brackets.

Table 8a: Spring 2013 Teacher Survey on So1 Program, by School

Survey Question:	K088	K381	R002	R049
School of One is better than other math curricula I have used in the past	0.81		0.14	0.83
	(0.24)		(0.20)	(0.14)
School of One was straightforward to teach	0.75		0.21	0.75
	(0.00)		(0.25)	(0.00)
We use School of One common planning time effectively	1.00		0.50	0.92
	(0.00)		(0.29)	(0.14)
Student transitions during School of One are typically smooth and efficient	0.81		0.00	0.75
	(0.13)		(0.00)	(0.00)
Student noise levels are typically appropriate during instructional time	0.69		0.00	0.50
	(0.13)		(0.00)	(0.25)
Students are typically quiet while taking exit slips	0.75		0.00	0.75
	(0.00)		(0.00)	(0.00)
When teachers or TRs were absent, my school provided adequate coverage	0.31		0.04	0.92
	(0.13)		(0.09)	(0.14)
So1 enables me to accelerate student learning more than a traditional model	0.75		0.14	0.92
	(0.20)		(0.20)	(0.14)
Students have a positive So1 experience working together collaboratively	0.75		0.11	0.67
	(0.00)		(0.13)	(0.14)
Students have a positive So1 experience working directly with a teacher	0.81		0.43	1.00
	(0.13)		(0.24)	(0.00)
Students have a positive So1 experience working on the computer	0.63		0.18	0.75
	(0.14)		(0.12)	(0.00)
I would participate as a School of One teacher again next year	0.94		0.43	1.00
	(0.13)		(0.31)	(0.00)
Number of Teachers	4	0	7	3

Note: Sample excludes teachers in K381, who were inadvertently not given these questions in spring 2013. Responses have been rescaled from 0 (least agreement) to 1 (most agreement). Standard deviations are shown in paretheses.

Table 8b: Spring 2014 Teacher Survey on So1 Program, by School

Survey Question:	K088	K381	R002	R049
School of One is better than other math curricula I have used in the past	0.88	0.85	0.33	0.75
	(0.14)	(0.22)	(0.14)	(0.00)
School of One was straightforward to teach	0.69	0.80	0.42	0.75
	(0.31)	(0.33)	(0.29)	(0.00)
We use School of One common planning time effectively	0.81	0.65	0.67	0.83
	(0.38)	(0.38)	(0.38)	(0.14)
Student transitions during School of One are typically smooth and efficient	0.94	0.65	0.50	0.75
	(0.13)	(0.22)	(0.43)	(0.25)
Student noise levels are typically appropriate during instructional time	0.88	0.55	0.33	0.67
	(0.14)	(0.33)	(0.14)	(0.14)
Students are typically quiet while taking exit slips	0.94	0.60	0.25	0.67
	(0.13)	(0.29)	(0.00)	(0.14)
When teachers or TRs were absent, my school provided adequate coverage	0.19	0.55	0.25	0.75
	(0.24)	(0.33)	(0.43)	(0.00)
So1 enables me to accelerate student learning more than a traditional model	0.75	0.75	0.33	0.75
	(0.00)	(0.18)	(0.14)	(0.25)
Students have a positive So1 experience working together collaboratively	0.75	0.80	0.25	0.75
	(0.00)	(0.21)	(0.00)	(0.00)
Students have a positive So1 experience working directly with a teacher	0.75	0.85	0.42	0.83
	(0.00)	(0.14)	(0.14)	(0.14)
Students have a positive So1 experience working on the computer	0.75	0.85	0.17	0.75
	(0.00)	(0.14)	(0.14)	(0.00)
I would participate as a School of One teacher again next year	0.81	0.85	0.25	0.83
	(0.13)	(0.22)	(0.25)	(0.14)
I would recommend School of One to a fellow math teacher	0.81	0.85	0.33	0.83
	(0.13)	(0.22)	(0.14)	(0.14)
Number of Teachers	4	5	3	3

Note: Responses have been rescaled from 0 (least agreement) to 1 (most agreement). Standard deviations are shown in paretheses.

