CONSORTIUM FOR POLICY RESEARCH IN EDUCATION
TEACHERS COLLEGE
COLUMBIA UNIVERSITY

BookNook in Rocketship Schools: Final Impact Results

Douglas Ready, Director
ddr2111@tc.columbia.edu

October 20, 2023

Consortium for Policy Research in Education (CPRE)

The Consortium for Policy Research in Education (CPRE) brings together education experts from renowned research institutions, including Teachers College, Columbia University; the University of Pennsylvania; Harvard University; Stanford University; the University of Michigan; the University of Wisconsin-Madison; and Northwestern University. CPRE was launched in 1985 as the first national federally funded R&D center for state and local education policy. Since then, CPRE has studied the design, implementation, and effects of hundreds of policies and programs. CPRE researchers and staff at Teachers College (CPRE-TC) are continuing this long tradition by conducting rigorous research and evaluation that aims to improve elementary and secondary education through increased educational effectiveness, equity, and access.

To learn more about CPRE-TC, visit our website at www.tc.columbia.edu/cpre

Abstract

The Consortium for Policy Research in Education at Teachers College, Columbia University (CPRE-TC), conducted a 12-week cluster randomized controlled trial to examine the efficacy of BookNook, a supplemental web-based tutoring platform focused on reading growth. Cohorts of first- through fourth-grade students attending six Rocketship public charter schools in Northern California were randomly assigned to receive tutoring via the BookNook platform. Control cohorts of students within the same grade and school continued to receive the reading supports and activities that Rocketship traditionally provides. We found evidence that BookNook tutoring supported student reading growth. Intent-to-Treat models indicate that students in cohorts assigned to receive BookNook outperformed their control-group peers by roughly 0.05 SDs. Given the substantial variability in usage rates among students enrolled in BookNook cohorts, we also leveraged Treatment-on-the-Treated approaches. These models suggest even larger positive effects for students with higher usage rates. Students who completed 10 or more BookNook sessions experienced a reading advantage of 0.08 SDs, while those who completed 20 or more sessions—the recommended dosage—experienced a 0.26 SD developmental advantage.

Introduction

Over the past several years, scholars, policymakers, and practitioners have touted the benefits of one-to-one and small-group tutoring for students' academic development.[1] The Consortium for Policy Research in Education at Teachers College, Columbia University (CPRE-TC), conducted a 12-week cluster randomized controlled trial to examine the efficacy of BookNook, a supplemental web-based tutoring platform focused on reading growth. Cohorts of first- through fourth-grade students attending six Rocketship public charter schools in Northern California were randomly assigned to receive tutoring via the BookNook platform. Control cohorts of students within the same grade and school continued to receive the reading supports and activities that Rocketship traditionally provides.

In the sections below we share more information about BookNook and its implementation in these six schools. We then describe our data and the analytic approaches we employed in our analyses. Our results section includes information on baseline equivalency between treatment and control groups, insights into fidelity of implementation, and the results of our primary models estimating the causal impact of BookNook on student learning. We close with a summary of our findings and discuss potential implications for future research on BookNook specifically and supplemental tutoring programs more broadly.

The Intervention

BookNook is a Tier 2 intervention designed for students that are struggling to learn and grow in their reading skills.[2] It uses a synchronous teaching platform and curriculum grounded in the science of reading to deliver high-dosage tutoring services tailored to meet students where they are and help them learn at an accelerated pace. Students meet with the same tutor and small group of students to build relationships and to allow the tutor to develop an understanding of each student's strengths and skill gaps. Students work through BookNook's virtual platform to participate in their lessons. At the core of every BookNook lesson is a text.

