



# Extended Learning Opportunities ESSER Tutoring Program Evaluation 2023–24

## Tutoring Evaluation Overview

The research and evaluation (R&E) division of the Department of Research, Assessment, and Data was asked to complete an evaluation of the in-school tutoring services provided by vendors during the 2023–24 school year. The vendors were selected through the request for proposal process by the extended learning opportunities office and paid for using ESSER funding to address learning loss accrued during the COVID-19 pandemic and the disruption that the pandemic caused to regular school operations. The ELO office also used ESSER funding to pay for after-school tutoring that was provided by school staff; however, the focus of this evaluation is on the tutoring services provided by external vendors with which the department had contracted for the 2023–24 school year.

A total of six vendors were selected to provide in-school tutoring to students. The vendors varied in their approach to tutoring, with some providing only one-on-one sessions while others provided a mix of one-on-one and group sessions as well as in the content areas covered during tutoring sessions. Details about the vendor’s method of delivery, tutoring session formats, and content areas is found in table 1 below. A list of the school locations for each vendor is found in table 2.

<b>Vendor</b>	<b>Session format(s)</b>	<b>Content area(s)</b>	<b>Method of delivery</b>
<b>Vendor One</b>	One-on-one	Reading	In person
<b>Vendor Two</b>	One-on-one, peer group, small group	Reading/ELA and math	In person
<b>Vendor Three</b>	One-on-one, peer group, small group	Reading/ELA and math	In person
<b>Vendor Four</b>	One-on-one	Reading/ELA and math	Virtual platform
<b>Vendor Five</b>	One-on-one, peer group, small group	Reading/ELA and math	In person
<b>Vendor Six</b>	One-on-one, peer group, small group	All core content areas	In person

Table 1



Vendor	School Locations
<b>Vendor One</b>	Barbee Montessori, Bryant, Forest Home, Greenfield, Keefe Avenue, Stuart, Vieau
<b>Vendor Two</b>	Browning, Carver, Douglas Middle School, Golda Meir, Goodrich, Holmes, Kilbourn, Metcalfe, Project STAY
<b>Vendor Three</b>	Alcott, Auer Avenue, Barton, Clement Avenue, Craig Montessori, Fernwood Montessori, Fifty-Third Street, Hartford University, Humboldt Park, IDEAL, James Madison Academic Campus, Kluge, Lancaster, MacDowell Montessori, Metcalfe, Milwaukee Academy of Chinese Language, Obama SCTE, Starms Discovery Center, Starms Early Childhood Center, Zablocki
<b>Vendor Four</b>	<i>Not applicable as it is a virtual tutoring service</i>
<b>Vendor Five</b>	Audubon High School, Congress, Franklin, Garland, Marshall High School, Obama SCTE
<b>Vendor Six</b>	Bay View High School, Bradley Tech, Holmes, Hopkins, Jackson, Milwaukee High School of the Arts, Pulaski High School, Reagan High School, Roosevelt Middle School, Spanish Immersion, Story

Table 2

Students were identified for tutoring if they met certain criteria, with the priority of identifying students performing below grade level but not currently receiving Tier 2 or Tier 3 interventions. Vendors were required to complete tutoring session forms after each session that they completed with a student or group of students. The form asked tutors to provide information about the content area covered, the number of minutes spent with the student(s), the time of day and date of the session, any resources or curricula used, and notes about how the session went. These forms were collected via the district’s survey platform and Qualtrics, and a dashboard was created to assist vendors in tracking and ensuring that their data were accurate and complete.

The goal of the evaluation was to provide a better understanding of whether in-school tutoring was an effective use of ESSER dollars, and, if so, which vendors demonstrated a greater impact on student outcomes. To measure this, the R&E team developed six driving questions for their analysis:

1. Do students who receive tutoring services have significantly different Star Math percentile ranking from fall to spring?
2. Do students who receive tutoring services have significantly different Star Reading percentile ranking from fall to spring?
3. Do students who receive tutoring services have significantly different Star Math percentile ranking from fall to spring among the tutoring vendors?
4. Do students who receive tutoring services have significantly different Star Reading percentile ranking from fall to spring among the tutoring vendors?
5. Do students who receive tutoring services have significantly different attendance if they had more than 10 tutoring sessions?
6. Do students who receive tutoring services have significantly different pass/fail course success rates if they had more than 10 tutoring sessions?



