Off to a Great Start: The Potential for Tutoring Paired with the Off2Class Foundational Literacy Curriculum to Boost English Proficiency Gains for Adolescent Newcomer English Learners

Havala Hanson, Ph.D.*
Evergreen Analytics

Lauren Bates
Evergreen Analytics

October 8, 2025

Abstract

Adolescent English learners with low literacy strive to learn a new language with minimal or no reading skills. Their efforts are often complicated by having special learning needs or limited experiences with formal education. Meanwhile, they need English literacy in middle and high school, where teachers expect students to read to learn rather than learn to read. Evidence on effective interventions to help these students quickly master both English and reading is scarce. This randomized controlled trial study evaluates the impact of a tailored curriculum, Off2Class Foundational Literacy, paired with high-impact tutoring (HIT), on English reading proficiency among English learners with low literacy. It finds that HIT had a large impact (0.35 SD) on English reading proficiency measured by WIDA ACCESS Reading scores relative to a control group that received the curriculum through teacher-led instruction only. Students received an average of 618 minutes of tutoring in groups of 1 to 3 students per tutor. At a district cost of $665 per learner, the findings offer promising evidence for school districts with similar English language development delivery models.

* Corresponding author: hhanson@evergreenanalytics.org

Background

One in every ten middle school students and one in 14 high school students are English learners (U.S. Department of Education, 2024). They represent diverse backgrounds and bring a wide range of academic and English skills to the schools they enroll in. Among them, adolescent English learners enroll in U.S. schools for the first time as middle or high school students or have not reclassified as English proficient by the time they reach middle or high school. These students often have different needs than students who began kindergarten as English learners. They are more likely to be dully classified as students with special needs and English learners (Morgan et al., 2018). They may also have had limited or interrupted formal education experiences (Short & Boyson, 2004). Those with low literacy in their first language are often behind grade level in other content areas. Nevertheless, they find themselves in middle and high school classes where literacy is expected rather than explicitly taught and assigned to teachers who are trained to teach content, not literacy or English language acquisition (Short & Fitzsimmons, 2007).

This sets high barriers to learning English and poses challenges for English as a second language (ESL) teachers in differentiating instruction for students who can read and understand grammar in their first language versus those whose abilities are limited due to minimal formal education or different learning needs. Research suggests that strong literacy skills in English learners' first language transfer to literacy in their second language (Garrison-Fletcher et al., 2022). In other words, students with low literacy in their first language and limited academic English proficiency often take longer to reclassify as English proficient. They are substantially more likely to linger in English learner programs for more than nine years (Thompson, 2015).

Low literacy and English skills raise the risk of adverse outcomes in secondary school and beyond. Compared to former English learners who reclassified as English proficient in elementary school, they are more likely to be retained in their grade, fail content area courses, and drop out of high school (Suárez-Orozco et al., 2010). The longer they remain English learners, the more challenging it becomes to keep pace academically, as the demands of English language proficiency compete with increasingly complex content in upper-grade levels (Valdéz et al., 2017; Suárez-Orozco et al., 2010). Access to courses that prepare them for college and the careers they aspire to may be limited due to time spent in remedial content area and English language development (ELD) courses (Callahan & Shrifer, 2016; Hanson et al., 2016).

As they enter the workforce, individuals with low literacy can expect to earn less than half the annual wages of people with high levels of literacy (Rothwell, 2020). Low literacy and English proficiency are also linked to poor physical and mental health outcomes (Shi et al., 2009; Zhang, 2021) and can even limit problem-solving abilities (Özenç & Çarkıt, 2021). Impacts extend into late life. A longitudinal study found that functionally illiterate older adults were nearly three times as likely to have dementia at baseline (age 64) and twice as likely to develop dementia during the study period (3.5 years) compared to their literate peers (Arce Rentería et al., 2019).

The individual impacts of limited literacy and English proficiency have broad socioeconomic consequences. Most states allocate thousands of dollars per student, in addition to federal aid, to cover the costs associated with ELD (Griffith & Burns, 2025). The longer students linger in ELD programs, the more costly their education becomes. According to the Organization for Economic Cooperation and Development (OECD) 2023 Survey of Adult Skills (2024), 28% of U.S. residents aged 16–64 are functionally illiterate. This translates to 59 million people

according to U.S. Census population estimates (U.S. Census Bureau, 2025), a disproportionate share of whom are first- and second-generation immigrants, along with individuals who speak a language at home other than English (OECD, 2024). Moving all U.S. adults to at least a moderate level of literacy could generate an additional $2.2 trillion in annual income for the country, equivalent to 10% of the country's gross domestic product (Rothwell, 2020).

Although there is abundant research describing the problem, few studies have identified effective strategies for improving literacy among adolescent English learners as they work toward achieving English proficiency. Generally, both English learners and those who are not can benefit from focusing on reading comprehension while gaining content knowledge (Vaughn et al., 2017). Some studies suggest that structured cooperative learning or peer tutoring can improve literacy for English learners. However, the evidence is often outdated, based on sample sizes of fewer than 100 students, and does not include high school students. Nevertheless, one study found peer tutoring effective for English learners in grades three through six (Fuchs et al., 1997). The randomized controlled trial (RCT) found positive effects on reading comprehension for all students in the treatment group (effect size = 0.6), and especially for English learners with learning disabilities (effect size = 1.01).

Combinations of reading interventions that targeted different elements of literacy also showed gains in oral reading accuracy, speed, and comprehension (effect sizes between 0.57 and 0.75 across various measures) for current and former English learners, according to two studies conducted by Graves et al. (2011) (n = 50–59). Gains in reading comprehension (effect size = 0.26 – 0.27) were observed in a small RCT (n = 26) that combined reciprocal teaching of comprehension strategies with cooperative grouping or cross-age tutoring, though they were not statistically significant (Klinger & Vaughn, 1996).

The present study contributes new evidence to this limited body of research. It adds post-pandemic evidence on the impacts of a reading intervention for English learners. It focuses explicitly on adolescent English learners with low literacy levels, including high school students who are underrepresented in existing experimental or quasi-experimental studies. The study provides a framework for larger-scale research. While its sample size is small compared to many education research studies (approximately 150 students), it exceeds that of many studies that inform the existing knowledge base of effective interventions for increasing literacy among adolescent English learners.

Program background

Off2Class is an online platform that offers objectives-based ELD curricula designed to support secondary ESL teachers, including a lesson library of over 1,500 lessons, scaffolding tools, assessments, and data dashboards. Lessons include instructional guidance using research-based strategies and integrated practice for all four language domains (listening, reading, speaking, and writing). Aligned with several ELD frameworks, including WIDA, it offers flexible implementation models to fit each district's English learner service model (e.g., in-person pull-out or push-in, virtual, whole-class, or small group instruction). Off2Class provides teacher training and guidance on selecting the most suitable instructional options for students based on their needs. The Off2Class Foundational Literacy curriculum is designed for secondary students who are developing basic literacy skills. It provides age-appropriate content backed by the science of reading.

Off2Class partners with Littera Education (Littera) to offer virtual tutoring using Off2Class curricula. Littera collaborates with K–12 districts to customize HIT in reading, math, and English language development. They hire tutors through a highly selective process and

ensure they are trained and certified in delivering Off2Class materials. Their model emphasizes strong relationships between tutors and students, aiming to keep students with the same tutor throughout the program. The online platform features audio, video, and interactive tools for students and teachers. Littera's tutoring management system provides real-time data to monitor program efficacy.

Study overview

This study aims to understand the impact of Off2Class Foundational Literacy with HIT on English reading skills among adolescent English learners in grades 6 through 12 who have low literacy skills. Off2Class Foundational Literacy lessons were delivered to adolescent English learners for an intended 1,080 minutes of lessons through twice-weekly 30-minute lessons for 18 weeks through one of two intervention models:

1. *Tutor + Teacher model:* Trained virtual tutors deliver lessons for groups of up to three students. During their ELD class, students participate in virtual lessons using a laptop with a camera, headphones, and a microphone. Their ELD classroom teacher monitors their behavior and engagement while they are online with their tutor. Students receive teacher-led instruction as usual on days when the class meets and no tutoring session is scheduled. In mixed-eligibility class sections, teachers lead instruction with ineligible students while students work with their tutors.

2. *Teacher-only model:* Teachers deliver Off2Class Foundational Literacy lessons to a small group of eligible students for 30 minutes twice weekly. In mixed-eligibility class sections, while teachers lead instruction for eligible students,

ineligible students work independently. Students receive instruction in their ELD

class as usual outside of the Off2Class Foundational Literacy lesson time.

The study compares the impact of the two Off2Class Foundational Literacy program

models—the "Teacher + Tutor" model and the "Teacher-only" model—on English proficiency

and literacy as measured by growth in WIDA ACCESS Reading scores. It further details how the

program models were implemented and describes the study's context to characterize how the

findings might apply in other settings or define areas for further research.

Evaluation questions

This study has seven evaluation questions:

1. What are the demographic and academic characteristics of the students who are

   offered tutoring and the schools where tutoring is offered?

2. What is the average dosage of tutoring students received?

   a. Does dosage vary meaningfully by student characteristics, grade band, or

      the variant of the Off2Class model that students received (i.e., Tutor +

      Teacher or Teacher-only)?

3. How are the usability and usefulness of the tutoring program perceived by tutors

   and teachers?

4. Were core components of the Off2Class models—including tutor and teacher

   training, completion of expected Off2Class lessons and homework, and tutor use

   of expected Littera platform tools—implemented as intended? What factors

   affected implementation quality?

5. How likely are students to achieve higher growth on WIDA ACCESS Reading scores when they are assigned to receive teacher-led Off2Class versus Off2Class with high-impact tutoring (HIT)?

    a. How does this likelihood vary across different grade bands and student demographics?

6. How does overall participation and engagement in each of the Off2Class program models correlate with student growth on the WIDA ACCESS Reading assessment?

7. What is the estimated per-pupil expense to the district for each of the Off2Class program models?

Sample and context

The study, a cluster randomized controlled trial (cRCT), takes place in Richland School District Two in Columbia, South Carolina. It includes 158 multilingual learners in grades 6 through 12 with January 2024 WIDA ACCESS scores between 1.0 and 2.5 in the Reading domain. WIDA ACCESS proficiency scores range from 1.0 to 6.0 and pertain to scale scores that range from 100 to 600. The assessment consists of four domains: Reading, Listening, Speaking, and Writing, which are combined to form a composite score. Students scoring in the 1.0 to 1.9 range are considered "Entering" proficiency. At this level, they know and can use minimal social and academic language when they have visual and graphic support. Students with proficiency scores between 2.0 and 2.5 are in the lower half of the "Emerging" proficiency level. At this level, they can identify main ideas and participate in discussions related to specific and familiar topics. In addition to having low levels of English proficiency, students in the "Entering" and early "Emerging" proficiency levels often have low literacy in their primary language.

Eligibility. Students were eligible to participate in the study if they had 2024 WIDA ACCESS Reading scores and were enrolled in the district at the start of the intervention (September 10, 2024). This means that 17 students who were newcomers to the district in the 2024–25 school year and took the WIDA Screener test instead of the ACCESS test were excluded from the study. Additionally, eight students who left the district before the intervention began on September 10, 2024, were excluded from the study, regardless of the program model to which their class section was assigned.

District context. The study was conducted in 11 middle and high schools. The district serves about 30,000 students with diverse demographic backgrounds. Sixty-three percent of students identify as Black, 15% identify as white, 14% identify as Hispanic or Latino/a, and 3% identify as Asian. Fifty-six languages are spoken among the student body, with English, Spanish, and Vietnamese being the most common. Thirteen percent of students are military-connected, owing to the local military base, which provides basic combat training for approximately 50% of all soldiers entering the U.S. Army each year.

Classroom context. English learners in Richland School District Two are enrolled in an ELD class with a certified ESL teacher until they reclassify as English proficient. Most ELD class sections enrolled students with different levels of English proficiency. Eligible students attended class with ineligible students (i.e., students who had 2024 WIDA ACCESS Reading scores above 2.5). Multilingual learner class sizes typically enrolled a small number of eligible students—often fewer than five. However, two class sections enrolled at least 10 eligible students and did not enroll ineligible students.

Data and methods

Experimental study. The experimental study is a randomized controlled trial designed to estimate the impact of assignment to the Tutor + Teacher model on WIDA ACCESS Reading scores compared to assignment to the Teacher-only model. Prior to receiving Off2Class Foundational Literacy lessons in September 2024, 158 students with baseline WIDA ACCESS Reading scores below 2.5 were randomly assigned to one of the two models by class section among 38 class sections that enrolled a mix of eligible and ineligible students and within class section for two class sections that enrolled only eligible students (81 in the Tutor + Teacher group and 77 in the Teacher-only group). The effect of being assigned to the Tutor + Teacher group is estimated using an "intent-to-treat" linear regression model with class section cluster robust standard errors adjusted for small sample sizes (Bell & McCaffrey, 2002). A Bayesian Markov Chain Monte Carlo (MCMC) approach complements the frequentist linear model to test the robustness of the results to modeling approaches.

Data. The study uses administrative records from the district, including student enrollment, demographics, student and teacher attendance, and assessment scores. It also includes detailed records of tutoring sessions from the Littera Tutoring Management System (TMS), such as the dates and times of tutoring sessions, attendance, content and skills covered, and tutor comments on student engagement.

Attrition. All class sections and most students (92%) remained in the study (i.e., they had a 2025 WIDA ACCESS Reading score), including 89% of students assigned to the Tutor + Teacher model and 96% assigned to the Teacher-only model (Table A7). Joint F-tests predicting attrition based on observable characteristics failed to reject the null hypothesis that there is no evidence that the baseline characteristics collectively have a significant relationship with attrition

among the entire sample and within the Tutor + Teacher and Teacher-only model samples. Individual observable characteristics were not predictive of attrition among the full sample and within samples of students assigned to each of the models (Table A10).

Baseline equivalency and compliance with random assignment. There were no significant differences between the remaining 72 students assigned to the Tutor + Teacher model and the 74 students assigned to the Teacher-only model across observable baseline characteristics (Tables A3 and A4). Additionally, compliance with initial random assignment was high. There was a 96% compliance rate overall when considering the offer of tutoring (i.e., students were scheduled for tutoring sessions). When considering students who received tutoring, the overall compliance rate was 94%. Sixty-seven of the 72 students assigned to the Tutor + Teacher model were offered tutoring (93%), and 64 received tutoring (89%). Only one student assigned to the Teacher-only model was offered and received tutoring, with lessons beginning in mid-January.

Implementation study. The implementation study provides insight into the extent to which the intervention was implemented as intended, helping to interpret the experimental study's results. It included four data sources: an online survey, online focus groups and interviews, training completion data, and implementation data supplied by tutors via the Littera online platform. The implementation study focused on adherence to the program model, the extent to which students received the intervention, and perceptions about the program.

Additional details about the study design, data, and methods are included in Appendix A.

Findings

This section presents findings for each of the study's evaluation questions. It begins with a summary of the students in the sample and the schools they attend. It then discusses the implementation study findings, detailing insights about the dosage of the intervention students

received, the usability and usefulness of Off2Class Foundational Literacy, and the implementation quality. Finally, it presents student outcomes, including the effects of assignment to the Tutor + Teacher model on WIDA ACCESS Reading scores relative to assignment to the Teacher-only model.

Characteristics of students and schools in the study

The program served the students Off2Class Foundational Literacy was designed for. Before the intervention, students who received Off2Class Foundational Literacy lessons were far behind national averages on the WIDA ACCESS Reading assessment. Compared with all WIDA ACCESS test takers nationally, these students scored nearly one standard deviation below the mean reading score in 2023–24 (Table 1). Most students identified as Latino/a (85%) and spoke Spanish at home (80%). About two-thirds (69%) were newcomers (i.e., they had been classified as English learners for fewer than five years). Similarly, 69% were economically disadvantaged. In addition, 16% were eligible for both ELD and special education services. Finally, about a quarter of students were enrolled in high school in 2023–24, and the remainder were enrolled in grades 5–8.[1] Among students receiving Off2Class Foundational Literacy lessons in 2024–25, about half (49%) were enrolled in high school. Of these students, 43% were assigned to the Tutor + Teacher model, and 54% were assigned to the Teacher-only model.

Table 1. Baseline characteristics of the analytic sample

Characteristic (2023–24)	Full sample		Tutor + Teacher model		Teacher-only model	
	N	Mean (SD)	N	Mean (SD)	N	Mean (SD)
WIDA ACCESS Reading scores (standardized)	146	-0.906 (0.536)	72	-0.883 (0.558)	74	-0.928 (0.517)
Female	67	0.459 (0.5)	32	0.444 (0.5)	35	0.473 (0.503)
Asian	12	0.082 (0.276)	5	a	7	a
Latino/a	124	0.041 (0.199)	62	0.028 (0.165)	62	0.054 (0.228)
Race other than Asian or Latino/a	22	0.151 (0.359)	10	0.139 (0.348)	12	0.162 (0.371)

[1] Students enrolled in grade 5 in the baseline year were enrolled in grade 6 during the intervention.