---

[1] For reviews, see White, S., Groom-Thomas, L., & Loeb, S. (2022). Undertaking complex but effective instructional supports for students: A systematic review of research on high-impact tutoring planning and implementation. (EdWorkingPaper: 22-652). Retrieved from Annenberg Institute at Brown University: https://doi.org/10.26300/wztf-wj14; and, Nickow, A., Oreopoulos, P., & Quan, V. (2020). *The impressive effects of tutoring on PreK-12 learning: A systematic review and meta-analysis of the experimental evidence* (NBER Working Paper No. 27476). National Bureau of Economic Research. doi:10.3386/w27476

[2] https://www.booknook.com/

BookNook's model features lessons that support students in building phonological awareness, phonics, fluency, vocabulary and comprehension skills. BookNook's phonological awareness/phonics lessons help students acquire and build foundational reading skills. These lessons follow the same flow each session. Lessons begin with a skill introduction, which teaches students the skill or standard the lesson is aligned to. Tutors present the skill and relevant examples for students to both see and hear, which provides the foundations needed for students to then practice the skill. For example, students may be introduced to consonant-vowel-consonant words, such as "cat" or "big," and will be prompted to break the words apart to hear the different sounds. Students then move to skill practice, where they have the opportunity to apply the skill or standard of the lesson through 2-5 different practice activity types depending on the lesson. During the foundational text reading section, students practice the focal skill with a text that incorporates a high number of words aligned to that skill. The texts in these lessons integrate the components of phonics, fluency, and text reading comprehension skills through the authentic practice reading. After finishing the text, students discuss what they have read. Tutors are encouraged to ask questions that aim to engage students in the text and form a deeper understanding of what they have just read. Lastly, each phonics lesson ends with a formative assessment, which provides data on student progress towards mastery of the skill through 4-5 aligned questions. These data are incorporated into the program's algorithm for determining student advancement to a new set of lessons.

During BookNook's fluency lessons, students engage in activities that support practice in oral reading fluency. Fluency lessons follow the same sequence each time: fluency introduction; modeled reading; fluency activity; fluency practice. During the introduction, students review the three components of fluency: accuracy, pace, and expression. Tutors then model or play a recording of the fluent reading of a passage for students. Students also have the opportunity to read the same passage aloud and practice the same techniques they heard in the modeled reading. Following the modeled reading, students engage in a fluency activity called "What's Wrong." In this activity, students monitor a tutor or audio reading and pay close attention to and evaluate the components of fluency. Finally, students engage in their own reading of a passage, while the tutor notes any errors. The BookNook algorithm uses these notes to calculate and help the tutor understand the student's accuracy and words per minute score.

Comprehension lessons are designed to support students in practicing and building skills in the vocabulary and comprehension components of the science of reading. Each comprehension lesson begins with vocabulary instruction, given that lack of vocabulary understanding can impede comprehension of a text. Vocabulary that are essential to the comprehension of the text that students will read later in the lesson are pre-taught in a scaffolded exposure that includes words, definitions, audio, images, and typing exercises. Following direct vocabulary instruction, students engage in an interactive matching activity that provides additional exposure to the lesson's vocabulary words and their meaning. At the conclusion of the vocabulary section of the lesson, students engage in a check-for-understanding activity that allows them to see and place vocabulary words in context in various sentences. As students transition from vocabulary to comprehension, they engage in a pre-read strategy session to help them build engagement and motivation around the upcoming text. These discussions prompt activation of pre-reading strategies such as activating prior knowledge and making predictions.

Tutors then lead a readthrough with group discussion questions. These questions are structured to not only support students in building the skills of the standard of the lesson, but also to develop reading comprehension skills more generally that can be used to comprehend the lesson text and other texts they encounter outside of the lesson. The discussion questions prompt students to think critically, make inferences, engage in vocabulary work, and focus on the lesson-aligned standard. Immediately following the readthrough, students discuss and synthesize what they have just read. Students then engage in a comprehension activity called Feed the Animals that focuses on work around students identifying main ideas, themes, summarizing, retelling and/or sequencing to build comprehension of the text they read, but again are skills that can be applied to texts they read outside of the lesson.

Lastly, students engage in individual thinking and analysis about the text through text-dependent questions aligned to the standard of the lesson. This formative assessment provides data on their progress towards mastery of the skill or standard. Schools can opt into virtual tutoring through BookNook (Tutoring Service Delivery) or in-person tutoring using their current staff (District-Led Delivery). In terms of our current study, Rocketship Schools opted for the Tutoring Service Delivery format. With this option, BookNook provides virtual tutors, most of whom have previously taught academic subjects to K-12 students, have 3+ years of experience with tutoring, and/or are currently enrolled in a teaching certification program.