## Methodology

To evaluate the effectiveness of in-school tutoring as a whole and by vendor, the R&E team consulted with the ELO office to determine the student outcome metrics. The teams decided to use Star Reading percentile rank (PR) change and Star Math PR change for students in grades 1–8, and average daily attendance (ADA) and ELA and math course passing rate for students in grades 9–12. Star was selected since it is used for district accountability and is given three times per year, which allows for within-year comparison, whereas the Wisconsin Forward Exam or ACTs are given only once per year and are subject to delays in test result availability for district personnel. High schools do not uniformly administer the Star for their student assessment, so coursework performance and ADA were used instead to ensure that there were enough data points for a robust analysis.

During the 2023–24 school year, 4,092 students had at least one tutoring session recorded with at least one vendor. However, a set of inclusion criteria was established to isolate the effects of tutoring as much as possible. Below is a list of the inclusion criteria:

- Students had to have either Star assessment results or course performance to be included in the analysis. Therefore, K4 and K5 students were not included in the evaluation since they neither take the Star test or another similar standardized assessment that could be used for measuring outcomes nor receive marks in their classes.
- For grades 1–8, students needed to have a fall and spring Star Reading test to be included in the Star Reading PR change analysis and a fall and spring Star Math test to be included in the Star Math PR change analysis. Additionally, only students who tested during the testing window were included in the PR growth data, and their first score is what was used in the PR change calculation. This approach of limiting the test data to the testing windows and the first score aligns with how the district reports and uses the Star test for accountability purposes. Other tests can appear in student data but are used for progress monitoring individual students rather than for district outcome and accountability measures.
- Any student who received tutoring services from multiple vendors was excluded from the analysis since we could not attribute any effects on the student's outcomes to exposure to a specific vendor. Sixty-six students were identified as receiving services from more than one vendor and thus were removed from the analysis.
- Sessions that were recorded in Qualtrics but had comments indicating that the student was absent from the session due to a range of factors, such as Wisconsin Forward or Star testing, a class field trip, or a schoolwide assembly, were also removed from the analysis.

For each student who received tutoring services, R&E team members calculated the number of sessions that they received from the vendor, regardless of the content area covered in the session. The team used this calculation in conjunction with the dosage information provided by each vendor to develop a treatment and control group for comparing student outcomes. Across vendors, the dosage requirement was 35 sessions or more before one could expect to see the effects of their tutoring program.

Consequently, the R&E team established the treatment group as all students with more than 10 tutoring



sessions with a vendor and the control group as all students with 10 or fewer sessions. This method was preferable to comparing students who received tutoring with students who did not receive tutoring (e.g., the rest of the school, other students in their grade, the district as a whole) since those who received 10 or fewer sessions met the program’s eligibility criteria and were selected for tutoring services, making it likely that they would have similar academic performance backgrounds to the treatment group.

The statistical test that the R&E team used was a logistic regression\* for each dependent variable (i.e., Star Reading PR change score, Star Math PR change score, ADA, and ELA/math course pass rate). To reduce any potential confounding effects from differences in the demographics of students served among vendors, the following demographic variables were controlled for: race/ethnicity, special education status, English language learner status, and gender. Additionally, some students (approximately 317 in the data set) received after-school tutoring services, which was also controlled for to isolate any effects to in-school tutoring services.

Benchmark scores were established to identify growth and performance above and beyond district averages or expected growth on each outcome. For the ADA, if the student had an ADA that was higher than the district’s average for all high school students (74.3%), that counted as an improvement (i.e., “1”), and if the ADA was the same or lower, that counted as no improvement (i.e., “0”). For the ELA and math course passing rate, the R&E team took the average of all passing (i.e., “1”) and failing (i.e., “0”) grades across all ELA and math courses that students took during the 2023–24 school year. If the average passing rate for the student was above the district’s average passing rate for all math and ELA courses (82%), that counted as an improvement; if the student had the same or a lower average passing rate, that counted as no improvement. See table 4 for details on each metric’s benchmarks.

