2022

Mark Miller Subaru/United Way One-to-One Tutoring Program in Mathematics: Evaluation Report for Academic Year 2021-2022

Prepared by the Utah Education Policy Center (UEPC)
July 2022

THE UNIVERSITY OF UTAH
UTAH EDUCATION POLICY CENTER

Bridging Research, Policy, and Practice

The Utah Education Policy Center (UEPC) is a an independent, non-partisan, not-for-profit research-based center at the University of Utah founded in the Department of Educational Leadership and Policy in 1990 and administered through the College of Education since 2007. The UEPC mission is to bridge research, policy, and practice in public schools and higher education to increase educational equity, excellence, access, and opportunities for all children and adults.

The UEPC informs, influences, and improves the quality of educational policies, practices, and leadership through research, evaluation, and technical assistance. Through our research, evaluation, and technical assistance, we are committed to supporting the understanding of whether educational policies, programs, and practices are being implemented as intended, whether they are effective and impactful, and how they may be improved and scaled-up, and become sustainable.

Please visit our website for more information about the UEPC.
http://uepc.utah.edu

Andrea K. Rorrer, Ph.D., Director
Phone: 801-581-4207
andrea.rorrer@utah.edu

Cori Groth, Ph.D., Associate Director
Phone: 801-581-4207
cori.groth@utah.edu

Ellen Altermatt, Ph.D., Assistant Director for Research and Evaluation
ellen.altermatt@utah.edu

Follow us on Twitter: @UtahUEPC

.

Acknowledgements

The Utah Education Policy Center (UEPC) thanks volunteers in the Mark Miller Subaru/United Way One-to-One Tutoring Program for sharing their experiences via tutor surveys. We are also grateful to the principals, teachers, staff, and students at Granite School District for their participation in and support of data collection efforts.

Citation: Altermatt, E. R. & Rorrer, A. K. (2022). *Mark Miller Subaru/United Way One-to-One Tutoring Program in Mathematics: Evaluation Report for Academic Year 2021-2022*. Salt Lake City, UT: Utah Education Policy Center.

Table of Contents

List of Tables

List of Figures

Introduction

Program Overview

The Mark Miller Subaru/United Way One-to-One Tutoring Program was designed to improve the mathematics outcomes of elementary-school-aged students by combining a tutoring intervention with a math personalized software intervention and, for some students, an incentive intervention.

In all, 74 tutors provided tutoring to students in two third-grade classrooms. Tutors were asked to volunteer two times per week for a total of 90 minutes per week. All tutoring occurred virtually via Zoom. For each session, tutors joined a Zoom meeting that included all students and the teacher and, after a brief time together, tutors were placed into a breakout room with a student. Students were asked to share their screens so that tutors could observe and provide support as they completed activities in i-Ready.

Before meeting with students, tutors participated in a virtual training session. During this session, tutors were provided logistical information (e.g., how to connect to the classroom via Zoom) and introduced to the i-Ready program. Tutors also received instruction on the importance of promoting positive attitudes among students (e.g., a "growth mindset") and building relationships with students.[1] During training, tutors were encouraged to watch an additional orientation video and to regularly communicate with teachers about students' progress.

Study Overview

The purpose of the current study is to evaluate the effectiveness of the One-to-One Tutoring Program. Participants in the study were third grade students, their tutors, and their teachers. Students were assigned to one of three "intervention" conditions or to a "comparison" condition during the 2021-2022 academic year. Descriptions of each condition are provided in Table 1.

Table 1. Descriptions of "intervention" and "comparison" study conditions

Intervention (School #1)	Software only (Classroom #1)	Students in this condition had access to i-Ready. Designed for K-8 students, i-Ready is a software program designed to provide personalized instruction and assessment to students.
	Software + tutoring (Classroom #2)	Students in this condition had access to i-Ready and to tutors who were trained to provide virtual support to students as they used program.
	Software + tutoring + incentives (Classroom #3)	Students in this condition had access to i-Ready and to tutors who were trained to provide virtual support to students as they used program. In addition, students had opportunities to earn incentives for meeting performance goals.
Comparison (School #2)	Comparison (Classroom #4)	Students in this condition did not have access to i-Ready, tutors, or incentives.

