Supplemental Exploratory Results from the StartSmart K-3 Plus Randomized Controlled Trial

Damon Cann, Mustafa Karakaplan, Margaret Lubke, and Cyndi Rowland
Utah State University

12-26-2015

General Project Background

The StartSmart K-3 Plus project began as an effort to evaluate the effectiveness of the innovative K-3 Plus model as implemented in the State of New Mexico. The K-3 Plus program lengthens the school year by providing an additional 25 days of education during the summer each year prior to grades K-3. Providing summer instruction could be helpful at many points in time. However, these grades may be uniquely important as research suggests that summer learning loss is greatest in the early grade levels, and accumulated summer learning loss over several years can amount to a substantial amount of under-achievement (Alexander, Entwisle, and Olsen, 2007). These losses are most pronounced for students from lower SES groups (Entwisle, Alexander, and Olsen 2001). The logic behind the K-3 Plus program is that it not only minimizes summer learning loss (by putting students in an enriching environment during the summer) but also helps students to experience gains during the summer to propel them forward in their achievement.

We do not intend to provide a detailed account here of the K-3 Plus program's history. Nor do we intend to provide a specific account of our exact research design for studying it through the StartSmart K-3 Plus randomized controlled trial. Specific details of the report and methodology can be found at http://startsmartk3plus.org. Our purpose here is to extend upon our core evaluation report and share additional information beyond want was delivered in our core evaluator's report. The first area where we offer additional information is the presentation of treatment-on-the-treated (TOT) effects for the beginning of 3rd grade time point. The second area where we seek to fill in results is by looking at the difference between intervention and control students at the end of the school year.

Treatment on the Treated Effects

In the first year of the program, students had a very high rate of compliance with their treatment assignment (the vast majority of intervention students indeed attended the summer session and only a handful of control students found some way to participate in a K-3 Plus summer program). Across the first four years of schooling, though, we found that intervention students became less likely to attend the summer session. This is natural—students beginning their school experience are typically quite enthusiastic about beginning Kindergarten while by 2nd or 3rd grade some children lose their enthusiasm for attending school in the summer time. At the same time, some of our schools began to offer state-funded K-3 Plus services and a handful of children in the Control group found a way to obtain the very 25-day program mirrored by StartSmart. In the presence of non-compliance with treatment assignment, intent to treat effects (such as those in our evaluators report) are potentially much smaller than the effectiveness of the program for individuals who actually complied with their treatment assignment. Accordingly, we offer treatment-on-the-treated estimates of program effectiveness here as a supplement to our main impact analysis.

Table 1: TOT Effects at the Final Assessment Point

Outcome Domain	Baseline Equivalence Supported?	ITT Treatment Effect (Std. Err.)	TOT Treatment Effect (Std. Err.)	*n*
Expressive Vocabulary	Yes	.400 (.602)	1.111 (1.537)	1293
Broad Reading	Yes	1.684* (.741)	4.070* (1.965)	1313
Broad Math	Yes	1.374* (.671)	3.540* (1.732)	1276
Basic Writing	Yes	2.037 * (.690)	5.022* (1.831)	1276
Social Skills	Yes	1.048 (.856)	2.490 (1.914)	1089
Receptive Language	Yes	.036 (.555)	.293 (1.568)	1021

*Note: Standard Errors appear below estimates in parentheses. Baseline Equivalence is supported if the difference between the average pretest scores is less than .25 standard deviations. * denotes p < .05, two-tailed.*

More technically, using students who attended rather than their treatment assignment introduces endogeneity as students who elect to attend could be systematically different from those who choose not to attend. This problem is fairly straightforward to address in that group assignment is a known exogenous variable that can serve as an instrumental variable for attendance. In the TOT models, we use a simple instrumental variables regression model where program participation (rather than intervention group membership) is the independent variable of interest. Program participation is coded as 1 for students who participated in the K-3 Plus program through all 4 years of the study and 0 for students who did not participate in all 4 years.[1] Experimental group assignment can readily serve as a valid instrumental variable because it was randomly generated. We also include the same control variables as we used in our other models previously: pre-test scores, gender, and maternal education. Table 1 below contains the treatment effects (the

[1] We also ran models that conceptualized program participation in terms of the number of summers the student attended services. For students who attended all four summers, the results are substantively similar in this model as in the results presented in Table 1 above. Naturally, the effects are smaller for students who attended a smaller number of summers.

coefficient on the treatment assignment) and their statndard errors for both the ITT effects presented in our full report and the TOT results for comparison.[2]

When looking at the effects of the K-3 Plus program on those who were actually treated, the effects are much more substantial. For Broad Reading, the effect is about four points on the test (which equates to an effect size of *d*=.27); for Broad math the effect of treatment on the treated is about 3.5 points (*d*=.24). The effect is strongest on the Basic Writing test, with an average effect of treatment on the treated of about 5 points (*d*=.36). The TOT effects still show no statistically significant effect of the program on Expressive Vocabulary, Social Skills, or Receptive Language.