For the Star outcomes, students who had a PR increase from fall to spring of 3 points or lower were coded as no effect/change (i.e., “0”) and 4 points or higher as a positive change/effect (i.e., “1”). The PR is normed to grade-level performance of students nationally, so the average change in PR is not readily available across grade levels. Additionally, lower grades tend to have greater growth in PR than higher grade levels. For example, in 2023–24, there was a high of 15.1 for first graders and low of 0.0 for eighth graders, with the averages beginning to drop precipitously around third grade for both math and reading. Given that more students in the tutoring sample were in grades 3–8 than in grades 1–2 and that all vendors worked with students across grade levels, the R&E team determined that using a number in between the district’s highest and lowest average change would be the most appropriate. See table 3 for details on the district’s average PR change by grade level for 2023–24.

*\*The Hosmer–Lemeshow test and model summary for the logistic regression showed that the model was not the best fit for the type of data; however, this is likely due to some of the variables entered as covariates (i.e., control variables) measuring similar things, leading to redundancies. This does not alter the key findings of the results.*



<b>Star Reading</b>				
DPI On Target PR	Grade	Average Fall PR	Average Spring PR	Average PR Change Fall to Spring
60	2nd	24.4	32.9	8.5
60	3rd	22.2	27.7	5.5
60	4th	23.9	27.4	3.5
60	5th	23.1	25.5	2.4
60	6th	22.6	22.5	-0.1
60	7th	23.0	23.9	0.9
60	8th	22.6	22.6	0.0
60	9th	16.5	16.6	0.1
60	10th	18.2	18.4	0.2
<b>Star Early Literacy</b>				
DPI On Target PR	Grade	Average Fall PR	Average Spring PR	Average PR Change Fall to Spring
60	1st	22.4	37.5	15.1
<b>Star Math</b>				
DPI On Target PR	Grade	Average Fall PR	Average Spring PR	Average PR Change Fall to Spring
60	1st	37.2	49.9	12.7
60	2nd	29.0	39.9	10.9
60	3rd	30.2	35.5	5.3
60	4th	26.9	33.4	6.5
60	5th	26.2	31.5	5.3
60	6th	23.8	25.5	1.7
60	7th	23.3	25.9	2.6
60	8th	23.5	25.4	1.9
60	9th	20.4	22.0	1.6
60	10th	27.3	26.8	-0.5

Table 3 (data as of 8/16/2024)

<b>Outcome metric</b>	<b>Improvement</b>	<b>No improvement</b>
Star Reading PR change score	>3 PR change = Improvement	≤3 PR change = No improvement
Star Math PR change score	>3 PR change = Improvement	≤3 PR change = No improvement
ELA and math course passing rate score	>82% = Improvement (“1”)	≤82% = No improvement (“0”)
Average daily attendance (ADA) score	>74.3% = Improvement (“1”)	≤74.3% = No improvement (“0”)

Table 4





Using the inclusion criteria, the sample size was reduced from 4,092 students to 3,756 when excluding the K4 and K5 students. In addition, 2,590 students in grades 1–8 had both a fall and spring Star Reading test, and 2,611 students had a fall and spring Star Math test. All high school students had attendance data, and 940 and 935 had course marks in at least one ELA or math course, respectively. The R&E team first ran the logistic regression test for all students who remained in the sample, regardless of vendor utilized, to determine whether there was (1) a significant effect observed for students who received more than 10 tutoring sessions regardless of vendor and (2) whether the vendor provider had a significant effect on the student outcome variables. If no significant effects were observed for either of these variables, the analysis would have ended; however, both the number of sessions and the vendor that was utilized demonstrated significant effects on the outcomes (i.e., dependent variables), so the team continued with a vendor-specific analysis for the Star metrics.

## Results

The R&E team ran the regression test in multiple iterations to answer the six driving research questions. The findings for each question are presented below.

*Research Question 1: Do students who receive tutoring services have significantly different Star Math percentile ranking from fall to spring?*

- Nonsignificant differences were observed for students who scored three PR points higher or lower ( $\beta = .059, p = .138$ ). Specifically, students who received more than 10 tutoring sessions, regardless of which vendor they received tutoring from, did not demonstrate a significant difference in the Star Math assessment outcome (i.e., change of 4 or more PR points). However, the test showed that the variable for which vendor used was significantly affecting the test results, which led the team to continue the investigation by analyzing this question for each vendor.