Table 9: Treatment Estimates for NY State Test

	Full Sample		
	Years in So1	NY Math Score	
	(1)	(2)	(3)
Potential Years of So1	0.93		
	(0.02)**		
Potential Years * R002	-0.40		
	(0.02)**		
Years in School of One		0.08	-0.02
		(0.14)	(0.09)
Baseline Controls			√
Number of Students	4,520	4,520	4,520
Number of Schools/Academies	8	8	8
	Excluding R002/K014		
	Years in So1	NY Math Score	
	(1)	(2)	(3)
Potential Years of So1	0.93		
	(0.03)**		
Years in School of One		0.09	-0.00
		(0.18)	(0.11)
Baseline Controls			√
Number of Students	2,755	2,755	2,755
Number of Schools/Academies	6	6	6

Note: Each column presents a treatment coefficient estimate (and standard error, in parentheses) from a separate regression using data collapsed to the treatment unit level (i.e. school or academy). Randomization block controls are fixed effects; student characteristics include demographics, prior year classifications (e.g. English language learner, special education), and prior year absence rate; prior year test controls including a cubic in prior math and English test scores, an indicator for missing prior English score, and an interaction of this indicator with a cubic in prior math scores. So1 treatment selection does not necessarily mean the student received the So1 program.

Table 10: Treatment Estimates for NWEA Test

	Years in So1	NWEA Score	
	(1)	(2)	(3)
Potential Years of So1	0.98		
	(0.01)**		
Years in School of One		0.15	0.06
		(0.12)	(0.06)
Baseline Controls			√
Number of Students	1,711	1,711	1,711
Number of Schools/Academies	6	6	6

Note: Each column presents a treatment coefficient estimate (and standard error, in parentheses) from a separate regression using data collapsed to the treatment unit level (i.e. school or academy). Randomization block controls are fixed effects; student characteristics include demographics, prior year classifications (e.g. English language learner, special education), and prior year absence rate; prior year test controls including a cubic in prior math and English test scores, an indicator for missing prior English score, and an interaction of this indicator with a cubic in prior math scores. So1 treatment selection does not necessarily mean the student received the So1 program.

Table 11: Treatment Estimates for Student Survey Responses on Self-Reported Attitudes/Behaviors

	Years in So1	Intrinsic Motivation		External Motivation		Amotivation (Reverse Scored)		Math Interest/ Competence		Self-Directed Learner		Tech. Confidence/ Knowledge	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)
Potential Years of So1	0.92												
	(0.03)**												
Years in School of One		0.04	0.07	0.05	0.05	-0.02	-0.04	-0.13	-0.12	0.05	0.05	-0.04	-0.08
		(0.06)	(0.05)	(0.04)	(0.03)	(0.01)	(0.03)	(0.05)+	(0.06)	(0.04)	(0.03)	(0.01)+	(0.02)**
Baseline Controls			√		√		√		√		√		√
Number of Students	2,936	2,180	2,180	2,216	2,216	2,207	2,207	2,192	2,192	2,223	2,223	2,189	2,189
Number of Schools/Academies	6	6	6	6	6	6	6	6	6	6	6	6	6

Note: Each column presents a treatment coefficient estimate (and standard error, in parentheses) from a separate regression using data collapsed to the treatment unit level (i.e. school or academy). Randomization block controls are fixed effects; student characteristics include demographics, prior year classifications (e.g. English language learner, special education), and prior year absence rate; prior year test controls including a cubic in prior math and English test scores, an indicator for missing prior English score, and an interaction of this indicator with a cubic in prior math scores. So1 treatment selection does not necessarily mean the student received the So1 program.

