ESSA Level II Report

February 2025

Contents

Executive Summary

To demonstrate effectiveness of their program and services, the impact of K12 Tutoring on student outcomes was examined. The study was also designed to satisfy Level II requirements (Moderate Evidence) of the Every Student Succeeds Act (ESSA).

Study Sample, Measures, and Methods

This study utilized a quasi-experimental design with propensity score matching to meet ESSA Level II evidence standards. It analyzed 354 fourth-grade students from two full-time virtual schools in the southwestern U.S. The treatment group included 118 students who participated in K12 Tutoring during the 2023–2024 school year, while the comparison group consisted of 236 demographically similar students with comparable fall achievement who did not receive tutoring. The sample was 55% female and predominantly Hispanic (43%), with 61% classified as economically disadvantaged. Additionally, 25% received special education services, and 9% were English Language Learners.

Administrative and standardized assessment data were used to assess the impact of tutoring on student outcomes. Descriptive statistics provided insights into participant characteristics and program implementation, while regression analyses measured tutoring effects on student outcomes. A matched sample was created to ensure baseline equivalence.

Program Implementation and Student Findings

Student Usage. On average, each fourth-grade student participating in K12 Tutoring received 729 minutes of instruction across 12 tutoring sessions between December 2023 and June 2024.

Student Outcomes. The impact of K12 Tutoring on student outcomes was assessed by comparing tutoring participants to a demographically and academically similar group of non-users. Students who received K12 Tutoring demonstrated significantly higher end-of-year NWEA MAP Growth math RIT scores than their peers who did not participate. Further regression analysis examined the effect of tutoring intensity on math MAP Growth. The findings revealed a positive and statistically significant relationship, with students who received high-dosage tutoring (> 12 hours) experiencing the greatest academic gains.

Conclusion

Given the positive outcome findings, this study meets ESSA Level II (Moderate Evidence) requirements. It was a properly designed and well-implemented quasi-experimental study that documented baseline equivalence, incorporated statistical controls, included over 350 students across multiple schools, and demonstrated at least one statistically significant positive result.

Introduction

High-impact (or high-dosage) tutoring, characterized by frequent and intensive instructional support, has been shown to significantly improve student academic outcomes, particularly in mathematics and reading. Research indicates that tutoring sessions of at least three times per week, delivered in small groups or one-on-one, yield substantial gains in student achievement (Kraft & Falken, 2021). A recent meta-analysis found that high-impact tutoring can produce effect sizes of 0.29 or higher, equivalent to months of additional learning (Nickow et al., 2024). Moreover, such tutoring is particularly beneficial for historically underserved students, helping to close achievement gaps and support learning recovery post-pandemic (Dietrichson et al., 2021). The personalized nature of tutoring allows for targeted interventions, immediate feedback, and stronger student engagement, making it one of the most effective educational strategies for accelerating learning (Robinson & Loeb, 2021).

K12 Tutoring provides students with tailored academic support tailored to their unique learning needs. Specializing in K-12 education, it offers targeted tutoring in core subjects such as math, English, science, and social studies, helping students build confidence and improve their skills. With state-certified tutors and a personalized online learning environment, K12 Tutoring ensures students receive high-quality instruction that aligns with their school curriculum. K12 Tutoring focuses on fostering academic growth, closing learning gaps, and preparing students for long-term success (see logic model [Long & Henschel, 2023], Appendix A).

To demonstrate effectiveness of their program and services, the impact of K12 Tutoring on student outcomes was examined. The study was also designed to satisfy Level II requirements (Moderate Evidence) of the Every Student Succeeds Act (ESSA).

The study had the following research questions:

Implementation Research Questions

1. For students participating in K12 Tutoring, how many total tutoring minutes and tutoring sessions were completed?

Effectiveness Research Questions

NWEA MAP Growth Math Assessment

2. What was the impact of K12 Tutoring on the end-of-year NWEA MAP Growth math assessment for students who received tutoring compared to those who did not?
3. How did the end-of-year NWEA MAP Growth math assessment scores vary based on the total amount of time students spend in tutoring?

Method

This section of the report provides a concise overview of the study's design, setting, participant details, measurement tools, analytical methods, and baseline equivalence procedures.

Study Design

This study employed a quasi-experimental design utilizing propensity score matching to meet ESSA Level II evidence standards. The treatment group consisted of students who participated in K12 Tutoring during the 2023–2024 school year, while the comparison group comprised demographically similar students with comparable fall achievement who did not use K12 Tutoring.

