



# The Impact of Tutoring on State Assessment Outcomes in Mathematics: Evidence from the Texas STAAR

ESSA Level II Report  
January 2026



## Contents

<b>Executive Summary</b> .....	3
Study Sample, Measures, and Methods .....	3
Results .....	3
Conclusion .....	3
<b>Introduction</b> .....	4
<b>Method</b> .....	5
Study Design .....	5
Setting and Participants .....	5
Measures .....	6
Data Analysis .....	6
Baseline Equivalence .....	6
<b>Results</b> .....	7
Implementation Research Findings .....	7
Impact Research Findings .....	7
<b>Conclusions</b> .....	8
<b>References</b> .....	10
<b>Appendix A. Additional Information on Study Design and Methods</b> .....	11
<b>Appendix B. Additional Details on Results</b> .....	13

## Executive Summary

To evaluate the effectiveness of its tutoring services on student academic achievement, K12 Tutoring conducted a comprehensive program evaluation. The study was designed to meet the Moderate Evidence (Level II) requirements of the Every Student Succeeds Act (ESSA) by examining the impact of tutoring participation on mathematics state assessment outcomes on the State of Texas Assessment of Academic Readiness (STAAR) mathematics assessment.

## Study Sample, Measures, and Methods

This study compared outcomes for students who participated in K12 Tutoring with those of a matched group of non-participating peers. The sample included 632 students in grades 4-8 enrolled in three full-time virtual schools in Texas during the 2023-2024 and 2024-2025 academic years. Of these students, 316 received K12 Tutoring services, while 316 served as matched comparison students selected based on similar academic, demographic, and enrollment characteristics. The sample was demographically diverse and included students with a range of socioeconomic backgrounds, special education status, and English learner designation.

Academic outcomes were assessed using scaled scores from the STAAR mathematics assessment, administered in the spring prior to tutoring implementation (baseline) and again at the end of the tutoring year (final). Regression analyses were conducted to evaluate program effects while controlling for baseline achievement and relevant covariates. Baseline equivalence between the treatment and comparison groups was established in accordance with What Works Clearinghouse (WWC) standards.

## Results

**Tutoring Participation.** Among students who participated in K12 Tutoring, the average cumulative instructional time was 13.60 hours. Analyses revealed significant differences in tutoring dosage by grade band, with middle school students receiving more tutoring than elementary students, and by academic year, with higher participation observed in 2023-2024 than in 2024-2025.

**Student Outcomes.** Students who participated in K12 Tutoring demonstrated significantly higher final STAAR mathematics scores compared with their matched peers. Regression results indicated a statistically significant positive effect of tutoring participation, corresponding to an estimated gain of approximately 3.7 additional percentile points relative to the comparison group. Additional analyses testing interactions between tutoring participation and demographic variables did not yield statistically significant effects, suggesting that program impacts were consistent across demographic subgroups.

## Conclusion

Students who participated in K12 Tutoring demonstrated significantly greater gains in mathematics than their matched peers. These results provide evidence that K12 Tutoring is an effective supplemental instructional intervention and meets ESSA Tier II (Moderate Evidence) standards, based on a rigorous quasi-experimental design, baseline equivalence, appropriate statistical controls, and a diverse sample.

## Introduction

Across the United States, mathematics proficiency among K-12 students remains a significant concern, with the 2024 Nation's Report Card indicating that only about 28% of eighth-grade students performed at or above the NAEP Proficient level in mathematics, a figure that remains below pre-pandemic benchmarks and highlights ongoing challenges in foundational mathematics learning (NCES, 2024).

These national patterns are reflected within Texas, where state assessment results show that a majority of students (57%) in Grades 3-8 do not meet grade-level expectations on the State of Texas Assessment of Academic Readiness (STAAR) mathematics assessments, with notable declines in proficiency as students progress from elementary to middle school (TEA, 2025). Such trends underscore persistent gaps in mathematics achievement that disproportionately affect learners as they advance through grade levels, reinforcing the need for effective, research-based instructional supports to promote meaningful academic gains.

