ANNENBERG
BROWN UNIVERSITY

EdWorkingPaper No. 24-938

Can Computer-Assisted Instruction Help Schools to Close the Achievement Gap: Evaluation of a District-Wide Reading Intervention

Yusuf Canbolat
Strategic Data Project

Rebeca Arndt
University of Central Florida

A concerning number of middle and high school students lack fundamental reading skills in the United States. One common way schools address this issue is by supporting those students with computer-assisted instruction. This study evaluates the causal effect of one such computer-assisted instruction intervention on English Language Arts achievement for middle and high school students in a large urban Southeast school district. The district uses a computer-based online learning platform as part of its multi-tiered system of support. The study benefits the usage data in the learning platform from about ten thousand students by exploiting difference in differences and event study estimations. Particularly, it offers a novel method by utilizing the time of initial platform usage and dates of within-year tests for each student. Our results indicate that, on average, the intervention increases test scores by 0.14 SD—a modest but important magnitude given the scale of the intervention. The magnitude of the effect is relatively larger for students who use the platform consistently and among English Language learners. Results are robust against several sensitivity tests including inverse probability weighting, and type of aggregated treatment effect parameter. These results suggest that effective computer-assisted instruction can help schools narrow the achievement gap among students, particularly for English Language learners.

VERSION: April 2024

Suggested citation: Canbolat, Yusuf, and Rebeca Arndt. (2024). Can Computer-Assisted Instruction Help Schools to Close the Achievement Gap: Evaluation of a District-Wide Reading Intervention. (EdWorkingPaper: 24-938). Retrieved from Annenberg Institute at Brown University: https://doi.org/10.26300/f6dd-v158

Can Computer-Assisted Instruction Help Schools to Close the Achievement Gap: Evaluation of a District-Wide Reading Intervention

Yusuf Canbolat
Strategic Data Project

Rebeca Arndt
University of Central Florida

A concerning number of middle and high school students lack fundamental reading skills in the United States. One common way schools address this issue is by supporting those students with computer-assisted instruction which has relatively lower cost than alternative policies such as tutoring. This study evaluates the causal effect of one such computer-assisted instruction intervention on English Language Arts achievement in a large urban Southeast school district. The district uses a computer-based online learning platform as part of its multi-tiered system of support. The study benefits the usage data in the learning platform from about 26,000 students by exploiting difference in differences and event study estimations. It offers a novel method by utilizing the time of initial platform usage and dates of within-year tests for each student. The results indicate that, on average, the intervention increases test scores by 0.14 SD—a relatively modest but still important magnitude given the scale of the intervention. The magnitude of the effect is larger for students who use the platform consistently and among English Language learners. Results are robust against several sensitivity tests including inverse probability weighting, and type of aggregated treatment effect parameter. These results suggest that effective computer-assisted instruction can help schools narrow the achievement gap among students, particularly for English Language learners.

Introduction

In school, the ultimate purpose for reading is comprehension because students need to comprehend the texts they read in order for learning to occur. While in lower-elementary the focus is placed on learning to read, in upper elementary and above, the focus is placed on reading to learn. In other words, reading performance is vital not only for lower elementary students but also for middle and high school students who are required to meet grade-level standards. However, according to the National Assessment of Educational Progress (NAEP), only 31% of eighth graders met or exceeded the NAEP reading proficiency in 2022 (NAEP, 2022a). The vast majority of eighth (69%) and twelve graders (63%) performed at or below the NAEP Basic level in reading (National Assessment of Educational Progress, 2019). An even bleaker picture is revealed when examining the reading achievement gap by socioeconomic and English learner (EL) status (National Assessment of Educational Progress, 2022b). As this data is concerning, many school districts nationwide began searching for ways to address the reading achievement gap of secondary-level (middle and high school) students.

In response to the achievement gap, schools have embraced a variety of approaches to support struggling students such as tutoring, small group instruction, and computer-assisted instruction (CAI) (Kraft et al., 2022; Nickow et al., 2020; Robinson et al., 2021). CAI, one of the school-based interventions under scrutiny, has been used interchangeably with other terms such as computer-aided learning, computer-assisted instruction, and intelligent/cognitive tutoring systems when describing interventions that may personalize learning experiences (Major et al., 2021). Because CAI potentially has a lower cost than some other interventions such as tutoring, it has been widely used across the country. One such technology-based intervention found under the umbrella of CAI is Lexia® PowerUp Literacy®, referred to from here on as PowerUp, a learning platform created to enhance middle and high school students' word recognition automaticity via word study, grammar, and comprehension skills.

Causal studies indicate that CAI can improve student learning among middle and high school students who are left behind by their peers. For instance, based on recent findings of a meta-analysis, CAI showed a positive and significant effect with an average of 0.17 SD on secondary-level student test scores with academic difficulties (Dietrichson et al. 2020). Although two randomized controlled trials with small sample sizes (n = 155 students and n = 122 students) suggested important gain in reading scores among students who used PowerUp compared to their peers who studied in conventional instructional settings (Hurwitz et al., 2022; Hurwitz & Macaruso, 2021), no research has been conducted so far with a large sample size. This study addresses the research gap around the effectiveness of PowerUp to improve student learning on a large scale.

The current study contributes to existing literature in several ways. First, this study uses an extensive sample from a large urban school district. Second, this study offers a novel methodological approach to examine the effect of PowerUp on test scores by exploiting the variation in the timing of the first log-in to PowerUp during the school year. Third, this study addresses potential sources of bias in the treatment through a quasi-experimental method of difference in differences. To achieve this, we integrate the log-in data with the timing of two progress monitoring tests and an end-of-year test to capture the treatment effect on ELA scores during the school year. Moreover, within this empirical framework, we also study the treatment effect of time spent on PowerUp to understand whether the longer use of PowerUp is associated with an increased effect. Such an approach further strengthens our method as it enables us to not only compare achievement among students who use and do not use PowerUp but also among those who use PowerUp more frequently than their peers. Lastly, we examine the effect of heterogeneity by English language and socioeconomic status.

Results indicate that, on average, PowerUp increases ELA scores by 0.14 SD. Longer use of PowerUp is associated with a larger effect on ELA scores. While the PowerUp platform has no significant effect among students who use less than half an hour weekly, the magnitude of the effect is about 0.20 SD for students who use more than an hour and a half weekly. Also, the magnitude of the effect is larger among ELL and free or reduced-price lunch (FRPL) students. At the extreme, the effect is 0.25 SD among ELL students.

The paper is structured as follows: The next two sections present the conceptual framework of PowerUp as well as specific information about the program. Then, we discuss the literature on computer-assisted instruction (CAI). To follow, we introduce the intervention context and discuss our method. In closure, we share the conclusions of this study and discuss the implications of our results for policy and research.

The Simple View of Reading

While reading researchers outline several components of reading abilities (Grabe & Stoller, 2019; Newton et al., 2018), Gough & Tunmer (1986) proposed a broad reading comprehension model referred to as the Simple View of Reading (SVR). The first reason for developing the SVR model was to potentially identify reading disabilities. A secondary reason for the development of this model was to better understand the role that decoding plays in reading comprehension. The SVR was anchored in two interdependent components, necessary and of equal importance: decoding (word recognition) and listening comprehension. On the one hand, this model outlines the importance of decoding to be paramount to fluency and text comprehension. On the other hand, word recognition is not the sole key ingredient to reading comprehension, as listening comprehension is also mandatory. Listening comprehension refers to the reader's ability to extract meaning from oral language and it stems from vocabulary knowledge (general academic, content-specific), grammatical knowledge, syntactic knowledge, and prior knowledge. It is possible that readers exhibiting inadequate levels of decoding and/or linguistic comprehension are struggling readers. Despite the complexity of each of these two components, positioning reading comprehension as a product of two elements has been embraced in reading pedagogy.