Characteristic (2023–24)	Full sample		Tutor + Teacher model		Teacher-only model	
	N	Mean (SD)	N	Mean (SD)	N	Mean (SD)
Spanish home language	117	0.801 (0.4)	59	0.819 (0.387)	58	0.784 (0.414)
Economic disadvantage	101	0.692 (0.463)	47	0.653 (0.479)	54	0.73 (0.447)
Special education	24	0.164 (0.372)	11	0.153 (0.362)	13	0.176 (0.383)
Newcomer	100	0.685 (0.466)	49	0.681 (0.47)	51	0.689 (0.466)
Average number of years classified as an English learner	146	3.852 (2.888)	72	3.643 (2.909)	74	4.056 (2.872)
Enrolled in high school vs. middle school	71	0.267 (0.444)	31	0.236 (0.428)	40	0.297 (0.46)
Grade level enrolled in	146	7.527 (1.666)	72	7.431 (1.555)	74	7.622 (1.773)

[a] *Suppressed due to small sample size.*
Source: Authors' analysis of Richland School District Two data, 2024–25.

The characteristics of students assigned to the Tutor + Teacher model and students assigned to the Teacher-only model were similar. There were a few notable differences, none of which were statistically significant (Tables A1 and A2). A higher percentage of students assigned to the Teacher-only model were economically disadvantaged (73%) relative to students assigned to the Tutor + Teacher model (65%). The baseline assessment scores of these students were also 0.045 standard deviations lower. Finally, they were more often enrolled in high school in the baseline year (30% vs. 24%).

The program was implemented in large, diverse secondary schools. In total, 11 schools were included in the study: six middle schools and five high schools (Table 2). In 2023–24, middle schools enrolled, on average, about 1,000 students, and high schools enrolled about 1,800. About 65% of students identified as Black, about 15% identified as white, and about 12% identified as Latino/a. Nearly 1,000 students (6.7%) were classified as English learners across the sample schools. About a quarter of these students were reclassified as English proficient at the end of the year, and a similar percentage met their progress toward proficiency target.

About half of all students enrolled in the study's sample of schools were proficient in English language arts, and a third were proficient in math. The 2023–24 graduation rate was 86%. Only 69% of graduates met the criteria for career readiness, and only 28% met the criteria

for college readiness. Teacher and student perceptions yielded average school climate indicator

scores of 5.7 out of 10 for middle schools and 2.8 out of 5 for high schools.

Table 2. Characteristics of the schools that students in the study attended, 2023–24

Characteristic	Sample middle and high schools	Sample middle schools	Sample high schools
Number of schools[a]	11	6	5
Mean student enrollment	1,355	947	1,823
Student-teacher ratio	19.3	19.2	19.1
Race/ethnicity			
African American or Black	64.7%	63.3%	64.9%
American Indian or Alaska Native	0.2%	0.2%	0.1%
Asian or Pacific Islander	2.8%	2.7%	2.8%
Hispanic or Latino/a	12.4%	12.5%	11.7%
White	14.4%	15.4%	15.3%
Two or more races	5.2%	5.5%	4.9%
Gender			
Female	50%	48.6%	51.1%
Federal program eligibility			
English learner	6.7%	6.3%	6.4%
Economically disadvantaged	63%	60.4%	63.3%
Special education	12.3%	12.2%	12.4%
English learners			
Total number of students	924	513	431
Percentage reclassified as English proficient	23.5%	28.5%	17.8%
Percentage not proficient who met progress toward proficiency target	26.8%	35.4%	20.9%
2024 middle and high school academic performance			
Percentage proficient in English language arts	53.9%	47.4%	64.3%
Percentage proficient in math	34.3%	29.7%	42.1%
Graduation and college and career readiness			
Percentage of first-year students on track to graduate	70%	N/A	70%
Graduation rate (2024)	85.9%	N/A	85.9%
Percentage of students in the graduation cohort who were college-ready	27.8%	N/A	27.8%
Percentage of students in the graduation cohort who were career-ready	69%	N/A	69%
Overall perceptions of school climate[b]	[b]	5.7 / 10	2.8 / 5

[a] *One middle school in the district did not have participating class sections.*
[b] *South Carolina administers school climate surveys to students, parents, and teachers annually as part of its school accountability measures. The School Climate indicator comprises three teacher perception factors and two student perception factors. Elementary and middle schools can achieve a maximum score of 10 points, and high schools can achieve a maximum score of 5 points. As such, the table does not report results for middle schools and high schools combined.*
*Source: Authors' analysis of 2023–24 South Carolina Department of Education data.*

Intervention dosages were lower than intended.

Intervention dosages varied, but they were consistently lower than the intended dosage of 1,080 minutes. Several factors influenced dosage, including school schedules, teacher absences, and scheduled session length. While tutored students typically received less than their intended dosage, they often mastered all skills in the Foundational Literacy curriculum, advancing to more challenging content.

Many students assigned to the Tutor + Teacher model received less than the intended dosage. About three-quarters of students were scheduled to receive the full dosage or more (Figure 1). However, only 10% of students received at least 1,080 minutes of tutoring by the end of the WIDA ACCESS assessment window (March 7, 2025). Further, 2 in 5 students received less than half of the intended dosage.

Figure 1. Percentage of intended dosage, scheduled vs. received
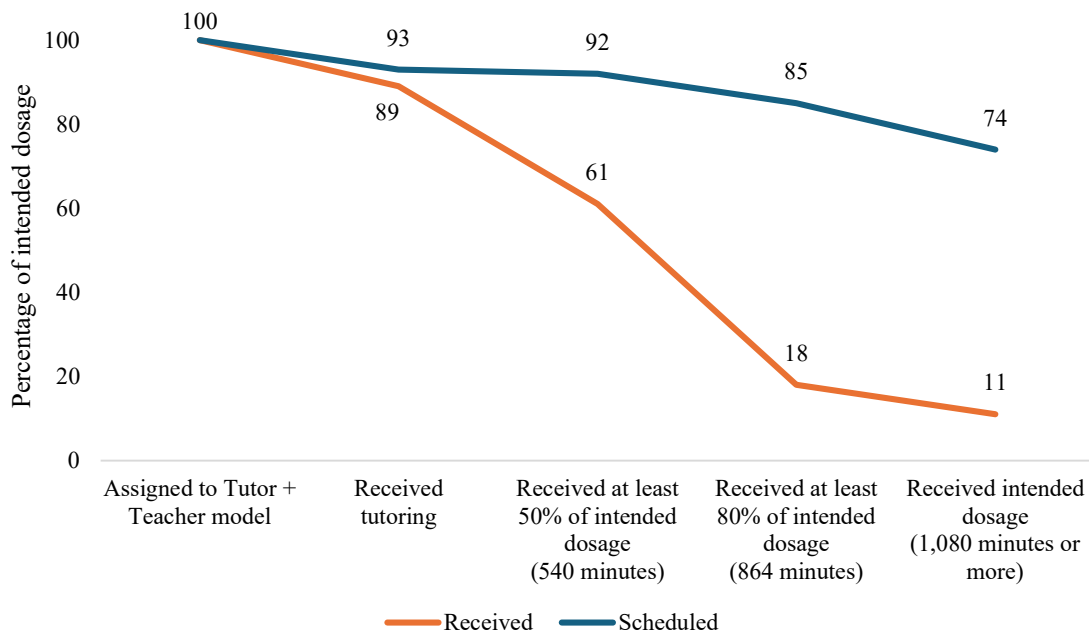
N = 72 students assigned to the Tutor + Teacher model
Source: Authors' analysis of Richland School District Two data, 2024–25.

Three-quarters of students advanced beyond the Foundational Literacy curriculum.
Although they typically did not receive the intended dosage, students often covered more than
one Foundational Literacy skill during the same lesson. Guidance for tutors and teachers was to
customize the learning experience for each student by spending little to no time on content that
students had already mastered. As a result, 77% of students assigned to the Tutor + Teacher
model had advanced to content beyond the Foundational Literacy curriculum before their last
lesson, often in the Off2Class Newcomer Step-by-Step curriculum. Additionally, 53% of students
received Off2Class WIDA ACCESS preparation lessons during tutoring sessions. On average,
students covered 17 unique Off2Class lesson skills in total across their tutoring sessions.

Most student groups received 500–700 minutes of tutoring. On average, students assigned
to the Tutor + Teacher model received 618 minutes of tutoring (Table 3). With two exceptions,
students assigned to this model received between 500 and 700 minutes of tutoring across
demographic and academic characteristics. Students with baseline WIDA ACCESS Reading
scores in the second quartile were the only group to receive over 700 minutes of tutoring.

There were some noteworthy differences in dosage across groups. Male students received
90 more minutes of tutoring than female students, and middle school students received 153 more
minutes of tutoring than high school students. Students who were economically disadvantaged
received an average of 130 more minutes of tutoring than students who were not economically
disadvantaged. Students with special needs received, on average, 52 fewer minutes of tutoring
than students without special needs.

Table 3. Dosage by demographic background characteristics

Group	N	Percentage of intended dosage	Average minutes received
All students	72	57%	618
Economically disadvantaged	47	61%	663

Group	N	Percentage of intended dosage	Average minutes received
Not economically disadvantaged	25	49%	533
Eligible for special education	11	53%	574
Not eligible for special education	61	58%	626
Newcomer English learner	49	58%	630
Long-term English learner	23	55%	592
Female	32	53%	568
Male	40	61%	658
Latino/a	62	56%	602
High school (grades 9–12)	31	49%	531
Middle school (grades 6–8)	41	63%	684
Baseline WIDA ACCESS Reading score quartile			
Bottom quartile	18	50%	536
2nd quartile	17	68%	735
3rd quartile	21	50%	543
Top quartile	16	63%	684

Notes: Grade bands reflect 2024–25 values. All other measures were taken at baseline (2023–24). WIDA ACCESS Reading score quartiles were calculated before dropping records without outcome scores (N = 158).
Source: Authors' analysis of Richland School District Two data, 2024–25.

Busy school schedules limited options for providing the full dosage of tutoring minutes.

As designed, Off2Class Foundational Literacy lessons were to be rescheduled for a different day when missed for any reason, such as school closures, emergency drills, and assemblies. This adaptive scheduling was not implemented for students assigned to the Tutor + Teacher model, and teachers did not specify how they prioritized content for lessons during weeks with school closures or class cancellations. Implementation was likely inconsistent across teachers.

In practice, options for rescheduling missed sessions were limited. During the intervention period (September 10, 2024, through March 7, 2025), the district identified 24 days when no tutoring could be scheduled. The reasons included holidays, a conference, and an e-learning day. In addition, middle schools were scheduled for half days on three days during the intervention period. Inclement weather disrupted an additional six days of learning. Middle school students also had six days of English language arts and math benchmark testing.

Students missed many tutoring sessions when they were present in class. On average, students assigned to the Tutor + Teacher model attended 52% of their scheduled tutoring minutes. Overall, 39% of scheduled minutes were missed due to student absences, 8% of the intended minutes were missed due to students logging in late or leaving early, and 1% were missed due to tutor absences. However, teachers marked students absent from class an average of only 4.2 times during the study period, and they were tardy an average of 0.6 class periods. For 90% of sessions where tutors indicated students were absent, teachers had marked them present in class.

Teacher attendance influenced dosage. In total, teachers were absent 69 days during the study period. Some teachers were absent more than others, with a range of 3 to 21 days (Figure 2). Teacher absences correlated with student absences from tutoring sessions. Among students assigned to the Tutor + Teacher model, 69% were absent from their tutoring session on days when their teacher was absent, compared to 40% on days when their teacher was present.

*Figure 2. Number of teacher absences, by teacher*

N = Nine teachers
Source: Authors' analysis of Richland School District Two data, 2024–25.

There is no documentation regarding whether substitute teachers were able to follow teacher notes to deliver Off2Class Foundational Literacy lessons to students assigned to the Teacher-only model—or if teachers provided Off2Class lesson instructions for substitutes at all. It is assumed students received few to no Foundational Literacy lessons when their teacher was absent, since substitutes received no Off2Class Foundational Literacy training and may have had limited to no training for teaching adolescent English learners in general (Reupert et al., 2023).

Class period length influenced dosage. Session length and frequency varied due to class period length. Four schools—two middle schools and two high schools—scheduled longer class periods that enabled 35- to 45-minute tutoring sessions. A fifth middle school scheduled 41-minute intervention periods for students in grades 7 and 8, but only a 30-minute period for students in grade 6. Overall, students whose sessions were scheduled for 35 to 45 minutes received 773 minutes of tutoring, on average, compared with 525 minutes for students with 30-minute sessions. Further, 30% of students assigned to 35- to 45-minute sessions received the intended dosage, and none of the students assigned to 30-minute sessions received the intended dosage (Figure 3).
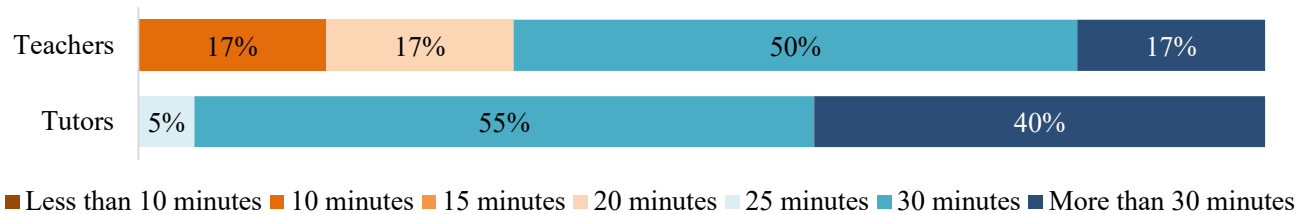
Figure 3. Percentage of intended dosage received, by scheduled session length
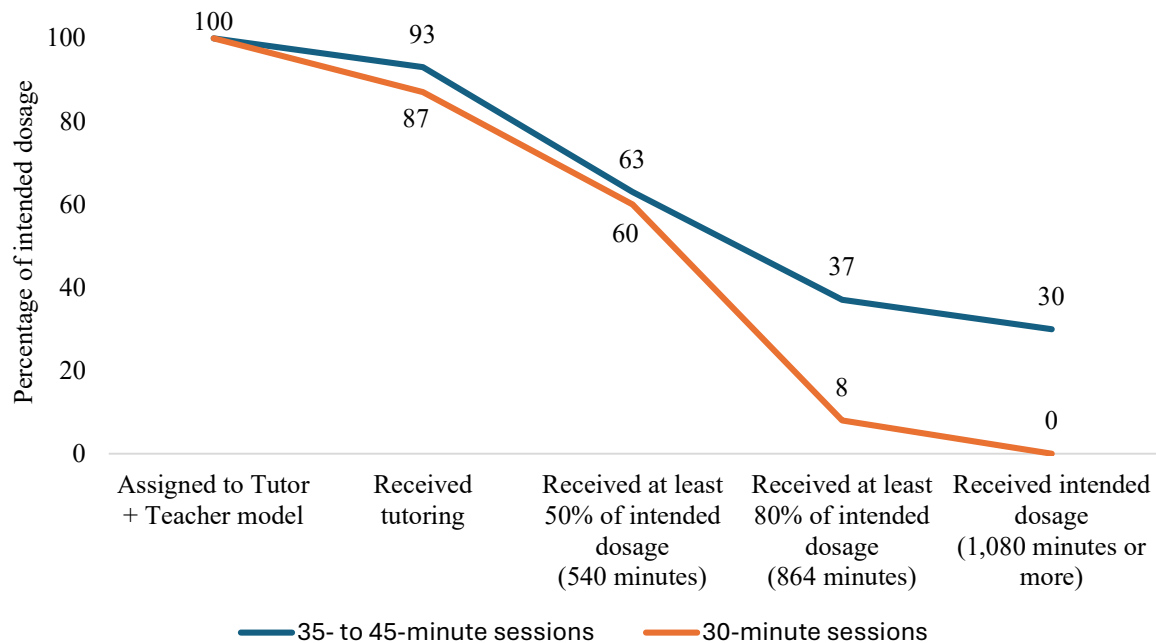
N = 72 students, 45 scheduled for 30-minute sessions and 27 scheduled for 35- to 45-minute sessions
Source: Authors' analysis of Richland School District Two data, 2024–25.