The Implementation

The BookNook implementation involved first- through fourth-grade students enrolled in six Rocketship public charter schools in Northern California. As part of its regular instructional programming, Rocketship organizes students into same-grade cohorts containing roughly 20-30 students each, usually resulting in three to four cohorts per grade, depending on enrollments. Rocketship students experience four content blocks each day: Humanities, STEM, Enrichment, and Learning Lab. Within each school and grade, we randomly assigned all cohorts to treatment or control groups. Students enrolled in treatment cohorts were to receive BookNook tutoring during their Learning Lab period two to three times per week for 30 minutes per session. Students in cohorts assigned to the control condition would continue with the regular reading supports provided during Learning Lab. The roughly 12-week implementation began in late January, 2023 and concluded in early May, 2023. During this period the Rocketship academic calendar included two week-long vacations. As such, the actual intervention period was 10 weeks, with a full treatment exposure calculated as the completion of 20-30 BookNook sessions.

Data and Methods

The data for the study include student-level reading test scores, academic and socio-demographic measures, and variables that link students to grade-level cohorts and schools. All cohorts completed the study in their original treatment and control states and the study experienced zero assignment-level attrition. Seven students declined to participate in the study prior to implementation; no students withdrew from the study during implementation. The initial sample included 1,900 first- through fourth-grade Rocketship students. Our analytic sample only includes students with full demographic and assessment data. No students were missing demographic data. There was, however, student-level missing data associated with the baseline and follow-up MAP assessments, with 6.5% of students missing data on one or both assessments. Missingness rates were virtually identical across students assigned to treatment (6.41%) and control cohorts (6.53%). This baseline and follow-up assessment restriction necessarily excludes students who enrolled in either treatment or control groups during the implementation (i.e., no joiners are included in the sample). No students with full data (and thus included in these analyses) switched treatment/control cohorts during the implementation.

Our final analytic sample includes 77 student cohorts (*n*=42 treatment, 35 control) containing 1,777 first- through fourth-grade students (*n*=959 treatment, 818 control), of whom 79% are Hispanic, 9% Black, 8.6% Asian, 2.4% white and 1.1% American Indian/Alaskan Native or Native Hawaiian/Pacific Islander. ELL students represent just over half of the sample, and 9.5% of students receive special education services. Roughly 48% of students are identified as female.

Measures

Outcome. Rocketship schools administer the MAP assessments three times each year—Fall, Winter, and Spring—as part of its regular assessment program. MAP is a computer-adaptive assessment that measures student academic growth, producing scores that are vertically equated using the Rasch unit (RIT) scale. We use reading results from the Winter administration as our baseline measure, and reading scores from the Spring administration as the follow-up (post-implementation) outcome. The Winter MAP administration occurred in December prior to the BookNook implementation, and the Spring administration took place in mid-May, after the conclusion of the study. Scores at each timepoint were standardized (z-scored) within grade.

Covariates. Because of the random assignment process, OLS estimation will provide unbiased treatment estimates and it is not necessary to control for other student characteristics. However, including pre-random assignment covariates that are correlated with the outcome into our models can improve impact estimate precision. As covariates, our models include dummy indicators of student race/ethnicity (Asian, Black, White and other race/ethnicity) with Hispanic students as the un-coded comparison group. Unfortunately, confidentiality concerns related to small sample sizes required us to organize American Indian/Alaskan Native and Native Hawaiian/Pacific Islander students into a single "other race/ethnicity" category. Our analyses also leverage data on student sex (female = 1, male = 0) and special education (IEP) and English language learner (ELL) status (yes = 1, no = 0).

Cohort Baseline Equivalency

To establish baseline equivalence across treatment and control cohorts and the analytic sample of students, we constructed a series of nine separate Ordinary Least Squares (OLS) regression models in which the cohort-average baseline MAP assessment score and aggregate means of the eight student demographic variables served as outcomes. These models, which parallel the impact models discussed below, can be described as,

$$Y_{cgs} = b_0 + b_1(BookNook) + \eta + e_i$$

where Y_{cgs} represents the average standardized baseline MAP reading assessment score or demographic variable for cohort *c,* in grade *g*, in school *s*. *BookNook* is an indicator of whether the cohort was randomly assigned to the treatment condition. School-by-grade fixed effects are indicated by η, while e_i indicates the cohort-level error term.