These results reinforce the notion that the summer intervention promotes improved performance among students who comply with their group assignment.

Longer-term Effects of the Program

Our previous efforts have looked at the effectiveness of the StartSmart K-3 Plus program at the beginning of Kindergarten (after one summer of exposure) and at the beginning of 3rd grade (after the full program has been completed with up to 4 summers of exposure). The final time point of the study (the beginning of 3rd grade) makes sense as a proper exploratory analysis comparison point. However, all of our time points come quite shortly after the intervention has taken place. Ultimately, the greatest interest in the program comes from its long-term effects. While we do not have evidence at this point in time as to whether differences between intervention and control groups differ at time points beyond the beginning of 3rd grade, we can approach the question of whether the effects of the summer intervention can still be observed at the end of a school year. We focus here at the end of the 2nd grade year, where students have had (potentially) 3 summers of intervention services, but then go through a school year where both intervention and control group students are receiving education.[3] We present treatment on the treated results here using an instrumental variables model with child achievement on various outcome domains as the dependent variable and attendance at all possible summer sessions (here, 3 sessions are possible in the summers prior to grades K, 1, and 2), gender, maternal education, and achievement scores at baseline. Experimental group assignment serves as an instrumental variable to address possible endogeneity of participation in the summer sessions. Results appear in Table 2.

---

[2] Full regression results are available upon request.

[3] We note here that a handful of students were retained in grade. These students continued to be assessed on the same schedule as other students and would have been at the end of second grade if they had made normal progress toward graduation. These students still had 3 potential years of receiving the summer intervention (if in the treatment group) as well as the regular school year services.

Table 2: TOT Effects at the End of 2nd Grade

Outcome Domain	TOT Treatment Effect (Std. Err.)
Expressive Vocabulary	1.174 (.988)
Broad Reading	1.673 (1.500)
Broad Math	1.003 (1.305)
Basic Writing	3.381* (1.461)
Social Skills	-1.234 (2.572)
Receptive Language	1.047 (1.118)

*Note: Standard Errors appear below estimates in parentheses. * denotes p < .05, two-tailed.*

Only Basic Writing shows statistically significant effects at the end of the second grade year. While its effect size is substantively meaningful (*d*=.237), the fact that it is the only outcome domain with a statistically significant result does not reflect well on the program. At this point, we have several hypotheses about possible reasons why the program could be moving students forward in the school year but that differences between intervention and control group students seem to diminish over the course of the school year. Among them, two stand out in ways that we can bring data to bear on them at this point: whether the student stays with the same teacher from the summer into the regular school year and whether summer language instruction took place in the students' best language.

Remaining with the Same Instructor

We suspect that the summer gains are most likely to be maintained if the student's teacher knows the status of the student and where he or she is at from summer school when the regular school year begins. If this is not the case, the regular school year teacher may simply restart instruction that students who received K-3 Plus services in the summer had just covered in their summer learning experience. There are a variety of ways that the school year teacher could have this information, but the most straightforward to test using our existing data is whether the student has the same regular school year teacher that he/she had in this summer.

We devise a simple new test looking at this issue. Here we consider a student to be "treated" only if the student attended school in all four years and if the student stayed with the same teacher from summer to the regular school year in each of those years. Since being treated requires attendance, there is the same potential for endogeneity in this scenario as in our foregoing TOT analyses. The remedy, though, is similar, using experimental group assignment as an instrumental variable in a standard instrumental variables regression model. We employ the same set of control variables as in our TOT models above to estimate the effectiveness of program participation when participation is defined as staying with the same teacher. TOT treatment effects appear in Table 3.

Table 3: TOT Effects at the End of 2nd Grade for Students with the Same Teacher

Outcome Domain	TOT Treatment Effect (Std. Err.)
Expressive Vocabulary	2.941 (2.475)
Broad Reading	4.321 (3.868)
Broad Math	10.124* (3.680)
Basic Writing	8.522* (3.698)
Social Skills	-2.952 (6.147)
Receptive Language	2.356 (2.523)

*Note: Standard Errors appear below estimates in parentheses. * denotes p < .05, two-tailed.*

When conceptualizing treatment as remaining with the same teacher, we find that the program has statistically significant effects on both math and writing skills. Perhaps more impressive, though, is the size of these effects. For math, achievement improves by over two thirds of a standard deviation (*d*=.687) and for writing the effect is *d*=.60 standard deviations. These are very substantial effects, though we see no statistically significant effects for expressive vocabulary, receptive language, social skills, or reading.