*Research Question 2: Do students who receive tutoring services have significantly different Star Reading percentile ranking from fall to spring?*

- For each one-unit increase in the number of tutoring sessions greater than 10, there was a 71% increase in the odds of demonstrating a 4+ PR change score on the Star Reading assessment. This result indicates that, contrary to the math outcomes, receiving at least 11 sessions (or more) increased the odds of students' PR increasing by 4 or more points, with these odds growing as students received more sessions. Consequently, no matter which vendor was used, receiving regular tutoring sessions in reading and/or ELA positively impacted students' performance on the Star Reading test. As with the Star Math assessment test, the vendor provider variable also had a significant effect on the dependent variable, so analyzing the data for each vendor was necessary to determine whether any vendors outperformed others in this area.



*Research Question 3: Do students who receive tutoring services have significantly different Star Math percentile ranking from fall to spring among the tutoring vendors?*

- Students who received tutoring services from Vendor Two had a 65% increase in the odds of increasing their PR by 4 points or more on the Star Math assessment as compared to the students who received tutoring services from the other vendors.
- Students who received tutoring services from Vendor Three did not demonstrate a significant change in the odds of increasing their PR by 4 points or more on the Star Math assessment as compared to the students who received tutoring services from the other vendors ( $\beta = .01, p = .527$ ).
- Students who received tutoring services from Vendor Four had an 80% increase in the odds of increasing their PR by 4 points or more on the Star Math assessment as compared to the students who received tutoring services from the other vendors.
- Students who received tutoring services from Vendor Five had a 78% decrease in the odds of increasing their PR by 4 points or more on the Star Math assessment as compared to the students who received tutoring services from the other vendors. In other words, students who received services from Vendor Five were less likely to increase their PR in math than students who received services from other vendors.
- Students who received tutoring services from Vendor Six had a 53% decrease in the odds of increasing their PR by 4 points or more on the Star Math assessment as compared to the students who received tutoring services from the other vendors. In other words, students who received tutoring services from Vendor Six were less likely to increase their PR in math than students who received services from other vendors.
- Students who received tutoring services from Vendor One were not included in this analysis since Vendor One tutors solely in reading.

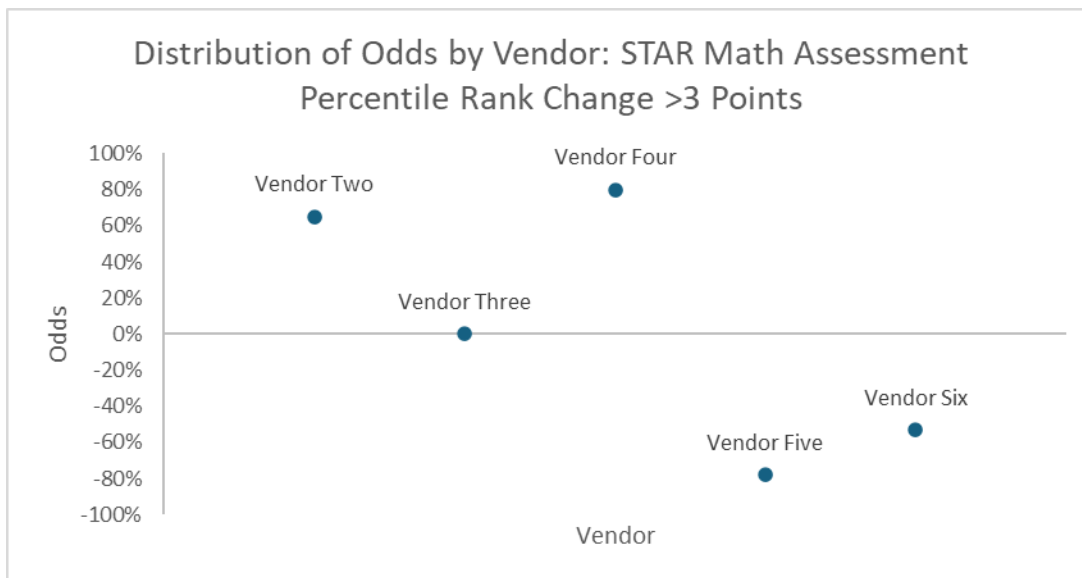


Figure 1



*Research Question 4: Do students who receive tutoring services have significantly different Star Reading percentile ranking from fall to spring among the tutoring vendors?*