Several studies examined the adequacy of the two-component model suggesting the SVR is a valid model, particularly for early grades (Language and Reading Research Consortium, 2015; Chiu, 2018; Lonigan et al., 2018; Tunmer & Chapman, 2012). For instance, Tunmer and Chapman (2012) examined the SVR with the purpose of expanding the model and found that the fundamental two-component structure of the model should remain intact, however, it appears that linguistic comprehension influences reading comprehension directly and indirectly through decoding. For adolescent readers, Cervetti (2020) suggests that "unpacking the infrastructure of the Simple View" (p. 58) might be necessary for supplementing it with skills aptitudes as background knowledge, vocabulary knowledge, use of strategies, and a far-reaching array of skills such as inference making, disciplinary reading, and reasoning.

PowerUp Design and Organization

In alignment with Cervetti's (2020) guidance, PowerUp is proposing to unpack the Simple View of Reading by delivering instruction across three strands, namely Word Study, Language Comprehension (Grammar), and Comprehension under the following formula: Word Recognition (WR) x Academic Language (AL) = Reading Comprehension (RC). Word Study is intended to increase the accurate recognition of words, which is an aspect of vocabulary knowledge (Nation, 2019) related to reading comprehension (Perfetti, 2007; Stæhr, 2008). As readers focus visually on about 80% of the content words and roughly on 50% of the function words in a text (Pressley, 2006; Stanovich, 2000), it is clear that the accurate recognition of words is crucial to comprehend a text. The Grammar strand, referred to as academic language comprehension includes skills such as syntactic parsing (grammar concepts), semantic proposition formation, awareness of text structure and organizational patterns (sentence construction and structure of different text genres) (Nagy & Townsend, 2012; Schleppegrell, 2012).

PowerUp is a blended learning solution with online and offline components. The online component starts with the placement test after which students can begin working on their individualized paths to reaching reading proficiency aligned with the three strands of the program. The offline component consists of lessons and resources that teachers can use to deliver one-on-one or small-group instruction based on individual students' needs. Teacher resources can be used in addition to the general curriculum (Grant, 2022).

Starting from the three strands, the PowerUp model is organized in a top-down fashion, continuing with zones of skill development within each strand (i.e., foundational, intermediate, advanced), levels (increased with complexity in each zone), activities (available at every level support instruction) and units (skill-specific exercises and tasks). Teachers can monitor students' progress via yearly and monthly progress reports. For instance, the monthly reports contain the range of units to be completed by strands in order to make meaningful progress. However, real-time performance data is available both at the individual and class level so students and teachers can track progress in the content and skills developed, identify struggling areas, and provide information about student usage of the program. The yearly progress provides a big picture by outlining the student's progress based on the results of the placement test, assuming that the students meet the weekly recommended usage.

Effectiveness of Computer-assisted Instruction (CAI) and PowerUP

The effectiveness of CAI has been amply examined in the last decade. For example**,** Dietrichson et al.`s (2020) international systematic review of targeted school interventions to improve reading and math achievement among secondary-level students identified a positive and significant effect size for CAI. The average of 49 effect sizes was 0.17. Compared to other targeted school interventions, CAI had a larger effect size than incentives and coaching personnel but had a smaller effect size than some other instructional components such as progress monitoring, peer-assisted, and small group instruction. When examining specifically the literacy-based component included in the interventions, it appears that decoding interventions revealed the highest effect size (average ES = 0.216), followed by spelling and writing (average ES = 0.167), comprehension (average ES = 0.155), and other domains (average ES = 0.154). It was also identified that fluency interventions (average ES = 0.153) were more effective than multiple reading (average ES = 0.140) and vocabulary interventions (average ES = 0.137).

Moreover, in their meta-analysis of the effectiveness of CAI embedding personalized learning, Major et al. (2021) found a similar effect size to Dietrichson and his colleagues (2020). Based on 16 randomized controlled trials, Major and his colleagues (2021) found that technology-supported personalized learning interventions showed a significant positive effect of 0.16 on students' literacy learning. One important finding from the study was that studies with high personalized features had significantly higher effect sizes than studies with medium personalization features, suggesting that highly adaptive learning approaches have a larger learning impact than lesser adaptive programs. When examining the conditions for delivery of instruction (technology or teacher and technology), it appears that it does not significantly impact the effectiveness of the intervention, however, it looks like technology-only delivery had a higher effect size (0.19) than teacher and technology (0.12). Interestingly, the results of this meta-analysis indicate no statistically significant difference between studies with moderate and strong intensity and duration suggesting that technology-supported personalized learning interventions implemented with strong intensity may yield similar results to personalized learning interventions implemented with moderate intensity.

A closer look at the summary of research from ten research studies evaluating the effectiveness of interventions for struggling adolescents in the United States, compiled by (Boulay et al., 2015), reveals that the READ 180®, a reading intervention providing adaptive, technology-based instruction, produced positive effects on students' reading achievement as revealed by three randomized controlled trials conducted by Loadman et al. (2011), Sprague et al. (2012), and Swanlund et al. (2012). The findings of these three studies involving struggling readers in Massachusetts, Ohio, and Wisconsin, indicate a statistically significant positive effect of READ 180® on students' reading achievement (ES = 0.14 to 0.19). Although READ 180® intervention was recommended to be implemented in 90-minute instructional blocks with the use of computer-assisted instruction for only 20 minutes, the mean attendance was about half of the intervention time (Swanlund et al. 2012).

A more recent study assessed the impact of a multicomponent, multi-strategy computer-assisted reading intervention called the Comprehension Circuit Training (CCT), on the reading comprehension of 228 middle school students completing an average of 13 lessons or approximately 39 hours of intervention (Fogarty et al., 2017). This intervention incorporated an instructional design addressing foundational skills (word identification, vocabulary knowledge, and reading fluency) and higher-level text processing skills (strategies for constructing meaning from text such as self-questioning and inference-making) and targeted instructional delivery with a focus on technology-mediated instruction (video modules). Performance trends estimated with pre- and post-test latent variables showed mixed impact for comprehension (ES = 0.14), sight word efficiency (ES= - 0.04), oral reading fluency (ES = - 0.08), and vocabulary matching (ES = 0.43). The authors asserted that the intervention appeared to be more effective for students with lower entry-level comprehension skills and less beneficial for students with lower-average comprehension skills as measured by the pretest (Fogarty et al., 2017).

Another intervention for struggling adolescents, namely Voyager Passport Reading Journeys® with a computer-based component (i.e., Strategic Online Learning Opportunities®) showed mixed effects on reading achievement depending on the domain assessed: (1) comprehension showed effect sizes of 0.00 to 0.27 and (2) general literacy achievement suggested effect sizes ranging from -0.12 to 0.06 (Dimitrov et al., 2012; Schenck et al., 2012; Vaden-Kiernan et al., 2012). The Voyager Passport Reading Journeys® intervention generally made use of computer-based components for vocabulary practice after whole group, teacher-led instruction. Similarly, Metsala & Kalindi (2022) evaluated the effectiveness of the Ooka Island computer program (i.e., Scholastic F.I.R.S.T.TM, 2017) as a supplement to classroom instruction. They found better outcomes for phonological blending, word reading, reading comprehension, and book-reading levels than the control group that used traditional instruction.

Cumulatively, based on research conducted thus far, there is evidence that implementing computer-assisted instructions (CAI) demonstrates a positive, significant, and often modest effect on standardized test scores (e.g., Dietrichson et al. 2020; Loadman et al., 2011; Loadman et al., 2011; Major et al., 202; Sprague et al., 2012, and Swanlund et al., 2012). Consistent with this literature, existing studies indicated that Lexia PowerUp can improve standardized test scores. Indeed, there is a large number of studies on Lexia Core5, a version of Lexia that has been used for earlier ages. For instance, Grant (2022) identified 23 studies on the effect of Lexia Core 5 on reading achievement. The majority of those studies find positive effects. However, aside from some of the correlational studies that found a positive relationship between PowerUp usage and test scores (Hurwitz & Macaruso, 2021; Macaruso & Rodman, 2009), only a single causal study evaluated PowerUp. In their clustered randomized trial, Hurwitz and Macaruso (2021) examined the PowerUp effect on test scores in a sample of 155 struggling readers enrolled in supplemental literacy classes in two middle schools. They found that the intervention improved the reading scores by 0.36 SD. However, the fairly small sample sizes in their studies limit the external validity. In the study at hand, we use a larger sample size to address the limitations in the existing literature.

