CPRE
CONSORTIUM FOR POLICY RESEARCH IN EDUCATION
TEACHERS COLLEGE
COLUMBIA UNIVERSITY

Amira Learning in Savannah-Chatham County Public Schools

October 2021

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 and is widely considered to be among the premier university-based research organizations focused on education policy and evaluation. 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

During the 2020-21 school year, Savannah-Chatham County Public School (SCCPSS) partnered with *Amira Learning*, an online assessment and supplemental practice software focused on early literacy development. This report examines the association between *Amira Learning* participation and student literacy development. We consistently found that increased usage was associated with stronger literacy development across all outcomes. These associations were strongest during the fall-to-winter period and much weaker during the winter-to-spring period (except for Kindergarteners, for whom the pattern reversed). Overall, these results consistently suggest that students who received a greater percentage of the recommended dosage of *Amira Learning* practice sessions gained more than their peers who received smaller dosages.

Introduction

This report describes student literacy outcomes associated with the implementation of *Amira Learning* in Savannah-Chatham County Public School (SCCPSS). *Amira Learning*, an online assessment and supplemental practice software, uses artificial intelligence (AI) to assess oral reading fluency and provide reading practice. During the 2020-21 academic year, kindergarten through third grade students participated in *Amira Learning*. Beginning in mid-September and continuing through May, schools and teachers were asked to encourage students to complete three to five tutor sessions per week, equivalent to approximately 30 minutes a week. Students initially engaged with *Amira Learning* in a remote setting due to the COVID-19 pandemic, but transitioned back to in-person, classroom settings mid-year. In this report, we address the following research questions:

1. What is the relationship between participation in *Amira Learning* tutor sessions and student literacy development during the 2020-21 school year?
2. How does the link between usage and literacy growth vary by student race/ethnicity, gender, special education status, and dyslexia risk status?
3. What is the relationship between student literacy growth and software usage patterns?

Data and Methods

Analytic Sample

Savannah-Chatham County Public Schools, the tenth largest district in Georgia, enrolls approximately 19,000 elementary students. District enrollment is 57% Black, 23% white, and 12% Hispanic. While approximately 5,000 first and second grade SCCPS students participated in *Amira Learning* at some point during the school year, there was significant variation in student usage, which we discuss in more detail below. Our analytic sample, which is limited to students who took a series of fall, winter, and spring assessments and engaged in at least one tutor session, includes 2,305 first and second grade students in 29 schools.[1] As shown in Table 1, our sample was over 50% Black, roughly one-quarter white, and 13% Hispanic, suggesting that *Amira Learning* participants were representative of the SCCPSS student population in terms of racial/ethnic background. Slightly fewer than 10% of *Amira Learning* students received special education services, and approximately 7% of *Amira Learning* students were at risk for Dyslexia.

[1] We also provide separate sub-analyses for students in kindergarten (n=217) and third grade (n=1,182).

Table 1. Characteristics of First and Second Grade Amira Learning Students (*n*=2,305)

Grade	
% First	47.9
% Second	52.1
Race/Ethnicity	
% American Indian	0.4
% Asian	2.4
% Black	50.9
% Hispanic	12.9
% Native Hawaiian/PI	0.4
% White	27.1
% Multiracial	5.9
% Female	51.6
% Students with Disabilities	9.9
% Dyslexia at Risk[2]	7.0

Measures

***Amira Learning* Usage.** To capture *Amira Learning* usage we use continuous measures of the number of weeks students completed at least one *Amira Learning* tutor session. We created separate indicators for the fall-to-winter, winter-to-spring, and fall-to-spring periods. We also created a categorical version of the measure that indicates low-usage (0-4 weeks), medium-usage (5-9 weeks), and high-usage (10 or more weeks), again with unique indicators for fall-to-winter, and winter-to-spring. Similarly, we explore *Amira Learning* usage with both a continuous measure of the number of tutor sessions a student completed and a categorical version of the same variable (0-4 tutor sessions, 5-9 tutor sessions, 10-19 tutor sessions, 20-29 tutor sessions, and 30 or more tutor sessions) with unique indicators across the three time periods. A tutor session is considered any day a student logs on to the platform and reads one or more tutor stories. Students were encouraged to complete multiple tutor stories in a given session and multiple sessions per week, from mid-September through May, for a total of approximately 30 weeks and 30 minutes per week. Actual usage rates, however, were lower than expected, with the modal student participating in a tutor session for 13 weeks throughout the school year. Roughly 60 percent of students participated in 14 or fewer weeks of tutor sessions, meaning the

[2] 2,282 of the 2,305-student sample completed the Dyslexia Screener. Students identified as either At Risk, Strong Signals of At Risk, or Weaker Signals are considered "At Risk." Students identified as "low risk" are not considered "at risk."

majority of students received less than half of the intended treatment. Ultimately, only 18 percent of student participated for 20 weeks or more. Further, the average number of minutes read per week was nine minutes, well below the recommended 30 minutes per week.

Outcomes. We examine the associations between *Amira Learning* usage and five literacy outcomes, including Oral Reading Fluency, Vocabulary Size, Sight Recognition, Phonological Awareness, and Lexile score. Oral Reading Fluency (ORF) is a student's ability to read aloud with natural ease. Oral Reading Fluency is measured through Words Correct Per Minute (WCPM), which captures both the accuracy of words read as well as the minutes spent reading aloud. We use the adjusted WCPM score, which accounts for differences in passage difficulty. Vocabulary Size estimates the number of words likely present in a student's expressive vocabulary. Sight Recognition uses the Estimated Sight Recognition Inventory (ESRI) to estimate the percentage of sight words a student has mastered. Phonological Awareness, measured through Phoneme Segmentation Fluency, captures a student's ability to accurately produce phonemes within words. Finally, *Amira Learning* produces a Lexile score based on the ORF. Lexile is an outcome of reading ability with a higher Lexile score indicating that a student is capable of reading and understanding more challenging texts. We standardized (z-scored) all outcomes within grade. All assessments were administered three times throughout the year (fall, winter, and spring).

