TutorSmart works:
An external validation of a community-based tutoring model

Scott Andrews[1]
David Ledrick[2]
Tiffany Adamski[3]

December 2024

1. Associate Director of Strategic Improvement, TutorSmart. *Corresponding Author:* sandrews@tutorsmart-toledo.com
2. Board Member, TutorSmart. djledrick@gmail.com
3. Executive Director, TutorSmart. tadamski@tutorsmart-toledo.com

The authors would like to acknowledge Nancy Buderer's work on the statistical analysis and assistance in reporting the results. Ms. Buderer holds a master's degree in Biostatistics from the University of Michigan and has over 33 years of experience.

ABSTRACT

Introduction: TutorSmart (TS) is a community-run nonprofit organization that partners with schools to implement a data-driven, high-impact tutoring model focused on closing gaps in student achievement ("gap closure"). In the 2023-24 school year, TutorSmart partnered with Toledo Public Schools to support Out-of-School-Time literacy interventions for 3rd-graders. This paper demonstrates the effectiveness of the TS model.

Methods: Toledo Public Schools provided scores from the i-Ready Test (iR) and Ohio's State Test (OST) for all 3rd-graders on a Reading Improvement and Monitoring Plan (RIMP) at schools where the TutorSmart model was implemented. Students who participated in TutorSmart were compared to non-TutorSmart students who attended the same schools. The TutorSmart group was further assessed based on the student's level of participation. Mean test scores between groups were compared for each exam session using two-tailed Student's *t*-tests.

Results: At schools where TutorSmart ran programming, 224 third graders were on RIMPs; fifty-four (54) were enrolled in TS, leaving one hundred seventy (170) as the comparison group (TPS). In the fall of 2023, the two groups had similar iR and OST scores. By Spring 2024, the TS group had significantly higher iR scores (mean score 493 vs 469, p=0.02; mean change 40 vs. 25, p=0.005). A subgroup analysis defined program compliance (pc) as attaining either 30 hours of tutoring or a 66% attendance rate and meeting model (mm) if both these milestones were obtained. This pcTS group comprised 43 students, while 21 students qualified as mmTS. By the Spring IR testing, both subgroups outperformed the comparison (mean change TPS 25 v. pcTS 41 p=.003 v. mmTS 49 p=.001). The removal rate from a RIMP did not reach statistical significance for any subgroup, though the TS students fared better than the TPS comparison group (TPS 25% v. pcTS 35% v. mmTS 38%).

Conclusion: The TutorSmart program model demonstrated significant gains on the iR diagnostic test and trended towards improved OST scores and RIMP removal rates. Student performance was enhanced in participants who met program requirements.

INTRODUCTION/BACKGROUND:

TutorSmart was established in 2017 to partner with local school districts to provide targeted academic support for underperforming students in the Greater Toledo area. The program gave classroom teachers (tutors) the resources to provide high-dose tutoring in small-group settings before or after school. In the 2023-24 academic year, TutorSmart served more than 350 students.

TutorSmart has traditionally tracked progress by estimating the number of months a student is behind their expected grade level (the student's "gap") and how many months they catch up while participating in TutorSmart (the degree to which that "gap" "closes"). TutorSmart's primary measure of success is the efficiency of this "gap closure". Gap closure directly aligns with TutorSmart's mission and informs the use of scarce resources. These internal measures, however, require validation. We chose Ohio's State Tests and the i-Ready exam, two validated, universally applied measures of proficiency within Toledo Public Schools. These measures illustrate the value of TutorSmart in terms familiar to parents, taxpayers, politicians, and donors.

The TutorSmart Program Model:

The TutorSmart program implements a comprehensive, four-phase educational model designed to optimize student achievement and teacher development. This systematic approach integrates evidence-based practices with technological innovation to deliver personalized academic support.