Dosage Effect

An important component drawing researchers' attention is the importance of length of time spent engaging with the CAI for student learning. One might expect that longer time spent interacting with CAI is associated with higher student achievement as students would possibly have extended opportunities to benefit from CAI. In their meta-analysis, Flynn et al. (2012) found that the number of sessions across interventions does not predict the treatment effect on reading skills. However, whether the effect of CAI increases with a longer time of use is not well understood (Hurwitz & Macaruso, 2021). While some find a positive relationship (e.g., Cheng et al., 2015; Graham et al., 2022) others find no significant relationship between CAI dosage and achievement (McMurray, 2013). Yet, evidence on this issue is limited partly due to the clustered nature of interventions that typically occur at the classroom or school levels. It appears that researchers typically do not have information about time spent on CAI (Grant, 2022). Instead, they use the duration of treatment across classrooms and schools. This analytical approach, unfortunately, obscured the variation across students. In this study, we address this specific research gap by leveraging a unique, student level log-file that captures time spent on PowerUp. We evaluate the dosage effect, above and beyond the treatment effect by exploiting the log (time) data.

Effect for English Language Learners

Another component receiving consideration from researchers is related to student characteristics (i.e., English Language and socioeconomic status) participating in reading interventions, including CAI. Presumably, CAI can benefit secondary-level ELL students more than non-ELL students perhaps because the adaptive technology-based interventions provide students with the needed support, mainly if those needs include systematic instruction on foundational word reading skills and instruction on word meaning and comprehension of texts which are generally primary level skills. This is because CAI enables ELL students to access additional individualized support that may be difficult to achieve in traditional secondary instructional settings such as foundational skills instruction (Keengwe & Hussein, 2014). CAI can also be beneficial because it allows students to progress at their own pace and strengthen their language skills because ELL students need to learn to speak and read at the same time (Richards-Tutor et al., 2016) meaning that they may need more flexibility when processing and acquiring information. Although CAI can help schools to close the achievement gap between ELL and non-ELL students in reading, evidence on the effect of the CAI among ELL students is limited. In their experimental study among English language learners (ELLs) in kindergarten and first-grade students in a Southwestern U.S. state, Cassady et al. (2018) found that students who received CAI outperformed those who received "business-as-usual" instruction in the standardized reading test. In their quasi-experimental study with 60 kindergarten ELL students, Macaruso and Rodman (2011) found that Lexia Early Reading improved test scores by an effect size of 0.36. In another small sample study with 26 students, Beechler and Williams (2012) found that ELL students with CAI had a greater increase in the number of sight words compared to their peers in traditional instructional settings. Although existing research indicates a positive CAI effect for ELL students, those studies are often small-scale and conducted on younger students, typically in kindergarten or first years of elementary school. This is probably because traditionally, literacy skills are a more important concern in earlier grades as those ages are critical for gaining these skills (e.g., Gersten et al., 2020). In this study, we evaluate the effectiveness of CAI for ELL students in middle and high school grades to contribute to the existing literature.

Effect for Socioeconomically Disadvantaged Students

At the same time, through the same scrutinizing lens, researchers looked at the socioeconomic status of students engaging in reading interventions. Theoretically, CAI can be an effective tool to dismantle systemic inequalities in educational outcomes by socioeconomic status. By enabling a large number of students to access individualized support and feedback via CAI, it can help schools to level the playing field. For instance, in their seminal meta-analysis on academic interventions for low-socioeconomic students (the majority of the student population from the U.S.), Dietrichson et al. (2020) examined the common intervention components to improve test scores. Although CAI was more effective than several other interventions such as after-school programs, coaching/mentoring, psychological/behavioral interventions, and personnel development, the effect size of 0.11 was not statistically significant. A closer look at the specific studies reveals heterogeneous CAI effects for socioeconomically disadvantaged students. For instance, in their randomized controlled trials with 65 middle-school teachers, Given et al., (2008) examined the effectiveness of a program called Fast ForWord finding no significant effect on reading, phonemic awareness, spelling, and language skills. In other experimental studies, however, researchers such as Macaruso et al. (2006) and Messer and Nash (2018) found positive effects among disadvantaged students. Our study contributes to this emerging literature by examining the PowerUp effect heterogeneity by socioeconomic status.

Overall, our study extends the above literature by focusing on the average treatment effect, dosage effect, and effect heterogeneity by ELL and socioeconomic status on English Language Arts (ELA) scores. In the third research question, we also examined the heterogeneity effect by grade. Specifically, we examine the following research questions:

- Research Question 1) What is the effect of PowerUp on ELA achievement?
- Research Question 2) Is time spent on the PowerUp associated with a larger increase in ELA achievement?
- Research Question 3) Does the PowerUp effect on ELA achievement vary by English Language Learner (ELL) status, FRLP status, and grade?

Intervention Context

We conducted the study in a large urban Southeast school district in the 2022-2023 school year. About half of the students are economically disadvantaged (48%); one-third of them are Hispanic (25%) or Black (8%). The district uses PowerUp within its multi-tiered system of support (MTSS). Tier 1 universal core instruction and prevention support is provided to all students. Students who need additional instruction to achieve grade-level academic standards and/or grade-band behavioral benchmarks are provided with access to supplemental Tier 2 interventions. Tier 3 interventions and remediation involve intensive systematic instruction in universal academic skills and academic and social behavior skills. The district leverages reading proficiency in state test scores and teacher discretion to identify Tier 3 students. All students in the lowest proficiency levels (Level 1 and Level 2) in state ELA tests are considered to be Tier 3 students. However, teachers have the autonomy to include other students in the Tier 3 group depending on students' needs and intervention type. At the district level, PowerUp is used as Tier 3 intervention for Tier 3 students (i.e., those who need the most extensive support) in grades 6 through 10. To improve their literacy skills, the district encourages schools to use PowerUp for their Tier 3 students. However, it does not prohibit other students from using PowerUp.

Data and Sample

We use the state ELA test to measure reading scores. This summative test is administered annually. In the school year of 2022/2023, in addition to the standardized summative assessment administered toward the end of the school year, the test offered two progress monitoring cycles, one in the fall and the other one in the winter. Because some students had already started using PowerUp even before the first progress monitoring test, we used the summative test taken in 2021/2022 as a baseline point. Thus, our dataset includes test scores from 2021/2022 (the first test cycle) and three test scores from 2022/2023 (the second, third, and fourth test cycles) for each student in our sample. We standardized test scores within each grade and test cycle with a mean zero and a standard deviation of one to ease the interpretation. The test score data includes the testing time for each student in each test. We also used demographic information from the district office. The demographics include free and reduced lunch (FRL) status, students with disabilities (SWD) status, English language learner (ELL) status, race, gender, and grade.

The sample consisted of all students in grades 6-10 enrolled in the district, excluding charter schools as the district does not oversee the implementation of PowerUp in those schools. The sample size of the study is 26,029. The numbers of students in the treatment and control groups are 9,792 and 16,237, respectively. On average, students in the control group have higher baseline reading scores than the treatment group by about 0.91 SD. The control and treatment groups differ in their demographic characteristics as well. The intervention group has more FRLP and English language learner students. Table 1 reports the characteristics of the control and treatment groups.