Setting and Participants

This quasi-experimental study was conducted during the 2023–2024 academic year and examined a matched sample of 354 fourth-grade students from two full-time virtual school districts across two southwestern states. While the districts operate independently, under separate administrations, the tutoring implementation was similar. The sample consisted of 118 students in the treatment group and 236 in the comparison group. Gender distribution was balanced, with 55% female students. The majority of participants identified as Hispanic (43%), followed by African American (19%), white (10%), American Indian (1%), Asian (1%), and 25% whose racial or ethnic background was unreported. Additionally, 61% of students were classified as economically disadvantaged based on free or reduced-price lunch eligibility, 25% received special education services, and 9% were identified as English Language Learners.

Measures

This study included the following measures to provide insights into the impact of K12 Tutoring on student outcomes.

K12 Tutoring Usage Metrics. Student-level tutoring usage data was collected to measure tutoring participation, including total minutes and session counts. This data was analyzed to assess variations in students' engagement with K12 Tutoring during the 2023–2024 school year and to examine its impact on student outcomes.

Student Outcomes. Administrative data was collected, such as assessment score and course data, to evaluate student outcomes.

Data Analysis

Descriptive statistics were used to summarize participant characteristics and inform implementation analyses. To evaluate the outcomes and impacts of K12 Tutoring, regression analyses were conducted, incorporating student-level covariates and nearest neighbor propensity score matching to mitigate potential selection bias. Additionally, standardized effect sizes (Hedges' *g*) were calculated to measure the magnitude of differences in outcomes between K12 Tutoring participants and non-users with similar demographic and academic profiles.

Baseline Equivalence

To ensure the validity of the study's findings and adhere to ESSA Level II standards, researchers conducted baseline equivalence tests on matched treatment and comparison student samples. These findings are discussed in the Results section and referenced in Appendix B.

Results

Implementation Research Findings

Tutoring usage metrics were analyzed to determine the extent to which students used K12 Tutoring during the 2023–2024 school year.

For students participating in K12 Tutoring, how many total tutoring minutes and tutoring sessions were completed?

Fourth-grade students who participated in K12 Tutoring completed an average of 12 sessions. Students (89%) generally completed 60-min sessions; however, due to scheduling, some students completed all their tutoring in 30-min sessions (6%) and others completed tutoring in a mixture of 30 and 60-min sessions (5%). On average, students completed 729 minutes of tutoring between December 2023 and June 2024. Tutoring sessions were completed as small groups sessions (of less than 4 students per tutor).

Effectiveness Research Findings

To answer effectiveness research questions, regression analyses were conducted using matched samples of K12 Tutoring users to non-users, as well as analyses for K12 Tutoring users only. The following section details (a) the impact of K12 Tutoring on learning outcomes, (b) the influence of time spent in K12 Tutoring on learning outcomes. Statistically significant findings are noted only if they meet the $p < .05$ threshold.

What was the impact of K12 Tutoring on the end-of-year NWEA MAP Growth math assessment for students who received tutoring compared to those who did not?

A matched sample of students who used K12 Tutoring and those who did not was constructed, incorporating MAP Growth assessment scores prior to tutoring, school, gender, and special education status. The baseline equivalence of the matched sample was confirmed with a Hedges' g of 0.01 and a standardized mean difference of 0.15, indicating comparable pre-tutoring MAP Growth assessment scores between groups.

Using this matched sample, a linear regression model was conducted to assess the effect of K12 Tutoring participation on end-of-year NWEA MAP Growth math score (as measured by RIT [Rasch Unit] points). The model controlled for pre-tutoring math MAP Growth scores, as well as school, gender, and special education status. Results indicated that students who participated in K12 Tutoring had significantly higher end-of-year math RIT scores than non-users. This difference was statistically significant ($p < .001$; $g = .21$, see Figure 1). For the full model, see Appendix C.

K12 Tutoring users had higher end-of-year NWEA MAP Growth math assessment RIT scores than non-users, and this difference was statistically significant.

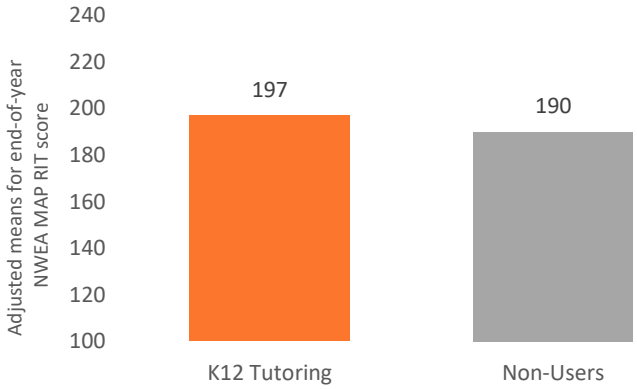


Figure 1. Adjusted means for end-of-year NWEA MAP Growth math assessment RIT scores for K12 Tutoring and non-users. Mean difference was statistically significant ($p < .001$, $g = .21$, $n = 354$).