A substantial body of research indicates that tutoring can produce meaningful improvements in student learning outcomes across diverse contexts, making it one of the more effective instructional interventions available in K-12 education (Dietrichson et al., 2017, Nickow et al., 2024; Nietzel et al., 2022). Meta-analyses of tutoring programs, including randomized controlled trials, find that students who receive one-on-one or small-group tutoring generally exhibit positive gains in academic achievement, with effect sizes in the range of approximately 0.20 to 0.40 standard deviations on standardized assessments, equivalent to several additional months of learning relative to comparison groups (Kraft et al., 2024; Nickow et al., 2024). These effects have been consistently observed across grade levels and subject areas, and tutoring interventions often outperform other common academic supports in terms of effect size. Moreover, research suggests that high-impact tutoring that is aligned with classroom instruction, delivered with sufficient frequency, and implemented by trained educators can yield even larger learning gains.

The state of Texas has responded to this evidence base through the enactment of House Bill 1416 (HB1416). Enacted by the Texas Legislature in 2023, HB1416 significantly revised the state's accelerated instruction requirements for students who do not meet grade-level expectations on the STAAR. Under this legislation, districts are required to provide supplemental instruction aligned with research-based high-impact tutoring practices, including small-group sessions of no more than four students and a prescribed minimum number of instructional hours dependent on student need. HB1416 also updated prior accelerated instruction mandates by prioritizing key content areas, adjusting tutor-to-student ratios, and clarifying implementation parameters for school systems.

Multiple schools implemented K12 Tutoring as a supplemental instructional service in response to the accelerated instruction requirements outlined in HB1416. K12 Tutoring provides targeted academic support delivered by certified teachers and is designed to complement core classroom instruction by aligning tutoring content with school curricula and state academic standards. Instruction is delivered primarily in small-group formats, enabling tutors to address students' identified learning needs while reinforcing key concepts introduced during regular coursework. This implementation model is intended to promote coherence between tutoring and classroom instruction and to support students who require additional academic assistance to meet grade-level expectations.

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

The study had the following research questions:

### *Implementation Research Question*

1. Among students who participated in K12 Tutoring, what was the cumulative amount of instructional time received?

### *Impact Research Question*

1. To what extent did participation in K12 Tutoring influence STAAR mathematics performance relative to matched non-participating students?
2. Were there differences in the impact of K12 Tutoring on STAAR mathematics performance among students with varying demographic characteristics?

## Method

This section describes the methodological framework used in the study, detailing the research design, study setting, sample characteristics, outcome variables, analytic strategy, and the steps taken to assess equivalence between groups prior to analysis.

### Study Design

A quasi-experimental design incorporating propensity score matching was employed to meet ESSA Tier II (Moderate Evidence) criteria. The study examined differences in academic outcomes between students who participated in K12 Tutoring during the 2023-2024 and 2024-2025 academic years and a statistically matched group of students who did not receive tutoring services (see Appendix A, Figure A1, for the study timeline). The matched comparison group was constructed using available demographic characteristics and state assessment performance from the spring immediately preceding the tutoring implementation year. These procedures were intended to support baseline equivalence between groups prior to the estimation of program-related differences in outcomes.

### Setting and Participants

The study was conducted during the 2023-2024 and 2024-2025 academic years and included a total of 632 students in grades 4 through 8 from three full-time virtual schools in Texas. Although each school maintained its own administrative framework, the tutoring program was implemented according to a standardized model across all sites.

The analytic sample consisted of 316 students who participated in the tutoring program and a matched cohort of 316 students who did not receive tutoring. Tutoring was implemented across two cohorts, comprising 166 students in the 2023-2024 academic year and 150 students in the 2024-2025 academic year. Gender was relatively balanced (54% Female). With respect to race and ethnicity, 31% of students identified as White, 24% as African American, 2% as Asian, 3% as other, and 40% did not report racial or ethnic background. Additional demographic and program eligibility information indicated that 44% of students qualified for free or reduced-price lunch, 30% received special education services, and 11% were classified as English Language Learners.