Covariates. One concern is that students who completed more *Amira Learning* weeks/sessions may have had additional social and academic background characteristics that also positively influenced their literacy development. To partially address this, we constructed a series of OLS regression models, which we describe below, that account for student sex, race/ethnicity (a series of dummy variables indicating whether the student identified as American Indian, Asian, Black, Hispanic, Native Hawaiian and Pacific Islander, or multiracial with white students serving as the comparison group), and special education status. We also account for the average time read per week and the number of days between assessments.

Analytic Approach

This study includes two broad types of analyses. First, we conducted a simple descriptive analysis of academic and socio-demographic differences between *Amira Learning* students who engaged with the platform to varying degrees. Second, we explored the link between *Amira*

Learning usage and student literacy development. We accounted for associations between usage and student characteristics through a series of OLS regression models that included the immediately prior, same-assessment literacy score as a covariate. These analysis of covariance (ANCOVA) or lagged-score models took the form

$$Y_{ij} = b_0 + b_1(\text{Amira Usage}) + X_{ij} + \delta_k + e_{ij} \quad (1)$$

where Y_{ij} is the end-of-period literacy outcome for student i in classroom j. The models employ four separate indicators of *Amira Learning* usage: (1) a continuous indicator of the number of weeks on the platform, (2) a categorical indicator of the number of weeks on the platform, (3) a continuous indicator of the number of tutor sessions completed, and (4) a categorical indicator of tutor sessions completed. A vector of student demographic characteristics (described above) as well as the baseline (beginning-of-period) literacy assessment score is indicated by X_{ij}, δ_k represents classroom fixed effects, and e_{ij} is the error term for student i in classroom j. We conducted these analyses separately for three time periods: fall-to-winter (mid-September to February 2021); winter-to-spring (February 2021 to May 2021), and; the entirety of the fall-to-spring intervention (Mid-September 2020 through May 2021).

Additionally, we performed sub-group analyses to identify the degree to which the associations between *Amira Learning* usage and literacy development differed across grade level, race/ethnicity, gender, special education status, and dyslexia risk status. We examine kindergarten and third graders separately, as their experiences with *Amira Learning* may differ from those of their first and second grade peers. For example, kindergartners are not able to start to engage *Amira Learning* until they are able to read connected texts, which influences their usage window and limits the number of kindergarteners who take the fall literacy assessment. Further, third graders may "test out" of *Amira Learning*, limiting their use of the platform.

Results

Descriptive Findings

We begin by exploring differences in minutes read per week and average number of tutor sessions between students who engaged *Amira Learning* to different degrees (see Table 2). Note first that there were twice as many high-usage students in the initial fall-to-winter period

compared to the winter-to-spring period. By definition, these high-usage students not only participated in the *Amira Learning* implementation for more weeks, but during those weeks also read for more minutes and completed more tutor sessions. More specifically, from the fall-to-winter period, high-usage *Amira Learning* students read on average for 13 minutes per week, while low-usage students read for seven and a half minutes per week ($p<.001$). Further, high-usage students completed on average 27 tutor sessions as compared to three tutor sessions among low-usage students ($p<.001$). We also find that medium-usage students read more minutes (+1.93) and completed more tutor sessions (+7.93) than low-usage students ($p<.001$).

We find similar trends for the winter-to-spring period, where high-usage students read 6.15 minutes more per week, and participated in approximately 36 tutor sessions more than their low-usage peers ($p<.001$); medium-usage students read two and a half more minutes per week, and completed approximately 12 tutor sessions more than their low-usage peers ($p<.001$). Further, we find that students, on average, read more minutes per week during the fall-to-winter period (+3 minutes) than during the winter-to-spring period. We are not able say what factor(s) drove the difference in implementation between the fall-to-winter and winter-to-spring periods; however, we know that implementation of *Amira Learning* shifted from remote at the start of the year to in-person by the end of the year.

Table 2. Average Mins Read/Week and Average Number of Tutor Sessions by Amira Usage (n=2,305)

	Fall-to-Winter			Winter-to-Spring		
	Low Usage 0-4 Weeks (n=758)	Med Usage 5-9 Weeks (n=807)	High Usage 10+ Weeks (n=740)	Low Usage 0-4 Weeks (n=992)	Med Usage 5-9 Weeks (n=921)	High Usage 10+ Weeks (n=392)
Avg. Read/Wk	7.53	9.46***	13.11***	4.61	7.20***	10.76***
Avg. # Sessions	2.97	10.90***	27.24***	6.47	18.34***	42.39***

***$p<0.001$. Note: all significance tests compared to low-usage category.

Next, we explore the socio-demographic differences between *Amira Learning* students who participated at varying usage levels. As indicated in Table 3, special education status is unrelated to *Amira Learning* participation ($p>.05$). However, there were important grade and racial/ethnic differences across usage categories ($p<.001$). High-usage students were more likely to be second graders (58% and 62%) than first graders (42% and 38%; $p<0.001$) in both fall-to-winter and winter-to-fall time periods, respectively. We also see a strong relationship between race/ethnicity and *Amira Learning* usage ($p<.001$). Black students were under-represented

among those in the high-usage category, while white students were over-represented. Table 3 further indicates gender differences in usage ($p<.05$), with a smaller proportion of females in the high-usage category during the fall-to-winter period. Students identified as being at risk for dyslexia were also progressively less prevalent among those in higher-use categories.