The initial phase centers on a student assessment utilizing the MobyMax digital learning platform. This technology enables educators to establish baseline measurements and identify specific learning gaps that require targeted attention. The second phase focuses on instructional delivery with interventions that are structured to promote academic achievement through sequential skill development and conceptual mastery.

TS also focuses on the instructors. In the third phase, TS provides specialized support through expert literacy coaching and data analysis assistance. This professional guidance ensures the maintenance of high instructional standards. By providing tutors with curated instructional resources, the program maintains pedagogical excellence. The final phase emphasizes learning integration through continuous improvement cycles. Tutors engage in ongoing data analysis and implement evidence-based teaching practices. This iterative approach enhances the effectiveness of tutoring sessions and improves classroom instruction more broadly.

Objectives

Our primary objective was to evaluate the efficacy of TutorSmart's tutoring model using standardized external performance measures. This will help validate our current internal method of evaluation. Our secondary objective was to analyze the differences between students with higher and lower levels of participation within TutorSmart. Significant differences would indicate effectiveness and help TutorSmart set programming goals for future students.

MATERIALS AND METHODS

Inclusion criteria

In the 2023-24 school year, the TutorSmart model was implemented at 10 schools in the Toledo Public Schools district, eight of which served 3rd-graders. All 3rd-grade students on a Reading Improvement and Monitoring Plan (RIMP) at these schools (n=224) were included in this study, 54 of whom participated in TutorSmart.

Data Collection and Analysis:

The primary data used in this analysis were student scores on two different standardized tests. The iReady (iR) is a criterion-referenced assessment administered three times each year, and the Ohio State Test (OST), a measure of attainment of state standards, is administered twice. The scores were obtained from Toledo Public Schools and analyzed with the help of a professional statistician.

Secondary data came from TutorSmart's records of student participation, which were collected site-by-site. From this data, two key TutorSmart metrics were calculated: Dosage (how many hours of TutorSmart intervention a student has received), and Engagement (how often the student attended tutoring, shown as a percentage).

Statistical Methods

The analyses compared the 54 3rd-grade students who participated in TutorSmart to the 170 who did not. Mean test scores for the two groups were compared at each testing interval using two-tailed Student's *t*-tests. The performance of each individual student was tracked, and a comparison was made across time periods for improvement. This 'within-student change' was calculated (Fall to Winter, Fall to Spring). Again, the mean change was tested for significance between groups using two-tailed Student's *t*-tests.

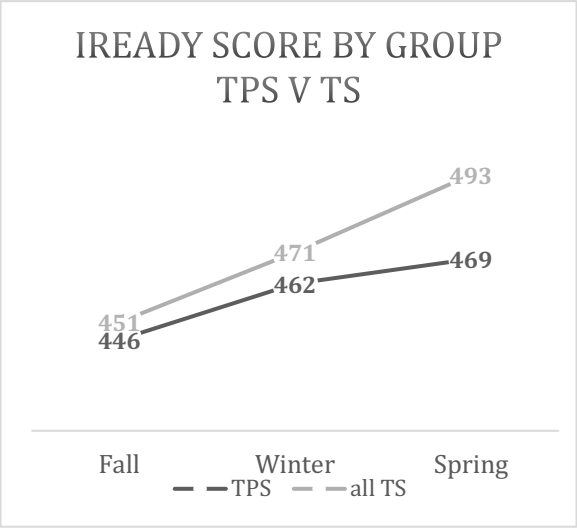
A further subgroup analysis was performed on the 54 TS students depending on the degree to which they met certain program requirements. Students were considered to be ‘program compliant’ (pc) if they either attended 30 hours of tutoring (Dosage) or if they came to at least two-thirds of the offered meetings during their period of enrollment (Engagement). Students who attained both these benchmarks were considered to ‘meet (the) model’ (mm). The mean, the standard deviation (SD), and the 95% confidence interval (CI) were described for each group. Because two different tests were conducted, we adjusted for multiple comparisons by using a critical value of $p<0.025$ to indicate statistical significance (i.e., type I error level of 0.05 divided by 2 tests).