Table 1:

Characteristics of treatment and control group students (Mean, (SD))

Variable	Treatment	Control	Difference
Standardized Baseline ELA Score	-0.56 (1)	0.35 (0.81)	-0.91
FRLP	0.62 (0.48)	0.44 (0.5)	0.18
ELL	0.077 (0.27)	0.029 (0.17)	0.05
SPED	0.11 (0.32)	0.088 (0.28)	0.03
Gender	0.44 (0.50)	0.51 (0.50)	-0.07
Asian	0.023 (0.15)	0.052 (0.22)	-0.03
Black	0.093 (0.29)	0.076 (0.26)	0.02
Hispanic	0.29 (0.45)	0.24 (0.43)	0.05
White	0.54 (0.50)	0.58 (0.49)	-0.03
Grade Level	7.7 (1.30)	8.2 (1.40)	-0.48
Observations	9,792	16,237	

In addition to the test score and student demographics data, we used student-level PowerUp data which includes information on weekly average usage, total times used in the school year, and number of days since first use. Table 2 reports the descriptive statistics of these variables.

In a typical week, students spent 57.64 minutes on PowerUp and an average of 16.24 hours on PowerUp during the school year. Examining the total time spent on PowerUp, we found that some students registered but never used the platform, therefore, we included students who never used PowerUp in the control group. The distribution of the usage variables is reported in Appendix 1 and 2.

Table 2:
Descriptive statistics of usage variables

	Mean	SD	Median	Min	Max
Weekly average usage (minutes)	57.64	32.23	59	0.01	716.00
Total usage (hours)	16.24	16.86	9.1	0.01	124.75
First log-in in the school year	59.83	71.42	22	1	298

Method

Estimation Strategy

We use the difference in differences (DID) approach to estimate the treatment effect of PowerUp on the ELA scores. We leverage variation in the timing of the students' first use of PowerUp, explained subsequently. Our DID estimation relies on the group-time average treatment effect (ATT (g, t)) proposed by Callaway and Sant'Anna (2021). "Time" is one of three test cycles in post-treatment period. "Group" denotes the first test cycle a student is treated. Our main ATT is the average of those ATT (g, t) estimates (Callaway & Sant'Anna, 2021; Goodman-Bacon, 2021).

Our first research question examines the effect of PowerUp on ELA scores. To answer our first research question, we use the following DID strategy.

$$y_{it} = \beta\, Treat_{it} * Post_{it} + X_{it} + \mu_i + \vartheta_t + \varepsilon_{it} \quad \text{(I)}$$

where y_{it} is within grade and test cycle standardized ELA test score for student *i* and test cycle *t*. β is the coefficient of interest, the effect of the PowerUp on the ELA score. *Treat* denotes whether student i has ever used PowerUp. $Post$ *is* a dummy variable that shows whether student i used PowerUp before test cycle *t*. Thus, $Treat_{it} * Post_{it}$ captures the treatment status. Students who never used PowerUp, including those who registered but not used it, were included in the control group in all test cycles (*t*). Those who used PowerUp at least once were in the treatment group in test cycle *t* if they used it before test cycle *t*. For instance, a student who started using PowerUp between the second and third test cycles will be in the control group in the first and second test cycles, and in the treatment group in the third and fourth test cycles. Thus, our specification leverages the variation in treatment timing. X_{it} is a vector of student demographics including disability status, English language learner status, free or reduced lunch status as a proxy for socioeconomic status, homeless status, race, and special education status. The specification exploits student-fixed effects (μ_i) to capture student-specific time-invariant characteristics. We also use the test-cycle fixed effect (ϑ_t) to control for factors changing each test cycle that are common to all students for a given cycle. We clustered the standard errors at the student level to account for serial correlation and general heteroskedasticity (Abadie et al., 2017).

Our second research question examines whether time spent on PowerUp is associated with larger increases in reading achievement. We operationalize the time spent on PowerUp with two variables. The first one is the weekly average use and the second one is the total time spent during the school year. We plotted the first day of sign-in across the school year in Appendix 2.

To estimate the effect of time spent on PowerUp, we fit regression (I) in different sub-samples that we created by 15-minute intervals of the weekly average use. We fit the regression (I) for each of those groups. In all models, the control group was the same: those who were never treated and those who were not yet been treated (i.e., those who will be in the treatment group in one of the following test cycles). By comparing the regression coefficients from those models, we examined whether the time spent on PowerUp is associated with a larger treatment effect. We followed the same approach for the total time spent on PowerUp. We created sub-groups by using 2-hour intervals. Our third research question examines whether the treatment effect on reading achievement varies by ELL status, FRLP status, and grade. We examined these heterogeneity effects by using interaction terms between treatment and each of these variables. We reported these results in Table 4.

Parallel Trend Assumption

The conventional DID estimator strictly relies on parallel trend assumption. That is, in the absence of treatment, the outcome variable would have a similar trend in control and treated units. Because we do not observe the same individual`s outcome both in treatment and control conditions, there is no statistical way to test the plausibility of the parallel trend assumption (Athey & Imbens, 2022). However, it is possible to check whether the parallel trend assumption holds in periods before treated units actually become treated. Although it is just a pre-test, evidence from this test offer a perspective on the credibility of parallel trend assumption (Callaway & Sant'Anna, 2021). The result from this pre-test imply that parallel trend assumption is credible. The *p*-value for the pre-test of parallel trends assumption is 0.64429 revealing that the pre-trend is not statistically different in treatment and control group. In addition, the event study results plotted in Figure 1 suggest that the parallel trend assumption holds in the pre-treatment period. Based on these results, we assumed that the parallel trend assumption is credible.

Robustness and Sensitivity

Group-time average treatment effect (ATT (g, t)) may be sensitive to the way it has been aggregated from group-time average treatment effect (Callaway & Sant'Anna, 2021). We examined whether ATT is sensitive to alternative aggregation methods. We employed three aggregation methods. First, we used "simple" aggregation which is the weighted average of all group-time average treatment effects with weights proportional to group size. Second, we used "group" aggregation which computes the average treatment effects across different groups. Third, we used "dynamic aggregation" similar to an event study estimation which computes average effects across different lengths of exposure to the treatment and is.

We examined whether the estimations are sensitive to the three alternative estimation methods. Inverse probability weighting uses the propensity to treatment and takes its inverse as a weight factor in analysis (Abadie, 2005). Another approach, outcome regression, imposes the conditional expectation of the outcome trend for the control group (Heckman et al., 1997: 1998). Sant'Anna and Zhao (2020) and Callaway and Sant'Anna (2021) propose "doubly robust" estimators that retrieve a consistent ATT if either a propensity score or outcome regression models are accurately identified. We rely on doubly robust estimators since they are based on less stringent assumptions. However, we evaluated whether the results were sensitive to the other two estimation methods. We reported inverse probability weighting, outcome regression, and doubly robust estimators to check the sensitivity of the results.

Finally, we explored whether our estimation is sensitive to sample attrition. Across four test cycles, 11.62% of students do not have test scores at least in one test cycle. To address this issue, we fit our difference in difference models first by "not allowing unbalanced panel" and then by "allowing unbalanced panel." Thus, we were able to compare the results from these models that use data from students who have test scores in all test cycles and those who have test scores in at least two cycles, respectively.

Overall, the combination of the three estimation methods, three types of aggregated treatment effect parameters, and balanced or unbalanced panel resulted in 18 differences in difference models. Our main estimate is Model 1 but we compared results across those models to check the robustness and sensitivity of the results.

Results

We report our difference in differences estimation for the PowerUp effect on ELA scores in Table 3. Our main estimate is reported in Model 1 which relies on doubly robust estimator, does not allow unbalanced panel, and uses "simple" aggregation method. Model 1 indicates that the average treatment effect of PowerUp on ELA score is 0.1389 SD (0.0084). The lower and upper bounds of 95% confidence interval are 0.1225 SD and 0.1552 SD, respectively.