Table 12: Treatment Estimates for Students' End-of-Year Survey Responses on their Math Class

	Panel A: Spring 2013			Panel B: Spring 2014		
Dependent variable:	Assigned to So1	Received School of One	Number of Students	Assigned to So1	Received School of One	Number of Students
Received School of One	0.99 (0.00)**		1,791	0.93 (0.06)**		1,681
I learned a lot in math this year		-0.17 (0.04)*	1,772		-0.05 (0.04)	1,673
I liked the way that math was taught this year		-0.23 (0.05)**	1,775		-0.09 (0.06)	1,670
I understood what the math teacher(s) wanted me to do in class this year		-0.11 (0.04)+	1,778		-0.04 (0.03)	1,672
All my needs were met in math class this year		-0.12 (0.03)*	1,761		-0.03 (0.02)	1,666
Pace of math class this year		-0.03 (0.04)	1,739		0.02 (0.03)	1,588

Note: Each row within each panel presents a coefficient estimate (and standard error, in parentheses) from a separate regression using data collapsed to the randomization unit level (i.e. school or academy). Neither R002 or K014 are included in this analysis. The coefficient on "Assigned to So1" is based on an OLS regression of So1 receipt on the randomized assignment, while the coefficients on "Received School of One" are two stage least squares estimate, where receipt of So1 is instrumented with randomized assignment. Results for "Assigned to So1" are based on the set of students who answered any of the six survey questions analyzed in the table during each of the spring surveys.

Table 13: Treatment Estimates for Teacher Survey Responses

	Latest Teacher Observation					2013		2014	
	Years in So1	Teacher Self-Efficacy		Use of Technology		Teaching So1	Eval. of Curriculum	Teaching So1	Eval. of Curriculum
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Potential Years Teaching So1	0.96								
	(0.04)**								
Years Teaching School of One		-0.17	-0.19	0.17	0.13				
		(0.20)	(0.17)	(0.09)	(0.09)				
Assigned to So1						1.00		0.92	
						(0.00)		(0.07)**	
Teaching School of One							0.84		0.74
							(0.30)*		(0.37)
Baseline Controls			√		√				
Number of Teachers	36	36	36	36	36	35	35	35	35
Number of Schools/Academies	6	6	6	6	6	6	6	6	6

Note: Each column presents a treatment coefficient estimate (and standard error, in parentheses) from a separate regression using data collapsed to the treatment unit level (i.e. school or academy). For regressions with baseline controls, the dependent variable is constructed from a teacher-level regression of the most recent survey response on baseline survey response and fixed effects for school/academy; we use the residual (inclusive of fixed effect) from this regression to aggregate to the school/academy level and estimate a second stage using weighted least squares, where weights are the number of teacher respondents.

Appendix Table 1a: End of Year Survey Questions on Math Class, by Grade Level

	Spring 2013					
	Grade 6		Grade 7		Grade 8	
Survey Question:	Non-So1	So1	Non-So1	So1	Non-So1	So1
I think I learned a lot in math this year	0.79	0.74	0.80	0.60	0.79	0.54
(0: Least Agreement; 1: Most Agreement)	(0.27)	(0.31)	(0.27)	(0.34)	(0.28)	(0.35)
	[322]	[268]	[350]	[269]	[324]	[256]
I liked the way that math was taught this year	0.71	0.64	0.72	0.49	0.72	0.35
(0: Least Agreement; 1: Most Agreement)	(0.31)	(0.36)	(0.31)	(0.37)	(0.31)	(0.35)
	[324]	[269]	[347]	[271]	[324]	[257]
I understood what the math teacher(s) wanted	0.74	0.71	0.74	0.64	0.73	0.52
me to do in class this year	(0.28)	(0.31)	(0.27)	(0.31)	(0.29)	(0.31)
(0: Least Agreement; 1: Most Agreement)	[326]	[267]	[349]	[270]	[327]	[256]
I thought that all my needs were met in math class this year	0.65	0.64	0.68	0.53	0.63	0.46
(0: Least Agreement; 1: Most Agreement)	(0.29)	(0.30)	(0.28)	(0.31)	(0.31)	(0.33)
	[325]	[265]	[346]	[265]	[323]	[252]
Math class this year was…	0.84	0.83	0.82	0.78	0.78	0.75
(0: way too easy/hard; 1: just right)	(0.28)	(0.29)	(0.29)	(0.32)	(0.30)	(0.32)
	[322]	[263]	[345]	[263]	[320]	[243]