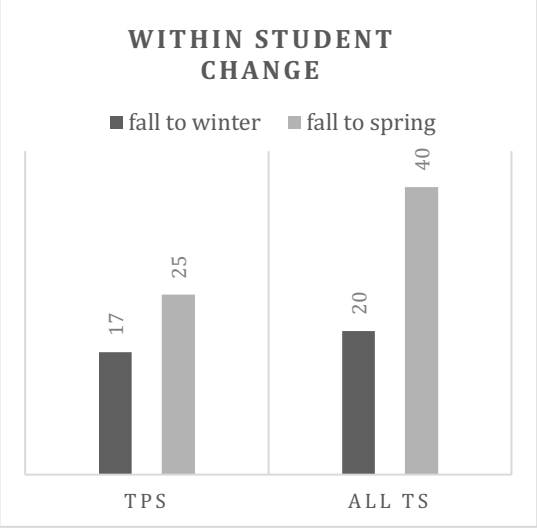
Students were classified as *removed from RIMP* if they had an Ohio Fall 2023 or Ohio Spring 2024 state test score of 700 or more. Group differences in the percentage of students removed from RIMP were tested using Chi-square tests. The odds of removal from RIMP between groups were expressed as an odds ratio with 95% confidence interval.

RESULTS

All TutorSmart vs. TPS (table 1)

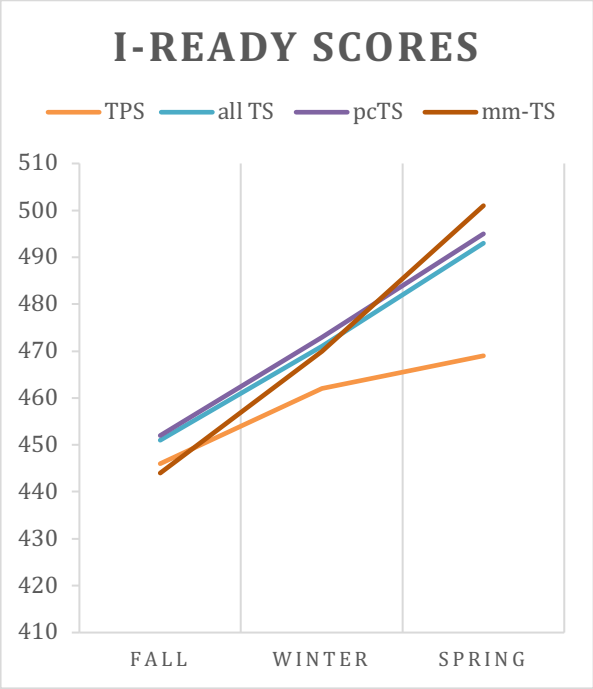
Fifty-four (54) 3rd-grade students were enrolled in the TutorSmart program. One hundred seventy (170) other 3rd-grade students on RIMPs from the same schools served as the comparison group. In Fall 2023, the two groups had similar i-Ready scores (mean 451 vs. 446, p=0.51) and similar Ohio’s State Test scores (mean 662 vs. 653, p=0.10) (Table 1). In Spring 2024, the TutorSmart students had significantly higher i-Ready scores compared to their TPS counterparts (mean 493 vs 469, p=0.02). i-Ready scores improved more in the TutorSmart students from Fall to Spring compared to other TPS students (mean change 40 vs. 25, p=0.005).