[1] A "growth mindset" is the belief that personality characteristics, including intellectual abilities, can be developed. In contrast, a "fixed mindset" is the belief that these characteristics are unchangeable. Growth mindsets are associated with myriad positive outcomes including a willingness to persist in the face of challenges (Yeager & Dweck, 2020).

As shown in Table 1, students who participated in the study attended one of two elementary schools in the Granite School District in Utah. In one of these schools, all three third-grade classrooms participated. One classroom was assigned to each of the three intervention conditions. In the second school, one of two third-grade classroom was assigned to the "comparison" condition. In both schools, math proficiency rates were quite low. For example, on statewide RISE mathematics tests administered in 2020-2021, fewer than 20% of third grade students scored at levels that indicated that they were "proficient" in mathematics. Students in the three intervention conditions used i-Ready software throughout the academic year. For students in the two tutoring conditions, tutoring began in late January 2022 and ended in May 2022.

To evaluate the program, the Utah Education Policy Center (UEPC) analyzed data from five sources. A description of each data source is provided in Table 2.

Table 2. Data sources

Source	Timing	Description
1. Acadience Math assessment scores	Administered at three time points: beginning of year, middle of year, and end of year.	Acadience Math is a standardized assessment designed to measure the acquisition of mathematics skills and to create benchmarks to identify students who may be at risk for mathematics difficulties.
2. i-Ready Diagnostic scores	Administered at three time points: fall, winter, and spring.	The i-Ready Diagnostic is an adaptive assessment used to identify student strengths and opportunities for growth in mathematics.
3. Tutor Survey	Administered in May 2022	All tutors were invited to complete the tutor survey to assess tutor perceptions of implementation and impact on student outcomes.
4. Teacher Survey	Administered in May 2022	The teacher survey was administered to teachers in the two tutoring conditions to assess teacher perceptions of program implementation and impact on student outcomes.
5. Student Survey	Administered in May 2022	All students in the intervention conditions were invited to complete the student survey to assess student attitudes toward math, i-Ready, and, where appropriate, tutors and incentives.

To ensure "apples-to-apples" comparisons, only students with achievement data at all three times points for a given assessment were included in analyses of achievement data. This resulted in sample sizes of 19, 21, 19, and 22 for Acadience Math and 19, 22, 19, and 22 for the i-Ready Diagnostic for the comparison, software only, software + tutoring, and software + tutoring + incentive conditions, respectively.

Report Organization

This report begins with a brief review of the literature on the promise of tutoring interventions and math personalized learning software interventions for improving students' mathematics achievement. This review is followed by a description of findings from each of the five data sources summarized in Table 2. The report ends with recommendations for future program and evaluation activities.

Literature Review

The Promise of Tutoring Programs

There is considerable evidence that tutoring programs can have large, positive effects on student learning outcomes (see Robinson, Kraft, Loeb, & Schueler, 2021, for a review). For example, a 2020 meta-analysis of 96 K-12 tutoring interventions in which students were randomly assigned to treatment or control conditions found consistent and positive effects on student learning outcomes as measured by standardized test scores (Nickow, Oreopoulos, & Quan, 2020). By one estimate, the effects of tutoring programs translate, on average, to between three and fifteen additional months of learning for students (Robinson et al., 2021). Although the mechanisms by which tutoring interventions contribute to learning are still being investigated, tutoring programs are frequently credited for providing students with opportunities to receive additional, focused, and customized instruction and for introducing alternative pedagogies (Nickow et al., 2020).

Tutoring programs appear to be especially effective when tutors receive adequate training and support, when the number of students paired with each tutor is small, when instruction is aligned with classroom learning, and when program implementation is informed by ongoing formative and summative assessments (National Student Support Accelerator, 2021; Nickow et al., 2020; Pellegrini, Neitzel, Lake, & Slavin, 2021; Robinson et al., 2021). Outcomes also appear stronger when tutors are consistently paired with the same students so that strong mentor-like relationships can be built (Robinson et al., 2021). Research indicates that timing of tutoring programs is important as well. For instance, tutoring programs that take place during the school day or right after school and are held at the school building appear to be more effective. In contrast, programs disconnected from school sites and hours suffer from lower attendance rates and a less academic culture (Robinson et al., 2021).