How did the end-of-year NWEA MAP Growth math assessment scores vary based on the total amount of time students spend in tutoring?

A linear regression model was conducted for K12 Tutoring students, exploring the influence of different intensities of K12 Tutoring usage on end-of-year NWEA MAP Growth math assessment scores. The model also included MAP Growth assessment score prior to beginning tutoring, school, gender, and special education status as covariates. As shown in Figure 2, students who spent more time in K12 Tutoring had higher end-of-year NWEA MAP Growth math assessment RIT scores. The difference was statistically significant, with students who completed more than 12 hours outperforming those with less than six hours ($p = .04$, $g = 0.38$) and those with 6–12 hours ($p = .03$, $g = 0.64$). This suggests that participants that completed tutoring in a high-dosage schedule saw the largest benefit from K12 Tutoring.

K12 Tutoring users who participated in a high-dosage schedule (more than 12 hours of tutoring) had higher end-of-year NWEA MAP Growth math assessment RIT scores, and this difference was statistically significant.

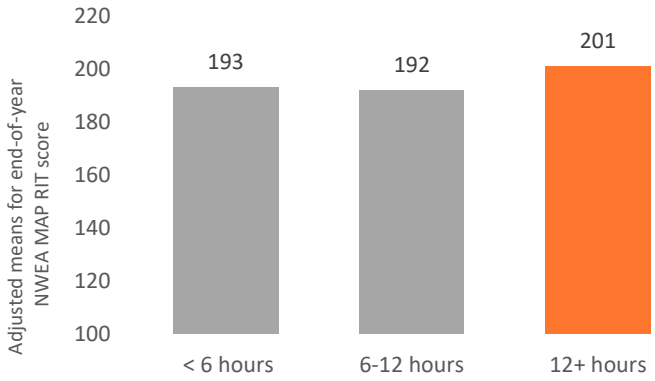


Figure 2. Influence of time spent in K12 Tutoring usage on adjusted means for end-of-year NWEA MAP Growth math assessment RIT scores. Mean difference between "12+ hours" and "< 6 hours" was statistically significant ($p = .04$, $g = 0.38$) as was the mean difference between "12+ hours" and "6-12 hours" ($p = .03$, $g = .64$).

Conclusions

Given multiple positive outcome findings, this study provides results to satisfy ESSA evidence requirements for Level II (Moderate Evidence). Specifically, this quasi-experimental study met the following criteria for Level II:

- ✓ Proper design and implementation
- ✓ Baseline equivalence for treatment and comparison groups
- ✓ Statistical controls through covariates
- ✓ At least 350 students in the analysis sample
- ✓ Representative, multi-site study
- ✓ At least one statistically significant, positive finding

References

Dietrichson, J., Bøg, M., Filges, T., & Klint Jørgensen, A.-M. (2017). Academic interventions for elementary and middle school students with low socioeconomic status: A systematic review and meta-analysis. *Review of Educational Research, 87*(2), 243-282.