Dosage likely varied for students assigned to the Teacher-only model. Teachers did not track the number of minutes each student assigned to the Teacher-only model received Off2Class Foundational Literacy lessons. Rather, they provided overall estimates of how many times per week lessons typically occurred and how many minutes they typically lasted. According to self-reports, most teachers implemented the program as designed. However, survey results do not capture lessons canceled due to teacher absences, school closures, and other reasons. Among survey respondents, 95% of tutors reported that lessons typically lasted at least 30 minutes, but 34% of teachers reported that they typically lasted only 10 or 20 minutes (Figure 4).

*Figure 4. Reported duration of typical Off2Class Foundational Literacy lessons*

Teachers	17%	17%	50%		17%
Tutors	5%	55%		40%	

■ Less than 10 minutes ■ 10 minutes ■ 15 minutes ■ 20 minutes ■ 25 minutes ■ 30 minutes ■ More than 30 minutes

N = Six teachers and 20 tutors
Source: Authors' analysis of Richland School District Two ELD teacher and Littera tutor Off2Class Foundational Literacy program surveys.

When sessions lasted fewer than 30 minutes, two of six teachers indicated that other tasks, instruction, or activities took priority (Figure 5). One teacher noted that problems with instructional technology often explained why lessons were shorter. For many tutors, classroom disruptions sometimes (25%), often (30%), or always (5%) factored into why lessons were shorter than 30 minutes. In open-ended comments, three tutors said that a lack of student participation sometimes shortened lessons. Another two tutors reported that students sometimes pointed their cameras at the wall or pretended their audio did not work.

Figure 5. Situations that prevented lessons from lasting at least 30 minutes
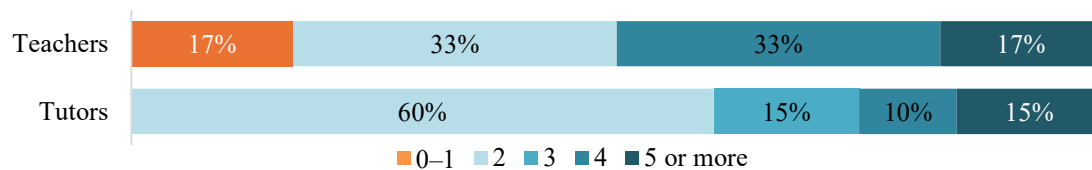

■ Never ■ Occasionally ■ Sometimes ■ Often ■ Always
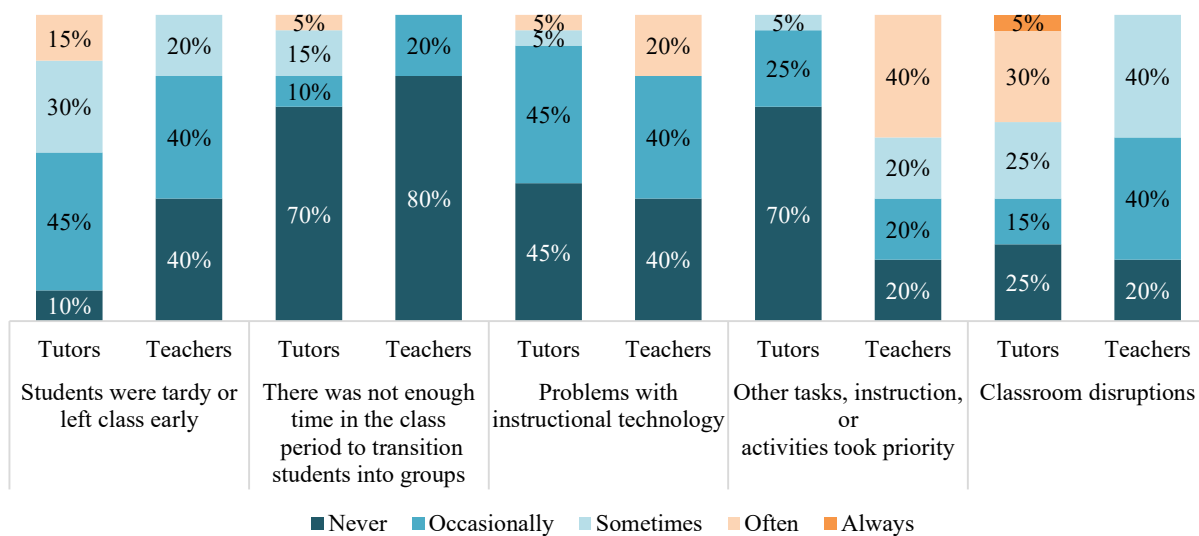
N = Five teachers and 20 tutors
Source: Authors' analysis of Richland School District Two ELD teacher and Littera tutor Off2Class Foundational Literacy program surveys.

Teachers and tutors reported that students typically received the minimum weekly dosage.
All tutors reported that there were typically at least two weekly sessions for each of their
Off2Class Foundational Literacy groups (Figure 6). One teacher indicated that students were
typically receiving no weekly lessons or just one.

Figure 6. Typical number of weekly lessons reported for students assigned to the Off2Class Foundational Literacy program
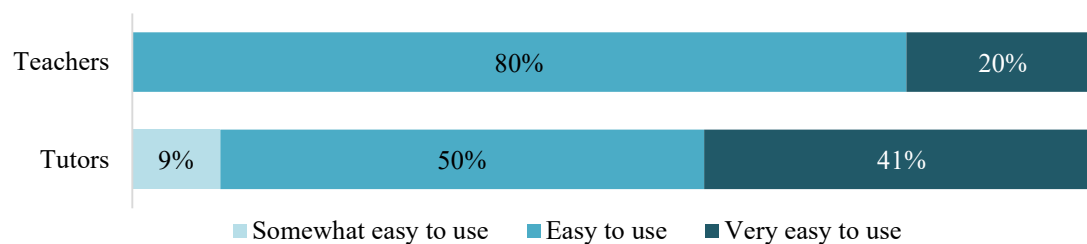
Teachers	17%	33%	33%	17%
Tutors	60%		15%	10%	15%

■ 0–1   ■ 2   ■ 3   ■ 4   ■ 5 or more

N = Six teachers and 20 tutors
Source: Authors' analysis of Richland School District Two ELD teacher and Littera tutor Off2Class Foundational Literacy program surveys.

Teachers and Tutors found Off2Class Foundational Literacy easy to use and helpful in teaching basic English skills.

Overall, surveyed teachers and tutors found the Off2Class Foundational Literacy program easy to use, with one teacher and nine tutors reporting that it was very easy to use (Figure 7). In focus groups, teachers unanimously said the program was easy to use. One teacher said, "I found the curriculum really easy, I mean, very self-explanatory to follow through. I didn't have any challenges with the curriculum."

Figure 7. Reported ease of using the Off2Class Foundational Literacy program

Teachers		80%	20%
Tutors	9%	50%	41%

■ Somewhat easy to use   ■ Easy to use   ■ Very easy to use
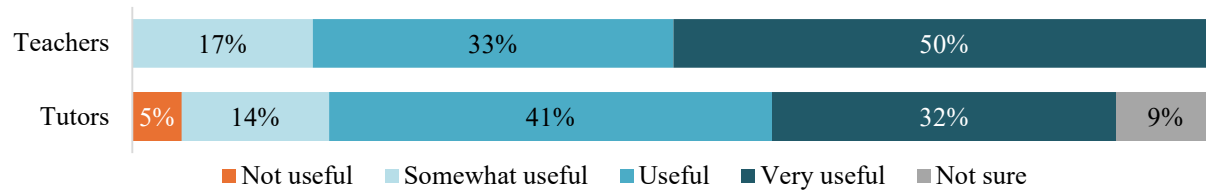
N = Five teachers and 22 tutors
Source: Authors' analysis of Richland School District Two ELD teacher and Littera tutor Off2Class Foundational Literacy program surveys.

One teacher described overcoming her initial trepidation about using the new curriculum. She said, "I'm learning it with my students. At first, I was stressed … But, you know, once I started, it was not as hard as I thought, and the short recordings they have are the best, best thing because really, even if you don't know how to do it, you just listen to the recording."

About half of the interviewed tutors held similar opinions, saying the curriculum was "very easy and fun." One said, "To me, it was just common sense because, apart from the training, with each lesson, you get that teacher slide where it just directs you to what you really have to do. So it was, for me—I don't want to say it was super easy, but it was. It wasn't too difficult." Four tutors recommended including additional instructor guidance or more varied content for different age groups and backgrounds to help customize lessons and support tutors who have little experience teaching English learners.

Most tutors and teachers believed the Off2Class Foundational Literacy program was useful for teaching students basic English skills. Most survey respondents—83% of teachers and 73% of tutors—considered the Off2Class Foundational Literacy program useful or very useful for teaching basic English skills (Figure 8). "I love this program," one tutor said. "I am sure it will help many students, especially those who are English learners." Another tutor said, "I like using this program, and I experienced how students communicate with confidence and feel comfortable participating in the groups." A teacher said the Off2Class Foundational Literacy program has "great lessons for all stages of English proficiency; I use it daily with my students."

Figure 8. Perceptions about the usefulness of the Off2Class Foundational Literacy program for teaching basic English skills

Teachers	17% · 33% · 50%	
Tutors	5% · 14% · 41% · 32% · 9%	

Legend: Not useful · Somewhat useful · Useful · Very useful · Not sure

N = six teachers and 22 tutors.
Source: Authors' analysis of Richland School District Two ELD teacher and Littera tutor surveys.

On the survey, only one tutor felt the program was not useful, and two were unsure. One of the tutors who felt unsure about the program's usefulness said they had just started in the last two months. Both said it was their first year using Off2Class materials of any kind. The tutor who said the program was not useful reported difficulties engaging students with the content.

During focus groups and interviews, tutors and teachers noted a few ways they felt Off2Class Foundational Literacy was useful for their students. One tutor appreciated the curriculum's inclusion of a lesson about work because it "is getting them [students] ready for the future" by supplying information about careers. The tutor viewed the work lesson as demonstrating high standards for students, regardless of their English skills. Another teacher said, "I think the lessons are sound. They're really good lessons and a good resource to work with ESL students … I think all students would benefit. It's a solid curriculum."

The components of the Off2Class Foundational Literacy intervention were implemented as intended

Implementation quality for training and use of the platform and curriculum were high. Tutors and teachers completed training that supported effective implementation. Littera and Off2Class provided basic training (which included a video and sample lessons) for tutors and teachers before they began using the program with students. All tutors who delivered Off2Class lessons completed online training through Littera. Seven of the eight teachers completed training

through Off2Class. The same seven teachers were also sent all five Off2Class training emails; the eighth teacher did not receive any of the training emails.

The majority of surveyed tutors agreed that the training they received oriented them to the program (Figure 9). Overall, 91% of tutors agreed or strongly agreed the training they received oriented them to how the online tutoring platform works. Fewer tutors felt their training oriented them to the structure of the program's lessons (76%), key topics in English foundational literacy instruction (66%), and student activities included in the program (62%). In the survey, one tutor mentioned a training gap, saying, "This was my first year. It would be nice if we were exposed to the scope and sequence for the program. It affords us the ability to scale back or advance if it is discovered that the students need more support or need to be more challenged."

Surveyed teachers were also generally positive about the training. Overall, 66% agreed that the training oriented them to how the online platform works, and all teachers agreed that the training oriented them to key topics in English foundational literacy instruction. In addition, 83% agreed that the training oriented them to the program's lesson structure and student activities.

Figure 9. Surveyed teachers' and tutors' agreement that the training oriented them to the Off2Class Foundational Literacy program
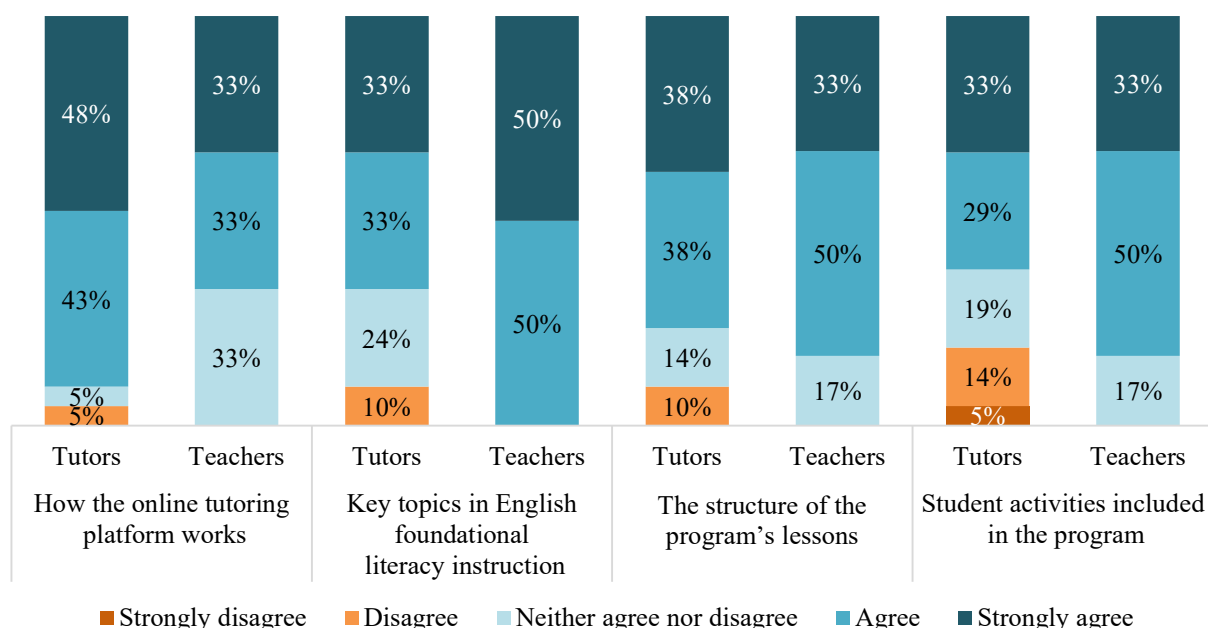
N = Six teachers and 21 tutors
Source: Authors' analysis of Richland School District Two ELD teacher and Littera tutor Off2Class Foundational Literacy program surveys.

During focus groups, teachers were generally satisfied with their Foundational Literacy training, noting they received an orientation via video or during a virtual introduction with Off2Class staff, which they considered sufficient, given the simplicity of the curriculum. However, they said short video clips of teachers providing Foundational Literacy instruction (especially word pronunciation with correct tongue and mouth placement) would also be helpful.

Interviewed tutors had varied reactions to the self-paced training provided through Littera. One said simply, "I think it was good," and described being prepared enough to teach—albeit with some practice. Another tutor said the training was "very brief. You kind of just jump in and kind of figure it out on your own." Another said, "It was very basic. It was basically introducing you to the program, how to use it, how to move, how to steer into getting into each of the programs we need to." A different tutor said, "The lack of training did make it a little

difficult at first to know how to approach it." Another tutor emphasized she did not feel prepared enough after the training; she said, "We're flying by the seat of our pants." This tutor felt she would have benefited from a background in ESL instruction. Tutors also said videos of instruction would be useful additions to the Foundational Literacy training offered by Littera. One tutor requested a "mock class, something where it's interactive" to get oriented to the curriculum.

Tutor data showed high levels of implementation fidelity. Tutors documented data about each session for each student, including student attendance, student engagement score, student skill acquisition, barriers to instruction (e.g., unresolved technical issues), and the number of skills taught. After randomly sampling 50% of recorded sessions, each of the five metrics was rated on a scale of 0 (i.e., no students attended the session) to 5 (e.g., at least one student attended, there were no barriers to instruction, etc.) (see Appendix A for details). Overall, 11 tutors had no completed sessions among the randomly selected sessions, and 52 tutors completed at least one session with a student. The average implementation score for the 52 tutors who worked with students was 4.67. The lowest average implementation score for a tutor was 3.8, and the highest was 5.0. In total, 21 tutors achieved a perfect 5.0 implementation fidelity score.

Instructional time for the Off2Class Foundational Literacy curriculum was dedicated as expected. Teachers were asked to deliver Off2Class Foundational Literacy lessons to students assigned to the Teacher-only model for 30 minutes twice a week (i.e., one hour per week). They could choose how to spend the remainder of their weekly time with students. For class sections that met four to five hours a week, 25% of instructional time for Off2Class Foundational Literacy was considered implemented as designed. As expected, teachers reported in the survey

that they spent at least 25% of their weekly instructional time delivering Off2Class Foundational Literacy lessons (Figure 10).

In focus groups, teachers generally did not report using other materials in their classes. One teacher supplemented Off2Class Foundational Literacy materials with software for students to practice talking to a computer, as they would for the WIDA ACCESS exam. Another used curriculum designed to prepare students for the assessment. Other activities included writing practice using slide software and writing in journals.