- Students who received tutoring services from Vendor One had a 55% increase in the odds of increasing their PR by 4 points or more on the Star Reading assessment as compared to the students who received tutoring services from the other vendors.
- Students who received tutoring services from Vendor Two had an 80% increase in the odds of increasing their PR by 4 points or more on the Star Reading assessment as compared to the students who received tutoring services from the other vendors.
- Students who received tutoring services from Vendor Three had a 92% increase in the odds of increasing their PR by 4 points or more on the Star Reading assessment as compared to the students who received tutoring services from the other vendors.
- Students who received tutoring services from Vendor Four did not have significantly greater odds of increasing their PR by 4 points or more on the Star Reading assessment as compared to the students who received tutoring services from the other vendors ( $\beta = -.043, p = .48$ ).
- Students who received tutoring services from Vendor Five had a 93% decrease in the odds of increasing their PR by 4 points or more on the Star Reading assessment as compared to the students who received tutoring services from the other vendors. In other words, students who received tutoring services from Vendor Five were less likely to increase their PR in reading than students who received services from other vendors.
- Students who received tutoring services from Vendor Six had a 42% decrease in the odds of increasing their PR by 4 points or more on the Star Reading assessment as compared to the students who received tutoring services from the other vendors. In other words, students who received tutoring services from Vendor Six were less likely to increase their PR in reading than students who received services from other vendors.

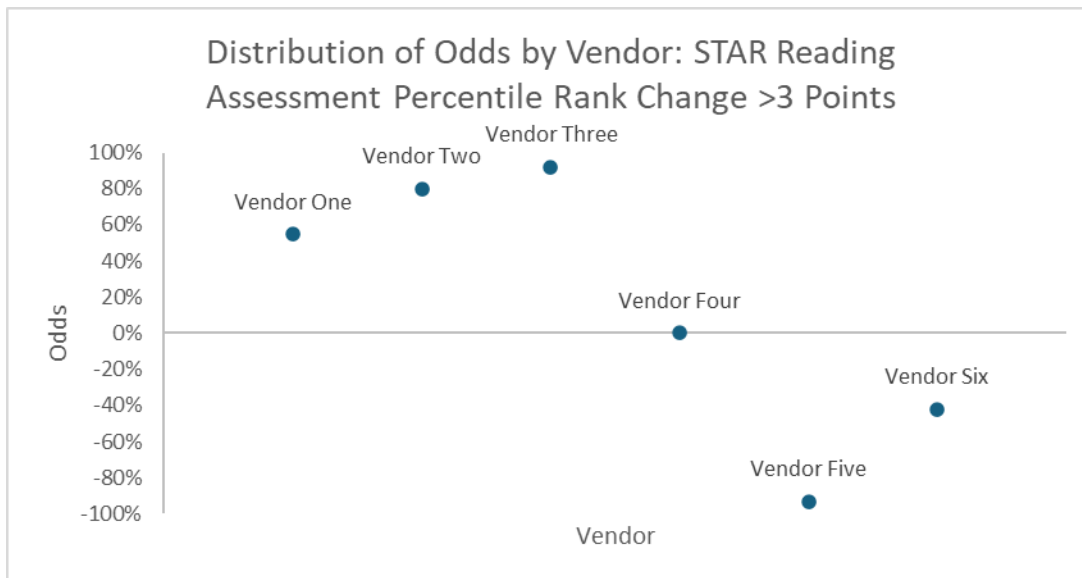


Figure 2





*Research Question 5: Do students who receive tutoring services have significantly different attendance if they had more than 10 tutoring sessions?*

- For each one-unit increase in number of tutoring sessions greater than 10, students have a 17 times increase in the odds of attending school at a rate above the district average for high school students (i.e., 74.3%).