Table 3. Socio-Demographic Characteristics of Amira Students by Number of Weeks Categories

	Fall-to-Winter			Winter-to-Spring		
	Low Usage 0-4 Weeks (*n*=758)	Med Usage 5-9 Weeks (*n*=807)	High Usage 10+ Weeks (*n*=740)	Low Usage 0-4 Weeks (*n*=992)	Med Usage 5-9 Weeks (*n*=921)	High Usage 10+ Weeks (*n*=392)
Grade***						
% First	53.4	48.2	42.0	51.1	48.7	38.0
% Second	46.6	51.8	58.0	48.9	51.3	62.0
Race/Ethnicity***						
% A. Ind	0.3	0.3	0.5	0.3	0.2	0.8
% Asian	1.7	2.1	3.5	1.7	2.6	3.8
% Black	59.2	53.4	39.7	56.9	50.7	36.5
% Hispanic	9.9	15.0	13.7	13.7	11.9	13.0
% NH/PI	0.4	0.4	0.3	0.4	0.4	0.0
% White	23.1	23.3	35.4	21.3	29.2	37.0
% Multiracial	5.4	5.6	6.9	5.7	4.9	8.9
% Female[1]*	52.6	55.3	46.6	52.2	51.4	50.8
% SWD[1]~	8.7	9.2	11.9	9.2	10.1	11.2
% Dyslexia[2] ***	9.3	7.5	4.1	6.7	8.3	4.9

~$p<.10$; *$p<0.05$; ** $p<0.01$; ***$p<0.001$.

[1] For female and students with disabilities (SWD), we find significant differences for Fall-to-Winter, but not Winter-to-Spring.

[2] For students at risk for dyslexia, we find significant differences ($p<.001$) for Fall-to-Winter, and $p<.10$ for Winer-to-Spring.

Table 4 displays the means and standard deviations for the five literacy outcomes across each of the three test administrations. Three quite different growth patterns are evident here, with obvious implications for our interpretations of the link between *Amira Learning* usage and student literacy development. First, with the WCPM and Phonetic outcomes, we see large gains during the fall-to-winter period, but no growth during the winter-to-spring period. Second, there is a mirror opposite pattern with the ERSI and Vocabulary outcomes: little to no growth during the fall, but substantial growth during the spring. The third pattern is with the Lexile outcome, the only outcome to suggest sustained literacy development during both periods.

Table 4. Fall, Winter, and Spring Test Score Outcomes among *Amira Learning* Students (*n*=2,305)

	Fall	Winter	Spring
WCPM, mean	37.2	48.7	47.5
SD	(31.5)	(34.6)	(32.1)
ERSI, mean	71.1	71.4	76.0
SD	(23.1)	(25.0)	(23.3)
Vocabulary, mean	3,574	3,565	3,889
SD	(1,792)	(1,664)	(1,719)
Phonetic, mean	70.8	76.1	77.1
SD	(24.7)	(24.3)	(23.8)
Lexile, mean	26.4	79.9	123.6
SD	(369.3)	(367.8)	(353.4)

The focus of this study are the test score patterns among students with varying *Amira Learning* usage rates, which are displayed in Table 5. Note first that across all outcomes, students who engaged with *Amira Learning* for ten or more weeks during the fall period *started* the academic year with considerably stronger literacy skills. For example, these high-usage students had higher initial WCPM scores (+8) compared to their low-usage peers ($p<.001$); fall differences between low- and high-usage across the other outcomes equate to roughly 0.25 SD ($p<.001$). It will be important for *Amira Learning* to consider *why* initially higher-ability students engaged more with the platform (and whether this is the case post-pandemic in other implementation sites). Recall that during the fall period schools were closed due to COVID-19. It is unclear whether parental influences and/or home contexts may have differentially influenced usage rates across the types of students who engaged to different degrees. Interestingly, we did not find significant initial skills differences between medium- and low-usage students.

With Table 5 we can also begin to explore the link between test-score growth and *Amira Learning* usage. Across every outcome, the initial skills difference that distinguished low- and high-usage students grew even wider during the fall; in other words, high-usage students gained more literacy skills during this period. For instance, the initial eight-point WCPM advantage among high-usage students grew to over 15 points by the winter assessment ($p<.001$). Similarly, with each of the other four outcomes, the initial skills gap between low- and high-usage students virtually doubled during the fall, to almost one-half standard deviation ($p<.001$). Note also that while the fall differences between low- and medium-usage students were non-significant, by the winter, medium-usage students were outperforming their low-usage peers across all outcomes (at least $p<.01$).

The links between test score growth and usage patterns during the winter-to-spring period are less clear.[3] Recall from Table 4 that students made smaller gains on the WCPM and Phonetic outcomes during the winter-to-spring period. This lack of literacy development as measured by these outcomes is evident here, with no usage category making appreciable gains. Conversely, we do see growth with the other three outcomes, again not surprising given the results in Table 4. With the ESRI, Phonological Awareness, and Lexile outcomes, medium-use students appear to have experienced somewhat stronger gains than students in the other two usage categories.

[3] Differences in winter scores between the fall-to-winter and winter-to-spring periods are due to the differing membership of the usage categories across the two periods (e.g., not all low-usage students in the fall were also low-usage in the spring).