The Promise of Personalized Learning Software

There is growing evidence that personalized learning software can also contribute to positive achievement outcomes, including positive impacts on learning and attitudes in mathematics for K-12 students. For example, personalized learning software use has been associated with heightened student engagement in the learning process, improved teacher-student interactions, enhanced higher-order problem solving techniques, and increased student math achievement (see Cheung & Slavin, 2013, Hillmayr, Ziernwald, Reinhold, Hofer, & Reiss, 2020, Ma, Adesope, Nesbit, & Liu, 2014, Steenbergen-Hu & Cooper, 2013, and Young, Gorumek, & Hamilton, 2018, for meta-analytic reviews). Researchers caution, however, that educational technology does not inevitably or independently produce these outcomes.

Math personalized learning software appears to be especially effective when student software usage levels are relatively high (Altermatt, Altermatt, Rorrer, & Moore, 2002; Owens, Rorrer, Ni, Onuma, Pecsok, & Moore, 2020; Su, Rorrer, Owens, Pecsok, Moore, & Ni, 2020), when educators have strong technological and pedagogical content knowledge (Koehler & Mishra, 2009), and when educators are provided with sufficient training and support to utilize educational technology with fidelity and to align technology use with other types of instruction to create strong blended learning environments (Pane, Steiner, Baird, & Hamilton, 2015; Pane Steiner, Baird, Hamilton, & Pane, 2017; REL Mid-Atlantic, 2017; Sarker, Wu, Cao, Alam, & Li, 2019). Because tutoring and mentoring programs often employ volunteers or paraprofessionals, training may be especially important for these individuals compared to experienced teachers (National Student Support Accelerator, 2021).

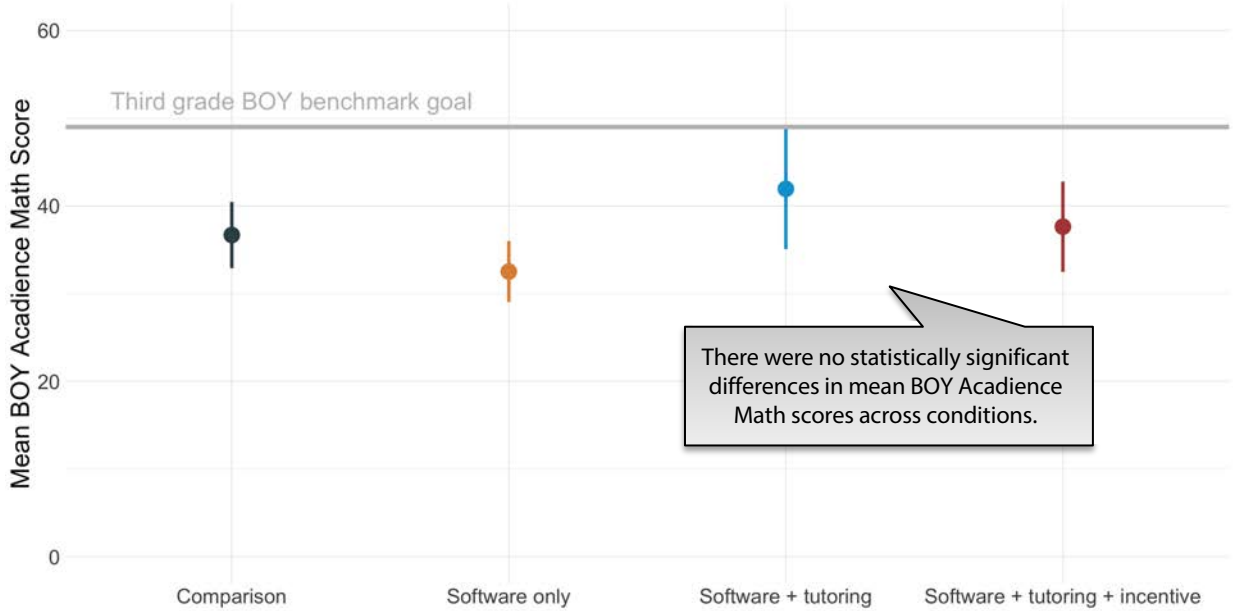
1 | Findings from Acadience and i-Ready Assessments

Mathematics Achievement at Baseline by Condition

Before examining the impact of the Mark Miller Subaru/United Way One-to-One Tutoring Program on students' mathematics achievement, we examined students' Acadience Math scores at the beginning of the school year (BOY) and middle of the school year (MOY) to ensure that students in the four conditions were relatively similar at baseline (i.e., before the tutoring program began). If substantial differences exist at baseline, it can be difficult to determine whether any differences in outcomes are due to the intervention(s) or to differences that existed before the intervention(s) began. For 2021-2022, the testing window for BOY Acadience Math was the beginning of the school year to September 30 and the testing window for MOY Acadience Math was December 1 to January 31. As a result, both assessments were completed or nearly completed before the tutoring program began in late January 2022.