Figure 10. Percentage of instructional time used for the Off2Class Foundational Literacy curriculum
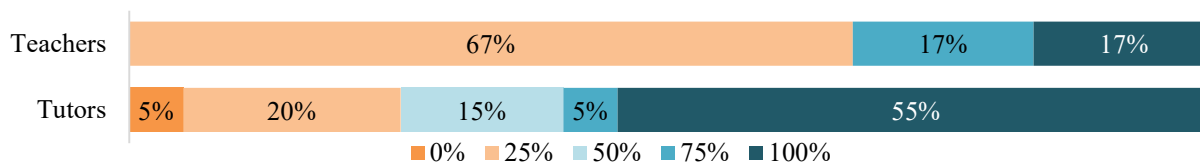
Teachers	67%			17%	17%
Tutors	5%	20%	15%	5%	55%

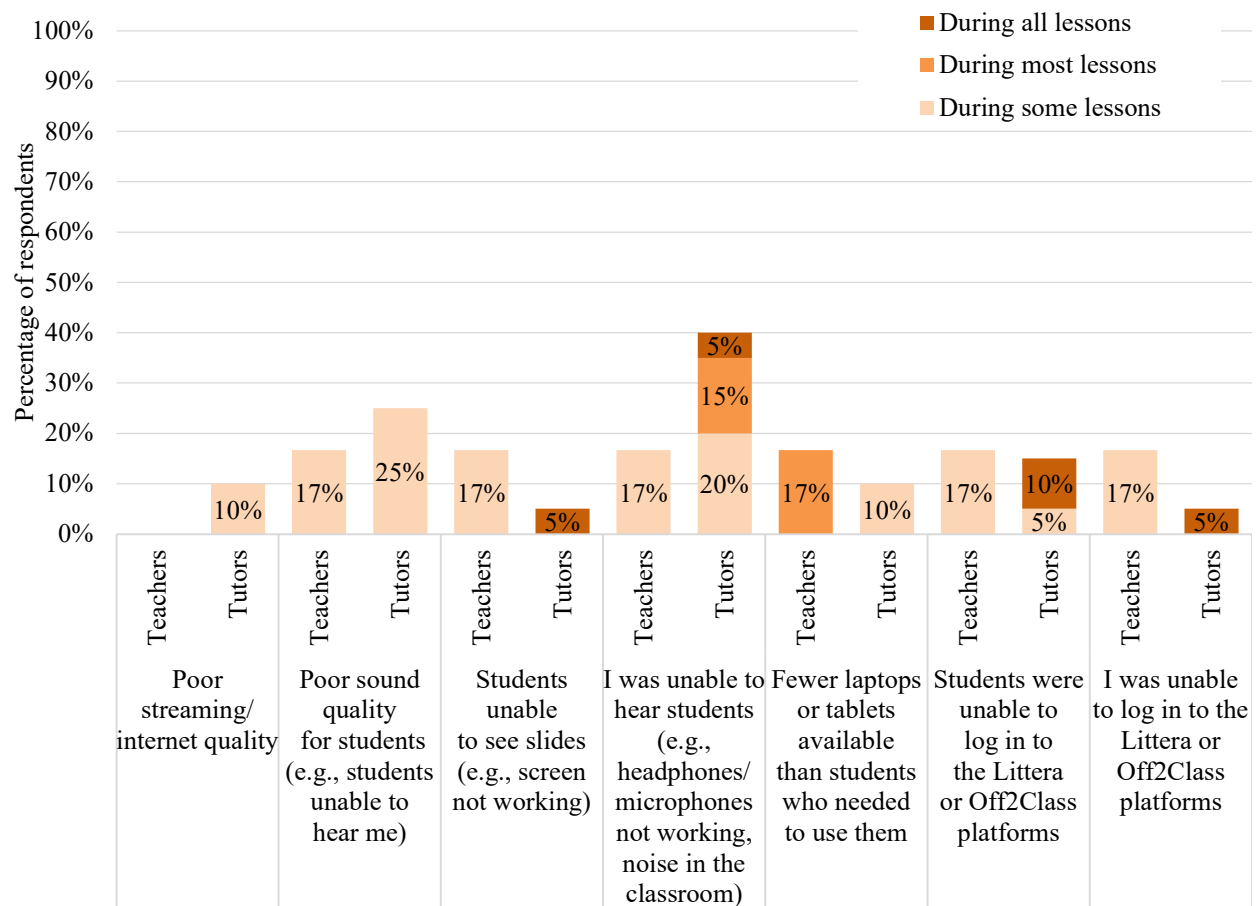■ 0% ■ 25% ■ 50% ■ 75% ■ 100%

N = Six teachers and 21 tutors
Source: Authors' analysis of Richland School District Two ELD teacher and Littera tutor Off2Class Foundational Literacy program surveys.

At each teacher's discretion, tutors were asked to advance students to a more challenging Off2Class curriculum who demonstrated readiness to move on. Most of the 25% of tutors who responded to the survey and indicated they did not spend all lessons on the Foundational Literacy curriculum reported using more advanced Off2Class curricula. However, one tutor who used Off2Class Foundational Literacy materials during only 25% of sessions reported using the Reading Horizons and Littera materials. In focus groups, tutors reported using only the Foundational Literacy curriculum. Two supplemented it with images from online searches and simple vocabulary definitions to help student comprehension, such as showing what a beagle is when students said they did not know. Another tutor said she used small whiteboards as a way for students to practice writing answers, rather than using the platform's embedded whiteboard feature.

Classroom conditions and other challenges to implementation quality. Few technical

problems interfered with Off2Class Foundational Literacy lessons. Sound and noise issues were

an exception for tutors. Seven tutors (35%) reported that the inability to hear students (e.g.,

headphones/microphones were not working, there was noise in the classroom) interfered with

lessons during some, most, or all lessons (Figure 11). Five tutors who reported being unable to

hear students also said students could not hear them due to poor sound quality.

Figure 11. Technical problems that interfered with Off2Class Foundational Literacy lessons

N = Six teachers and 20 tutors
Source: Authors' analysis of Richland School District Two ELD teacher and Littera tutor Off2Class Foundational
Literacy program surveys.

Interviews with teachers and tutors revealed the types of classroom disruptions affecting instruction. Some teachers described disruptions stemming from tutoring. One teacher said she had a class with multiple groups receiving tutoring at the same time—and all the groups were working on different lessons. "And then it would start to echo … it was kind of frustrating at times because there was a lot going on in one room," she said. "I feel like if they all would have had the same tutor, that would have been helpful."

Another teacher said, "I actually started sending a couple of my kids in the hallway to do it because it was so much. And then some kids, their voices were a lot louder too [when they were] answering questions. I mean, they all had on headphones, but it was like they were screaming. Some of them, I mean, it was so loud. And the other kids were, like, 'Oh my goodness.'" One teacher said she "wasn't happy" about tutors using Spanish with students because "the students that they use Spanish with are the ones that I fight the most with for getting off Google Translate and trying to learn English."

In interviews, tutors also mentioned distracting conditions in classrooms. One said, "There's a whole lot of disruption going on. There's also this situation where the whole class is doing something different, and these [tutoring] students are sandwiched between other classmates … Put yourself in the shoes of a high schooler who is challenged, and you're trying to teach them, and you need them to engage, and they've got their peers on the right and left of them. It's very influential to them, peer pressure, and that's difficult … Oftentimes, I am writing back to our platform and saying, 'This situation is going on. I can't even hear myself talk, and you can hear the other lessons sometimes.'"

Students receiving Off2Class Foundational Literacy lessons made notable gains in their WIDA ACCESS Reading scores

On average, students in the study gained half a standard deviation in WIDA ACCESS Reading scores between the 2023–24 and 2024–25 academic years (Table 4). Students assigned to the Tutor + Teacher model gained 0.656 standard deviations between these school years, and students assigned to the Teacher-only model gained 0.34 standard deviations.

*Table 4. Mean standardized WIDA ACCESS Reading scores in 2023–24 and 2024–25*

Off2Class Foundational Literacy model assignment	N	Mean 2023–24 standardized WIDA ACCESS reading score	Mean 2024–25 standardized WIDA ACCESS reading score	Mean difference between 2023–24 and 2024–25 standardized scores
Full sample	146	-0.872	-0.374	0.498
Tutor + Teacher	72	-0.838	-0.182	0.656
Teacher-only	74	-0.905	-0.565	0.340

Note: WIDA ACCESS reading scores were standardized within grade level using national 2024 WIDA means and standard deviations (Center for Applied Linguistics, 2025). Means were weighted by class section and strata random blocks.
Source: Authors' analysis of Richland School District Two data, 2023–24 through 2024–25.

Gains far exceeded national averages for students assigned to the Tutor + Teacher model. For comparison, a study using nationwide WIDA ACCESS data found middle school students gained, on average, 7.5 scale score points (0.22 standard deviations) in reading annually between 2020–21 and 2022–23 (Poole & Sahakan, 2024) (Figure 12). High school students gained 11 scale score points (0.29 standard deviations). This means growth for high school students assigned to the Teacher-only model was comparable to the national average. The gains made by high school students assigned to the Tutor + Teacher model were nearly three times the national average gain. Middle school students assigned to the Teacher-only model also grew about 1.6 times as much as the national average. Their peers assigned to the Tutor + Teacher model gained over twice the national average.

Figure 12. WIDA ACCESS Reading score growth compared to the national average

Annual WIDA ACCESS Reading domain scale score growth (standardized)
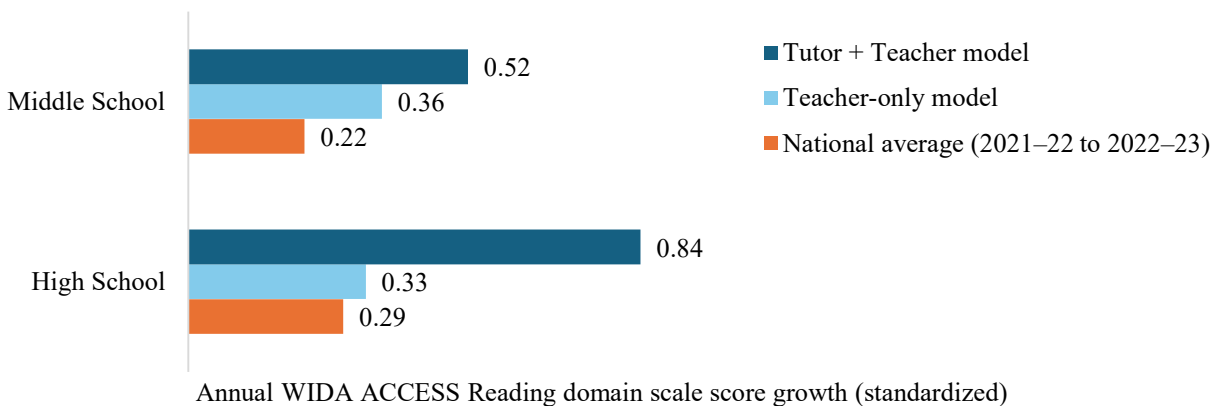
Note: Middle school students were in grades 6–8 and high school students were in grades 9–12 in 2024–25. Means are weighted by class section and strata random blocks. Standardized scores are calculated based on national means and standard deviations for the 2020–21 through 2023–24 WIDA ACCESS administrations (Center for Applied Linguistics, 2023; 2024; 2025).
Sources: Poole & Sahakan, 2024; authors' analysis of Richland School District Two data, 2023–24 through 2024–25.

  Growth resembled the large gains typically found among only first-year newcomer English learners. National averages mask variation in the pace of English language acquisition among different groups of English learners. In their first years, students typically make large gains toward English proficiency that taper off the longer they remain English learners (Thompson, 2015). One state-level study found that adolescents classified as English learners for fewer than three years achieved WIDA ACCESS Composite score gains of approximately 15 to 22 scale score points (approximately 0.26–0.38 standard deviations) between the first and second years they were observed in the data (Umansky et al., 2022). Annual growth slowed in subsequent years (Figure 13).

  The authors did not publish domain-level results. However, Reading domain growth was comparable to first-year composite score gains for students assigned to the Teacher-only model. It was substantially higher among students assigned to the Tutor + Teacher model (see Figure 13). This is notable, given that many students in the sample were not newcomers; they had been English learners for an average of five years during the 2024–25 school year (see Table 1).

Figure 13. Growth in WIDA ACCESS Composite scores by grade span and number of years as an English learner in Year 1, compared to study findings for Reading domain score growth

WIDA ACCESS Composite score growth by grade span and number of years as an English learner in Year (Umansky et al., 2022)
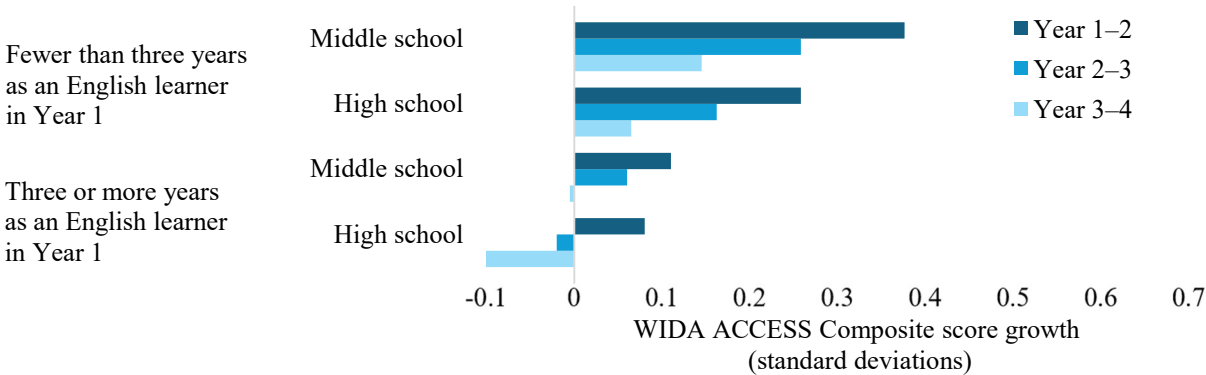
Study findings: WIDA ACCESS Reading score growth by Off2Class Foundational Literacy model assignment
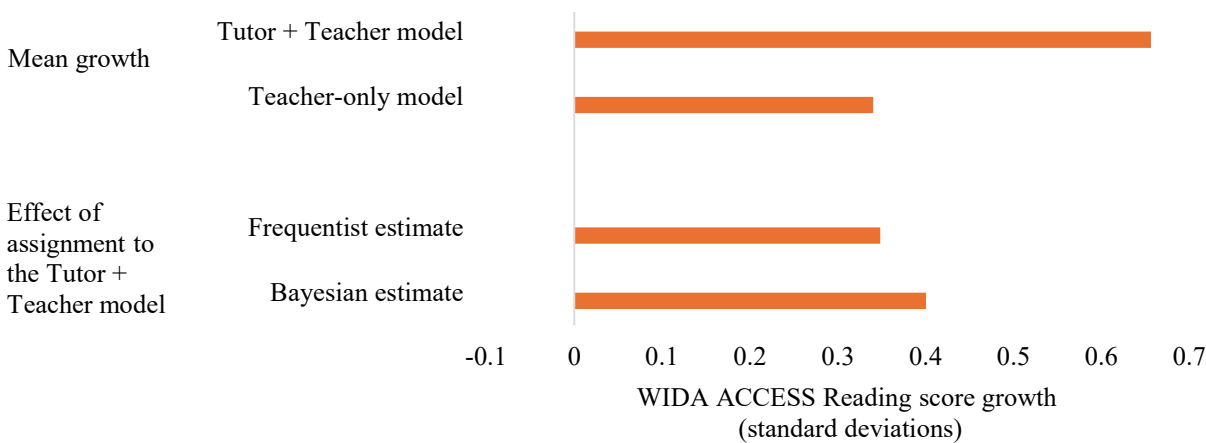
Note: The study findings sample includes 146 students, 72 of whom were assigned to the Teacher-only model and 74 of whom were assigned to the Tutor + Teacher model. The Umansky et al. (2022) study includes 60,778 English learners in grades 6–12, including 10,837 students with fewer than three years classified as an English learner in Year 1 of the study.
Sources: Umansky et al., 2022; authors' analysis of Richland School District Two data, 2023–24 through 2024–25.

Students assigned to the Tutor + Teacher model significantly outpaced the reading score gains of students assigned to the Teacher-only model

Students assigned to the Tutor + Teacher model demonstrated higher growth in the reading domain on the WIDA ACCESS assessment compared to students assigned to the Teacher-only model (see Table 4). After accounting for baseline WIDA ACCESS Reading scores and demographic characteristics, students assigned to the Tutor + Teacher model outpaced the

growth of students assigned to the Teacher-only model by 0.348 standard deviations, a statistically significant difference ($p = 0.02$) (Figure 13, Table 5).

A Bayesian modeling approach estimates a mean treatment effect of 0.4 with a 94% certainty that the true effect of assignment to the Tutor + Teacher model lies between 0.101 and 0.692 standard deviations (Figure 14). Because the most credible values for the true effect size are all above 0.1, we can be confident that the effect of being assigned to the Tutor + Teacher model is positive and nontrivial.