Table 5. Fall, Winter, and Spring Test Scores by Number of Weeks Used Categories

	Fall-to-Winter			Winter-to-Spring		
	Low Usage 0-4 Weeks (*n*=758)	Med Usage 5-9 Weeks (*n*=807)	High Usage 10+ Weeks (*n*=740)	Low Usage 0-4 Weeks (*n*=992)	Med Usage 5-9 Weeks (*n*=921)	High Usage 10+ Weeks (*n*=392)
Adj. WCPM						
Fall	34.37	34.77	42.60***	---	---	---
(SD)	(30.52)	(29.97)	(33.52)			
Winter	41.90	47.32**	57.23***	46.49	46.96	58.50***
(SD)	(33.68)	(32.49)	(35.95)	(33.92)	(33.36)	(37.36)
Spring	---	---	---	43.97	47.29~	57.18***
(SD)				(31.67)	(30.85)	(34.36)
ESRI						
Fall	68.52	69.91	74.98***	---	---	---
(SD)	(23.30)	(23.15)	(22.40)			
Winter	65.69	71.46***	77.17***	69.51	70.72	77.76***
(SD)	(26.59)	(24.42)	(22.48)	(25.61)	(24.67)	(23.11)
Spring	---	---	---	73.79	76.45*	80.80***
(SD)				(24.35)	(22.54)	(21.81)
Vocab Size						
Fall	3,363	3,488	3,886***	---	---	---
(SD)	(1,696)	(1,770)	(1,871)			
Winter	3,235	3,512**	3,963***	3,436	3,478	4,097***
(SD)	(1,595)	(1,600)	(1,720)	(1,623)	(1,607)	(1,795)
Spring	---	---	---	3,712	3,857	4,410***
(SD)				(1,675)	(1,660)	(1,863)
Phonological Awareness						
Fall	68.29	69.83	74.57***	---	---	---
(SD)	(25.39)	(24.66)	(23.46)			
Winter	70.40	76.38***	81.64***	74.51	75.44	81.69***
(SD)	(27.18)	(23.60)	(20.36)	(25.29)	(24.25)	(20.99)
Spring	---	---	---	74.81	77.52*	81.88***
(SD)				(25.15)	(22.86)	(21.62)
Lexile score						
Fall	-15.63	6.48	91.18***	---	---	---
(SD)	(359)	(368)	(372)			
Winter	1.82	75.66***	164.53***	53.67	62.87	186.34***
(SD)	(366)	(359)	(362)	(364)	(362)	(372)
Spring	---	---	---	86.69	121.91~	221.06***
(SD)				(351)	(345)	(361)

~p<.10; *p<0.05; ** p<0.01; ***p<0.001. Significance tests compared to low-usage students.

Analytic Results

The descriptive findings above may obscure links between *Amira Learning* participation and student literacy development given differences in the students who engaged more or less in the program. To address this, the analyses in this section employ methods that seek to adjust these descriptive patterns for student characteristics and their classrooms. We explore the link between *Amira Learning* usage and literacy development, with comparisons made to other *Amira Learning* students enrolled in the same classroom. All models adjust for the average number of minutes student read per week.

Model 1 in Table 6 indicates the relationship between *Amira Learning* usage rates and WCPM growth during the fall-to-winter period. We find that each additional week of usage was associated with a 0.030 SD increase in WCPM ($p<.001$). Rather than a continuous indicator of usage, Model 2 uses a categorical measure that compares literacy growth between medium- and high-usage students and their low-usage peers. We find that students in the medium-usage category (5-9 weeks) gained approximately 0.178 SD more than low-usage students (0-4 weeks), holding all else constant ($p<.001$). This linear trend continues with high-usage students gaining, on average, 0.304 SD more than their low-usage counterparts ($p<.001$).

Model 3 indicates the relationship between the number of *Amira Learning* tutor sessions completed and fall-to-winter WCPM development. We find that each additional tutor session completed was associated with a 0.012 SD increase in literacy development ($p<.001$). Model 4 uses a categorical measure to compare literacy growth between students who completed 5-9 tutor sessions, 10-19 tutor sessions, 20-29 tutor sessions, and 30 or more tutor sessions to those students who completed 0-4 tutor sessions. Overall, we continue to see a positive relationship with number of tutor sessions and adjusted WCPM. Students who completed 5-9 tutor sessions experienced a 0.105 SD advantage over their low-usage peers ($p<.05$). This advantage increased for students in the 10-19 sessions category (0.239 SD), the 20-29 sessions category (0.274 SD), and the 30+ sessions category (0.448; $p<.001$).

Table 6. Amira Usage and Fall to Winter Adjusted WCPM Development

	(1)	(2)	(3)	(4)
# Weeks	0.030***	--	--	--
Med Usage	--	0.178***	--	--
High Usage	--	0.304***	--	--
# Tutor Sessions	--	--	0.012***	--
5-9 Sessions	--	--	--	0.105*
10-19 Sessions	--	--	--	0.239***
20-29 Sessions	--	--	--	0.274***
30+ Sessions	--	--	--	0.448***
Time Read/Wk	-0.004*	-0.003	-0.007**	-0.006**
Constant	0.127	0.123	0.144	0.115

~p<.10; *p<.05; **p<.01; ***p<.001. Outcome is the Winter WCPM (z-scored). All models include the Fall WCPM score, classroom fixed effects and account for student sex, race/ethnicity, special education status, and the number of days between assessments.

Next, we explore the relationship between *Amira Learning* usage and our other literacy outcomes: vocabulary size, ESRI, Phonetical Awareness, and Lexile scores. We find similar and consistent results for these additional outcomes. As displayed in Model 1 in Tables 7-10, each additional week of tutor sessions during the fall-to-winter period was associated with a 0.025 to 0.029 SD advantage (p<0.001). Additionally, in Model 2 in Tables 7-10, we find that high-usage students gained 0.252 to 0.299 SD more than low-usage students. We continue to see these trends when defining usage in terms of number of tutor sessions completed. In Model 3 in Tables 7-10, each additional tutor session completed during the fall-to-winter period was associated with a 0.009 to 0.011 SD increase in that respective outcome. In Model 4 in Tables 7-10, we find that on average, students who completed 30 or more tutor sessions gained 0.376 to 0.393 SD more than students who completed only 0-4 tutor sessions. In sum, higher-usage students experienced consistently stronger literacy development across all outcomes during the first half of the school year. These findings are consistent with the descriptive results presented in Table 5.

Table 7. Amira Usage and Fall to Winter ESRI Development

	(1)	(2)	(3)	(4)
# Weeks	0.027***	--	--	--
Med Usage	--	0.208***	--	--
High Usage	--	0.276***	--	--
# Tutor Sessions	--	--	0.009***	--
5-9 Sessions	--	--	--	0.195***
10-19 Sessions	--	--	--	0.262***
20-29 Sessions	--	--	--	0.293***
30+ Sessions	--	--	--	0.376***
Time Read/Wk	-0.000	0.001	-0.002	-0.001
Constant	0.209	0.189	0.233	0.181

~p<.10; *p<.05; **p<.01; ***p<.001. Outcome is the Winter ESRI (z-scored). All models include the Fall ESRI score, classroom fixed effects, and account for student sex, race/ethnicity, special education status, and the number of days between assessments.