Table 3:

Difference in differences estimations for the treatment effect on ELA scores

Model	Estimation Method	Unbalanced Panel Allowed	Type of aggregated treatment effect parameter	Average Treatment Effect on Treated (ATT)	Std. Error.	[95% Conf. Interval]	
1	Doubly Robust Estimator	No	Simple	0.1389***	0.0084	0.1225	0.1552
2			Group	0.1215***	0.0079	0.106	0.137
3			Dynamic	0.1443***	0.0088	0.1269	0.1616
4		Yes	Simple	0.1687***	0.0089	0.1513	0.1861
5			Group	0.1466***	0.0086	0.1297	0.1636
6			Dynamic	0.1756***	0.0093	0.1573	0.1939
7	Inverse Probability Weighting	No	Simple	0.1379***	0.0087	0.1208	0.155
8			Group	0.1207***	0.0084	0.1043	0.1371
9			Dynamic	0.1433***	0.0092	0.1252	0.1613
10		Yes	Simple	0.1919***	0.0100	0.1723	0.2116
11			Group	0.1663***	0.0089	0.1488	0.1837
12			Dynamic	0.1998***	0.0099	0.1804	0.219
13	First Step Regression Estimator	No	Simple	0.1392***	0.0073	0.125	0.1535
14			Group	0.1217***	0.0078	0.1064	0.1370
15			Dynamic	0.1444***	0.0078	0.1292	0.1596
16		Yes	Simple	0.1705***	0.0085	0.1539	0.1871
17			Group	0.1480***	0.0079	0.1326	0.1634
18			Dynamic	0.1772***	0.009	0.1596	0.1948

The “group” and “dynamic” approaches to aggregate the group-time treatment effect reveal a similar average treatment effect. The magnitude of the effect in Model 2 and Model 3 are 0.1215 and 0.1443, respectively which are similar to the main results reported in Model 1 (i.e., 0.1389). These results suggest that the average treatment effect is not sensitive to the way group-time treatment effects are aggregated.

Models 4, 5, and 6 indicate that when unbalanced panel is allowed, the magnitude of the treatment effect increases marginally. For instance, Model 4 that uses the similar estimation method and type of aggregated treatment effect approach with the main estimates in Model 1 reveal an average treatment effect of 0.1687 SD (0.0089), about 0.03 SD increase relative to the main estimation.

Model 7 through 12 uses inverse probability weighting. The average treatment effect in Model 7 that uses the same approach with Model 1 except for the estimation method, is 0.1379 SD (0.0087). This estimation is similar to the estimates in Model 1 suggesting that the results from doubly robust and inverse probability weighting are similar. In terms of panel data structure, the magnitude of the effect is marginally high when unbalanced panel is allowed as reported in Model 11, 12, 13.

Models 13 through 18 reveal similar average treatment effects. For instance, the effect in Model 13 is 0.1392 (0.0073) which is similar to the results in Model 1. Comparable to the other estimation methods, the unbalanced panel results have marginally larger treatment effect in first step regression estimator.

Overall, the results across 18 difference in differences estimation are consistent. These results suggest that the main estimate reported in Model 1 is strong against the alternative estimation method, panel data structure, and the type of aggregated treatment effect parameters, even though the magnitude of the effect changes slightly across models.

Our next estimation relies on the event-study approach which is essentially the same approach as the "dynamic" approach reported in Table 3 (Callaway & Sant'Anna, 2021). As the estimates across the estimation method and panel data structure in Table 3 were quite similar, we focused on the event study estimation that relies on doubly robust estimation and balanced panel (i.e., Model 3 in Table 3). Figure 1 shows the event-study estimation for PowerUp effect on ELA scores. This estimation indicates that the ELA scores increase in the first test cycle that follows treatment. Then, it continues to increase in the later test cycles. The statistically insignificant difference in ELA scores in pre-treatment period between control and treatment groups became about 0.11 SD, 0.13 SD and 0.18 SD over three test cycles, respectively. This "dynamic" treatment effect from the event study estimation led to 0.1443 average treatment effect across three test cycles which is similar to our main "simple" aggregated average treatment effect of 0.1389 reported in Model 1.

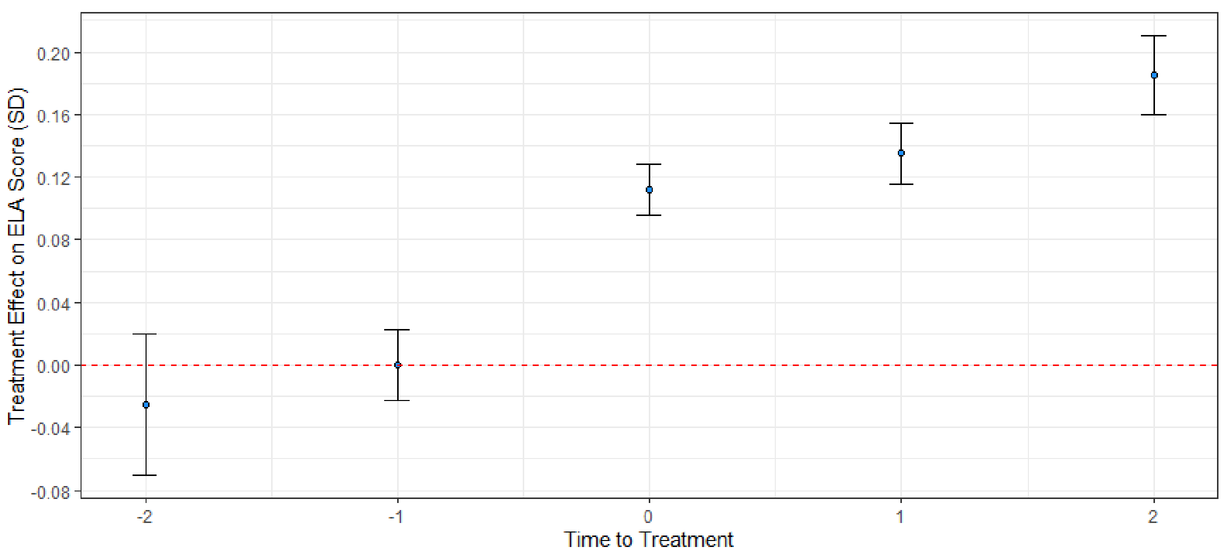


Figure 1: *Event study estimation for treatment effect on ELA scores*

In the second research question, we examined whether time spent on the PowerUp was associated with a larger increase in ELA scores. We first examined the treatment effect by average weekly usage. We illustrated the point estimates and 95% confidence intervals in Figure 2.

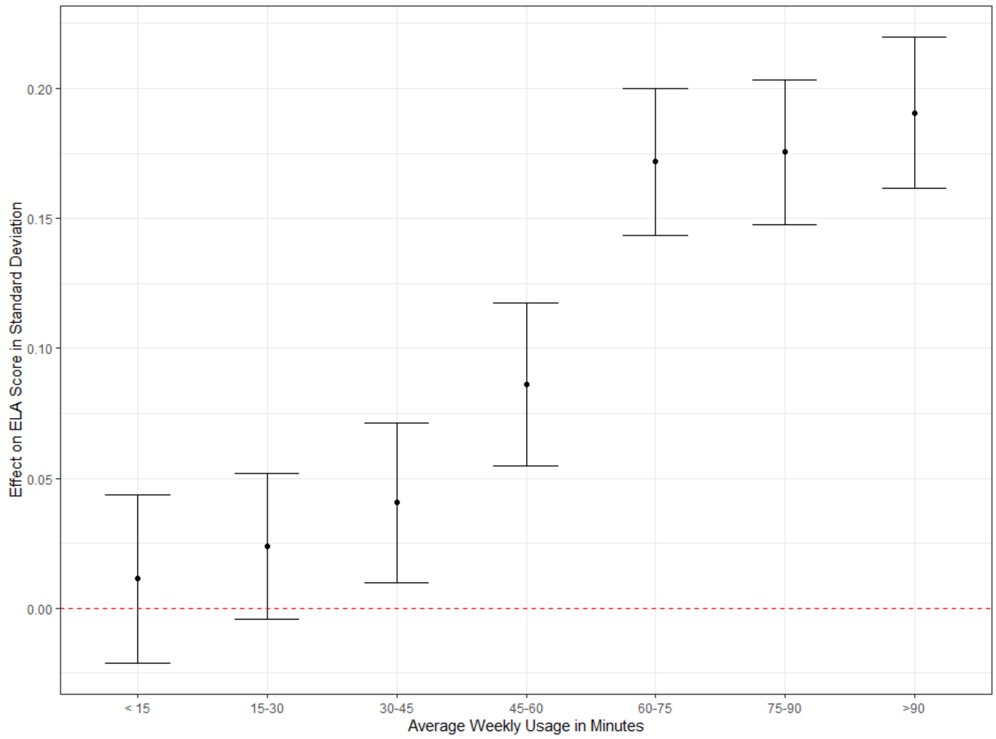


Figure 2: *Treatment effect on ELA scores by average weekly usage*

The magnitude of the positive effect on ELA scores generally increases as average weekly usage rises. There is no significant effect for students who use less than half an hour (i.e., < 15 minutes and 15-30 minutes in Figure 2) whereas the effect is close to 0.20 SD for students who use more than 90 minutes a week. These results suggest that to benefit from PowerUp, on average, students should spend at least half an hour a week, and the more they use it the more their ELA scores improve.