Note: Sample excludes students in K014 (who did not take surveys) and R002 (its matched pair), as well as students who switched out of their original schools. Responses are divided based on receipt of School of One. Response scales to questions about learning in math class differed across years (4 point scale in 2013 from "Not at al true" to "Very True" and 5 point scale in 2014 from "Strongly Disagree" to "Strongly Agree") and therefore responses have been rescaled from 0 to 1. For the pace of math class, students in both years provided a response on a 5 point scale ranging from "Way too Easy" to "Way too Hard" with "Just Right" as the middle of the scale. Answers here are normalized so that "Just Right" is 1 and either "Way too" response is equal to 0. Standard deviations are shown in paretheses and number of student responses are shown in brackets.

Appendix Table 1b: End of Year Survey Questions on Math Class, by Grade Level

	Spring 2014					
	Grade 6		Grade 7		Grade 8	
Survey Question:	Non-So1	So1	Non-So1	So1	Non-So1	So1
I think I learned a lot in math this year	0.83	0.78	0.73	0.70	0.76	0.67
(0: Least Agreement; 1: Most Agreement)	(0.21)	(0.22)	(0.25)	(0.27)	(0.25)	(0.25)
	[312]	[270]	[295]	[253]	[344]	[192]
I liked the way that math was taught this year	0.75	0.70	0.66	0.61	0.68	0.50
(0: Least Agreement; 1: Most Agreement)	(0.25)	(0.27)	(0.29)	(0.30)	(0.28)	(0.30)
	[311]	[269]	[294]	[254]	[343]	[192]
I understood what the math teacher(s) wanted	0.76	0.72	0.71	0.71	0.72	0.62
me to do in class this year	(0.21)	(0.21)	(0.24)	(0.23)	(0.24)	(0.25)
(0: Least Agreement; 1: Most Agreement)	[311]	[270]	[293]	[255]	[344]	[192]
I thought that all my needs were met in math class this year	0.69	0.65	0.63	0.64	0.66	0.56
(0: Least Agreement; 1: Most Agreement)	(0.25)	(0.23)	(0.23)	(0.24)	(0.25)	(0.24)
	[310]	[269]	[292]	[254]	[343]	[191]
Math class this year was…	0.82	0.85	0.76	0.80	0.81	0.78
(0: way too easy/hard; 1: just right)	(0.30)	(0.26)	(0.31)	(0.27)	(0.29)	(0.31)
	[296]	[255]	[278]	[247]	[323]	[183]

Note: Sample excludes students in K014 (who did not take surveys) and R002 (its matched pair), as well as students who switched out of their original schools. Responses are divided based on receipt of School of One. Response scales to questions about learning in math class differed across years (4 point scale in 2013 from "Not at al true" to "Very True" and 5 point scale in 2014 from "Strongly Disagree" to "Strongly Agree") and therefore responses have been rescaled from 0 to 1. For the pace of math class, students in both years provided a response on a 5 point scale ranging from "Way too Easy" to "Way too Hard" with "Just Right" as the middle of the scale. Answers here are normalized so that "Just Right" is 1 and either "Way too" response is equal to 0. Standard deviations are shown in paretheses and number of student responses are shown in brackets.

Appendix Table 2: End of Year Survey Questions on Math Class, by Special Education Status