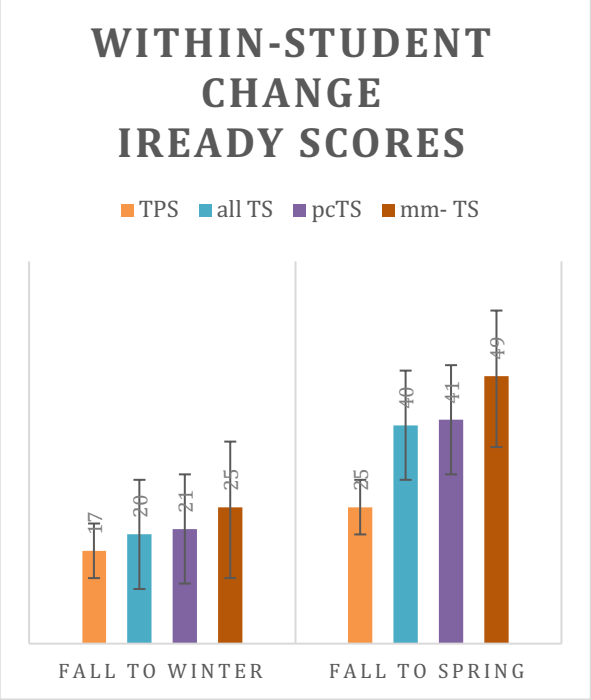




Program-Compliant and Meets-Model subgroup performance (table 2)

Students were defined as program-compliant (pcTS) if they either completed 30 hours of tutoring or attended two-thirds of the sessions offered. Dosage levels ranged from 10 to 76 hours, with a median of 43 hours. Thirty-two (32) students qualified as program-compliant. Engagement levels ranged from 33% to 100%, with a median of 72%. Again, thirty-two (32) students (not the same 32) had enough engagement to be considered 'program compliant.' All told, forty-three of the 54 students obtained at least one of these benchmarks and designated "program-compliant" (pcTS). The twenty-one students meeting both requirements were classified as the 'meets model' subgroup (mmTS). By the Spring IR testing, both subgroups outperformed the comparison (mean change TPS 25 v. pcTS 41 p=.003 v. mmTS 49 p=.001). The removal rate from a RIMP did not reach statistical significance for any subgroup, though the TS students fared better than the TPS comparison group (TPS 25% v. pcTS 35% v. mmTS 38%).





TPS: Toledo Public School students* not involved inTutorSmart
all TS: ANY TutorSmart participation
pcTS: program compliant (>30hrs dosage OR >66% engagement)
mmTS: meets model (>30hrs dosage AND >66% engagement)

**all students in the study are enrolled in TPS and on a Reading Improvement and Monitoring Plan*

Discussion

The efficacy of the TutorSmart high-dosage tutoring model extends beyond its primary focus on gap closure in reading and mathematics. While the program was not specifically designed to raise standardized test scores or to facilitate removal from Reading Improvement and Monitoring Plans (RIMPs), analysis of these secondary outcomes provides valuable insights into its broader educational impact.

To validate the TutorSmart model against external benchmarks, this study examined data from two standardized assessments: the Ohio State Test (OST) and i-Ready (iR). The OST serves as the State of Ohio's primary tool for measuring student achievement against state learning standards and facilitating school performance comparisons. The iR, a state-approved diagnostic assessment, offers more frequent evaluation opportunities and provides both normative and criterion-referenced data.

The study prioritized iR results due to its ability to measure individual student progress compared to the binary nature of RIMP removal status. At the study's inception, the 224 participating students demonstrated limited proficiency with a mean OST score of 655, significantly below the career readiness threshold of 725. The established grade-level benchmark of 700 on the iR assessment determined RIMP placement status.

Given that Ohio state law mandates tutoring for students on RIMPs, students within the control group (TPS) likely participated in alternative tutoring programs or other inventions. Rather than differentiating between specific interventions, these students were aggregated to represent a "business-as-usual" comparison group. Despite this conservative approach, TutorSmart participants demonstrated markedly superior academic gains.