We also evaluated whether the total time spent on PowerUp during the school year is associated with the larger increase in ELA scores. We plotted the point estimates and confidence intervals in Figure 3. There is no significant effect among students who used PowerUp for less than five hours during the school year. The effect is statistically significant for all other student groups (i.e., those who use more than five hours in the school year). The magnitude of the effect generally increases as the number of units completed rises. For the students who used the platform for more than 30 hours during the school year, the magnitude is about 0.30 SD.

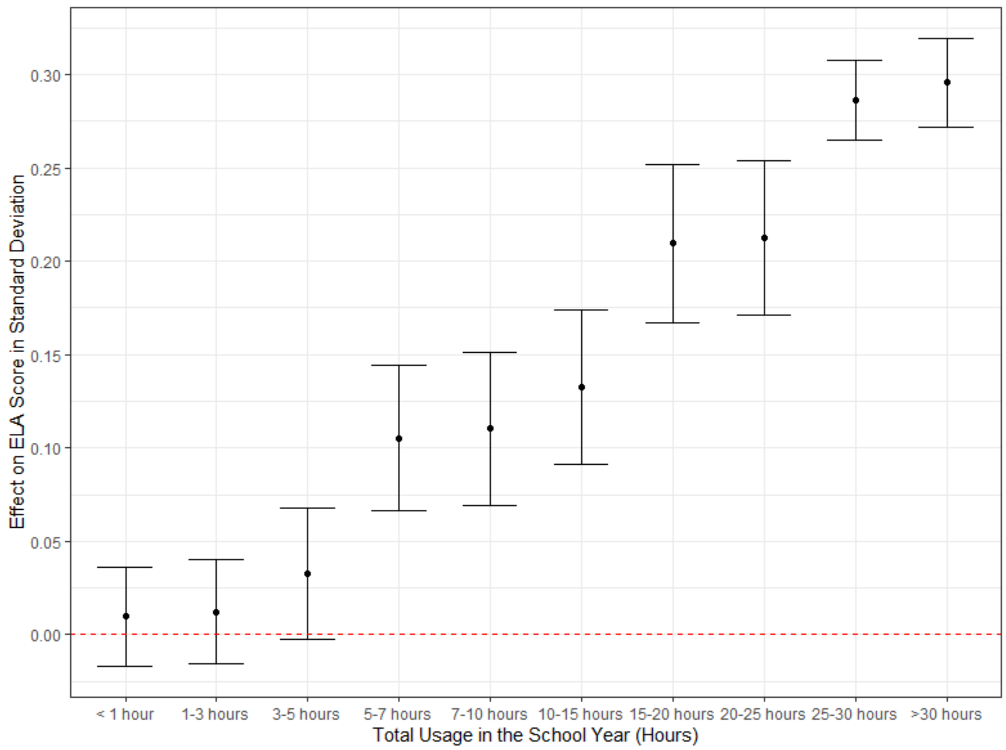


Figure 3: *Treatment effect on ELA score by total usage in the school year*.

Generally, the results from the total time spent show a similar trend to that of the average weekly time spent on PowerUp. However, the positive trend in weekly time spent appears to be saturated after an hour whereas the positive trend continues to increase across all categories of total usage in the school year. These results imply that the effect accrues with the longer days of use suggesting that students who start using PowerUp early in the school year and use consistently more than an hour in a week can experience higher levels of gain in test scores.

In research question 3, we examined whether the effect varies by ELL and FRLP status, and by grade. The information in Table 4 indicates that the effects are larger for ELL students and socioeconomically disadvantaged students, however, there are no indications of heterogeneity across grades.

Table 4:
Treatment effect heterogeneity by sub-groups

	ELL	FRLP	Grade
Treatment	0.1157*** (0.0067)	0.0762*** (0.0094)	0.1522*** (0.0351)
Treatment* ELL	0.1341*** (0.0254)		
Treatment* FRLP		0.0791*** (0.0116)	
Treatment* Grade			-0.0035 (0.0045)
Observation	101,091	101,091	101,091
R^2	0.83869	0.83871	0.83860

Importantly, there is a large difference in the magnitude of the effect by ELL status. The statistically significant treatment and interaction effects lead to an effect of about 0.25 SD for ELL students (0.1157 SD + 0.1341 SD = 0.2498 SD). Likewise, the statistically significant treatment and interaction effects indicate a treatment effect of about 0.16 SD (0.0762 SD + 0.0791 SD) among FRLP students. This effect is slightly higher than the average treatment effect. Overall, these results suggest that PowerUp is relatively more effective in improving test scores among ELL and economically disadvantaged students.

Discussion and Conclusion

Due to heightened concerns about the low reading proficiency of middle and high school students (NAEP, 2022a; NAEP, 2022b), school districts around the country made increasing efforts to address the literacy gap for secondary-level students using computer-assisted instruction (CAI). Technology-based literacy interventions implemented with secondary-level students such as Lexia® PowerUp Literacy® showed positive effects on student learning. In our study, DID results reveal that the average treatment effect of PowerUp is 0.14 SD. These results are consistent with the results published by Dietrichson et al. (2020) who looked at 14 CIA studies with 49 effect sizes to determine an average effect size of 0.17. Likewise, the results are aligned with the findings of Major et al., (2021) determining that studies examining literacy computer-assisted learning (CAL) interventions produced a significant positive effect of 0.16 on students' literacy learning. However, the magnitude of the effect in this study is lower than that obtained by Hurwitz & Macaruso (2021) whose study yielded an overall PowerUp treatment effect of 0.36 in their cluster randomized controlled trial with 155 struggling middle school students. These results suggest that the magnitudes of the positive PowerUp effect on test scores may be smaller when it is scaled up to larger student populations. More evidence will be needed to answer these questions. Future research can look at the PowerUp effect in other school districts to contribute to this growing literature.

A more in-depth look at the time spent on PowerUp showed increased growth in ELA scores in addition to the treatment effect. Although there was evident growth at 60 minutes per week, students who spent more than an hour on PowerUp had higher ELA scores than students who spent less time. This evidence contradicts Major et al.'s (2021) findings suggesting there is no statistically significant difference in the CAI effect by the duration of treatment. The results of this study showed that time spent on PowerUp was associated with an additional increase in effect per week to the treatment effect at different threshold levels (i.e., below 15 minutes, 30 minutes, etc.). The difference may be due in part to the way usage is measured across studies. Our study offered a nuanced measure of usage as we were able to link the log-in and test score data whereas most studies exploit usage data at the school or classroom level by clustered intervention design (Major et al., 2021). We encourage researchers to exploit log files to improve the emerging understanding of time spent and intervention effect.