Student Language and Language or Instruction

New Mexico is a diverse state on many dimensions. Among the many diversities in the state is language. In the regular school year, many New Mexico schools offer instruction in Spanish in addition to standard English language instruction. In a school where 1/3 of students are Spanish-speakers with 60 students in a given grade level, this is easy to administer—20 students are placed in

a Spanish instruction classroom and the 40 English-speaking students are divided between two additional classrooms. In the summer, though, this can be a challenge. If 20 students at the school participate in the K-3 Plus program, one classroom would be warranted by the number of students. However, at this school one might have 7 Spanish-speaking students and 13 English-speaking students. As implemented in the state, many schools would offer a single classroom instructed in English, which could result in outcomes that are suboptimal in terms of the effects. These incongruities in language of instruction could result poorer outcomes for Spanish-speaking students than for English-speaking students.

To test this hypotheses, we re-run the TOT instrumental variables models as was done for Table 2 above with a simple adjustment: we include a variable for the best language of the student at the time of randomization into the study (1=Spanish, 0=English) and the interaction of student language with the variable measuring program participation. This interaction term allows us to evaluate whether the program is more effective for English-speaking students (for whom language congruity is a non-issue) than for Spanish-speaking students. Because our receptive language measure was only for English-speaking students, we do cannot run a model with that outcome measure that includes language as a variable.

Because the program participation variable is likely endogenous, the interaction of program participation and student language is also likely endogenous. However, just as experimental group assignment is a valid instrument for program participation, so also is the interaction of experimental group assignment and language a valid instrument for the interaction of program participation and student language. This results in our ability to estimate a treatment effect for students whose best language was English at randomization as well as a treatment effect for students whose best language was Spanish at randomization. We present those effects in Table 4.

Table 4: TOT Effects at the End of 2nd Grade for Students by Language

Outcome Domain	TOT Treatment Effect-English (Std. Err.)	TOT Treatment Effect-Spanish (Std. Err.)
Expressive	2.086#	-1.126
Vocabulary	(1.080)	(2.103)
Broad	3.631*	-3.998
Reading	(1.624)	(3.226)
Broad Math	1.592	-.757
	(1.593)	(2.139)
Basic Writing	5.469*	-3.018
	(1.703)	(2.743)
Social Skills	-2.380	1.803
	(2.989)	(4.979)

Note: Standard Errors appear below estimates in parentheses.
** denotes p < .05, two-tailed, # denotes p < .10, two-tailed.*

Here, we see that gains are being maintained through the school year for English-speaking students on some outcome domains (especially those that are the most language intensive). Expressive vocabulary gains just miss the standard .05 cutoff for statistical significance for English speaking students (p = .052). We statistically significant effects for Broad Reading and Basic Writing that endure through the school year for English speaking students, and the effect sizes are substantively meaningful (d = .24 for reading and .38 for writing). There are no statistically significant effects for math and social skills for English-speaking students. For Spanish-speaking students, the program shows no statistically significant effects on any of our outcome domains.

While we do not know the specific reason why the program seems to be more effective for English-speaking students than for Spanish-speaking students, the results here suggest that something related to language is influencing the ability of students to maintain gains from the summer program through the duration of the full school year. We posit that mismatches between the language of instruction and the student's best language may be at least a part of the explanation.

Discussion

Our initial full report shows that the K-3 Plus program generates strong and statistically significant outcomes for students in the core academic domains of reading, writing, and math when the students are tested shortly after the beginning of the school year (and, by consequence, shortly after the summer session ends). In this addendum, we add depth to those findings by exploring how TOT estimates differ from ITT estimates, showing that when estimates of program effectiveness are quite strong when one considers actual attendance at the program rather than simply looking at assignment to treatment.

In this addendum, we also explore the extent to which the gains realized during the summer dissipate during the school year. We find that for some groups, the benefits derived from the summer session diminish or even disappear over the course of the summer while for other groups the benefits of program participation are maintained during the course of the school year. Specifically, for English-speaking students, the benefits of the summer program are maintained across the length of the school year; for students who were Spanish-speakers at randomization, the benefits are not maintained. We also find that for students who stayed with the same teacher, substantial gains in achievement in math and writing were realized and maintained throughout the school year.

Taken together, these findings provide a strong amount of evidence of the prospective effectiveness of the K-3 Plus program for boosting achievement. The extent to which these gains are maintained in the long-term, however, appears to be contingent on a variety of factors. Future research should investigate the conditions under which summer learning gains can be maintained in order to maximize the effectiveness of extended school year programs.