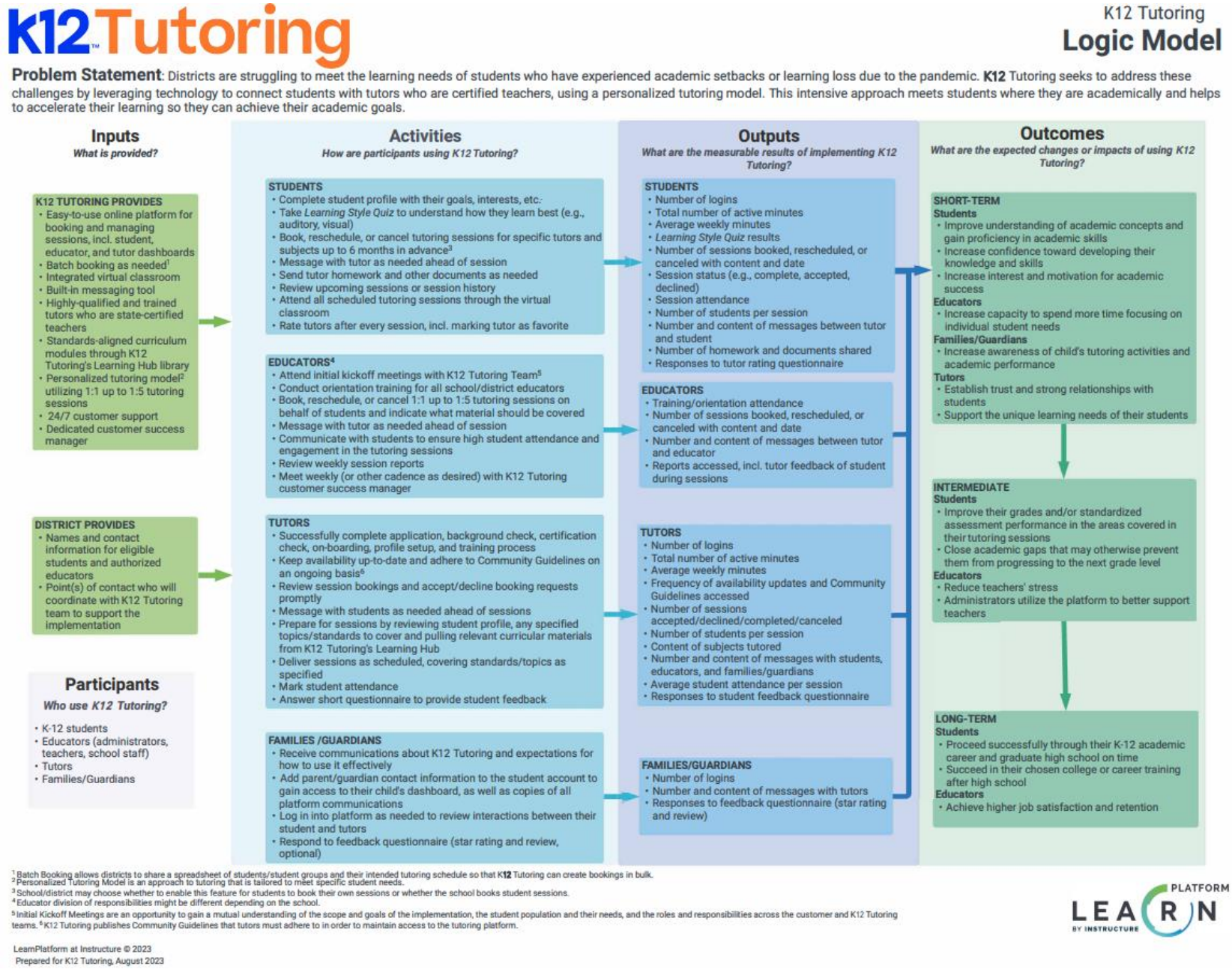
Kraft, M. A., & Falken, G. (2021). A blueprint for scaling tutoring across public schools. *Brown University Annenberg Institute.*

Long, C. & Henschel, M. (2023). K12 Tutoring Logic Model. *LearnPlatform by Instructure.*

Nickow, A., Oreopoulos, P., & Quan, V. (2024). The promise of tutoring for PreK–12 learning: A systematic review and meta-analysis of the experimental evidence. *American Educational Research Journal, 61*(1), 74-107.

Robinson, C. D., & Loeb, S. (2021). High-impact tutoring: State of the research and priorities for future learning. *National Student Support Accelerator, 21*(284), 1-53.

Appendix A. K12 Tutoring Logic Model (Long & Henschel, 2023)



Appendix B. Additional Information on Study Design and Methods

A total of 118 fourth-grade K12 Tutoring students had complete demographic, achievement, and usage data. In addition, a sample of 824 comparison students with complete demographic and achievement data was obtained. Using this dataset of 118 K12 Tutoring students and 824 non-users, propensity score matching (i.e., nearest neighbor matching without replacement) was used to create a matched study sample of K12 Tutoring students and 236 non-users with similar academic and demographic profiles. Consequently, the matched study sample included 354 students. Additional demographic information on participating students is included in Table B1.

Table B1. Student demographics by group for matched sample.