*Research Question 6: Do students who receive tutoring services have significantly different pass/fail course success rates if they had more than 10 tutoring sessions?*

- For each one-unit increase in number of tutoring sessions greater than 10, students have a 4 times increase in the odds of passing more ELA and math courses than their district peers.

Based on the logistic regression tests, the R&E team was able to draw the following broad conclusions:

- Students who received more than 10 sessions of tutoring were significantly more likely to have higher attendance rates compared to those who received fewer than 10 sessions across vendors.
- Students who received more than 10 sessions of tutoring were significantly more likely to pass their math and ELA courses compared to those who received fewer than 10 sessions across vendors.
- The vendor that had significant positive effects on both students' Star Reading and Star Math assessments was Vendor Two.
- Without factoring in the specific vendor, students who received more than 10 sessions of tutoring did not have significantly increased odds of gaining 4 or more PR points on the Star Math assessment. Furthermore, the specific vendor analyses indicated that only two vendors demonstrated positive effects on student math outcomes.
- Without factoring in the specific vendor, students who received more than 10 sessions of tutoring had significantly increased odds of gaining 4 or more PR points on the Star Reading assessment. When analyzing the data by vendor, three vendors demonstrate significant positive impacts on student reading outcomes.

## Limitations

There were a number of limitations that should be considered when interpreting the results of the tutoring program evaluation. Below is not an exhaustive list but one that includes the factors that the R&E team felt were most relevant when drawing conclusions about a vendor's overall impact on student outcomes.

- The team did not collect individual tutor characteristics that could play a role in the quality of their sessions, such as the years of experience in tutoring, having a background or degree in teaching/education, or tutor demographics.
- For the analysis, the team used the total number of sessions rather than the total number of minutes spent in the tutoring session. This method was selected to better align with the dosage criteria provided by the vendors.



- Due to time constraints, the team looked at the overall number of sessions that the students had with the tutor rather than breaking the number of sessions down by each content area and how the number of sessions in each content area affected their Star PR change, attendance rate, and course pass rate. This method was chosen since many students received tutoring sessions in both ELA/reading and math—sometimes within the same session. Additionally, this made the Star analysis more aligned with the high school pass rates analysis since the average pass rate was taken across any ELA and math course completed during the year (versus separating out the pass rates for subject areas). The exception to this analysis rule was for Vendor One, which tutored solely in reading, thus the team analyzed only Star Reading data.
- The quality of the tutoring sessions between and within vendors likely varied widely; however, the district did not have a feasible method for observing or evaluating the quality of the session beyond the data collected in the Qualtrics form.
- Since we did not have the exact dosage information from all vendors, we approximated the number of sessions that students should have received if tutoring services were provided at least once per week for most weeks during the school year (approximately 35 sessions). Those vendors that did not provide exact dosage information had listed on their websites that they provide services “every week,” which would require a minimum of 35 sessions across the school year.
- It is likely that the group size varied between and within vendors for the group sessions (e.g., 2–3 students, 4–6 students), which could play a role in the effectiveness of the session. This information was not readily or reliably available in the data collected in the Qualtrics form.
- While we did control for several student-level characteristics (i.e., race/ethnicity, gender, special education status, and English language learner status), we did not control for other potential factors such as the student’s average daily attendance or mobility between schools. Additionally, school-level characteristics that could have impacted the tutoring program’s implementation, such as the school’s mobility rate, its culture and climate, and the organizational environment, were also not controlled for in the analysis.

## Conclusion

Notwithstanding the limitations described above, the results demonstrate that the use of ESSER funding to provide in-school tutoring services to students significantly improved the students’ chances of improving on a variety of metrics, namely their Star test performance, attendance rates, and course passing rates. To be sure, some vendors outperformed others in terms of student outcomes observed, and others did not demonstrate significant effects for specific content areas, but the overall effect observed regardless of vendor was a significant increase in the odds that the students who received at least 11 tutoring sessions made notable improvements in the Star Reading test, attendance, and/or coursework performance. Math tutoring was not quite as consistently effective across vendors, but specific providers did demonstrate impressive results for students. The R&E team hopes that this evaluation provides clarity on the value and effectiveness of the tutoring program more generally as well as guidance for the ELO office should the staff decide to pursue tutoring through an external vendor in the future.