Table 8. Amira Usage and Fall to Winter Vocabulary Development

	(1)	(2)	(3)	(4)
# Weeks	0.027***	--		
Med Usage	--	0.148***	--	--
High Usage	--	0.258***	--	--
# Tutor Sessions	--	--	0.011***	--
5-9 Sessions	--	--	--	0.106*
10-19 Sessions	--	--	--	0.211***
20-29 Sessions	--	--	--	0.272***
30+ Sessions	--	--	--	0.393***
Time Read/Wk	-0.003	-0.002	-0.006*	-0.005*
Constant	0.265	0.265	0.279	0.254

~p<.10; *p<.05; **p<.01; ***p<.001. Outcome is the Winter vocabulary size (z-scored). All models include the Fall vocabulary size score, classroom fixed effects, and account for student sex, race/ethnicity, special education status, and the number of days between assessments.

Table 9. Amira Usage and Fall to Winter Phonological Awareness Development

	(1)	(2)	(3)	(4)
# Weeks	0.029***	--		
Med Usage	--	0.203**	--	--
High Usage	--	0.299***	--	--
# Tutor Sessions	--	--	0.009***	--
5-9 Sessions	--	--	--	0.182***
10-19 Sessions	--	--	--	0.253***
20-29 Sessions	--	--	--	0.327***
30+ Sessions	--	--	--	0.380***
Time Read/Wk	-0.001	0.000	-0.002	-0.001
Constant	0.415~	0.402	0.442~	0.394

~p<.10; *p<.05; **p<.01; ***p<.001. Outcome is the Winter phonological awareness (z-scored); All models include the Fall phonological awareness score, classroom fixed effects, and account for student sex, race/ethnicity, special education status, and the number of days between assessments.

Table 10. Amira Usage and Fall to Winter Lexile Development

	(1)	(2)	(3)	(4)
# Weeks	0.025***	--		
Med Usage	--	0.194***	--	--
High Usage	--	0.252***	--	--
# Tutor Sessions	--	--	0.009***	--
5-9 Sessions	--	--	--	0.162***
10-19 Sessions	--	--	--	0.247***
20-29 Sessions	--	--	--	0.260***
30+ Sessions	--	--	--	0.380***
Time Read/Wk	-0.002	-0.001	-0.004	-0.003
Constant	0.191	0.172	0.209	0.164

~p<.10; *p<.05; **p<.01; ***p<.001. Outcome is the Winter Lexile (z-scored). All models include the Fall Lexile z-score, classroom fixed effects, and account for student sex, race/ethnicity, special education status, and the number of days between assessments.

Winter-to-Spring, and Fall-to-Spring Time Periods

We also estimated the associations between *Amira Learning* usage and literacy development for the winter-to-spring and fall-to-spring periods. Table 11 suggests that, on average, each week of usage during the winter-to-spring period was associated with 0.011 SD additional WCPM growth ($p<.05$), and a 0.019 SD advantage during the full fall-to-spring period ($p<.05$). It is important to note that these gains are much smaller than the fall gains reported in Model 1 Table 6 (0.030; $p<.001$) suggesting that the links between usage and literacy development were much stronger during the fall period. Recall from Table 4 above that on average, WCPM growth was quite weak (to non-existent) during the winter-to-spring period.

We find this trend across the other literacy outcomes as well during the winter-to-spring period and fall-to-spring periods: usage estimates that are considerably smaller than those from the initial fall-to-winter period. We do not know what factor(s) drove these differences; however, we know that students started the school year fully remote before returning to in-person classes by the mid-point.

Table 11. Amira Usage and WCPM Development in Winter-to-Spring and Fall-to-Spring

	Winter-to-Spring	Fall-to-Spring
WCPM		
# Weeks	0.011*	0.019***
Time Read/Wk	0.001	-0.001
ESRI		
# Weeks	0.011~	0.016***
Time Read/Wk	0.003	0.004
Vocab Size		
# Weeks	0.013*	0.018***
Time Read/Wk	0.001	0.001
Phonological Awareness		
# Weeks	0.021**	0.019***
Time Read/Wk	0.004	0.003
Lexile		
# Weeks	0.013*	0.016***
Time Read/Wk	0.001	0.002

~$p<.10$; *$p<.05$; **$p<.01$; ***$p<.001$. Outcome is the Winter or Spring outcome (z-scored). All models include the same-subject baseline score, classroom fixed effects, and account for student sex, race/ethnicity, special education status, and the number of days between assessments.

Subgroup Analyses

We also explored usage estimates by grade (kindergarten and 3rd grade, separately), gender, race/ethnicity, special education status, and dyslexia risk status. As indicated in Table 12, the fall-to-winter usage estimates with WCPM for kindergarteners are non-significant ($p>0.05$), but bear in mind the quite small sample. For kindergarteners during the winter-to-spring period, however, each additional week of usage was associated with 0.044 SD additional WCPM growth ($p<0.05$). We continue to find these associations for all other literacy outcomes from winter-to-spring. In other words, we find the opposite trends for kindergarteners as we did with first and second graders, for whom *Amira Learning* benefits accrued largely in the fall rather than the spring. This trend may be because kindergarteners were not able to use *Amira Learning* until they could read text, which may not have occurred until later in the school year. Supporting this hypothesis, the third-grade patterns were quite similar to the first and second grade patterns reported above. More specifically, we find that for each additional week of usage during the fall, third graders gained 0.034 SD *(*$p<.001$), similar to the first and second grade findings (0.030 SD; see Model 1 in Table 6). We continue to find that third graders gained 034-0.038 SD ($p<.001$) for each additional week of usage across all other literacy outcomes. However, unlike first and second grade, we find no association between number of weeks of usage and WCPM among third graders from fall-to-spring ($p>0.05$).