Upon this examination, it was revealed that the effects of PowerUp were larger for students identified as ELL (0.25 SD) and socioeconomically disadvantaged students (0.16 SD). The results of this study examining in more detail the ELL group are consistent with Cassady et al., (2018), Macaruso & Rodman (2011), and Beechler & Williams (2012) who studied the CAI effect among younger ELL students. Our study extends findings from those studies to middle and high school students implying that CAI offers an opportunity for ELL students to catch up with their peers even in later grades. Also, we did not find any heterogeneity effect across grades. These results are promising for schools that continue supporting their ELL students with literacy interventions (including foundational skills) regardless of their grades. The magnitude of the effect for economically disadvantaged students is only marginally larger than the overall treatment effect.

Overall, the findings of this study suggest that depending upon usage, PowerUp supports secondary-level students in improving their reading achievement, particularly among ELL students. Even though the average PowerUp effect of 0.14 SD is not large in magnitude, this effect is educationally meaningful, especially at a large scale and with a larger magnitude for ELL students. In his recent review of empirical distributions of effect-size estimates for educational interventions, Kraft (2023) shows that about one-third of the educational interventions evaluated through randomized controlled trials have less than 0.05 SD effect on standardized achievement outcomes. Even though comparing findings from our quasi-experimental study to randomized controlled trials has some limitations in interpreting the effect size, Kraft's (2023) results suggest that most studies on larger scales have marginal effect sizes. Therefore, the identified effect size of the PowerUp platform deserves attention and needs to be studied in other contexts to understand whether such an effect is consistent across different educational settings.

As previously described, PowerUp is designed as a blended learning model that incorporates online and offline components. The online component enables learners to be independent in their learning as they continue on their paths towards reading proficiency. The offline component involves teacher-led instruction using Lexia lessons. One acknowledged limitation is that this study measured the effectiveness of PowerUp on student reading achievement by examining usage data of the online component and did not evaluate the offline component. The offline component of PowerUp is yet to be explored.

References

Abadie, A. (2005). Semiparametric Difference-in-Differences Estimators. *The Review of Economic Studies*, *72*(1), 1–19. https://doi.org/10.1111/0034-6527.00321

Beechler, S., & Williams, S. (2012). Computer Assisted Instruction and Elementary ESL Students in Sight Word Recognition. *International Journal of Business and Social Science*, *3*(4).

Boulay, B., Goodson, B., Frye, M., Blocklin, M., & Price, C. (2015). Summary of Research Generated by Striving Readers on the Effectiveness of Interventions for Struggling Adolescent Readers. NCEE 2016-4001. *National Center for Education Evaluation and Regional Assistance*.

Callaway, B., & Sant'Anna, P. H. C. (2021). Difference-in-Differences with multiple time periods. *Journal of Econometrics*, *225*(2), 200–230. https://doi.org/10.1016/j.jeconom.2020.12.001

Cassady, J. C., Smith, L. L., & Thomas, C. L. (2018). Supporting emergent literacy for English language learners with computer-assisted instruction. *Journal of Research in Reading*, *41*(2), 350–369. https://doi.org/10.1111/1467-9817.12110

Cervetti, G. N. (2020). The nature and development of reading for understanding. *Reaping the Rewards of the Reading for Understanding Initiative*, 41–66.

Cheng, M.-T., Lin, Y.-W., & She, H.-C. (2015). Learning through playing Virtual Age: Exploring the interactions among student concept learning, gaming performance, in-game behaviors, and the use of in-game characters. *Computers & Education*, *86*, 18–29. https://doi.org/10.1016/j.compedu.2015.03.007

Chiu, Y. D. (2018). The Simple View of Reading Across Development: Prediction of Grade 3 Reading Comprehension From Prekindergarten Skills. *Remedial and Special Education*, *39*(5), 289–303. https://doi.org/10.1177/0741932518762055

Dietrichson, J., Filges, T., Klokker, R. H., Viinholt, B. C. A., Bøg, M., & Jensen, U. H. (2020). Targeted school-based interventions for improving reading and mathematics for students with, or at risk of, academic difficulties in Grades 7–12: A systematic review. *Campbell Systematic Reviews*, *16*(2). https://doi.org/10.1002/cl2.1081

Dimitrov, D., Jurich, S., Frye, M., Lammert, J., Sayko, S., & Taylor, L. (2012). Year One Evaluation Report/Impact Study: Illinois Striving Readers. *RMC Research Corporation*.

Flynn, L. J., Zheng, X., & Swanson, H. L. (2012). Instructing Struggling Older Readers: A Selective Meta-Analysis of Intervention Research. *Learning Disabilities Research & Practice*, *27*(1), 21–32. https://doi.org/10.1111/j.1540-5826.2011.00347.x

Fogarty, M., Clemens, N., Simmons, D., Anderson, L., Davis, J., Smith, A., Wang, H., Kwok, O., Simmons, L. E., & Oslund, E. (2017). Impact of a Technology-Mediated Reading Intervention on Adolescents' Reading Comprehension. *Journal of Research on Educational Effectiveness*, *10*(2), 326–353. https://doi.org/10.1080/19345747.2016.1227412

Gersten, R., Haymond, K., Newman-Gonchar, R., Dimino, J., & Jayanthi, M. (2020). Meta-Analysis of the Impact of Reading Interventions for Students in the Primary Grades. *Journal of Research on Educational Effectiveness*, *13*(2), 401–427. https://doi.org/10.1080/19345747.2019.1689591

Given, B. K., Wasserman, J. D., Chari, S. A., Beattie, K., & Eden, G. F. (2008). A randomized, controlled study of computer-based intervention in middle school struggling readers. *Brain and Language*, *106*(2), 83–97. https://doi.org/10.1016/j.bandl.2007.12.001

Goodman-Bacon, A. (2021). Difference-in-differences with variation in treatment timing. *Journal of Econometrics*, *225*(2), 254–277. https://doi.org/10.1016/j.jeconom.2021.03.014

Gough, P. B., & Tunmer, W. E. (1986). Decoding, reading, and reading disability. *Remedial and Special Education*, *7*(1), 6–10.

Grabe, W., & Stoller, F. L. (2019). *Teaching and Researching Reading* (3rd ed.). Routledge. https://doi.org/10.4324/9781315726274

Graham, S., Silva, M., & Restrepo, M. A. (2022). Reading intervention research with emergent bilingual students: A meta-analysis. *Reading and Writing*. https://doi.org/10.1007/s11145-022-10399-8

Grant, G. V. C. (2022). *A Causal-Comparative Study of the Supplemental Lexia Core5 Reading Computer-Assisted Instruction Program Intervention for Improving the Reading Achievement of Elementary School Students with Disabilities* [PhD Thesis]. Northcentral University.

Hurwitz, L. B., & Macaruso, P. (2021). Supporting struggling middle school readers: Impact of the Lexia® PowerUp Literacy® program. *Journal of Applied Developmental Psychology*, *77*, 101329. https://doi.org/10.1016/j.appdev.2021.101329

Hurwitz, L. B., Macaruso, P., Thang, S., & Studwell, J. (2022). Bolstering Middle School Students' Component Reading Skills: An Evaluation of the Lexia® PowerUp Literacy® Blended Learning Program. *Computers in the Schools*, *39*(1), 80–97. https://doi.org/10.1080/07380569.2022.2037298

Kraft, M. A. (2023). The Effect-Size Benchmark That Matters Most: Education Interventions Often Fail. *Educational Researcher*, 0013189X231155154.

Kraft, M., Edwards, D., & Cannata, M. (2022, November 17). *Taking High-Dosage Tutoring to Scale: Program Implementation, Causal Effects, and Lessons Learned*. 2022 APPAM Fall Research Conference. https://appam.confex.com/appam/2022/meetingapp.cgi/Paper/45795

Language and Reading Research Consortium. (2015). Learning to Read: Should We Keep Things Simple? *Reading Research Quarterly*, *50*(2), 151–169. https://doi.org/10.1002/rrq.99

Loadman, W. E., Moore, R. J., Ren, W., Zhu, J., Zhao, J., & Lomax, R. (2011). Striving Readers Year 5 Project Evaluation Report: Ohio. An Addendum to the Year 4 Report. *Ohio State University*.