Data Analytic Plans

We employed two primary analytic techniques with these data to measure the impact of BookNook on student reading growth. The first approach provides the average causal effect of being assigned to the treatment group, often referred to as the "Intent-to-Treat" (ITT) estimate. This approach is thought of as producing the most policy relevant indicator of program impact given the typical constraints faced by social interventions implemented in the field.[3] Individuals or groups assigned to a treatment may not comply—hence the phrase, "*intent* to treat." To estimate the average effect of being randomly assigned to a BookNook cohort, relative to the outcomes of students assigned to control cohorts, we estimate an ITT regression model of the following form:

$$Y_{icg} = b_0 + b_1(BookNook) + X_i + \eta + e_i$$

where Y_{icg} represents the standardized MAP follow-up reading assessment score for student *i*, in cohort *c,* in grade *g*. *BookNook* is an indicator of whether the student's cohort was randomly assigned to participate in BookNook. X_i represents a vector of student-level covariates, including the baseline MAP reading assessment score, race/ethnicity, gender, and IEP and ELL status. School-by-grade fixed effects are indicated by η, while e_i indicates the student-level error term. In all models robust standard errors are clustered at the cohort level. We also constructed a multilevel model with random effects that produced estimates virtually identical to those resulting from the model described here (see Appendix A).

Our second analytic approach entailed a two-stage least squares or instrumental variable approach that explored whether increased BookNook usage among students in treatment cohorts

[3] see Glennerster, R., & Takavarasha, K. (2013). *Running randomized evaluations: A practical guide*. Princeton University Press.

was associated with increased reading learning. Recall that students assigned to treatment cohorts were to complete two to three, 30-minute sessions per week. However, as we discuss in more detail below, student usage rates were generally below what was expected. Because student cohorts were randomly assigned to BookNook, we can conceptualize the treatment of being assigned to a BookNook cohort as an "instrument" for participation in the program. Instrumental variable analysis is feasible in this case because we have met the "exclusion restriction," in which random assignment to the treatment group can only affect student test scores through actual participation in BookNook, or compliance with the prescribed treatment.[4] This type of analysis is considered the "Treatment-on-the-Treated" (TOT) approach, revealing the complier average causal effect of BookNook. We are confident this assumption is met, as students assigned to control cohorts were not provided BookNook accounts or logins during the implementation. Further, we know that random assignment at the cohort level was the only mechanism inducing student participation in the treatment, as again, control cohorts were not provided access to the platform.

With this approach, the first-stage model took the form,

$$BookNook\ Usage_{icg} = b_0 + b_1(Treatment\ Status)_{cg} + X_i + \eta = e_i \qquad \text{(First Stage)}$$

where *Treatment Status*$_{cg}$ is an instrument for *BookNook Usage*$_{icg}$. The second-stage model can then be expressed as,

$$Y_{icg} = b_0 + b_1(\widehat{BookNook\ Usage}) + X_i + \eta + e_i \qquad \text{(Second Stage)}$$

where Y_{icg} is the standardized MAP follow-up reading assessment score for student *i*, in cohort *c,* in grade *g*. The model uses the *BookNook Usage* estimates from the first-stage model. We estimate two separate parameters based on treated students' BookNook usage: 1) a binary indicator of students who completed 10 or more sessions during the implementation period, and; 2) a binary indicator of students who completed 20 or more sessions, which is the minimum recommended treatment dosage based on two sessions per week for the 10 weeks of

[4] See Angrist, J. D., & Pischke, J. S. (2009). *Mostly harmless econometrics: An empiricist's companion*. Princeton University Press.

actual instruction. X_i represents the vector of student-level background covariates described above, as well as the standardized baseline MAP scores. School-by-grade fixed effects are indicated by η, while e_i indicates the student-level error term. In all models robust standard errors are again clustered at the cohort level. It is important to stress that the "local average treatment effects" resulting from these models are relevant only for the types of students who would use BookNook at these higher rates given the opportunity to do so. These effects would not necessarily result if all treatment students had engaged at these levels.