Figure 14. Posterior distribution of treatment effect (assignment to the Tutor + Teacher model)
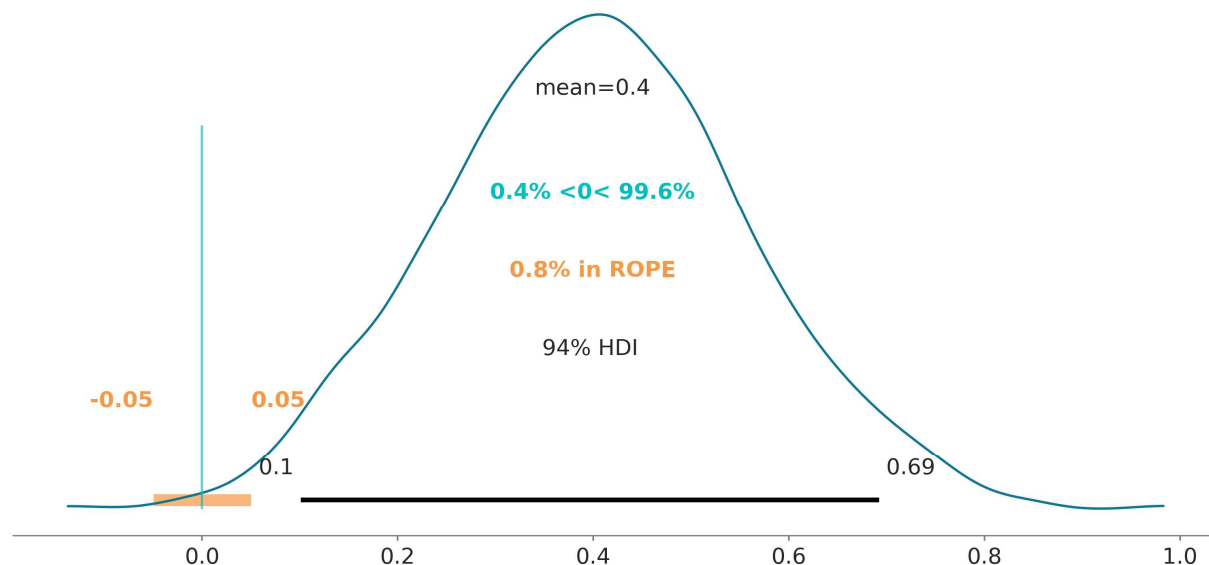
N = 146
Source: Authors' analysis of Richland School District Two data, 2023–24 through 2024–25.

Gains for students assigned to the Tutor + Teacher model were meaningfully different from those observed for students assigned to the Teacher-only model

In the Bayesian model's results, most of the distribution of possible effects (99.2%) lies outside the Region of Practical Equivalence (ROPE). The ROPE represents the range in which there is no meaningful difference in reading growth between the two Off2Class models. Teacher-led tutoring interventions tend to have effect sizes that are 0.05 standard deviations higher than

34

tutoring interventions led by paraprofessionals; thus, the ROPE was set at 0.05 (Nickow et al., 2020). The results suggest there is strong evidence that assignment to the Tutor + Teacher model has a meaningful effect on WIDA ACCESS Reading scores. This conclusion holds up when expanding the size of the ROPE. When setting the ROPE at 0.1, 97.6% of the distribution lies outside the ROPE, and when the ROPE is 0.15, 93.8% of the distribution is outside the ROPE.

The effect of assignment to the Tutor + Teacher model is larger than a full year of progress toward English proficiency, according to national averages. The effect of assignment to the Tutor + Teacher model vs. the Teacher-only model (0.348 to 0.4 standard deviations) represents about **1.2 to 1.8 years of progress towards proficiency** according to the national average single-year gains displayed in Figure 12 (0.22 and 0.29 standard deviations). They are also comparable to the first-year composite score gains for newcomer English learners shown in Figure 13. In addition, an effect of 0.348 to 0.4 standard deviations—depending on the chosen model—is approximately 0.15 to 0.3 standard deviations larger than treatment effects found among math and English language arts tutoring interventions for adolescents (Nickow et al., 2020).

The effect of assignment to the Tutor + Teacher model varied across demographic groups, but it was consistently positive. The effect of assignment to the Tutor + Teacher model on WIDA ACCESS Reading scores ranged from 0.09 for middle school students to 0.8 for students with special needs, according to frequentist estimates (Table 5). Effects were substantially larger for high school students than they were for middle school students, as well as for students who were not economically disadvantaged compared with economically disadvantaged students, regardless of analytic approach (i.e., frequentist or Bayesian). In addition, assignment to the Tutor + Teacher model had larger effects on reading scores for male students, students with

special needs, and long-term English learners than for female students, students without special needs, and newcomer English learners.

Table 5. Effects of being assigned to the Tutor + Teacher model on WIDA ACCESS Reading scores by demographic background characteristics

Characteristic	Frequentist model treatment effect (standard error)	Bayesian model mean treatment effect (SD) [94% HDI]
All students	0.348* (0.144)	0.400 (0.158) [0.101, 0.692]
High school (grades 9–12)	0.616 (0.279)	0.489 (0.244) [-0.016,0.88]
Middle school (grades 6–8)	0.088 (0.188)	0.095 (0.197) [-0.285,0.455]
Female	0.296 (0.199)	0.212 (0.205) [-0.127,0.633]
Male	0.395* (0.154)	0.378 (0.222) [-0.011,0.796]
Asian	0.363 (0.492)	0.497 (0.241) [0.019,1.031]
Latino/a	0.294 (0.446)	0.408 (0.144) [0.117,0.651]
Economically disadvantaged	0.164 (0.314)	0.281 (0.168) [-0.036,0.558]
Not economically disadvantaged	0.762* (0.310)	0.696 (0.206) [0.338,1.091]
Student with special needs	0.801 (0.327)	0.427 (0.358) [-0.134,1.103]
Student without special needs	0.258 (0.159)	0.147 (0.155) [-0.14,0.435]
Newcomer	0.276 (0.242)	0.402 (0.152) [0.077,0.645]
Long-term English learner	0.509** (0.165)	0.551 (0.176) [0.235,0.937]

* p < 0.05, ** p < 0.01
Note: Results represent marginal effects estimated from the full analytic sample. Models include covariates for grade band, race/ethnicity, economic disadvantage, special education, and newcomer status. Standard errors are clustered by class section, adjusted for small sample sizes. p-values are based on stratified wild bootstrap t-statistics with 2,000 draws. Newcomer English learners have fewer than five years as an English learner, and long-term English learners have been English learners for at least five years.
Source: Authors' analysis of Richland School District Two data, 2023–24 through 2024–25.

Effects were largest for students receiving 550 to 700 minutes of tutoring. Compared with students assigned to the Teacher-only model, students assigned to the Tutor + Teacher model experienced higher growth when they received 550–700 minutes of tutoring (roughly 18 to 20

lessons) than when they received less than 550 minutes or more than 700 minutes of tutoring (Table 6). The average dosage students received—618—falls within this range.

Table 6. Treatment effects varied by dosage

Dosage	N	Frequentist model treatment effect (standard error)	Bayesian model mean treatment effect (SD) [94% HDI]
Fewer than 400 minutes	20	0.118 (0.144)	0.059 (0.150) [-0.205, 0.373]
400 to 549 minutes	15	0.178 (0.122)	0.067 (0.272) [-0.400, 0.690]
550 to 699 minutes	18	0.237 (0.199)	0.112 (0.264) [-0.333, 0.696]
700 or more minutes	19	0.191 (0.666)	0.025 (0.246) [-0.423, 0.549]

Source: Authors' analysis of Richland School District Two data, 2023–24 through 2024–25.

Per-pupil program costs total approximately a third of state and federal funding for English learners in the district

The program's estimated per-pupil cost to the district is $664.52. The estimate includes costs for implementing both Off2Class Foundational Literacy delivery models described in the study (Tutor + Teacher and Teacher-only) for 158 students. It also includes costs for 81 students who received tutoring and 77 who received only Teacher-led instruction—including those who could not be included in the analytic sample—along with the 83 tutors who worked with students during the study and nine teachers. This estimate takes personnel, technology, and in-kind donations from the school district, such as staff time and school building space, into account (Table 7)

For context, the per-pupil cost of the intervention is 15% to 67% of the $1,000 to $4,300 per-pupil costs of HIT interventions estimated in prior research (Kraft & Falken, 2021; Guryan et al., 2021), and it represents about a third of the per-pupil funds the district received from federal Title III grants and the state to support English learners.[2]

---

[2] South Carolina allocates 20% above base funding for each English learner student (Griffith & Burns, 2025), which amounts to $1,807.80 in 2024–25 for Richland School District Two (South Carolina Department of Education, 2025a). The district also received $86.84 per student in Title III funds in 2024–25 (South Carolina Department of Education, 2025b).

Table 7. Components of program costs to the district

Component	Description	Unit	Quantity	Price per unit	Total
Personnel					
Tutors (N = 83)	Professional tutors deliver lessons to groups of 1–3 students virtually in their classrooms for 30–45 minutes twice a week, depending on school and classroom schedules. Cost assumes 81 students receive the intended 18-hour dosage.	Hours	1,458	$18	$26,244
HR personnel	Recruit and hire tutors	Hours	100	$90	$9,000
IT personnel	Littera, Off2Class, and district IT support	Hours	120	$60	$7,200
Off2Class customer success manager	Supports rostering, teacher training, and other activities to ensure implementation fidelity	Hours	40	$80	$3,200
Data specialist	Supports district and internal staff	Hours	40	$60	$2,400
District coordinator	A district coordinator organizes scheduling, teacher onboarding, administrative tasks, and contract management	Hours	100	$60	$6,000
Training and support					
Teacher training (N = 9)	Grades 6–12 teachers receive five hours of training on using Off2Class instructional materials, resources, and associated technology	Hours	45	$52	$2,340
Tutor training (N = 83)	Tutors receive one hour of training for Off2Class Foundational Literacy	Hours	83	$18	$1,494
Tutor coaches or supervisors	Supervises tutors	Hours	180	$60	$10,800
Facilities for teacher training	School meeting room or classroom	Hours	2	$10	$20
Travel for training and support	Off2Class site visit to support implementation and training, including airfare, lodging, and meals	Trips	1	$3,000	$3,000
Equipment and materials					
Software licenses	Tutoring platform licenses	Fixed price	81	$36	$2,916
	Off2Class platform subscription	Fixed price	158	$60	$9,480
Student laptops (N = 158)	Student use of their school-assigned laptop with headphones and microphone	Hours	2,844	$0.05	$142.20
Teacher laptops (N = 9)	Teacher use of their school-assigned laptop	Hours	162	$0.05	$8.10
Other					
HR Processing	Fingerprinting and background checks	Fixed price	83	$250	$20,750
Total					$104,994.30

Note: Hourly rates for personnel include benefits at 1.5 times the base hourly rate.
Source: Authors' analysis of Off2Class, Littera Education, and Richland School District Two data, 2025.

Implications for practice and further inquiry

This study found that Off2Class Foundational Literacy lessons, especially when paired with a tutor, can accelerate the path to English reading proficiency for adolescent English learners with low literacy levels. Estimating the causal impact of tutoring on English reading proficiency has not been previously attempted for this population. The substantial difference in WIDA ACCESS Reading growth that students assigned to the Tutor + Teacher model achieved relative to those assigned to the Teacher-only model suggests that tutoring combined with specialized curricula is a promising intervention for reducing the time students spend classified as English learners and supporting their ability to read and learn content on grade level.

When adolescent English learners can linger in ELD programs for more than nine years or never reclassify (Thompson, 2015), interventions that speed English learners to proficiency and boost their reading skills can help individual students advance in their academic careers and improve their postsecondary education and work opportunities (Callahan & Shrifer, 2016; Hanson et al., 2016). At scale, interventions that can shorten the length of time it takes English learners to achieve proficiency can reduce public education costs while improving the economic prospects of the communities they join as adults (Griffith & Burns, 2025; Rothwell, 2020).

This study offers promising evidence for school districts with similar ELD delivery models—that is, they enroll English learners in an ELD course with an ESL-certified teacher. It also provides a platform for further inquiry. Little is known about the individual and societal short- and long-term impacts of interventions that accelerate English proficiency and literacy among students who have little of either. Valuable areas for further research include the following:

- *Improve understanding of the individual and combined impacts of specialized curriculum and tutoring*. This study's sample was not large enough to create a control group that did not receive Off2Class Foundational Literacy instruction or tutoring. Future studies should estimate the impact of Off2Class Foundational Literacy with and without HIT relative to a control group that receives "business-as-usual" ELD instruction.

- *Observe the number of years it takes adolescent English learners with low levels of literacy to reclassify as English proficient and estimate the impact of Off2Class Foundational Literacy with and without HIT on years to reclassification*. There is a lack of research and no consensus on how long it takes for adolescent English learners with low literacy levels to reclassify as English proficient. Longitudinal studies can demonstrate whether and by how much Off2Class Foundational Literacy with HIT can reduce time to proficiency.

- *Test the effectiveness of the intervention in different contexts and within different ELD delivery models*, such as pull-out, push-in, and multi-tiered system of support (MTSS). Importantly, many English learners are not enrolled in ELD programs, including roughly half of English learners in Illinois and Georgia (National Center for Education Statistics, 2024). Future studies should quantify the potential benefits of a specialized ELD curriculum paired with a professionally trained tutor for students with limited or no access to ELD programs.

References

Arce Rentería, M., Vonk, J. M. J., Felix, G., Avila, J. F., Zahodne, L. B., Dalchand, E., Frazer, K. M., Martinez, M. N., Shouel, H. L., & Manly, J. J. (2019). Illiteracy, dementia risk, and cognitive trajectories among older adults with low education. *Neurology, 93*(24), 2247–2256. https://doi.org/10.1212/WNL.0000000000008587

Bell, R., & McCaffrey, D. (2002). Bias reduction in standard errors for linear regression with multi-stage samples. *Survey Methodology*, *28*, 169–182. https://www150.statcan.gc.ca/n1/en/pub/12-001-x/2002002/article/9058-eng.pdf?st=O89ftgzd

Callahan, R. M., & Shifrer, D. (2016). Equitable Access for Secondary English Learner Students: Course Taking as Evidence of EL Program Effectiveness. *Educational Administration Quarterly*, *52*(3), 463–496. https://doi.org/10.1177/0013161X16648190

Center for Applied Linguistics. (2022). *Annual technical report for ACCESS for ELLs online English language proficiency test: Series 503, 2021–2022 administration.* [Annual Technical Report No. 18A].

Center for Applied Linguistics. (2023). *Annual technical report for ACCESS for ELLs online English language proficiency test: Series 601, 2022–2023 administration.* [Annual Technical Report No. 19A].

Center for Applied Linguistics. (2025). *Annual technical report for ACCESS for ELLs online English language proficiency test: Part 2: Technical results.* [Annual Technical Report No. 20A]. https://wida.wisc.edu/resources/annual-technical-report-access-ells-online-elp-test-series-602

Finlay, K. & Magnusson, L. M. (2019). Two applications of wild bootstrap methods to improve

 inference in cluster-IV models. *Journal of Applied Economics*, *34*(6), 911–933.

 https://doi.org/10.1002/jae.2710

Fuchs, D., Fuchs, L. S., Mathes, P. G., & Simmons, D. C. (1997). Peer-assisted learning

 strategies: Making classrooms more responsive to diversity. American Educational

 Research Journal, 34, 174–206. https://doi.org/10.2307/1163346

Garrison-Fletcher, L., Martohardjono, G., & Chodorow, M. (2019). Impact of home language

 reading skills' on L2 reading comprehension: a study of newcomer emergent

 bilinguals. *International Journal of Bilingual Education and Bilingualism*, *25*(1), 287–

 300. https://doi.org/10.1080/13670050.2019.1671307

Graves, A. W., Duesbery, L., Pyle, N. B., Brandon, R. R., & McIntosh, A. S. (2011). Two studies

 of tier II literacy development. The Elementary School Journal, 111(4), 641–660.

 https://www.researchgate.net/publication/288558381_Two_Studies_of_Tier_II_Literacy_

 Development_Throwing_Sixth_Graders_a_Lifeline

Griffith, M. & Burns, D. (2025). *Funding student needs: A review of state funding policies for*

 English learners and students from low-income backgrounds. Learning Policy Institute.

 https://files.eric.ed.gov/fulltext/ED670964.pdf

Guryan, J., Ludwig, J., Bhatt, M. P., Cook, P. J., Davis, J. M. V., Dodge, K., Farkas, G., Fryer, R.