Characteristic	K12 Tutoring (n = 118)		Non-users (n = 236)		Total Sample (N = 354)	
	Percent	*n*	Percent	*n*	Percent	*n*
Gender						
Female	54%	64	55%	131	55%	195
Male	46%	54	45%	105	45%	159
Ethnicity						
Hispanic	39%	46	44%	105	43%	151
Black	23%	27	18%	42	19%	69
White	13%	15	9%	21	10%	36
American Indian	3%	3	1%	2	1%	5
Asian	1%	1	1%	2	1%	3
Other/Unreported	22%	26	18%	64	25%	90
English Language Learner (ELL Status)						
Yes	10%	12	9%	21	9%	33
No	90%	106	91%	215	91%	321
Special Education Status						
Yes	28%	33	23%	55	25%	88
No	72%	85	77%	181	75%	266
Economically Disadvantaged						
Yes	59%	70	61%	145	61%	215
No	41%	48	39%	91	39%	139

Baseline Equivalence

After propensity score matching, regression analyses confirmed no statistically significant differences between student groups. Baseline equivalency was assessed to determine whether students who participated in K12 Tutoring differed from those who did not participate in tutoring on key characteristics. For the primary matching variable, pre-tutoring NWEA MAP RIT scores, the matched sample met What Works Clearinghouse (WWC) baseline equivalence standards, with a Hedges' g effect size of 0.01 and a standardized mean difference (SMD) below 0.25. No other demographic variables were found to be significant predictors of group assignment and were therefore not included in the matching model. See Table B2 for full baseline equivalency results.

Table B2. Baseline Equivalence.

Variable	K12 Tutoring		Non-users			
	Adjusted Mean (*SD*)	*n*	Adjusted Mean (*SD*)	*n*	Hedges' *g*	*p*-value
NWEA MAP RIT Score	187.05 (16.12)	118	187.20 (16.03)	236	0.01	.93
	%	*n*	%	*n*	Chi-square	*p*-value
School District (TX)	85	100	84	199	.00	1.00
Gender (Female)	54	64	55	131	.01	.91
Special Education (Yes)	28	33	23	55	.68	.41

Appendix C. Additional Information on Results

The following section provides additional details on the analyses conducted for the Effectiveness Research Questions.

What was the impact of K12 Tutoring on the end-of-year NWEA MAP Growth math assessment for students who received tutoring compared to those who did not?

A linear regression model was conducted to evaluate the impact of K12 Tutoring participation on end-of-year NWEA MAP Growth math scores, measured in RIT (Rasch Unit) points. Tutoring participation was included as a dummy-coded predictor, with non-users serving as the reference group. The model also controlled for pre-tutoring MAP Growth math scores (continuous predictor), as well as school, gender, and special education status (nominal predictors). Tutoring was a significant positive predictor of end-of-year NWEA MAP Growth math scores, with the full model accounting for 33% of the variance in growth scores, $F(5, 348) = 35.40$, $p < .001$. See Table C1 for the full model.

Table C1. End-of-year NWEA MAP Growth math scores as predicted by Tutoring.

	Unstandardized Beta (B)	Standard Error	*t*-value	*p*-value
Intercept	70.38	11.25	6.26	< .001
Tutoring	6.28	1.74	3.62	< .001
Pre-tutoring MAP Growth	0.65	0.06	11.37	< .001
School	-8.20	2.25	-3.65	< .001
Gender	1.66	1.66	1.00	.32
SE	2.39	2.39	1.18	.24

Note. The reference group for Tutoring is non-users.

How did the end-of-year NWEA MAP Growth math assessment scores vary based on the total amount of time students spend in tutoring?

A linear regression model was conducted to examine the impact of K12 Tutoring intensity on end-of-year NWEA MAP Growth math assessment scores among tutoring participants. Tutoring usage was included as an ordinal predictor, with students categorized into three groups: less than six hours, between six and 12 hours, and more than 12 hours of tutoring between December 2023 and June 2024. The reference group was set to students who completed more than 12 hours of tutoring. The model also controlled for pre-tutoring MAP Growth math scores (continuous predictor), as well as school, gender, and special education status (nominal predictors). As shown in Table C2, students who completed more than 12 hours of K12 Tutoring had higher end-of-year NWEA MAP Growth math RIT scores compared to those who received less than six hours and between six and 12 hours, with the full model explaining 30% of the variance in growth scores, $F(6, 111) = 9.37$, $p < .001$.

Table C2. End-of-year NWEA MAP Growth math scores as predicted by tutoring usage.

	Unstandardized Beta (B)	Standard Error	*t*-value	*p*-value
Intercept	97.22	19.93	4.88	< .001
Tutoring Usage				
< 6 hours	-7.86	3.86	-2.04	.04
6-12 hours	-8.90	4.01	-2.22	.03
Pre-tutoring MAP Growth	0.57	0.10	5.81	< .001
School	-5.56	4.40	-1.27	.21
Gender	4.64	2.99	1.55	.12
SE	-3.57	3.53	-1.01	.31

Note. The reference group for Tutoring Usage is students who completed 12+ hours of tutoring.