Results

We begin with results from the models establishing pre-treatment equivalency between treatment/control cohorts (see Table 1). Fortunately, we find no statistically significant or substantively meaningful differences in terms of baseline student academic and socio-demographic characteristics. This increases our confidence that the impact estimates we discuss below stem from engagement with BookNook tutoring and not pre-existing differences between students who did and did not experience BookNook.

Table 1. Baseline Equivalency of Students in the Analytic Sample (Cohort-Level Averages)

Characteristics	Treatment Cohorts (*n*=42)	Control Cohorts (*n*=35)	Difference (*SE*)
Baseline MAP	-0.034	0.039	-0.073 (0.050)
Asian	0.086	0.082	0.004 (0.009)
Black	0.099	0.086	0.013 (0.014)
Hispanic	0.783	0.796	-0.013 (0.018)
Other	0.008	0.016	-0.008 (0.006)
White	0.025	0.021	0.004 (0.008)
Female	0.483	0.479	0.004 (0.022)
IEP	0.086	0.106	-0.020 (0.017)
ELL	0.515	0.499	0.016 (0.024)

No differences significant at the $p<.10$ level. Baseline MAP scores are z-scored within grades.

Implementation Fidelity

As with many supplemental tech-enabled interventions, we found considerable variability in BookNook usage rates among students in cohorts assigned to the treatment condition (see

Table 2). Of the 959 students in treatment cohorts, only 196 (20.4%) met even the lower-bound threshold of BookNook engagement, calculated as two completed sessions per week. A plurality of students (45.2%) completed between 10 and 19 tutoring sessions, and 34.4% completed fewer than 10 sessions. Overall, treated students completed an average of thirteen sessions during the implementation period. We explored the extent to which these usage rates were associated with other baseline student background characteristics. Low-usage students began the study with baseline MAP scores roughly 0.19 SDs below those of their moderate- ($p<.05$) and high-usage peers ($p<.10$). In other words, initially higher-achieving students engaged BookNook to somewhat higher degrees. However, we found no associations between BookNook usage rates and student race/ethnicity, sex, or IEP and ELL status.

One important question is the extent to which this variability in student usage flowed from the motivations and interests of individual students, or from the motivations and interests of school staff. With the current implementation, the relevant adults were those staffing the Learning Labs, where treatment cohorts were to have received BookNook tutoring. One way to explore this question is to partition variance in usage into its within-Learning Lab and between-Learning Lab components. We found that almost half (45.5%) of the variability in usage rates exists across Learning Labs, with the remainder (54.5%) occurring within Learning Labs. This suggests that efforts to increase participation rates will have to target both students and staff. Clearly, some staff did not have appropriate expectations for student participation. However, even within the same Learning Labs, student participation rates varied substantially.

Table 2. BookNook Usage Among Students in Treated Cohorts (*n*=959)

	Student-Level Usage Rates		
	Low: 0-9 Sessions (*n*=330)	Moderate: 10-19 sessions (*n*=433)	High: 20+ Sessions (*n*=196)
Sessions Completed	4.76	14.57***	22.89***
SD	2.92	2.78	2.59
Baseline MAP Score	-0.125	0.063*	0.070~
SD	0.997	1.015	0.954
Asian	8.5	9.7	13.3
Black	7.9	9.9	3.1
Hispanic	80.0	77.6	81.6
Other	0.3	0.5	0.4
White	3.3	2.3	1.5
Female	47.9	46.2	52.6
IEP	10.3	8.3	8.2
ELL	56.4	49.4	55.6

*p<0.05; **p<.01; ***p<.001. Sessions completed and MAP scores compared to low-usage category. Associations between usage and race/ethnicity, sex, IEP and ELL status are non-significant (p>.05).