## Measures

**K12 Tutoring Participation Metrics.** Tutoring was delivered primarily in small groups, typically with fewer than four students assigned to each tutor. Student-level participation data were collected to capture the total minutes spent in tutoring sessions. These metrics were used to describe overall engagement with the K12 Tutoring program.

**Student Outcomes.** Mathematics achievement was assessed using administrative data from the State of Texas Assessment of Academic Readiness (STAAR). Students completed the assessment in the spring preceding the tutoring implementation year (baseline assessment) and again in the spring of the tutoring year (final assessment). Scaled scores from these two time points were employed to measure academic growth and served as the primary outcome for evaluating the impact of K12 Tutoring.

## Data Analysis

Descriptive analyses were first conducted to summarize the demographic and academic characteristics of the sample and to examine patterns of program implementation. Propensity score matching, incorporating student-level demographics and baseline academic performance and employing nearest neighbor selection, was used to reduce potential selection bias and create comparable groups of tutored and non-tutored students. To evaluate the association between K12 Tutoring participation and mathematics achievement, regression models were then conducted. In addition, standardized effect sizes were calculated to provide interpretable estimates of differences in outcomes between the intervention and matched comparison groups.

## Baseline Equivalence

Demonstrating equivalence between the matched treatment and comparison groups at baseline was essential to support the internal validity of the study and to satisfy ESSA Tier II (Moderate Evidence) requirements. The procedures used to assess baseline equivalence, along with the corresponding results, are presented in the Results section and are described in greater detail in Appendix A.

## Results

### Implementation Research Findings

Among students who participated in K12 Tutoring, what was the cumulative amount of instructional time received?

Among students who participated in K12 Tutoring, the average cumulative instructional time was 13.60 hours (see Table 1). Analysis of variance (ANOVA) revealed a significant main effect of grade band, with middle school students spending more time in tutoring than their elementary school peers,  $p = .001$ . There was also a significant main effect of academic year, with students in the 2023-2024 academic year receiving more time in tutoring than those in 2024-2025,  $p < .001$ . The interaction between grade and academic year was not statistically significant,  $p = .75$ .

Given these observed differences in time spent in tutoring, it is possible that variations in academic growth may arise. Therefore, both grade band and academic year were included as covariates in subsequent analyses examining program outcomes.

**Table 1**

*Average Instructional Time Spent in Tutoring (in Hours) by Grade Band and Academic Year*

Grade Band	Academic Year	
	2023-2024	2024-2025
Elementary	15.36	9.47
Middle School	17.98	12.81

### Impact Research Findings

To address the study’s impact questions, regression analyses were conducted comparing matched groups of students who participated in K12 Tutoring with non-participating peers. This section reports findings related to the effect of K12 Tutoring on student learning outcomes. Statistical significance is reported for results with  $p < .05$  only.

#### To what extent did participation in K12 Tutoring influence STAAR mathematics performance relative to matched non-participating students?

To examine the effect of K12 Tutoring on STAAR mathematics performance, a matched sample was created by pairing students who received tutoring with non-participants. Matching criteria included baseline STAAR scores, cohort (based on academic year of participation), school, grade band, instructional days enrolled, average weekly course logins, and demographic characteristics including gender, ethnicity, special education status, free or reduced-price lunch eligibility, and English learner status. Baseline equivalence between the two groups was confirmed, with a Hedges’  $g = .002$ , indicating negligible differences in STAAR scores prior to tutoring.

Within this matched sample, a linear regression analysis was conducted to estimate the effect of tutoring on final STAAR scores. Tutoring participation was included as a binary predictor, while controlling for baseline STAAR scores, academic year, school, and grade band. Results indicated that students who participated in K12 Tutoring achieved significantly higher end-of-year STAAR scores compared with their matched peers (Hedges'  $g = .10$ ,  $p = .002$ ; see Table 2), corresponding to an estimated gain of approximately 3.7 additional percentile points than students who did not receive tutoring.

Given observed differences in cumulative tutoring time across grade bands and academic years, additional interaction terms were tested. These interaction models did not improve model fit, and no statistically significant interactions were observed, suggesting that the effect of tutoring on STAAR outcomes was consistent across grade levels and academic years.