Lonigan, C. J., Burgess, S. R., & Schatschneider, C. (2018). Examining the Simple View of Reading With Elementary School Children: Still Simple After All These Years. *Remedial and Special Education*, *39*(5), 260–273. https://doi.org/10.1177/0741932518764833

Macaruso, P., Hook, P. E., & McCabe, R. (2006). The efficacy of computer-based supplementary phonics programs for advancing reading skills in at-risk elementary students. *Journal of Research in Reading*, *29*(2), 162–172. https://doi.org/10.1111/j.1467-9817.2006.00282.x

Macaruso, P., & Rodman, A. (2009). Benefits of computer-assisted instruction for struggling readers in middle school. *European Journal of Special Needs Education*, *24*(1), 103–113. https://doi.org/10.1080/08856250802596774

Macaruso, P., & Rodman, A. (2011). Benefits of Computer-Assisted Instruction to Support Reading

Acquisition in English Language Learners. *Bilingual Research Journal*, *34*(3), 301–315. https://doi.org/10.1080/15235882.2011.622829

Major, L., Francis, G. A., & Tsapali, M. (2021). The effectiveness of technology-supported personalised learning in low- and middle-income countries: A meta-analysis. *British Journal of Educational Technology*, *52*(5), 1935–1964. https://doi.org/10.1111/bjet.13116

McMurray, S. (2013). An evaluation of the use of Lexia Reading software with children in Year 3, Northern Ireland (6-to 7-year olds). *Journal of Research in Special Educational Needs*, *13*(1), 15–25.

Messer, D., & Nash, G. (2018). An evaluation of the effectiveness of a computer-assisted reading intervention. *Journal of Research in Reading*, *41*(1), 140–158.

Metsala, J. L., & Kalindi, S. C. (2022). The Effects of a Computer-Based Early Reading Program on the Literacy Skills of Kindergarten Students. *Computers in the Schools*, *39*(4), 373–393. https://doi.org/10.1080/07380569.2022.2127344

Nagy, W., & Townsend, D. (2012). Words as Tools: Learning Academic Vocabulary as Language Acquisition. *Reading Research Quarterly*, *47*(1), 91–108. https://doi.org/10.1002/RRQ.011

Nation, P. (2019). The different aspects of vocabulary knowledge. In *The Routledge Handbook of Vocabulary Studies* (pp. 15–29). Routledge.

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* (Working Paper No. 27476). National Bureau of Economic Research. https://doi.org/10.3386/w27476

Perfetti, C. (2007). Reading Ability: Lexical Quality to Comprehension. *Scientific Studies of Reading*, *11*(4), 357–383. https://doi.org/10.1080/10888430701530730

Pressley, M. (2006). *Reading instruction that works: The case for balanced teaching, 3rd ed* (pp. x, 469). Guilford Press.

Richards-Tutor, C., Baker, D. L., Gersten, R., Baker, S. K., & Smith, J. M. (2016). The Effectiveness of Reading Interventions for English Learners: A Research Synthesis. *Exceptional Children*, *82*(2), 144–169. https://doi.org/10.1177/0014402915585483

Robinson, C. D., Kraft, M. A., Loeb, S., & Schueler, B. E. (2021). Accelerating Student Learning with High-Dosage Tutoring. EdResearch for Recovery Design Principles Series. *EdResearch for Recovery Project*.

Schenck, A., Jurich, S., Frye, M., Lammert, J., & Sayko, S. (2012). Evaluation Report/Impact Study: Virginia Striving Readers Intervention Initiative (VSRII). *RMC Research Corporation*.

Schleppegrell, M. J. (2012). Academic Language in Teaching and Learning: Introduction to the Special Issue. *The Elementary School Journal*, *112*(3), 409–418. https://doi.org/10.1086/663297

Sprague, K., Zaller, C., Kite, A., & Hussar, K. (2012). Springfield-Chicopee School Districts Striving Readers (SR) Program. Final Report Years 1-5: Evaluation of Implementation and Impact. *Education Alliance at Brown University*.

Stæhr, L. S. (2008). Vocabulary size and the skills of listening, reading and writing. *Language Learning Journal*, *36*(2), 139–152.

Stanovich, K. E. (2000). *Progress in understanding reading: Scientific foundations and new frontiers*. Guilford Press.

Swanlund, A., Dahlke, K., Tucker, N., Kleidon, B., Kregor, J., Davidson-Gibbs, D., & Hallberg, K. (2012). Striving Readers: Impact Study and Project Evaluation Report–Wisconsin Department of Public Instruction (with Milwaukee Public Schools). *American Institutes for Research*.

Tunmer, W. E., & Chapman, J. W. (2012). The simple view of reading redux: Vocabulary knowledge and the independent components hypothesis. *Journal of Learning Disabilities*, *45*(5), 453–466. https://doi.org/10.1177/0022219411432685

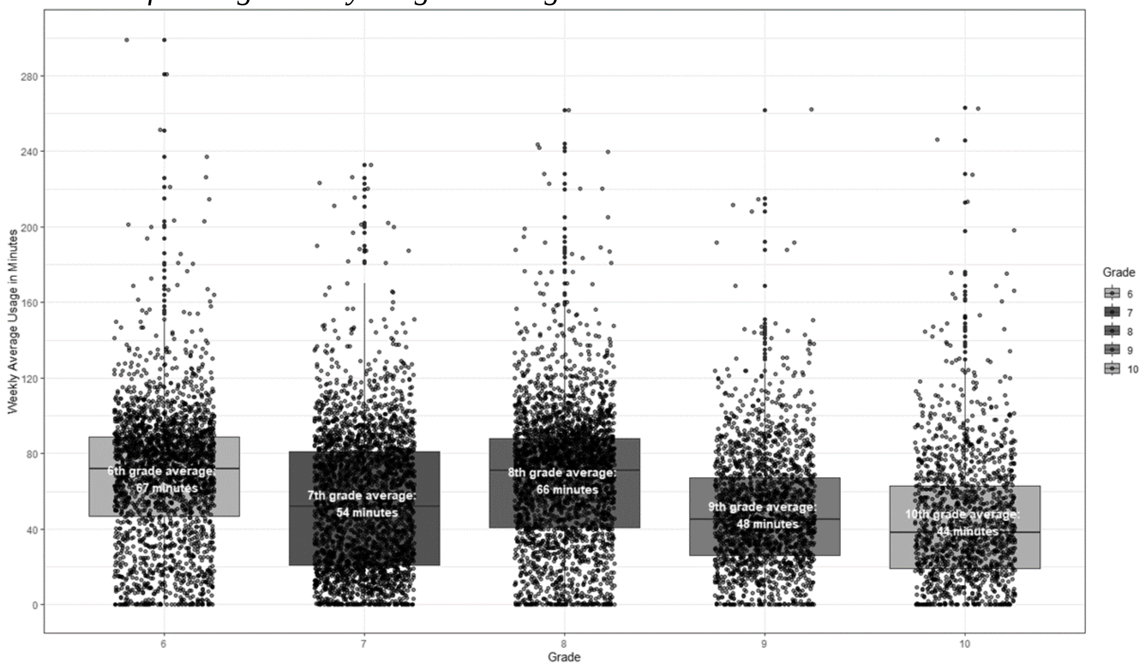
Vaden-Kiernan, M., Caverly, S., Bell, N., Sullivan, K., Fong, C., Atwood, E., Borman, G., Park, S. J., & Jones, D. H. (2012). Louisiana Striving Readers: Final Evaluation Report.

Vandergrift, J. M. N., Dana R. Ferris, Christine C. M. Goh, William Grabe, Fredricka L. Stoller, Larry. (2018). *Teaching English to Second Language Learners in Academic Contexts: Reading, Writing, Listening, and Speaking*. Routledge. https://doi.org/10.4324/9781315626949

Appendices

Appendix 1:

Distribution of average weekly usage across grades



Appendix 2:

Distribution of date first sign-in across the school year