As shown in Figure 1, BOY Acadience Math scores were similar across the four conditions, with mean scores ranging from 32.52 to 41.98. A one-way analysis of variance confirmed that there were no statistically significant differences in BOY scores across conditions, $F(3, 77) = 0.60$, $p > .05$.[2] As indicated by the horizontal line, students in each condition were, as a group, "below benchmark" at the beginning of the school year. The BOY benchmark goal for third grade students is 49.

Figure 1. Mean beginning of school year (BOY) Acadience Math scores by condition

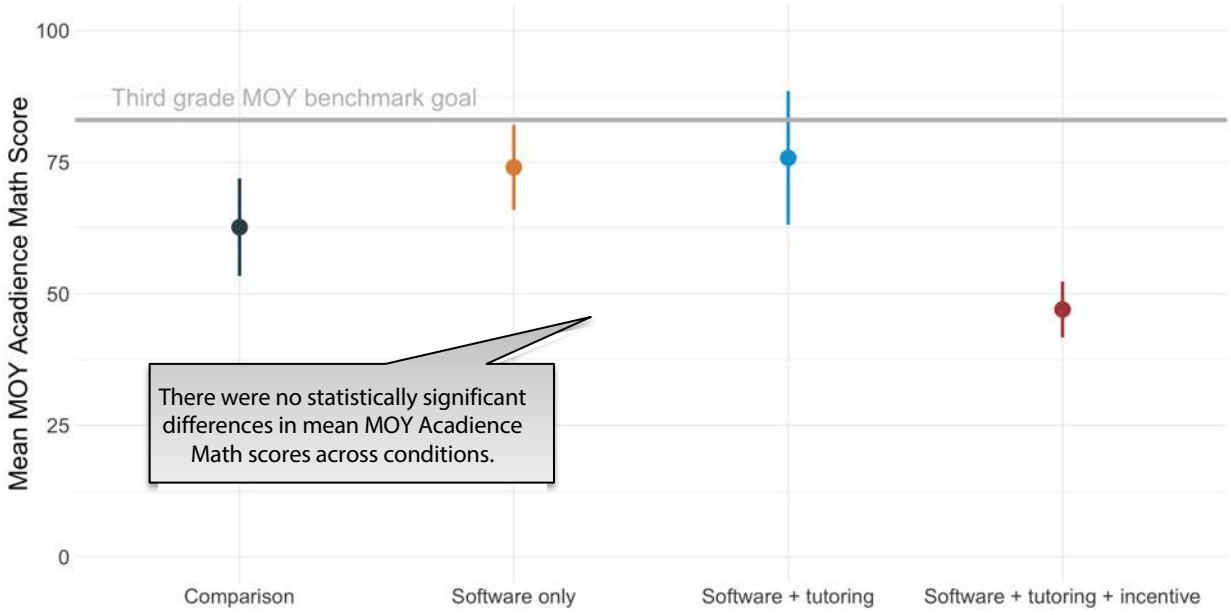


Note. Dots represent means. Bars represent the standard error of the mean.[3]

[2] A statistically significant result is one that is unlikely to occur by chance. When *p* values are less than .05, there is less than a 5% chance of finding a difference this large or larger if the null hypothesis is true. Here, the null hypothesis is that there are no differences in mean math scores across conditions. When *p* values are greater than .05, this chance is greater than 5% which is traditionally considered unacceptable for rejecting the null hypothesis.
[3] The standard error of the mean is a measure of how far the sample mean of the data is likely to be from the population mean.

As shown in Figure 2, MOY Acadience Math scores were also similar across the four conditions, with mean scores ranging from 47.00 to 75.79. A one-way analysis of variance confirmed that there were no statistically significant differences in MOY scores across conditions, $F(3, 77) = 2.28$, $p > .05$. As indicated by the horizontal line, students in each condition were, as a group, still "below benchmark" at the middle of the school year. The MOY benchmark goal for third grade students is 83.