Table 12. Amira Usage and WCPM Development, Kindergarten and Third Grade

	Kindergarteners (n=217)	Third Graders (n=1,182)
Fall-to-Winter		
# Weeks	0.020	0.034***
Time Read/Wk	0.000	-0.003
Constant	-2.077~	0.580
Winter-to-Spring		
# Weeks	0.044*	0.015
Time Read/Wk	-0.000	-0.006
Constant	-1.190*	-0.310

~$p<.10$; *$p<.05$; **$p<.01$; ***$p<.001$. Outcome is the winter or spring adjusted WCPM z-score. All models include the prior WCPM score, classroom fixed effects, and account for student sex, race/ethnicity, special education status, and the number of days between assessments.

Socio-demographic and academic characteristics. As indicated in Tables 13-14, estimates for every outcome among non-special education students and those not at-risk for dyslexia were significant during the fall period, reflecting the full-sample estimates reported above. In stark contrast, no usage estimates for special education students and students at-risk for dyslexia were significant for any outcome. But this difference is largely a reflection of the very small sub-group sample sizes. Moreover, none of the subgroup estimates were statistically different from each other. For example, the 0.031 number-of-weeks WCPM estimate in Table 13 for non-special education students, while *itself* significant, is not significantly different from the 0.025 estimate for special education students. Similarly, we found that the usage estimates did not differ across male and female students, as indicated in Table 15.

Table 13. Amira Usage and Fall-to-Winter Literacy Development by Special Education Status

	Special Education Students (n=228)	Non-Special Education Students (n=2,077)
WCPM		
# Weeks	0.025	0.031***
Time Read/Wk	-0.000	-0.004~
ESRI		
# Weeks	0.031	0.027***
Time Read/Wk	0.018	-0.000
Vocab Size		
# Weeks	0.005	0.028***
Time Read/Wk	0.013	-0.003
Phonol. Awareness		
# Weeks	0.052	0.029***
Time Read/Wk	0.016	-0.000
Lexile		
# Weeks	0.019	0.026***
Time Read/Wk	0.009	-0.001

~p<.10; ***p<.001. Outcome is the Winter outcome (z-scored). All models include the Fall outcome, classroom fixed effects, and account for student sex, race/ethnicity, and the number of days between assessments.

Table 14. Amira Usage and Fall-to-Winter Literacy Development by Dyslexia Risk Status

	Dyslexia at Risk (*n*=160)	Non-Dyslexia (*n*=2,122)
WCPM		
# Weeks	0.006	0.029***
Time Read/Wk	-0.005	-0.005*
ESRI		
# Weeks	0.018	0.023***
Time Read/Wk	-0.012	-0.001
Vocab Size		
# Weeks	0.002	0.026***
Time Read/Wk	-0.004	-0.004
Phonol. Awareness		
# Weeks	0.024	0.024***
Time Read/Wk	-0.023	-0.001
Lexile		
# Weeks	0.004	0.023***
Time Read/Wk	-0.002	-0.002

~*p*<.10; ***p*<.001. Outcome is the Winter outcome (z-scored). All models include the Fall outcome, classroom fixed effects, and account for student sex, race/ethnicity, and the number of days between assessments.

Table 15. Amira Usage and Fall-to-Winter Literacy Development by Gender

	Female (*n*=1,190)	Male (*n*=1,115)
WCPM		
# Weeks	0.033***	0.030***
Time Read/Wk	-0.004	-0.003
ESRI		
# Weeks	0.030***	0.027***
Time Read/Wk	0.004	-0.003
Vocab Size		
# Weeks	0.029***	0.028***
Time Read/Wk	-0.001	-0.003
Phonol. Awareness		
# Weeks	0.030***	0.031***
Time Read/Wk	0.002	-0.004
Lexile		
# Weeks	0.027***	0.026***
Time Read/Wk	0.000	-0.001

~*p*<.10; ***p*<.001. Outcome is the Winter outcome (z-scored). All models include the Fall outcome, classroom fixed effects, and account for student sex, race/ethnicity, nd the number of days between assessments.

However, we did find important racial/ethnic differences in the usage estimates. As indicated in Table 16, Hispanic students gained less for each additional week of usage across all literacy outcomes compared to their white peers (the estimates differ statistically from each other). Black students gained 0.019 SD less, and 0.015 SD less in WCPM and Vocabulary size, respectively, than white students for each additional week of *Amira Learning* usage (again, the estimates differ significantly). However, we do not find a significant difference in gains between Black and white students for ESRI, phonological awareness, or Lexile score.

Table 16. Amira Usage and Fall to Winter Literacy Development by Race/Ethnicity

	Asian (*n*=56)	Black (*n*=1,174)	Hispanic (*n*=297)	Multiracial (*n*=137)	White (*n*=625)
WCPM					
# Weeks	0.050	0.027***	0.010	0.004	0.046***
Time Read/Wk	0.030	-0.002	0.007	0.002	-0.010*
ESRI					
# Weeks	0.011	0.028***	0.015	0.036	0.031**
Time Read/Wk	0.030	0.002	0.005	0.015	-0.008
Vocab Size					
# Weeks	0.024	0.023***	0.001	0.009	0.038***
Time Read/Wk	0.085	-0.000	0.006	0.016	-0.011*
PA					
# Weeks	-0.003	0.035***	0.018	0.034	0.033**
Time Read/Wk	0.030	0.002	0.005	0.011	-0.011*
Lexile					
# Weeks	0.013	0.024***	0.007	0.031	0.029***
Time Read/Wk	0.056	0.001	0.007	0.003	-0.008~

~p<.10; *p<.05; **p<.01; ***p<.001. Outcome is the Winter outcome (z-scored). All models include the Fall outcome, classroom fixed effects, and account for student sex, special education status, and the number of days between assessments.