 J., Mayer, S., Pollack, H., & Steinberg, L. (2021, March). *Not too late: Improving*

 academic outcomes among adolescents (Working Paper 28531). National Bureau of

 Economic Research. https://www.nber.org/papers/w28531

Hanson, H., Bisht, B., & Greenberg Motamedi, J. (2016). *Advanced course enrollment and*

 performance among English learner students in Washington State (REL 2017–187).

Washington, DC: U.S. Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance, Regional Educational Laboratory Northwest. https://files.eric.ed.gov/fulltext/ED570326.pdf

Kahan, B. C., & Morris, T. P. (2013). Analysis of multicentre trials with continuous outcomes: when and how should we account for centre effects? *Statistics in medicine*, *32*(7), 1136–1149. https://doi.org/10.1002/sim.5667

Kahan, B. C., & Harhay, M. O. (2015). Many multicenter trials had few events per center, requiring analysis via random-effects models or GEEs. *Journal of Clinical Epidemiology*, *68*(12), 1504–1511. https://doi.org/10.1016/j.jclinepi.2015.03.016

Kraft, M. A. & Falken, G. T. (2021). *A blueprint for scaling tutoring across public schools*. (EdWorkingPaper: 20-335). Annenberg Institute at Brown University. https://edworkingpapers.com/ai20-335

Modugno, L., & Giannerini, S. (2015). The Wild Bootstrap for Multilevel Models. *Communications in Statistics - Theory and Methods*, *44*(22), 4812–4825. https://doi.org/10.1080/03610926.2013.802807

Morgan, P. L., Farkas, G., Cook, M., Strassfeld, N. M., Hillemeier, M. M., Pun, W. H., Wang, Y., & Schussler, D. L. (2018). Are Hispanic, Asian, Native American, or Language-Minority Children Overrepresented in Special Education? *Exceptional Children*, *84*(3), 261–279. https://doi.org/10.1177/0014402917748303

Nickow, A., Oreopoulos, P., & Quan, P. (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://www.nber.org/papers/w27476

Organization for Economic Cooperation and Development. (2024, December 10). *Survey of adult skills 2023: United States.* OECD Publishing, Paris. https://www.oecd.org/en/publications/survey-of-adults-skills-2023-country-notes_ab4f6b8c-en/united-states_427d6aac-en.html

Özenç, M., & Çarkıt, C. (2021). The relationship between functional literacy and problem-solving skills: A study on 4th-grade students. *Participatory Educational Research*, *8*(3), 372–384. https://doi.org/10.17275/per.21.71.8.3

Poole, G. A. & Sahakyan, N. (2024). *Examining English learner testing, proficiency, and growth: Continued trends since the COVID-19 pandemic*. [Research Report No. RR-2024-1]. Madison, WI: Center for Applied Linguistics. https://wida.wisc.edu/sites/default/files/resource/Research-Report-Examining-English-Learner-Testing-Proficiency-Growth-2024.pdf

Reupert, A., Sullivan, A., Tippett, N., White, S., Woodcock, S., Chen, L., & Simons, M. (2023). An Exploration of the Experiences of Substitute Teachers: A Systematic Review. *Review of Educational Research*, *93*(6), 901–941. https://doi.org/10.3102/00346543221149418

Rothwell, J. (2020). *Assessing the economic gains of eradicating illiteracy nationally and regionally in the United States.* Gallup. https://www.barbarabush.org/wp-content/uploads/2020/09/BBFoundation_GainsFromEradicatingIlliteracy_9_8.pdf

Shi, L., Lebrun, L. A., & Tsai, J. (2009). The influence of English proficiency on access to care. *Ethnicity & Health*, *14*(6), 625–642. https://doi.org/10.1080/13557850903248639

Short, D. J., & Boyson, B. A. (2004). *Creating access: language and academic programs for secondary school newcomers.* Washington, D.C: Center for Applied Linguistics and Delta Systems Co., Inc.

Short, D. J. & Fitzsimmons, D. (2007). *Double the work: Challenges and solutions to acquiring language and academic literacy for adolescent English language learners*. Washington, DC: Alliance for Excellent Education. https://web.stanford.edu/~hakuta/Courses/Ed330X%20Website/Ed%20330X_2012/DoubleWork.pdf

Suárez-Orozco, C., Gaytán, F. X., Bang, H. J., Pakes, J., O'Connor, E., & Rhodes, J. (2010). Academic trajectories of newcomer immigrant youth. *Developmental Psychology*, *46*(3), 602–618. https://pubmed.ncbi.nlm.nih.gov/20438174/

South Carolina Department of Education. (2025a). *Revenue per pupil reported by school district – fiscal year 2024–2025*. https://ed.sc.gov/finance/financial-services/student-data/revenue-per-pupil-reported-by-school-district-fiscal-year-2024-2025/

South Carolina Department of Education. (2025b). *Title III, Part A/MLP South Carolina Guidance: FY25 ML Funding*. https://docs.google.com/spreadsheets/d/1OZl6QiNsiSKAgOTXp9j2JOhZV-JB0JASZi7i9rPUQMU/edit?gid=309772389#gid=309772389

Thompson, K. D. (2015). English learners' time to reclassification: An analysis. *Educational Policy*, *31*(3), 330–363. https://doi.org/10.1177/0895904815598394

Umansky, I. M., Thompson, K. D., Soland, J., & Kibler, A. K. (2022). Understanding newcomer English learner students' English language development: Comparisons and predictors. *Bilingual Research Journal*, *45*(2), 180–204. https://doi.org/10.1080/15235882.2022.2111618

U.S. Census Bureau. (2025). *Age and sex composition in the United States: 2023.* Current

Population Survey Tables, *Table 1*. https://www.census.gov/data/tables/2023/demo/age-

and-sex/2023-age-sex-composition.html

U.S. Department of Education, Institute of Education Sciences, National Center for Education

Statistics. (2024). English Learners in Public Schools. *Condition of Education*.

https://nces.ed.gov/programs/coe/indicator/cgf

U.S. Department of Education, National Center for Education Statistics, Institute of Education

Sciences. (2025). *What works clearinghouse procedures and standards handbook,*

version 5.0. https://ies.ed.gov/ncee/WWC/Docs/referenceresources/Final_WWC-

HandbookVer5_0-0-508.pdf

U.S. Department of Education, National Center for Education Statistics, Institute of Education

Sciences. (n.d.a.). *Assessing attrition bias*.

https://ies.ed.gov/ncee/wwc/Docs/ReferenceResources/wwc_attrition_v2.1.pdf

U.S. Department of Education, National Center for Education Statistics, Institute of Education

Sciences. (n.d.b.). *Assessing attrition bias – addendum*.

https://ies.ed.gov/ncee/WWC/Docs/ReferenceResources/wwc_attrition_v3.0.pdf

Vaughn, S., Martinez, L. R., Wanzek, J., Roberts, G., Swanson, E., & Fall, A.-M. (2017).

Improving content knowledge and comprehension for English language learners:

Findings from a randomized control trial. *Journal of Educational Psychology, 109*(1),

22–34. https://doi.org/10.1037/edu0000069

Zhang, Q. (2020). The cost of illiteracy: A causal inference study on how illiteracy affects

physical and mental health. *Health Education Journal*, *80*(1), 54–66.

https://doi.org/10.1177/0017896920949894

Appendix A: Study design and methodology details

This appendix includes additional details about the study design and methodology, including baseline equivalency, attrition, and compliance with random assignment.

Study design

Eligible students were randomly assigned to receive Off2Class Foundational Literacy lessons through either the Teacher + Tutor or the Teacher-only model. The random assignment approach, described below, preserves a small group size for students assigned to the Teacher-only model, which represents the intended real-world implementation outside the study (i.e., teachers assign some students to tutors while teaching others in a small group).

Class sections were stratified by whether they were mixed eligibility, enrolling both eligible and ineligible students based on their baseline WIDA ACCESS Reading domain assessment scores or whether they enrolled only eligible students. Eligible students in mixed-eligibility classrooms were randomly assigned to one of the treatment models at the class section level (i.e., cluster random assignment). Within two class sections, all students were eligible, and they were randomly assigned to either the Tutor + Teacher or the Teacher-only intervention model at the student level.

This random assignment approach ensured that students assigned to the Teacher-only model would consistently receive instruction in small groups. Assigning eligible students in mixed-eligibility classrooms to either receive tutors or teacher-led instruction enabled teachers to divide the classroom into no more than two groups at a time—eligible and ineligible groups. Random assignment at the student level within all class sections would require teachers to simultaneously assign ineligible students to independent work, monitor students assigned to tutors, and teach Off2Class Foundational Literacy lessons to a third group of students. Random

assignment within the large class sections with only eligible students means that students experience the intervention in the same way: teachers are always occupied with another group of students while students are tutored and students assigned to the Teacher-only model experience small group instruction.

Sample exclusions. The study randomly assigned 183 students by class section cluster to one of the two intervention models in August 2024. Students were excluded from the sample if they left the district before the intervention began on September 10, 2024 (N = 8) or if they were missing a baseline assessment score (17). The latter condition applies primarily to students who were in their first year in United States schools; the results do not generalize to such students. Exclusion criteria were applied to students in both intervention models in the same way. Students who are not in the sample due to these exclusions are not considered attrition. After exclusions, the initial study sample comprised 158 students, with 81 assigned to the Tutor + Teacher model and 77 assigned to the Teacher-only model.

Analytic approach

Cluster-robust linear models accommodate this study's random assignment approach, incorporating class section cluster-robust standard errors with a correction for small sample sizes (Bell & McCaffrey, 2002) (Equation A1).

$$Y_{ij} = \alpha + \beta(Tutor_{ij}) + \theta X_{ij} + e_{ij} \qquad \text{(A1)}$$

where Y_{ij} is 2025 WIDA ACCESS Reading scores for student i in class section j, Tutor captures the effect of being randomly assigned to the Tutor + Teacher model relative to assignment to the Teacher-only model, X_{ij} represents a vector of student-level covariates including 2024 WIDA ACCESS Reading scores and indicators for gender, racial/ethnic group, grade band (high school or middle school), newcomer status, and eligibility for free and reduced-price lunch, and special

education programs. *e* represents class section cluster-robust residual error adjusted for small sample sizes.

Additionally, the test statistics and *p*-values are based on a stratified wild cluster bootstrapping procedure with 2,000 replicates and Rademacher weights for finite-sample *p*-values (Finlay & Magnusson, 2019; Modugno, 2015). Wild weights are drawn within each stratum for each cluster. Wild cluster bootstrapping t-statistics and *p*-values improve Type I error control in studies with many small clusters. Adjusted cluster-robust *p*-values were compared with *p*-values from the wild cluster bootstrapped analysis to ensure agreement in findings of statistical significance.

Bayesian models accompanied frequentist cluster-robust linear models. Bayesian analysis estimated a range of probable treatment effects and assessed the likelihood that WIDA ACCESS Reading score growth for students assigned to the teacher-led Off2Class model is outside a +/- 0.05 standard deviation region of practical equivalence (ROPE) for students assigned to the Tutor + Teacher model. According to Nickow et. al (2020), teacher-led tutoring interventions tend to have effect sizes that are 0.05 standard deviations higher than tutoring interventions led by paraprofessionals, accounting for other program features. The ROPE, therefore, represents the lack of a meaningful difference between the two Off2Class models.

The model is specified as follows:

$$Y_{ijk} = \alpha + m_{ijk} + w_{ijk} + \theta^R_{r[ijk]} + \theta^F_{f[ijk]} + \theta^C_{c[j]} + u_k + \varepsilon_{ijk}$$

(A2)

where Y_{ij} is the 2025 WIDA ACCESS Reading score for student *i* in class section *j* and strata *k*, as a linear function of an overall intercept α, and the effects of the intervention (m_{ijk}), 2024 WIDA ACCESS Reading scores (w_{ijk}), a set of indicators for race/ethnicity, gender, and grade

band (θ^R), θ^F, a set of indicators for enrollment in federal programs and newcomer status, a random class section effect (θ^C), and a random strata effect (u_j). ε_{ijk} is a residual error term. Note that random effects replace clustering in Bayesian models to capture hierarchical structures since clustered errors are specific to frequentist models. Prior distributions for model parameters are specified below.

$$\alpha \sim N(0,1)$$

$$m_j \sim N(0,\sigma_m^2)$$

$$w_j \sim N(0,1)$$

$$\theta^R \sim N(0, \sigma^R) \text{ with } \sum \theta^R = 0$$

$$\theta^F \sim N(0, \sigma^F) \text{ with } \sum \theta^F = 0$$

$$\theta^C \sim N(0, \sigma^C) \text{ with } \sum \theta^C = 0$$

$$u_j \sim N(0, \tau^2)$$

$$\tau \sim \text{half-Normal}(0, 0.5)$$

The remaining variance parameters—σ_m, σ_w, σ_R, σ_F, and σ_C—are assigned a half-Normal prior distribution centered at zero with standard deviation φ, which is estimated φ from the data, assuming the variance parameters come from a half-Normal distribution with mean 0 and variance φ^2. Estimating φ rather than specifying it a priori allows for an empirical determination of the similarity of variance parameters.

Results are repeated using ROPEs of +/- 0.5, +/- 0.1, and +/- 0.15 standard deviations to estimate the likelihoods that growth in WIDA ACCESS Reading scores differs by a wider margin and improve understanding of the posterior distribution.

Baseline equivalency

After presenting full sample means without accounting for random block effects, an assessment of baseline equivalency is presented. Frequentist and Bayesian models estimate differences in baseline characteristics based on assignment to the Tutor + Teacher model, including class section cluster-robust standard errors adjusted for small sample sizes (Bell & McCaffrey, 2002). Additionally, another set of frequentist and Bayesian Markov Chain Monte Carlo (MCMC) models predict assignment to the Tutor + Teacher group given a baseline characteristic and adjusted class section robust standard errors (frequentist model) or class section and strata random effects (Bayesian model). The Bayesian models assess the robustness of the frequentist results. Bayesian methods facilitate more accurate inferences and quantify uncertainty for studies with small sample sizes and complex model structures. Each Bayesian model includes 1,000 sample draws and 2,000 tuning steps with the same set of covariates as frequentist models.

Observable characteristics used as covariates include the following:

- Baseline assessment scores: 2024 WIDA ACCESS Reading scores, standardized within grade level according to 2024 national WIDA ACCESS Reading score means and standard deviations (Center for Applied Linguistics, 2025).

- Gender: a binary indicator for whether the student is female

- Race/ethnicity: binary indicators for Asian and Latino/a, with students from other racial/ethnic groups as the referent category

- Newcomer: an indicator for whether the student has been an English learner for fewer than five years

- Special education: an indicator for whether the student is eligible for special education services

- Economic disadvantage: an indicator for whether the student is economically disadvantaged

- Home language: an indicator for whether the student's home language is Spanish

- Grade level: measured as a binary indicator for whether the student was in high school vs. middle school or as a categorical variable

Additional model components include:

- Class section: The cluster random assignment variable, comprised of class name, class period, and teacher name. Class section is used to cluster standard errors in frequentist models and as a random effect in Bayesian models.

- Strata: an indicator equal to zero for students who enrolled in a mixed-eligibility class section and were assigned to a treatment model at the class section level and one for students enrolled in a non-mixed-eligibility class section and were randomly assigned to a treatment condition within the class section. Strata is a fixed effect in frequentist models and a random effect in Bayesian models.

Sample means

Ignoring clusters and strata, baseline means were similar between groups (Table A1). There were no noteworthy changes in means between the initial and analytic samples within groups.

Table A1. Sample baseline means and standard deviations without random class section or strata weights

	Initial sample Mean (SD)			Analytic sample Mean (SD)		
Characteristic	**Overall N = 158**	**Tutor + Teacher N = 81**	**Teacher only N = 77**	**Overall N = 146**	**Tutor + Teacher N = 72**	**Teacher only N = 74**
Baseline WIDA ACCESS Reading scores (standardized)	-0.906 (0.536)	-0.891 (0.556)	-0.922 (0.517)	-0.906 (0.536)	-0.883 (0.558)	-0.928 (0.517)
Female	0.449 (0.499)	0.432 (0.498)	0.468 (0.502)	0.459 (0.5)	0.444 (0.5)	0.473 (0.503)
Asian	0.089 (0.285)	0.074 (0.264)	0.104 (0.307)	0.082 (0.276)	0.069 (0.256)	0.095 (0.295)
Latino/a	0.842 (0.366)	0.852 (0.357)	0.831 (0.377)	0.849 (0.359)	0.861 (0.348)	0.838 (0.371)
Race other than Asian or Latino/a	0.158 (0.366)	0.148 (0.357)	0.169 (0.377)	0.151 (0.359)	0.139 (0.348)	0.162 (0.371)
Spanish home language	0.804 (0.398)	0.827 (0.38)	0.779 (0.417)	0.801 (0.4)	0.819 (0.387)	0.784 (0.414)
Economic disadvantage	0.677 (0.469)	0.63 (0.486)	0.727 (0.448)	0.692 (0.463)	0.653 (0.479)	0.73 (0.447)
Special education	0.165 (0.372)	0.16 (0.369)	0.169 (0.377)	0.164 (0.372)	0.153 (0.362)	0.176 (0.383)
Newcomer	0.696 (0.461)	0.704 (0.459)	0.688 (0.466)	0.685 (0.466)	0.681 (0.47)	0.689 (0.466)
High school vs. middle school	0.285 (0.453)	0.247 (0.434)	0.325 (0.471)	0.267 (0.444)	0.236 (0.428)	0.297 (0.46)
Grade level enrolled in	7.557 (1.702)	7.407 (1.579)	7.714 (1.82)	7.527 (1.666)	7.431 (1.555)	7.622 (1.773)

Note: Results for Black and white students are suppressed due to small N.
Source: Authors' analysis of Richland School District Two data, 2023–24.