Impact Results

Table 3 provides estimates of the causal impact of BookNook on student reading growth. The Intent-to-Treat (ITT) estimates, displayed in the far left column, indicate that students enrolled in cohorts assigned to the treatment condition modestly outperformed their same-grade, same-school peers enrolled in control cohorts (ES = 0.052; p<.05). Recall that the ITT analytic approach does not account for actual BookNook usage, but instead considers only whether students were offered the treatment, in this case via membership in a cohort that was randomly assigned to the treatment condition. The Treatment-on-the-Treated (TOT) approach, however, allows us to explore the extent to which BookNook efficacy is associated with increased usage. The TOT results presented in the middle column indicate that students who completed 10 or more and sessions also gained somewhat more reading skills compared to their control group peers (ES = 0.080; p<.05). Note that this estimate is slightly larger than the ITT estimate, though the two estimates are not significantly different from one another. However, we find a substantially larger effect for students who complied with the recommended BookNook dosage of at least two completed sessions per week, for a total of 20 or more sessions. As displayed in

the far right column, these high-usage students outperformed their peers assigned to control cohorts by over one-quarter standard deviation (ES = 0.257; $p<.05$).

Table 3. BookNook Effects on Student Reading Growth

	Intent-to-Treat	Treatment-on-the Treated	Treatment-on-the Treated
BookNook Cohort	**0.052***	--	--
	(0.024)	--	--
BookNook: 10+ Sessions	--	**0.080***	--
	--	**(0.034)**	--
BookNook: 20+ Sessions	--	--	**0.257***
	--	--	**0.122**
Baseline MAP Score[1]	0.827***	0.826***	0.827***
	(0.170)	(0.017)	0.016
Asian[1]	0.111**	0.113**	0.115**
	(0.035)	(0.034)	0.034
Black	-0.020	-0.018	-0.009
	(0.045)	(0.045)	0.045
White	-0.130	-0.129	-0.127
	(0.098)	(0.095)	0.097
Other Race/Ethnicity	0.029	0.033	0.020
	(0.102)	(0.100)	0.097
Female	-0.025	-0.025	-0.029
	(0.025)	(0.025)	0.026
ELL	-0.148***	-0.146***	-0.148***
	(0.033)	(0.032)	0.032
IEP	-0.001	0.001	0.001
	(0.055)	(0.025)	0.055
Constant	-0.024	-0.036	-0.079
	(0.031)	(0.030)	0.041
R^2	0.759***	0.760***	0.757***

*$p<0.05$; **$p<.01$; ***$p<.001$. Robust standard errors (indicated in parentheses) are clustered at the cohort level. All models include school-by-grade fixed effects.
[1] Outcome and baseline MAP scores are standardized (z-scored) within grades.
[2] All racial/ethnic groups compared to Hispanic students.

Conclusion and Discussion

This study examined the implementation and efficacy of BookNook with first- through fourth-graders in six Rocketship schools in Northern California. We found evidence that BookNook tutoring supported student reading growth. Our ITT models suggest that students in cohorts assigned to receive BookNook tutoring outperformed their control-group peers by roughly 0.05 SDs. Our TOT analyses indicate even larger positive effects among students with higher usage rates. Treatment students who completed 10 or more BookNook sessions experienced a reading skills advantage of 0.08 SDs, while those who completed 20 or more sessions—the recommended dosage—experienced a 0.26 SD developmental advantage. It is important to bear in mind the caveats associated with the TOT estimates. Namely, these effects are relevant for the types of students who completed more sessions. We cannot claim that providing all students the same levels of BookNook tutoring would have produced similar levels of reading learning.

The presence of these significant effects is somewhat surprising given the relatively short ten-week implementation period. Students and staff in these schools did not have prior experience working with the BookNook platform. We typically assume that new interventions take time for both students and staff to become comfortable with a given approach and its procedures. This suggests that perhaps the BookBook platform is structured such that the start-up and launch efforts we associate with many interventions are reduced. In short, given the relatively brief treatment period, these findings are particularly encouraging.

One concern that accompanies many supplemental ed-tech implementations is weak usage among students assigned to treatment conditions. This is indeed what we found with the current study, where few students received BookNook tutoring at the expected levels. Only 20% of students enrolled in treatment cohorts completed 20 or more tutoring sessions, the recommended dosage. BookNook and other developers must continue to consider how to increase fidelity of implementation. This will likely require deeper conversations with school staff, more meaningful professional development activities, and consistent and ongoing communications throughout the implementation period. With non-core instructional strategies such as BookNook, school staff will likely need to be convinced of the potential benefits for their students. If these critical issues of usage are not addressed, promising interventions such as BookNook are unlikely to fully achieve their aims of improving student academic outcomes.