	Spring 2013				Spring 2014			
	General Ed.		Special Ed.		General Ed.		Special Ed.	
Survey Question:	Non-So1	So1	Non-So1	So1	Non-So1	So1	Non-So1	So1
I think I learned a lot in math this year	0.80	0.62	0.77	0.63	0.78	0.72	0.77	0.72
(0: Least Agreement; 1: Most Agreement)	(0.27)	(0.35)	(0.29)	(0.34)	(0.24)	(0.25)	(0.24)	(0.24)
	[743]	[622]	[228]	[169]	[688]	[575]	[262]	[138]
I liked the way that math was taught this year	0.72	0.48	0.71	0.53	0.70	0.60	0.69	0.66
(0: Least Agreement; 1: Most Agreement)	(0.31)	(0.38)	(0.32)	(0.36)	(0.28)	(0.30)	(0.27)	(0.29)
	[742]	[623]	[227]	[172]	[686]	[577]	[261]	[136]
I understood what the math teacher(s) wanted	0.74	0.62	0.72	0.63	0.74	0.69	0.71	0.70
me to do in class this year	(0.27)	(0.32)	(0.30)	(0.32)	(0.23)	(0.23)	(0.24)	(0.24)
(0: Least Agreement; 1: Most Agreement)	[747]	[621]	[229]	[170]	[688]	[580]	[259]	[135]
I thought that all my needs were met in math class this year	0.67	0.54	0.61	0.55	0.67	0.62	0.65	0.61
(0: Least Agreement; 1: Most Agreement)	(0.29)	(0.32)	(0.30)	(0.31)	(0.24)	(0.24)	(0.25)	(0.24)
	[741]	[613]	[229]	[167]	[684]	[576]	[260]	[136]
Math class this year was…	0.82	0.80	0.80	0.75	0.81	0.82	0.77	0.81
(0: way too easy/hard; 1: just right)	(0.28)	(0.30)	(0.31)	(0.34)	(0.29)	(0.28)	(0.33)	(0.29)
	[738]	[602]	[225]	[165]	[649]	[556]	[247]	[127]

Note: Sample excludes students in K014 (who did not take surveys) and R002 (its matched pair), as well as students who switched out of their original schools. Responses are divided based on receipt of School of One. Response scales to questions about learning in math class differed across years (4 point scale in 2013 from "Not at al true" to "Very True" and 5 point scale in 2014 from "Strongly Disagree" to "Strongly Agree") and therefore responses have been rescaled from 0 to 1. For the pace of math class, students in both years provided a response on a 5 point scale ranging from "Way too Easy" to "Way too Hard" with "Just Right" as the middle of the scale. Answers here are normalized so that "Just Right" is 1 and either "Way too" response is equal to 0. Standard deviations are shown in paretheses and number of student responses are shown in brackets.

Appendix Table 3a: End of Year Survey Questions on Math Class, by School

	Spring 2013							
	K088		X368	K381	R002		R049	
Survey Question:	Non-So1	So1	Non-So1	So1	Non-So1	So1	Non-So1	So1
I think I learned a lot in math this year	0.83	0.71	0.77	0.59	0.81	0.44	0.78	0.59
(0: Least Agreement; 1: Most Agreement)	(0.25)	(0.33)	(0.28)	(0.35)	(0.28)	(0.36)	(0.28)	(0.34)
	[327]	[249]	[413]	[324]	[286]	[576]	[250]	[220]
I liked the way that math was taught this year	0.76	0.57	0.69	0.47	0.68	0.33	0.70	0.45
(0: Least Agreement; 1: Most Agreement)	(0.30)	(0.39)	(0.32)	(0.37)	(0.32)	(0.37)	(0.32)	(0.36)
	[326]	[249]	[413]	[324]	[288]	[578]	[250]	[224]
I understood what the math teacher(s) wanted	0.78	0.70	0.71	0.57	0.75	0.45	0.73	0.61
me to do in class this year	(0.26)	(0.28)	(0.29)	(0.32)	(0.26)	(0.33)	(0.28)	(0.33)
(0: Least Agreement; 1: Most Agreement)	[328]	[250]	[414]	[321]	[288]	[575]	[254]	[222]
I thought that all my needs were met in math class this year	0.69	0.59	0.61	0.54	0.69	0.38	0.67	0.51
(0: Least Agreement; 1: Most Agreement)	(0.28)	(0.31)	(0.30)	(0.32)	(0.29)	(0.31)	(0.30)	(0.33)
	[324]	[243]	[413]	[319]	[287]	[578]	[252]	[220]
Math class this year was…	0.84	0.85	0.80	0.80	0.74	0.69	0.80	0.71
(0: way too easy/hard; 1: just right)	(0.27)	(0.26)	(0.30)	(0.32)	(0.30)	(0.36)	(0.31)	(0.34)
	[328]	[243]	[415]	[312]	[278]	[558]	[238]	[214]