The analysis encompassed all TutorSmart enrollees regardless of participation levels, reflecting real-world implementation conditions where student engagement varies due to external factors. Students who met both expectations (mmTS group) demonstrated nearly double the growth rate of the control group over the academic year. This group’s performance demonstrates the potential of the TS program. Notably, the larger pcTS group (42 students) achieved comparable results, indicating that even consistent, if not maximum, participation can yield significant benefits. These findings support the program's ability to maintain effectiveness even at scale. While improvements in OST scores and RIMP removal rates were less pronounced, both metrics showed positive trends favoring TutorSmart participants.

Limitations: Several methodological constraints warrant consideration when interpreting this study's findings. First, the statistical power was insufficient to detect certain subgroup differences with optimal precision, primarily due to the size of the study. Additionally, the study was performed within a single school district which may restrict the generalizability of results in dissimilar districts.

Selection bias represents another significant consideration. The TutorSmart program's intensive structure requires three 75-minute sessions over a semester of school which may be less accessible for students without strong family support systems. Despite the program's commitment to universal accessibility through teacher and principal referrals and the absence of financial barriers, the time commitment is substantial.

Conclusion

Compared to the control group, students participating in TutorSmart programming have test scores that are significantly improved. Students meeting the TutorSmart model demonstrated nearly twice the growth rate as those in the control group. Students who met either the engagement or dosage specifications also showed enhanced gains. This study demonstrates that the TutorSmart model is effective in improving standardized test scores.

Appendix 1: Tables

TABLE 1

Third Grade RIMP Students' Outcomes with and without TutorSmart

	TPS Students		**All TutorSmart Participants**		**p-value**
No. Students	170		54		
	Mean (SD)	**95% CI**	**Mean (SD)**	**95% CI**	
iReady					
Fall 2023	446 (50)	438 - 454	451 (47)	438 - 464	0.51
Winter 2024	462 (57)	453 - 471	471 (54)	456 - 485	0.33
Spring 2024	469 (64)	459 - 480	493 (53)	478 - 508	0.02
Within-student change from					
Fall to Winter	17 (31)	12 - 22	20 (25)	13 - 27	0.59
Fall to Spring	25 (29)	20 - 30	40 (33)	30 - 49	0.005
Ohio State Test					
Fall 2023	653 (35)	648 - 659	662 (33)	653 - 672	0.10
Spring 2024	673 (35)	667 - 679	676 (34)	667 - 686	0.54
Within-student change from					
Fall to Spring	20 (31)	14 - 25	14 (31)	5 - 23	0.27
Removed from RIMP, n (%)	43 (25%)		16 (30%)		

SD is standard deviation. CI is confidence interval.

P-values for test scores are from two-tailed Student's T-tests.

P-value for removed from RIMP is from a Chi-square test.

Removed from RIMP is defined as Ohio State test score of 700 or more.

TABLE 2

TPS versus TutorSmart Protocol Compliant and Meets Model

	TPS Students		**TS Protocol Compliant >30 hours dosage OR >66% engagement**		**TS Meets Model >30 hours dosage AND >66% engagement**	
No. Students	170		43		21	
	Mean (SD)	**95% CI**	**Mean (SD)**	**95% CI**	**Mean (SD)**	**95% CI**
iReady						
Fall 2023	446 (50)	438 - 454	452 (49)	436 – 468	444 (57)	418 – 471
Winter 2024	462 (57)	453 - 471	473 (58)	455 – 490	470 (66)	440 – 500
Spring 2024	469 (64)	459 - 480	495 (56)[a]	478 – 513	501 (55)[c]	475 – 526
Within student change from:						
Fall to Winter	17 (31)	12 - 22	21 (26)	12 – 29	25 (32)	10 – 40
Fall to Spring	25 (29)	20 - 30	41 (28)[b]	32 – 50	49 (26)[d]	36 – 62
Ohio State Test						
Fall 2023	653 (35)	648 - 659	663 (35)	652 – 673	656 (35)	640 – 672
Spring 2024	673 (35)	667 - 679	681 (35)	669 – 692	686 (36)	669 – 703
Within-student change from Fall to Spring	20 (31)	14 - 25	17 (31)	7 - 27	27 (35)	11 - 44
Removed from RIMP, n (%)	43 (25%)		15 (35%)		8 (38%)	