When weighted by class section and strata (i.e., an indicator for class section by strata combination), baseline means are also similar between the initial and analytic samples overall and within assigned treatment groups (Table A2).

Table A2. Sample baseline means and standard deviations with random class section and strata weights

	Initial sample Weighted mean (SD)			Analytic sample Weighted mean (SD)		
Characteristic	**Overall N = 158**	**Tutor + Teacher N = 81**	**Teacher only N = 77**	**Overall N = 146**	**Tutor + Teacher N = 72**	**Teacher only N = 74**
Baseline WIDA ACCESS Reading scores (standardized)	-0.888 (0.51)	-0.877 (0.546)	-0.9 (0.469)	-0.872 (0.501)	-0.838 (0.533)	-0.905 (0.464)
Female	0.473 (0.499)	0.489 (0.5)	0.456 (0.498)	0.48 (0.5)	0.499 (0.5)	0.462 (0.499)

Characteristic	Initial sample Weighted mean (SD)			Analytic sample Weighted mean (SD)		
	Overall N = 158	Tutor + Teacher N = 81	Teacher only N = 77	Overall N = 146	Tutor + Teacher N = 72	Teacher only N = 74
Asian	0.091 (0.287)	0.07 (0.254)	0.113 (0.316)	0.089 (0.285)	0.078 (0.268)	0.101 (0.301)
Latino/a	0.855 (0.352)	0.886 (0.318)	0.823 (0.381)	0.858 (0.349)	0.885 (0.319)	0.832 (0.374)
Race other than Asian or Latino/a	0.145 (0.352)	0.114 (0.318)	0.177 (0.381)	0.142 (0.349)	0.115 (0.319)	0.168 (0.374)
Spanish home language	0.829 (0.376)	0.854 (0.354)	0.804 (0.397)	0.825 (0.38)	0.838 (0.369)	0.813 (0.39)
Economic disadvantage	0.679 (0.467)	0.644 (0.479)	0.715 (0.451)	0.681 (0.466)	0.644 (0.479)	0.718 (0.45)
Special education	0.222 (0.415)	0.236 (0.425)	0.206 (0.405)	0.215 (0.411)	0.211 (0.408)	0.22 (0.414)
Newcomer	0.653 (0.476)	0.632 (0.482)	0.676 (0.468)	0.645 (0.479)	0.614 (0.487)	0.676 (0.468)
High school vs. middle school	0.315 (0.464)	0.264 (0.441)	0.368 (0.482)	0.298 (0.457)	0.259 (0.438)	0.337 (0.473)
Grade level	7.584 (1.82)	7.29 (1.713)	7.893 (1.878)	7.539 (1.794)	7.289 (1.717)	7.789 (1.834)

Note: Results for Black and white students are suppressed due to small N.
Source: Authors' analysis of Richland School District Two data, 2023–24.

Tests for baseline equivalency

Linear or linear probability models were used to test for significant differences across treatment groups on baseline assessment scores and other observable baseline characteristics. Each characteristic is a function of assignment to the Tutor + Teacher model with class section clustered errors adjusted for small sample sizes (Equation A3).

$$Y_{ij} + \alpha + \beta_1 Tutor_{ij} + e_{ij} \qquad\qquad (A3)$$

where Y_{ij} represents a baseline characteristic for student i in class section j; *Tutor* is an indicator for whether the student was assigned to the Tutor + Teacher model, and e is residual error, clustered at the class section level and adjusted for small sample sizes.

The results show there are no significant differences between groups based on treatment assignment (Tables A3 and A4). Although not statistically significant, some differences are worth noting. These include baseline assessment scores, which are 0.04 standard deviations higher in

the Tutor + Teacher model within the analytic sample, according to the frequentist approach, and 0.07 standard deviations higher, according to the Bayesian approach. Students assigned to the Tutor + Teacher model also had lower percentages of economically disadvantaged students (7.7 percentage points in both analytic approaches) and high school students (6.1 percentage points with the frequentist approach but only 2.5 percentage points with the Bayesian approach). At a minimum, these factors should be included as covariates in analytic models. None of the characteristics has an effect size larger than 0.25, but most fall within the range where statistical adjustments are required to satisfy baseline equivalence (an effect size between 0.05 and 0.25 in absolute value) (U.S. Department of Education, 2022, p.53).

Table A3. Tests for baseline equivalency (frequentist approach)

Characteristic	Initial sample N = 158			Analytic sample N = 146		
	Coef. (SE)	p-value	Hedges' g	Coef. (SE)	p-value	Hedges' g
Baseline WIDA ACCESS Reading scores (standardized)	0.032 (0.1)	0.756	0.059	0.044 (0.108)	0.689	0.082
Female	-0.035 (0.068)	0.612	-0.07	-0.029 (0.066)	0.674	-0.057
Asian	-0.03 (0.053)	0.59	-0.104	-0.025 (0.051)	0.636	-0.091
Latino/a	0.021 (0.063)	0.747	0.056	0.023 (0.062)	0.714	0.064
Spanish home language	0.048 (0.066)	0.482	0.12	0.036 (0.069)	0.614	0.088
Economic disadvantage	-0.098 (0.093)	0.31	-0.208	-0.077 (0.095)	0.432	-0.165
Special education	-0.008 (0.069)	0.906	-0.022	-0.023 (0.069)	0.746	-0.061
Newcomer (< 5 years as a multilingual learner)	0.015 (0.12)	0.9	0.033	-0.009 (0.12)	0.944	-0.018
High school vs. middle school	-0.078 (0.127)	0.548	-0.171	-0.061 (0.131)	0.648	-0.137
Grade level	-0.307 (0.539)	0.578	-0.18	-0.191 (0.545)	0.731	-0.114

Note: Results for Black and white students are suppressed due to small N.
Source: Authors' analysis of Richland School District Two data, 2023–24.

Table A4. Tests for baseline equivalency (Bayesian approach)

Characteristic	Initial sample N = 158 Mean (SD)	94% HDI	Analytic sample N = 146 Mean (SD)	94% HDI
Baseline WIDA ACCESS Reading scores (standardized)	0.062 (0.112)	[-0.154, 0.271]	0.071 (0.116)	[-0.134, 0.3]
Female	-0.04 (0.086)	[-0.182, 0.129]	-0.026 (0.088)	[-0.195, 0.135]
Asian	-0.002 (0.07)	[-0.132, 0.131]	-0.016 (0.061)	[-0.129, 0.099]
Latino/a	0.018 (0.07)	[-0.122, 0.137]	0.019 (0.067)	[-0.108, 0.142]
Spanish home language	0.038 (0.077)	[-0.112, 0.177]	0.024 (0.076)	[-0.131, 0.158]
Economic disadvantage	-0.102 (0.09)	[-0.276, 0.062]	-0.076 (0.09)	[-0.243, 0.095]
Special education	-0.013 (0.072)	[-0.145, 0.12]	-0.026 (0.069)	[-0.153, 0.108]
Newcomer (< 5 years as a multilingual learner)	0.079 (0.109)	[-0.134, 0.27]	0.058 (0.113)	[-0.157, 0.263]
High school vs. middle school	-0.008 (0.108)	[-0.207, 0.199]	0.025 (0.112)	[-0.183, 0.236]
Grade level	-0.030 (0.075)	[-0.153, 0.118]	-0.014 (0.072)	[-0.144, 0.122]

Note: Results for Black and white students are suppressed due to small N.
Source: Authors' analysis of Richland School District Two data, 2023–24.

Probability of assignment to the Tutor + Teacher group given baseline characteristics

The reverse relationship was also tested for consistency with the above results. The purpose is to show that having a particular characteristic is not a significant predictor of selection into the Tutor + Teacher group. The probability of being assigned to the Tutor + Teacher model was estimated with a cluster-robust linear probability model adjusted for small sample sizes:

$$Y_{ij} = \beta_1 x_{ij} + e_{ij} \qquad\qquad \text{(A4)}$$

where Y_{ij} represents assignment to the Tutor + Teacher model for student i in class section j, predicted by a baseline characteristic, x. e represents class section cluster robust residual error adjusted for small sample sizes.

None of the observable characteristics had a statistically significant relationship with assignment to the Tutor + Teacher model (Table A5). Effect sizes were less than 0.25 for all baseline characteristics included in the model.

Table A5. Probability of being assigned to the Tutor + Teacher intervention model by baseline characteristic (frequentist approach)

Characteristic	Initial sample N = 158			Analytic sample N = 146		
	Coef. (SE)	*p*-value	Hedges' g	Coef. (SE)	*p*-value	Hedges' g
Baseline WIDA ACCESS Reading scores (standardized)	0.028 (0.087)	0.087	0.055	0.039 (0.093)	0.685	0.077
Female	-0.036 (0.069)	0.069	-0.071	-0.029 (0.066)	0.675	-0.057
Asian	-0.092 (0.171)	0.171	-0.183	-0.083 (0.179)	0.657	-0.165
Latino/a	0.039 (0.118)	0.118	0.077	0.045 (0.123)	0.717	0.090
Race other than Asian or Latino/a	-0.039 (0.118)	0.118	-0.077	-0.045 (0.123)	0.717	-0.090
Spanish home language	0.076 (0.106)	0.106	0.151	0.056 (0.109)	0.619	0.111
Economic disadvantage	-0.112 (0.108)	0.108	-0.222	-0.09 (0.113)	0.439	-0.179
Special education	-0.015 (0.127)	0.127	-0.030	-0.042 (0.127)	0.747	-0.082
Newcomer	0.018 (0.142)	0.142	0.036	-0.01 (0.14)	0.944	-0.020
High school vs. middle school	-0.095 (0.157)	0.157	-0.189	-0.078 (0.168)	0.651	-0.155
Grade level	-0.027 (0.047)	0.047	-0.053	-0.017 (0.05)	0.733	-0.034

Note: Results for Black and white students are suppressed due to small N.
Source: Authors' analysis of Richland School District Two data, 2023–24.

Joint probability of assignment to the Tutor + Teacher intervention model

A joint F-test was conducted to predict assignment to the treatment group as a function of observable characteristics with standard errors clustered by class section and strata. Observable characteristics included baseline WIDA ACCESS Reading scores and indicators for female, Asian, Latino/a, Black, white, Spanish home language, economic disadvantage, special education, newcomer status, and enrollment in high school vs. middle school. For both the initial sample and the analytic sample, we fail to reject the null hypothesis that there is no evidence of baseline differences among observable characteristics collectively (Table A6).

Table A6. Joint F-test predicting assignment to the Tutor + Teacher intervention model

Sample	F-Statistic	*p*-value
Initial sample (N = 158)	0.127	0.722
Analytic sample (N = 146)	0.096	0.758

Source: Authors' analysis of Richland School District Two data, 2023–24.

Attrition

This section reports findings about the characteristics of the 12 students missing outcome scores. It describes the extent to which attrition is related to baseline characteristics and summarizes how treatment assignment is related to attrition. Finally, it presents an analysis of estimated attrition bias.

Sample attrition overview

There was no cluster-level attrition. All class sections that began the intervention with students had students with outcome measures. Twelve students were missing outcome scores— nine from the group assigned to the Tutor + Teacher model and three from the group assigned to the Teacher-only model (Table A7). The overall attrition rate was 7.6% and the differential attrition rate was 7.2%.

Table A7. Sample attrition

Metric	Tutor + Teacher	Teacher-only	Overall	Differential
Initial sample N	81	77	158	4
Left sample	9	3	12	6
% Attrition	11%	4%	7.6%	7.2%

Source: Authors' analysis of Richland School District Two data, 2023–24 through 2024–25.

Characteristics of students missing outcome scores

The initial and analytic samples had consistent student compositions across observable characteristics (Table A8). There were no noteworthy differences overall and within each treatment group. When comparing means between the sample of students with missing outcome scores and those included in the study, it is important to note that sample sizes are small (N = 12

students missing outcome scores). Treatment group-level results are not shown due to the small sample sizes of students missing outcome scores.

Potentially notable differences include the proportion of students who were in high school in the initial, analytic, and leaver samples. Six of the twelve students who left the sample were enrolled in high school in the baseline year, compared to about a quarter of students in the initial sample. However, differences in average grade level were not remarkable. Additionally, there were higher shares of newcomers among the students missing outcome scores relative to the initial and analytic samples. Yet, the 15-percentage point difference represents two students.

Table A8. Characteristics of students in the initial sample, students in the analytic sample, and students missing outcome scores

	Initial sample N = 158	Analytic sample Mean (SD) N = 146	Missing outcome Mean (SD) N = 12
Baseline WIDA ACCESS Reading scores (standardized)	-0.906 (0.536)	-0.906 (0.536)	-0.908 (0.556)
Female	0.449 (0.499)	0.459 (0.5)	0.333 (0.492)
Asian	0.089 (0.285)	0.082 (0.276)	0.167 (0.389)
Latino/a	0.842 (0.366)	0.849 (0.359)	0.75 (0.452)
Race other than Asian or Latino/a	0.158 (0.366)	0.151 (0.359)	0.25 (0.452)
Spanish home language	0.804 (0.398)	0.801 (0.4)	0.833 (0.389)
Economic disadvantage	0.677 (0.469)	0.692 (0.463)	0.5 (0.522)
Special education	0.165 (0.372)	0.164 (0.372)	0.167 (0.389)
Newcomer	0.696 (0.461)	0.685 (0.466)	0.833 (0.389)
High school vs. middle school	0.285 (0.453)	0.267 (0.444)	0.5 (0.522)
Grade level	7.557 (1.702)	7.527 (1.666)	7.917 (2.151)

Note: Results for Black and white students are suppressed due to small N.
Source: Authors' analysis of Richland School District Two data, 2023–24.

Attrition timing

Nearly half of the students with missing outcome scores left the district within the first five weeks of the intervention (Table A9). Five of the 12 had exited the district by October 14, 2024. Two of the four students assigned to the Tutor + Teacher model who left in the first five weeks never attended a tutoring session. Eleven of the 12 students had left the district by February 17, 2025.

Table A9. Week of attrition after the intervention began

Week exited the district relative to the intervention start (9/10/2024)	Number of students who left the sample		
	Tutor + Teacher (N = 9)		Teacher only (N = 3)
	Number of students missing outcome scores	Minutes of tutoring received	Number of students missing outcome scores
1	0	N/A	0
2	1	0	1
3	0	N/A	0
4	2	260, 303	0
5	1	0	0
6+	5	105, 406, 416, 427, 1,407	2

Source: Authors' analysis of Richland School District Two data, 2024–25.

Predicting attrition by observable baseline characteristics

Linear probability models estimated the probability of missing an outcome score by each observable baseline characteristic (Equation A5):

$$y_{ij} = \beta_1 x_{ij} + e_{ij} \qquad (A5)$$

where *y* represents the probability of a missing outcome score (i.e., attrition from the sample) for student *i* in class section *j,* predicted by a baseline characteristic, *x. e* represents class section cluster robust residual error adjusted for small sample sizes.

The analysis was conducted within three samples: students assigned to the Tutor + Teacher model, students assigned to the Teacher-only model, and the full sample. There were no statistically significant predictors of attrition among characteristics within any of the three samples. Some coefficients seem notable, such as those for high school students in the Teacher-only group (0.12). Larger shares of high school students missing outcome scores compared to middle school students aligns with this population's elevated risk of high school dropout (Suárez-Orozco et al., 2010).