Note: Sample excludes students who switched out of their original schools. Responses are divided based on receipt of School of One. Response scales to questions about learning in math class differed across years (4 point scale in 2013 from "Not at al true" to "Very True" and 5 point scale in 2014 from "Strongly Disagree" to "Strongly Agree") and therefore responses have been rescaled from 0 to 1. For the pace of math class, students in both years provided a response on a 5 point scale ranging from "Way too Easy" to "Way too Hard" with "Just Right" as the middle of the scale. Answers here are normalized so that "Just Right" is 1 and either "Way too" response is equal to 0. Standard deviations are shown in paretheses and number of student responses are shown in brackets.

Appendix Table 3b: End of Year Survey Questions on Math Class, by School

	Spring 2014							
	K088		X368	K381	R002		R049	
Survey Question:	Non-So1	So1	Non-So1	So1	Non-So1	So1	Non-So1	So1
I think I learned a lot in math this year	0.78	0.80	0.74	0.68	0.63	0.45	0.81	0.69
(0: Least Agreement; 1: Most Agreement)	(0.24)	(0.23)	(0.27)	(0.26)	(0.31)	(0.33)	(0.21)	(0.26)
	[342]	[264]	[331]	[280]	[526]	[209]	[278]	[171]
I liked the way that math was taught this year	0.69	0.71	0.67	0.59	0.55	0.33	0.74	0.50
(0: Least Agreement; 1: Most Agreement)	(0.28)	(0.28)	(0.29)	(0.29)	(0.32)	(0.33)	(0.26)	(0.30)
	[340]	[267]	[330]	[278]	[524]	[209]	[278]	[170]
I understood what the math teacher(s) wanted	0.73	0.74	0.71	0.66	0.63	0.46	0.75	0.67
me to do in class this year	(0.24)	(0.22)	(0.24)	(0.24)	(0.27)	(0.28)	(0.22)	(0.22)
(0: Least Agreement; 1: Most Agreement)	[341]	[265]	[331]	[280]	[525]	[208]	[276]	[172]
I thought that all my needs were met in math class this year	0.67	0.66	0.64	0.61	0.57	0.41	0.67	0.59
(0: Least Agreement; 1: Most Agreement)	(0.23)	(0.24)	(0.26)	(0.24)	(0.27)	(0.27)	(0.25)	(0.24)
	[341]	[267]	[328]	[276]	[525]	[206]	[276]	[171]
Math class this year was…	0.81	0.86	0.78	0.81	0.70	0.61	0.81	0.75
(0: way too easy/hard; 1: just right)	(0.29)	(0.24)	(0.32)	(0.28)	(0.32)	(0.38)	(0.30)	(0.31)
	[328]	[252]	[302]	[270]	[516]	[197]	[267]	[163]

Note: Sample excludes students who switched out of their original schools. Responses are divided based on receipt of School of One. Response scales to questions about learning in math class differed across years (4 point scale in 2013 from "Not at al true" to "Very True" and 5 point scale in 2014 from "Strongly Disagree" to "Strongly Agree") and therefore responses have been rescaled from 0 to 1. For the pace of math class, students in both years provided a response on a 5 point scale ranging from "Way too Easy" to "Way too Hard" with "Just Right" as the middle of the scale. Answers here are normalized so that "Just Right" is 1 and either "Way too" response is equal to 0. Standard deviations are shown in paretheses and number of student responses are shown in brackets.