Conclusions

This report explored the implementation of *Amira Learning* in Savannah-Chatham County Public Schools during the 2020-21 academic school year. The results described above provide promising evidence of the effectiveness of *Amira Learning*. More specifically, we consistently find that students who completed 10 or more weeks of *Amira Learning* tutor sessions from fall-to-winter experienced greater gains across all literacy outcomes as compared to students who only completed 0-4 weeks, holding all else even. Further, we found that medium-usage (5-9 weeks) students, despite starting in a similar place in the fall, had higher winter and spring scores than their low-usage peers. These results were consistent across usage measures and literacy outcomes.

These promising trends are present across the three time periods we examined, although the *Amira Learning* estimates are largest during the fall, and for additional grades (kindergarten and third grade, separately). Moreover, the *Amira Learning* advantage appears to be constant across gender. Unfortunately, we did find that Hispanic students did not benefit as much for each additional week of *Amira Learning* usage as white students across all literacy outcomes, and that Black students did not benefit as much for each additional week of usage as white students in terms of oral reading fluency and vocabulary size.

It is important to keep in mind that these may be conservative estimates of *Amira Learning's* potential, as the majority of students did not receive the recommended *Amira Learning* dosage. Roughly 60 percent of students received less than half of the intended treatment. In fact, even students who participated in *Amira Learning* tutor sessions for 20 or more weeks from fall-to-spring, on average, read 13 minutes per week as compared to the 30 minutes per week recommendation. In sum, these results tell a consistent (although not causal) story; students who received a greater percentage of the recommended dosage gained more than their peers who received smaller dosages.

Limitations

The obvious caveat with these analyses is that the implementation and resulting data cannot identify the causal impact of *Amira Learning* on student literacy development. First and foremost, we do not have any comparison students who did not receive the treatment, meaning the results provided here are suggestive only. There were also several limitations with the

implementation in SCCPS. The 2020-21 school year represents one of the most disruptive periods in the history of U.S. education. Students began the school year fully remote before returning to in-person classrooms midyear. We do not know how schools and teachers encouraged students to use *Amira Learning* while remote, and how they incorporated it into their classroom once they were back in classrooms. Further, as previously discussed, there is significant variation in usage among *Amira Learning* users. It is unclear what *Amira Learning*'s impact might have been if students completed the recommended dosage while in-person in school for a full academic year. However, the OLS regressions did adjust for many of the student characteristics associated with *Amira Learning* usage, notably initial literacy scores and student socio-demographic characteristics.

Given these limitations, more robust experimental and quasi-experimental studies are clearly warranted to provide contemporary evidence of the extent to which *Amira Learning* helps all students achieve reading proficiency. That being said, these results clearly suggest that fidelity of *Amira Learning* implementation is critical: students who received a greater percentage of the recommended dosage consistently gained more than those students who received smaller dosages.

Appendix - Wawasee Community School Corporation Findings

In this appendix, we describe student literacy outcomes associated with the implementation of *Amira Learning* in Wawasee Community School Corporation (WCSC) during the 2020-21 academic year. More specifically, we address the following research questions:

1. To what extent is use of *Amira Learning* tutor sessions associated with student literacy growth during the 2020-21 school year?
2. What is the relationship between student literacy growth and software usage patterns?

Data and Methods

Analytic Sample. Wawasee Community School Corporation (WCSC), located in north-central Indiana, enrolls approximately 750 K-3 elementary students attending one of three elementary schools. District student enrollments are 92% white and 6% Hispanic. While the district intended for all K-3 grade WCSC students to have the opportunity to participate in *Amira Learning* at some point throughout the school year, there was significant variation in student usage. The district encouraged teachers to use *Amira Learning* in their classrooms; however, it was not mandated. We selected students who completed the fall, winter, and spring assessments and engaged in at least one tutor session, which produced an analytic sample of only 88 K-3 students.

***Amira Learning* Usage.** Our primary measure of *Amira Learning* usage is the number of weeks students completed at least one *Amira Learning* tutor session. In addition to this continuous indicator, we use a series of categorical indicators: very low-usage (1-4 weeks), low-usage (5-9 weeks), medium-usage (10-14 weeks), and high-usage (15 or more weeks). We also explore *Amira Learning* usage in terms of the number of tutor sessions a student completed, as a continuous indicator as well as a series of categorical indicators: 1-9 tutor sessions, 10-19 tutor sessions, and 20 or more tutor sessions. Students were recommended to complete multiple tutor stories in a given session, multiple sessions per week, from mid-September through May for a total of approximately 30 weeks and 30 minutes per week. Actual usage rates, however, were lower than expected. Only 22 percent of students participated in 15 or more weeks of tutor sessions, meaning the vast majority of students received less than half of the intended treatment. Further, the average number of minutes read per week is 9 minutes, well below the recommended 30 minutes per week.

Outcomes. Our primary literacy outcome is Oral Reading Fluency (ORF), or a student's ability to read aloud with natural ease. Oral Reading Fluency is measured through Words Correct Per Minute (WCPM), which captures both the accuracy of words read as well as the minutes spent reading aloud. More specifically, we use the adjusted WCPM score as it is equated to be a directly comparable measure since it removes any variation from passage difficulty.

Analytic Approach. This study includes two broad types of analyses. First, we conducted a simple descriptive analysis of academic differences between *Amira Learning* students who engaged with the platform to varying degrees. Second, we explored the link between *Amira Learning* usage and their literacy development. We constructed an OLS regression model that allowed us to account for a student's grade-level, school, and their fall Oral Reading Fluency score.

Results

Descriptive Findings. As indicated in Appendix Table 1, we find that high-usage (15+ weeks) students read more minutes (+6 minutes; $p<0.01$) and complete more tutor sessions (+35 sessions; $p<.001$) as compared to their very low-usage peers. We also find that medium-usage (10-14 weeks) students read slightly more minutes (+4 minutes; $p<0.10$) and complete more tutor sessions (+19 sessions; $p<.001$) compared to our very low-usage students. These findings indicate that not only do our high- and medium-usage students participate in the platform more frequently by definition, but that when they are on the platform, they read more and complete more sessions than their very low-usage peers.