**Students who participated in K12 Tutoring  
achieved higher STAAR mathematics scores than their non-tutored peers.**

**Table 2**

*Adjusted Means for STAAR Mathematics Assessment Scale Scores for K12 Tutoring and Non-Users*

Grade Band	Baseline STAAR Score	Final STAAR Score	Δ Score (Final – Baseline)
No Tutoring	1415	1479	64
K12 Tutoring	1415	1499	84

**Were there differences in the impact of K12 Tutoring on STAAR mathematics performance among students with varying demographic characteristics?**

To test whether the observed outcomes above differed across student populations, an additional demographic variable was included in the model as an interaction term with tutoring. The following variables were tested in individual models and resulted in no significant interactions ( $p > .05$ ): gender, ethnicity, English language status, special education status, free or reduced lunch status. This suggests that the K12 Tutoring impacts were consistent across all demographic sub-groups.

## Conclusions

The findings of this study indicate that participation in K12 Tutoring was associated with higher final STAAR mathematics scores relative to matched peers, after controlling for baseline performance and relevant demographic characteristics. These results highlight the potential of structured, sustained tutoring interventions as an effective strategy for supporting student learning and addressing academic disparities.

Based on these positive outcomes, this study meets the ESSA Tier 2 (Moderate Evidence) standards. Specifically, the quasi-experimental design satisfied key criteria, including:

- ✓ 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., Schueler, B. E., & Falken, G. (2024). What impacts should we expect from tutoring at scale? Exploring meta-analytic generalizability. *Brown University Annenberg Institute*. EdWorkingPaper No. 24-1031.
- National Center for Education Statistics. (2024). *The Nation's Report Card: Mathematics 2024*. U.S. Department of Education, Institute of Education Sciences. <https://www.nationsreportcard.gov>
- Neitzel, A. J., Lake, C., Pellegrini, M., & Slavin, R. E. (2022). A synthesis of quantitative research on programs for struggling readers in elementary schools. *Reading Research Quarterly, 57*(1), 149-179.
- 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.
- Texas Education Agency (2025). 2025 STAAR Results. *TEA*. Retrieved from <https://tea.texas.gov/student-assessment/student-assessment-results/spring-2025-staar-all-combined-assessment-results.pdf>

## Appendix A. Additional Information on Study Design and Methods

A total of 316 students in grades 4 through 8 who participated in K12 Tutoring and had complete demographic, achievement, and program participation records were included in the study. Third-grade students were excluded because prior-year (grade 2) STAAR scores were not available, preventing the establishment of a baseline. A comparable sample of non-tutored students with full demographic and achievement information was identified for matching. Propensity score matching using nearest neighbor pairing without replacement (1:1) was employed to align tutored students with academically and demographically similar peers. Matching criteria included baseline STAAR scores, cohort (based on academic year of participation), school, grade band, instructional days enrolled, average weekly course logins, and demographic characteristics including gender, ethnicity, special education status, free or reduced-price lunch eligibility, and English learner status. To ensure an appropriate comparison, the matched comparison group was constructed using exact matching on cohort year and grade level. As a result, Cohort 1 included 166 students receiving tutoring and 166 matched non-users, and Cohort 2 included 150 students receiving tutoring and 150 matched non-users. Baseline performance was defined as the end-of-year STAAR score from the grade prior to tutoring, corresponding to the year immediately before each student’s participation (Figure A1). This process yielded 316 matched pairs (632 students total). Further details on participant demographics can be found in Table A1.