Appendix Table 4a: Spring 2013 Teacher Survey on Math Curriculum, by School

	Spring 2013							
	K088		X368	K381	R002		R049	
Survey Question:	Non-So1	So1	Non-So1	So1	Non-So1	So1	Non-So1	So1
I liked the math curriculum that I used this year	0.69	0.75	0.63	0.79	0.56	0.21	0.68	1.00
	(0.21)	(0.00)	(0.14)	(0.25)	(0.43)	(0.27)	(0.19)	(0.00)
The students' learned a lot in math this year	0.83	0.81	0.75	0.71	0.44	0.11	0.68	1.00
	(0.22)	(0.13)	(0.00)	(0.19)	(0.52)	(0.20)	(0.12)	(0.00)
Most of the students' needs were met in math class this year	0.67	0.88	0.63	0.67	0.44	0.04	0.68	0.75
	(0.18)	(0.14)	(0.21)	(0.20)	(0.52)	(0.09)	(0.12)	(0.00)
Math teaching materials are effective in helping students…								
...improve in math	0.69	0.81	0.67	0.83	0.63	0.11	0.71	0.83
	(0.17)	(0.24)	(0.13)	(0.20)	(0.43)	(0.13)	(0.09)	(0.14)
...on the statewide test	0.78	0.75	0.67	0.75	0.56	0.14	0.79	0.67
	(0.20)	(0.00)	(0.13)	(0.22)	(0.43)	(0.20)	(0.09)	(0.14)
Number of Teachers	9	4	6	6	4	7	7	3

Note: Sample excludes teachers in K014 (who did not take surveys) as well as teachers who entered the schools after the school year 2012-2013. Responses are divided based on actual work in the School of One program. Response scales have been rescaled from 0 (least agreement) to 1 (most agreement). Standard deviations are shown in paretheses.

Appendix Table 4b: Spring 2014 Teacher Survey on Math Curriculum, by School

	Spring 2014							
	K088		X368	K381	R002		R049	
Survey Question:	Non-So1	So1	Non-So1	So1	Non-So1	So1	Non-So1	So1
I liked the math curriculum that I used this year	0.70	0.75	0.54	0.80	0.69	0.42	0.54	0.75
	(0.16)	(0.00)	(0.19)	(0.11)	(0.26)	(0.14)	(0.30)	(0.00)
The students' learned a lot in math this year	0.75	0.88	0.63	0.75	0.63	0.25	0.71	0.67
	(0.12)	(0.14)	(0.14)	(0.18)	(0.23)	(0.00)	(0.22)	(0.14)
Most of the students' needs were met in math class this year	0.70	0.75	0.58	0.70	0.53	0.25	0.61	0.67
	(0.23)	(0.00)	(0.20)	(0.21)	(0.21)	(0.00)	(0.28)	(0.14)
Math teaching materials are effective in helping students…								
...improve in math	0.73	0.69	0.54	0.75	0.66	0.42	0.64	0.67
	(0.08)	(0.13)	(0.19)	(0.18)	(0.19)	(0.14)	(0.20)	(0.14)
...on the statewide test	0.78	0.56	0.46	0.75	0.59	0.42	0.57	0.75
	(0.08)	(0.13)	(0.25)	(0.18)	(0.23)	(0.14)	(0.24)	(0.00)
Students were aware of their learning goals	0.83	0.81	0.75	0.80	0.81	0.67	0.75	0.75
	(0.12)	(0.24)	(0.16)	(0.21)	(0.18)	(0.38)	(0.00)	(0.00)
Students who started below grade level	0.65	0.81	0.42	0.80	0.53	0.67	0.54	0.67
were well served by the math curriculum	(0.21)	(0.13)	(0.29)	(0.21)	(0.28)	(0.29)	(0.17)	(0.14)
Students who started above grade level	0.75	0.88	0.60	0.75	0.59	0.67	0.64	0.67
were well served by the math curriculum	(0.17)	(0.14)	(0.22)	(0.25)	(0.23)	(0.38)	(0.13)	(0.14)
Number of Teachers	10	4	5	5	8	3	7	3

Note: Sample excludes teachers in K014 (who did not take surveys) as well as teachers who entered the schools after the school year 2012-2013. Responses are divided based on actual work in the School of One program. Response scales have been rescaled from 0 (least agreement) to 1 (most agreement). Standard deviations are shown in paretheses.