Two-tailed p-value: TPS versus TS Protocol Compliant: [a] p=0.013, [b] p=0.003
Two-tailed p-value: TPS versus TS Meets Model: [c] p=0.039, [d] p=0.001

References

Butt, R.D., Litten, J., McClellan, P., Ryley, W., & Thomas, K. (2017, May). *Framework for tutoring services: a technical report submitted to the Toledo Community Foundation*. https://www.toledocf.org/wp-content/uploads/2019/09/A-Framework-for-Tutoring-Services_May-2017.pdf

Robinson, Carly D., Kraft, Matthew A., Loeb, S., & Schueler, Beth E. (2021, February). *Accelerating student learning with high-dosage tutoring.* EdResearch for Recovery; Annenberg Institute at Brown University; University of Virginia. https://files.eric.ed.gov/fulltext/ED613847.pdf

Higgins, S., Kokotsaki, D., & Coe, R. (2012, July). *The teaching and learning toolkit.* Education Endowment Foundation; The Sutton Trust. http://uk.renaissance.com/wp-content/uploads/2012/10/teachingandlearningtoolkit-july2012.pdf

High-dosage tutoring FAQ. (2024). Ohio Department of Education & Workforce. https://education.ohio.gov/getattachment/Topics/Learning-in-Ohio/Literacy/Third-Grade-Reading-Guarantee/High-Dosage-Tutoring-and-TGRG-FAQ-10-22-2024.pdf.aspx?lang=en-US

High-impact tutoring: district playbook. National Student Support Accelerator, in collaboration with BlueEngine. https://studentsupportaccelerator.org/sites/default/files/High_Impact_Tutoring_District_Playbook.pdf

i-Ready: Assessment for the Purpose of Instruction. (2024). Curriculum Associates, LLC. https://ebooks.curriculumassociates.com/story/assessment-for-instruction/page/1

Kraft, Matthew A. (2018, August). Interpreting effect sizes of education interventions. *EdWorkingPaper, 19(10).* Annenberg Institute; Brown University. https://scholar.harvard.edu/sites/scholar.harvard.edu/files/mkraft/files/kraft_2019_effect_sizes.pdf

Nikow, A., Oreopolus, P., & Quan, V. The promise of tutoring through pre-K-12 learning: a systematic review and meta-analysis of the experimental evidence. *American Educational Research Journal, 61(1)*, 74-107. https://doi.org/10.3102/00028312231208687

Reading Improvement and Monitoring Plan template. Ohio Department of Education & Workforce. https://education.ohio.gov/getattachment/Topics/Learning-in-Ohio/Literacy/Third-Grade-Reading-Guarantee/RIMP-Template-K-5-REVISED.docx.aspx

Third Grade Reading Guarantee guidance. (2024, October). Ohio Department of Education & Workforce. https://education.ohio.gov/getattachment/Topics/Learning-in-Ohio/Literacy/Third-Grade-Reading-Guarantee/Third-Grade-Reading-Guarantee-Guidance-October-2024.pdf.aspx?lang=en-US

Third Grade Reading Guarantee reading diagnostic scoring information. (2024). Ohio Deptartment of Education & Workforce. https://education.ohio.gov/getattachment/Topics/Learning-in-Ohio/Literacy/Third-Grade-Reading-Guarantee/2024-2025-Reading-Diagnostic-Scoring-Information-1.pdf.aspx?lang=en-US

Understanding Ohio's State Tests score reports 2023-2024. Ohio Department of Education & Workforce. https://oh-ost.portal.cambiumast.com/content/contentresources/en/Understanding_State_Tests_Reports_2023-2024_Final.pdf