**Figure A1**

*Timeline of Student Participation and Assessments*

	Academic Year		
	2022-2023	2023-2024	2024-2025
<b>Cohort 1</b>	Baseline STAAR	TUTORING <i>n</i> = 166	Final STAAR
<b>Cohort 2</b>		Baseline STAAR	TUTORING <i>n</i> = 150 Final STAAR

**Table A1**

*Student Demographics*

Characteristic	K12 Tutoring ( <i>n</i> = 316)		Non-users ( <i>n</i> = 316)		Total Sample ( <i>N</i> = 632)	
	Percent	<i>n</i>	Percent	<i>n</i>	Percent	<i>n</i>
Gender						
<i>Female</i>	53.14%	168	55.06%	174	54.11%	342
<i>Male</i>	46.84%	148	44.94%	142	45.89%	290
Ethnicity						
<i>White</i>	29.43%	93	32.59%	103	31.01%	196
<i>Black</i>	24.37%	77	23.42%	74	23.89%	151
<i>Asian</i>	2.53%	8	1.90%	6	2.22%	14
<i>Other</i>	3.16%	10	3.48%	11	3.32%	21
<i>Unreported</i>	40.51%	128	38.61%	122	39.56%	250
English Learner						
<i>Yes</i>	11.08%	35	11.08%	35	11.08%	70
<i>No</i>	88.92%	281	88.92%	281	88.92%	562
Special Education Status						
<i>Yes</i>	31.96%	101	27.85%	88	44.30%	280
<i>No</i>	68.04%	215	72.15%	228	55.70%	352
Economically Disadvantaged						
<i>Yes</i>	46.20%	146	42.41%	134	46.9%	201
<i>No</i>	53.80%	170	57.59%	182	53.1%	228

Analyses on the propensity score-matched sample were conducted to ensure the treatment and comparison groups did not differ significantly at baseline. As shown in Table A2, the matched sample met WWC equivalence criteria, with a Hedges’ *g* of 0.002, indicating an essentially negligible difference in pre-intervention characteristics.

**Table A2**

*Baseline Equivalence on STAAR Mathematics Assessment Scale Score (Completed Prior to Tutoring)*

Outcome Variable	K12 Tutoring		Non-users		Hedges’ <i>g</i>	<i>p</i> -value
	Adjusted Mean ( <i>SD</i> )	<i>n</i>	Adjusted Mean ( <i>SD</i> )	<i>n</i>		
Scale Score	1414.59 (131.44)	316	1415.02 (223.08)	316	0.002	.98

## Appendix B. Additional Details on Results

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

### To what extent did participation in K12 Tutoring influence STAAR mathematics performance relative to matched non-participating students?

A linear regression model was used to assess the effect of K12 Tutoring participation on STAAR mathematics scale scores. Tutoring status was entered as a binary predictor, with non-participants serving as the reference group. The model controlled for baseline STAAR mathematics scale scores (continuous). Because STAAR employs a vertical scale, grade band was included to adjust for systematic score differences associated with grade-level variation in test difficulty; students were categorized as Elementary (Grades 4-5) or Middle School (Grades 6-8). Cohort and school were also incorporated to account for potential systematic differences attributable to the year of participation or the school context.

Tutoring participation was associated with higher final STAAR scores, with tutored students performing on average 20 points higher than their non-tutored peers. As expected, baseline STAAR scores were strongly predictive of final STAAR performance, and middle school students had higher posttest scores than elementary students. No significant differences were observed between schools or cohorts. Overall, these results indicate that K12 Tutoring participation is linked to meaningful gains in mathematics achievement with growth patterns largely consistent across grade bands. The model accounted for 83% of the variance in posttest scores,  $F(6, 625) = 448.10, p < .001$ . Full details of the regression model are presented in Table B1.

**Table B1**

*End-of-year STAAR Mathematics Scores as Predicted by Participation in K12 Tutoring*

	Unstandardized Beta (B)	Standard Error	t-value	p-value
Intercept	152.61	28.22	5.41	< .001
Tutoring	19.74	6.45	3.06	.002
Baseline STAAR Score	.93	0.02	44.64	< .001
Grade Band	44.83	8.98	4.99	< .001
School				
School 2	-14.99	22.89	-0.66	.51
School 3	-7.37	12.03	-0.61	.54
Cohort	4.51	6.89	0.65	.51

*Note.* Reference groups were as follows: students who did not participate in tutoring (*Tutoring*), elementary (*Grade Band*), School 1 (*School*), students participating in tutoring in SY23-24 (*Cohort*).