Table A10. Predicted probability of missing an outcome score by baseline characteristic

Characteristic	Full sample N = 158		Tutor + Teacher N = 81		Teacher only N = 77	
	Coef. (SE)	*p*-value	Coef. (SE)	*p*-value	Coef. (SE)	*p*-value
Baseline WIDA ACCESS Reading scores (standardized)	-0.001 (0.042)	0.990	-0.021 (0.07)	0.784	0.020 (0.039)	0.668
Female	-0.036 (0.034)	0.342	-0.045 (0.063)	0.51	-0.021 (0.023)	0.502
Asian	0.073 (0.092)	0.496	a	a	a	a
Latino/a	-0.052 (0.064)	0.455	-0.065 (0.108)	0.591	-0.046 (0.077)	0.64
Race other than Asian or Latino/a	0.052 (0.064)	0.455	0.065 (0.108)	0.591	0.046 (0.077)	0.64
Spanish home language	0.014 (0.052)	0.798	0.048 (0.085)	0.627	-0.025 (0.065)	0.741
Economic disadvantage	-0.062 (0.049)	0.242	-0.088 (0.067)	0.226	-0.012 (0.059)	0.859
Special education	0.001 (0.057)	0.985	0.051 (0.098)	0.652	-0.047 (0.036)	0.322
Newcomer	0.049 (0.035)	0.208	0.099 (0.053)	0.158	-0.004 (0.03)	0.908
High school vs. middle school	0.080 (0.058)	0.230	0.052 (0.095)	0.648	0.120 (0.089)	0.308
Grade level	0.009 (0.015)	0.563	-0.008 (0.024)	0.747	0.027 (0.019)	0.258

[a] Suppressed due to small sample size.
Note: Results for Black and white students are suppressed due to small sample sizes.
Source: Authors' analysis of Richland School District Two data, 2023–24 through 2024–25.

Joint F-tests predicting attrition based on all the characteristics in Table A11 (minus "Race other than Asian or Latino/a") fail to reject the null hypotheses that there is no evidence that the baseline characteristics collectively have a significant relationship with attrition.

Table A11. Joint F-tests predicting attrition based on baseline characteristics

Sample	F-statistic (*p*-value)
Full sample	0.036 (0.851)
Assigned to Tutor + Teacher model	0.087 (0.771)
Assigned to Teacher-only model	0.310 (0.585)

Source: Authors' analysis of Richland School District Two data, 2023–24 through 2024–25.

Attrition bias

Estimating attrition bias begins with estimating an individual's propensity to respond (i.e., their predicted probability of not missing an outcome score given a set of baseline

characteristics). Next, the correlation between those predicted probabilities and the outcome is measured. Finally, bias is calculated using the correlation and predicted probability, as defined in U.S. Department of Education (n.d.a).

The results are sensitive to modeling choices when estimating the probability of being in the sample (Table A12). The probability of being in the analytic sample is close to one for both groups. Results vary when estimating probabilities using a linear probability model, a logistic model, and a Bayesian model with a logit link function for a Bernoulli distribution. Each approach has limitations, including the non-exhaustive list that follows.

- The linear probability model has the potential to exceed 1.0.

- Logistic models are susceptible to overparameterization, particularly with small sample sizes, and may not perform well when there are few events within a sample, such as three out of 77 in the Teacher-only group.

- Bayesian models handle high-dimensional data and small datasets well, but the mean value does not capture the uncertainty that Bayesian models quantify. The "true" probability lies within a range. Selecting a value from that range other than the mean could alter downstream conclusions.

Table A12. Attrition bias

Metric	Overall	Tutor + teacher	Teacher-only	Difference between treatment groups
Linear probability model				
Predicted probability of not missing the outcome score	0.924	0.920	0.928	0.008
Correlation between outcome and predicted probability of not missing the outcome score	-0.013	-0.011	-0.005	0.006
Bias (correlation between outcome and predicted probability * ((1 – predicted probability) / predicted probability)	-0.001	-0.001	-0.0004	0.0005
Logistic model				
Predicted probability of not missing the outcome score	0.924	0.919	0.929	-0.01
Correlation between outcome and predicted probability of not missing the outcome score	0.047	0.086	0.005	0.081
Bias (correlation between outcome and predicted probability * ((1 – predicted probability) / predicted probability)	0.004	0.008	0.0004	0.007
Bayesian model				
Predicted probability of not missing outcome score	0.924	0.921	0.928	-0.008
Correlation between outcome and predicted probability of not missing outcome score	0.038	0.047	0.024	0.023
Bias (correlation between outcome and predicted probability * ((1 – predicted probability) / predicted probability)	0.003	0.004	0.002	0.002

Source: Authors' analysis of Richland School District Two data, 2023–24 through 2024–25.

According to the results in Table A12, the sensitivity of the expected bias is likely to fall below 0.06 (Table A13). The linear probability model and Bayesian model suggest that the sensitivity of expected bias is less than 0.04.

Table A13. Sensitivity of expected bias at the optimistic attrition boundary with respect to attrition model parameter values

Difference in correlation between treatment and control groups	Overall correlation between outcomes and attrition				
	0.1	**0.2**	**0.3**	**0.4**	**0.5**
0.05	0.04	0.05	0.05	0.06	0.07
0.10	0.06	0.07	0.08	0.09	0.10
0.15	0.09	0.1	0.11	0.12	0.13
0.20	0.12	0.13	0.13	0.14	0.15
0.25	0.14	0.15	0.16	0.17	0.18

Source: U.S. Department of Education, n.d.b.

Compliance with random assignment

Overall compliance with assignment to the Teacher + Tutor or Teacher-only model was high (Table A14). Most students (88%) assigned to receive tutoring attended at least one tutoring session. Only one student assigned to the Teacher-only model received tutoring. This student began receiving tutoring on January 16, 2025, and received 183 minutes of tutoring over eight sessions before taking the WIDA ACCESS assessment.

Table A14. Compliance with random assignment

Assignment	Offered tutoring (N)	Received tutoring (N)	Not offered tutoring (N)	Did not receive tutoring (N)	Compliance rate (scheduled)	Compliance rate (received)
All assigned students						
Tutor + Teacher (N = 81)	75	71	6	10	92.6%	87.7%
Teacher-only (N = 77)	1	1	76	76	98.7%	98.7%
All students (N = 158)	76	72	82	86	95.6%	93.0%
Analytic sample						
Tutor + Teacher (N = 72)	67	64	5	8	93.1%	88.9%
Teacher-only (N = 74)	1	1	73	73	98.6%	98.6%
All students (N = 146)	68	65	78	81	95.9%	93.8%

Source: Authors' analysis of Richland School District Two data, 2024–25.

Implementation study data and methods

The implementation study used data from four sources: an online survey, online focus groups and interviews, training completion data, and implementation data supplied by tutors via the Littera online platform.

In March 2025, surveys were administered to eight teachers and 50 tutors who delivered Off2Class Foundational Literacy lessons between September 2024 and February 2025. Six teachers (75%) and 22 tutors (44%) responded to the survey. Participation was voluntary, and

respondents were allowed to skip questions. The minimum number of respondents for any question was five teachers or 20 tutors. Sample sizes are noted throughout the findings.

The teacher survey included 39 items (31 closed-ended and eight open-ended questions). The tutor survey included 36 items (28 closed-ended and eight open-ended questions). The surveys included questions about the following topics:

- Perceptions about the usefulness and usability of the Off2Class Foundational Literacy program

- Perceptions about the Off2Class Foundational Literacy program training and instructional delivery

- Perceptions about the effectiveness of the Off2Class Foundational Literacy program

- For analysis, the evaluation team calculated frequencies for each survey item for teachers versus tutors and read open-ended response items to identify themes.

The evaluation team held focus groups or interviews between March and April 2025: two tutor focus groups, one teacher focus group, two interviews with individual teachers (as they were the only teachers to attend the scheduled focus group session), one focus group with Off2Class project staff, one focus group with Littera project staff, and one interview with a district administrator. Focus group/interview participants were free to skip questions throughout the discussion.

The focus group/interview protocol was the same for tutors and teachers, but with questions targeting the Off2Class Foundational Literacy Curriculum delivery via Littera's online platform (for tutors) and the curriculum delivery in class (for teachers). The protocol included 13 open-ended items questions covering training, typical use of the curriculum, student learning and engagement, usefulness, usability, and challenges encountered. Interviewers used a semi-

structured interviewing approach, adding follow-up and clarifying questions as needed to fully probe the topics and delve into new ones as they arose. Interviews/focus groups lasted approximately 60 minutes. Six tutors participated in focus groups and five teachers participated in a focus group or interview. The evaluation team used NVivo qualitative analysis software to code all focus group data using emergent codes and then developed themes to identify patterns in the data. A single coder completed the focus group analysis to ensure rater reliability.

Training data differed for tutors versus teachers. For tutors, Littera provided a list of tutors who completed Off2Class training in the Littera platform, which the evaluation team filtered to identify study participant tutors. For teachers, Off2Class provided an Excel workbook of training data, including lists of teachers who completed training and training emails sent to teachers. The evaluation team filtered the lists to identify study participant teachers and then identified teachers who had and had not completed the training or received the training email messages.

Finally, the Littera online platform provided data on the following:

- Session time

- Session length

- Student attendance

- Tutor absence (meaning no session was held)

- Tutor notes about the student during the session

- Lesson name

- Skills included in the lesson

- Number of skills actually taught during the lesson

- Tutors' ratings of student skill acquisition and engagement during the session

Tutors completed these data fields for each student in a session. The evaluation team reviewed and scored a random sample of 1,457 records, reflecting 50% of each tutor's records. The evaluation team then calculated an average implementation score for each tutor based on the following:

- Student attendance

- Student engagement score

- Student skill acquisition

- The "barriers to instruction" document in the tutor note field (e.g., unresolved technical issues, student behavior issues, students leaving sessions early or arriving late)

- The number of skills taught

- Implementation scores ranged from 0 (e.g., no students attended the session) to 5 (e.g., at least one student attended, there were no barriers to instruction, the student acquired at least one skill, etc.).

Appendix B. Additional findings

This appendix presents descriptive findings from a subsample of students who remained within the same grade band (middle school or high school) during the baseline and outcome years. This allows for comparisons based on scale scores, which helps put the study's findings in context.

The WIDA ACCESS assessment is scaled within grade band. Table B1 shows scaled score growth for middle school and high school students in each treatment group who remained in the same grade band in the baseline and outcome years. Middle school students were those in grades 6–8 in both the baseline and outcome years (N = 57), and high school students were those in grades 9–12 in both baseline and outcome years (N = 39). Note that students in grades 6 and 9 were in the sample only if they repeated grade 6 (N = 5) or 9 (N = 21), respectively, in 2024–25.

Within the limited sample, on average, middle school students gained 20.8 scale score points, and high school students gained 17.1 scale score points. Among middle schoolers, those assigned to the Tutor + Teacher model gained 23.9 scale score points compared with 14.6 for students assigned to the Teacher-only model, on average. Similarly, high school students assigned to the Tutor + Teacher model gained 25.2 scale score points compared with 10.9 for students assigned to the Teacher-only model.

Table B1. WIDA ACCESS Reading domain scaled score growth for students in the same grade band in the baseline and outcome years

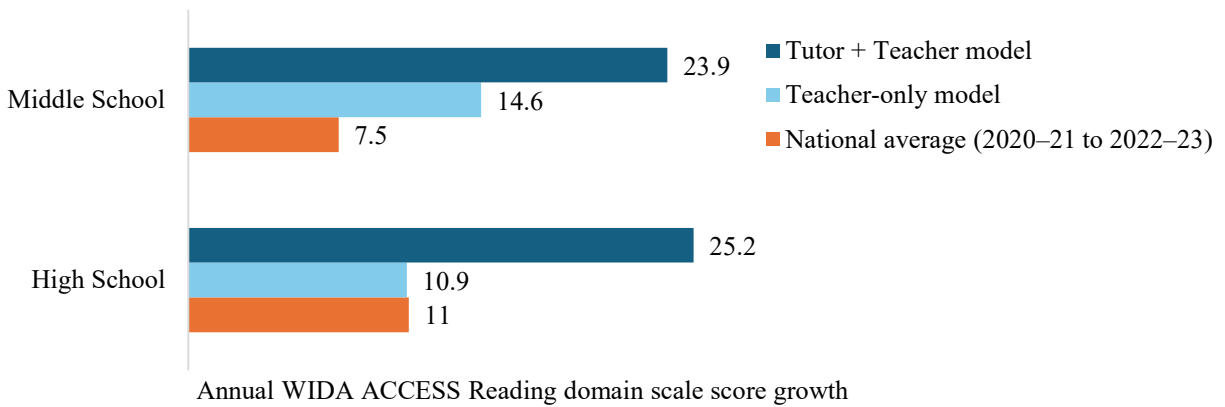| Grade band	Sample	N	Average number of years as an English learner	Score type	Mean 2023–24 WIDA ACCESS Reading score	Mean 2024–25 WIDA ACCESS Reading score	Mean difference between 2023–24 and 2024–25
Middle school	Full sample	57	5.2	Scaled	320.7	341.5	20.8
				Standardized	-0.784	-0.173	-0.611
	Tutor + Teacher	34	5.8	Scaled	321.1	345.0	23.9
				Standardized	-0.772	-0.070	-0.703
	Teacher-only	23	4.1	Scaled	320.0	334.6	14.6
				Standardized	-0.805	-0.375	-0.429
High school	Full sample	39	2.6	Scaled	340.6	357.7	17.1
				Standardized	-1.022	-0.552	-0.470
	Tutor + Teacher	17	2.5	Scaled	346.8	371.9	25.2
				Standardized	-0.851	-0.161	-0.690
	Teacher-only	22	2.7	Scaled	335.8	346.8	10.9
				Standardized	-1.154	-0.851	-0.302

Note: Means are weighted by class section and strata random blocks. Standardized scores are calculated based on national means and standard deviations for the 2023–24 WIDA ACCESS administration (Center for Applied Linguistics, 2025).
Source: Authors' analysis of Richland School District Two data, 2023–24 through 2024–25.
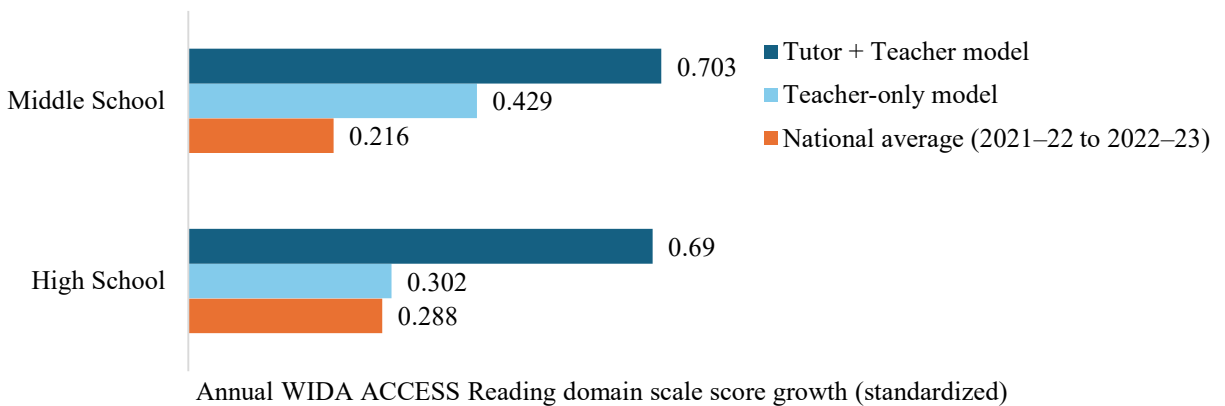
The findings for this limited sample are similar to those for the full sample (see Figure B1). High school students assigned to the Teacher-only model made gains in line with the national average, while their peers assigned to the Tutor + Teacher model achieved gains that were twice the national average. Middle school students assigned to the Teacher-only model also grew about twice as much as the national average. Their peers assigned to the Tutor + Teacher model gained more than three times the national average. These findings are consistent when scale scores are standardized according to national distributions within grade band and test administration year.

Figure B1. WIDA ACCESS Reading score growth compared to the national average

Scale score growth

Annual WIDA ACCESS Reading domain scale score growth

Middle School
- Tutor + Teacher model: 23.9
- Teacher-only model: 14.6
- National average (2020–21 to 2022–23): 7.5

High School
- Tutor + Teacher model: 25.2
- Teacher-only model: 10.9
- National average (2020–21 to 2022–23): 11

Standardized scale score growth

Annual WIDA ACCESS Reading domain scale score growth (standardized)

Middle School
- Tutor + Teacher model: 0.703
- Teacher-only model: 0.429
- National average (2021–22 to 2022–23): 0.216

High School
- Tutor + Teacher model: 0.69
- Teacher-only model: 0.302
- National average (2021–22 to 2022–23): 0.288

Note: Sample is limited to students who were in grades 6–8 for both 2023–24 and 2024–25 or in grades 9–12 for both years. Means are weighted by class section and strata random blocks. Standardized scores are calculated based on national means and standard deviations for the 2020–21 through 2023–24 WIDA ACCESS administrations (Center for Applied Linguistics, 2023; 2024; 2025).
Sources: Poole & Sahakan, 2024; authors' analysis of Richland School District Two data, 2023–24 through 2024–25.