Appendix

We also constructed a multilevel random effects model which analytically nested students within cohorts, which were nested within grade-by-school clusters.[5] The model can be described as,

$$\textit{Level 1 (Students):}\ Y_{icg} = \pi_{0cg} + \pi_{1cg}(\textit{Baseline MAP}_{icg}) + \pi_{2cg}(\textit{Asian}_{icg}) + \pi_{3cg}(\textit{Black}_{icg}) + \pi_{4cg}(\textit{White}_{icg}) + \pi_{5cg}(\textit{Other}_{icg}) + \pi_{6cg}(\textit{Female}_{icg}) + \pi_{7cg}(\textit{IEP}_{icg}) + \pi_{8cg}(\textit{ELL}_{cg}) + e_{icg}$$

$$\textit{Level 2 (Cohort):}\ \pi_{0cg} = \beta_{00g} + \beta_{01g}(\textit{BookNook}_{cg}) + r_{0cg}$$

$$\textit{Level 3 (Grade-by-School):}\ \beta_{00g} = \gamma_{000} + u_{00g}$$

where, at the student level (Level 1), Y_{icg} is the follow-up (spring) MAP reading score for student *i* in cohort *c* in grade-by-school cluster *g*; π_{0cg} is the intercept for cohort *c* in grade-by-school cluster *g*; $\pi_{1cg} \ldots \pi_{8cg}$ represent the coefficients for the student-level covariates, which include the baseline MAP reading score, a series of race/ethnicity indicators (Asian/Pacific Islander, Black, white, and other race/ethnicity, with Hispanic students as the un-coded comparison group), female, and special education and ELL status (1=yes, 0=no). The Level-1 error term for student *i* in cohort *c* in grade-by-school cluster *g* is represented by e_{icg}, assumed to be normally distributed with a mean of 0 and variance σ^2. At the cohort (treatment) level, β_{00g} is the intercept for grade-by-school cluster *g*. At the cohort level, we find our treatment indicator (*BookNook*), given that cohorts within grade-by-school clusters were randomly assigned. *BookNook* is group-mean centered (centered around grade-by-school cluster means). Often referred to as adaptive centering with random effects,[6] this approach reproduces the more traditional grade-by-school cluster fixed effects approach used in the Intent-to-Treat model described above on page 7. As with the fixed-effects model, this approach compares treatment and control cohorts within the same grade-by-school cluster. The Level-2 error term is represented by r_{0cg}. The model is unconditional at level 3, and simply represents the clustered standard errors at the grade-by-school level, with the Level-3 error term indicated by $u_{00g.}$

The findings of this alternate model specification are indicated in Table A below. The BookNook treatment estimate here (*ES* = 0.054; $p<.05$) is virtually identical to that found in Table 3 (*ES* = 0.052; $p<.05$).

[5] Raudenbush, S.W., & Bryk, A.S. (2002). *Hierarchical linear models: Applications and data analysis methods*. Thousand Oaks, CA: Sage.

[6] Raudenbush, S.W. (2009). Adaptive centering with random effects: An alternate to the fixed effects model for studying time-varying treatments in school settings. *Education Finance and Policy*, *4*(4), 468-491.

Table A. BookNook Effects on Student Reading Growth (Multilevel Model)

	Intent-to-Treat
BookNook Cohort	**0.054***
	(0.025)
Baseline MAP Score[1]	0.828***
	(0.022)
Asian[1]	0.119**
	(0.039)
Black	-0.031
	(0.055)
White	-0.121
	(0.099)
Other Race/Ethnicity	0.012
	(0.087)
Female	-0.021
	(0.025)
ELL	-0.148***
	(0.038)
IEP	0.002
	(0.055)
Constant	0.075
	(0.033)

*p<0.05; **p<.01; ***p<.001. Robust standard errors are indicated in parentheses.
[1] Outcome and baseline MAP scores are standardized (z-scored) within grades.
[2] All racial/ethnic groups compared to Hispanic students.