Next, the focus of this study are the test score patterns among *Amira Learning* students with varying usage rates, which are displayed in Appendix Table 2. We do not find a significant relationship between fall literacy outcomes and usage. However, by spring, we find that the high-usage students have higher adjusted WCPM ($p<.10$), vocab size ($p<.05$), and Lexile scores ($p<.05$) than their very low-usage peers. Additionally, we find that our medium-usage students have higher spring vocabulary size ($p<.05$) and Lexile scores ($p<.01$) compared to their very low-usage peers.

Analytic Results. We find a relationship between *Amira Learning* usage and Oral Reading Fluency. As indicated in Model 1 in Appendix Table 3, we find that students gain 1.75 adjusted WCPM (0.057 SD) for each additional week of usage ($p<.01$). Rather than a continuous

measure, Model 2 uses a categorical measure that compares low-, medium-, and high-usage to very low usage-students. In this model, we find that high-usage students gain 27.3 adjusted WCPM (0.892 SD) more than their very low usage peers ($p<.05$). We also find that medium-usage students gain 17.4 adjusted WCPM (0.568 SD) more than their very low-usage peers ($p<.10$). In Models 3 and 4, which use number of tutor sessions to capture usage, we find a similar relationship.

Conclusion

This report explored the implementation of *Amira Learning* in Wawasee Community School Corporation during the 2020-21 academic school year. Overall, we find high-usage students experienced greater gains as compared to very low-usage students. These results suggest that frequent usage of *Amira Learning* is important for literacy development. However, these analyses are greatly limited given the small sample sizes and low usage among WCSC students.

Appendix Table 1. Average Mins Read/Week, Average Number of Tutor Sessions, and Fall Outcomes by Number of Weeks Categories in WCSC

	Very Low Usage 1-4 Weeks (*n*=12)	Low Usage 5-9 Weeks (*n*=17)	Medium Usage 10-14 Weeks (*n*=40)	High Usage 15+ weeks (*n*=19)
Avg. Mins Read/Week	5.13	7.780	9.056~	11.519**
Avg # of Tutor Sessions	2.167	10.117	21.28***	36.895***
Fall Literacy Outcomes				
Adjusted WCPM	43.60	46.59	57.26	61.55
(SD)	(31.79)	(32.45)	(25.89)	(36.28)
ESRI	75.79	81.687	85.01	82.95
(SD)	(23.30)	(21.09)	(20.18)	(22.77)
Vocab Size	3722.58	4540.24	4786.23	4946.68
(SD)	(1550.01)	(2093.14)	(1777.82)	(2269.29)
Phonological Awareness	75.51	78.68	83.36	78.89
(SD)	(21.93)	(23.25)	(20.56)	(24.91)
Lexile	119.58	232.06	311.88	315.79
(SD)	(335.26)	(400.71)	(338.33)	(391.33)

~$p<.10$; *$p<.05$; **$p<.01$; ***$p<.001$. Note: all significance tests compared to Very Low-Usage.

Appendix Table 2. Fall and Spring Scores by Number of Weeks Categories in WCSC

	Very Low Usage 1-4 Weeks (*n*=12)	Low Usage 5-9 Weeks (*n*=17)	Medium Usage 10-14 Weeks (*n*=40)	High Usage 15+ weeks (*n*=19)
Adjusted Words Correct Per Minute (WCMP)				
Fall WCPM Score	43.60	46.59	57.26	61.55
(SD)	(31.79)	(32.45)	(25.89)	(36.28)
Spring WCPM Score	56.16	62.30	76.04	88.47~
(SD)	(40.48)	(31.38)	(28.50)	(30.87)
ESRI				
Fall ESRI Score	75.79	81.687	85.01	82.95
(SD)	(23.30)	(21.09)	(20.18)	(22.77)
Spring ESRI Score	71.88	81.83	89.55	92.37
(SD)	(26.52)	(28.68)	(19.11)	(18.12)
Vocab Size				
Fall Vocab Size	3722.58	4540.24	4786.23	4946.68
(SD)	(1550.01)	(2093.14)	(1777.82)	(2269.29)
Spring Vocab Size	3711.08	4853.77	6182.25*	6146.05*
(SD)	(1913.39)	(2402.75)	(2302.4)	(1786.53)
Phonological Awareness (PA)				
Fall PA	75.51	78.68	83.36	78.89
(SD)	(21.93)	(23.25)	(20.56)	(24.91)
Spring PA	68.82	79.00	86.01	90.28
(SD)	(28.81)	(31.39)	(22.56)	(22.46)
Lexile score				
Fall Lexile Score	119.58	232.06	311.88	315.79
(SD)	(335.26)	(400.71)	(338.33)	(391.33)
Spring Lexile Score	92.92	321.76	506.25**	531.84*
(SD)	(314.94)	(376.97)	(365.92)	(285.28)

~p<.10; *p<0.05; ** p<0.01; ***p<0.001. Significance tests compared to Very Low-Usage students.

Appendix Table 3. Amira Usage and Fall to Spring Adjusted WCPM among Amira Students in WCPC

	(1)	(2)	(3)	(4)
# Weeks	1.754**	--	--	--
Low Usage	--	7.433	--	--
Med Usage	--	17.363~	--	--
High Usage	--	27.266*	--	--
# Tutor Sessions	--	--	0.493*	--
10-19 Tutor Sessions	--	--	--	11.880~
20+ Tutor Sessions	--	--	--	16.138*
Time Read/Wk	-0.275	-0.362	--	--
Constant	12.355	15.709	24.516***	19.014*

~p<.10; *p<.05; **p<.01; ***p<.001. Outcome is the Spring Adjusted WCPM; all models include the Fall WCPM score and grade-level as a covariate. All models include school fixed effects.