Figure 2. Mean middle of school year (MOY) Acadience Math scores by condition



Note. Dots represent means. Bars represent the standard error of the mean.

Gains in Acadience Math Scores by Condition

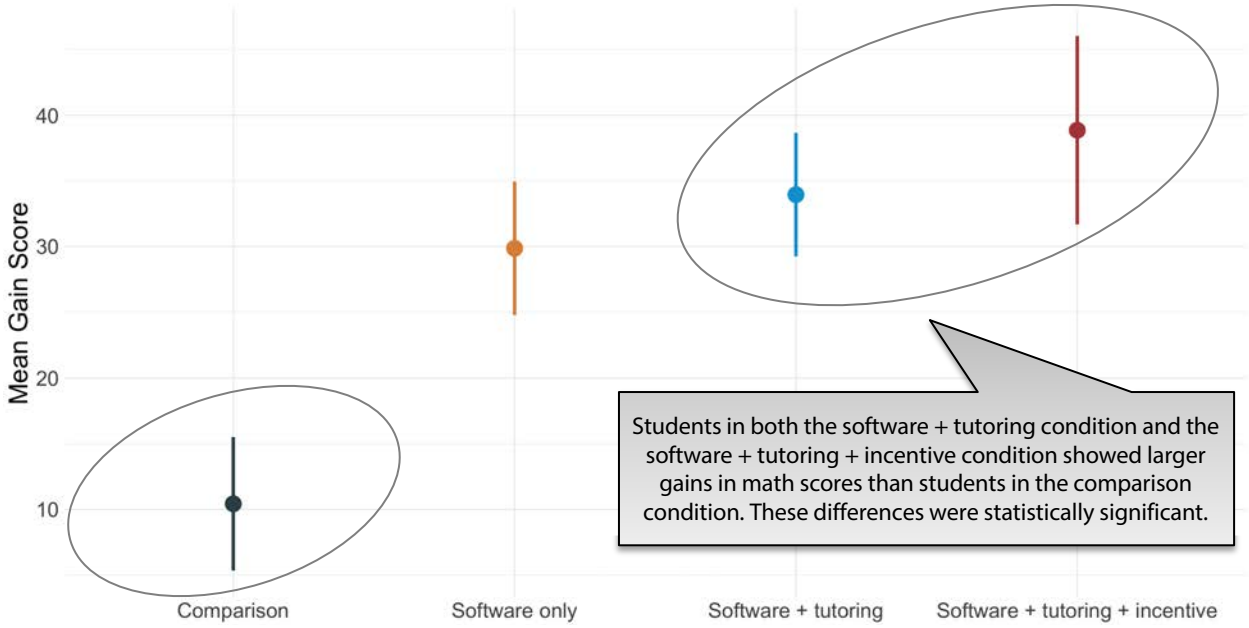
To explore the impact the tutoring program may have had on gains in performance on the Acadience Math test during the time period that tutors were present in classrooms (i.e., January to May 2022), we took two approaches. First, we conducted a one-way analysis of variance (ANOVA) to examine whether gains in Acadience Math scores from MOY to EOY differed by condition. Second, we conducted a repeated measures analysis of covariance (ANCOVA) to determine whether changes in scores across time (i.e., from MOY to EOY) differed by condition, controlling for Acadience Math scores at BOY. [4]

We calculated **gain scores** by subtracting MOY math scores from EOY math scores such that higher gain scores indicate greater increases in Acadience Math scores from the middle of the school year – that is, just as the tutors were beginning to work with students – to the end of the school year. As shown in Figure 3, there were differences in gain scores across conditions, with mean gain scores ranging from 10.42 to 38.86. A one-way ANOVA confirmed that these differences were statistically significant, $F(3, 77) = 4.60$, $p < .01$. Post-hoc Tukey tests were used to examine for which conditions differences were statistically significant. These analyses revealed that students in both tutoring conditions showed larger gains in Acadience Math scores than students in the comparison condition,

[4] See https://homes.ori.org/keiths/Tips/Stats_GainScores.html for a discussion of the strengths and weaknesses of these two approaches.

*p*s < .05. The difference in gain scores between students in the software only condition and the comparison condition was not statistically significant, $p > .05$. Likewise, the difference in gain scores between students in the two tutoring conditions was not statistically significant, $p > .05$.

Figure 3. Mean gains in Acadience math scores by condition

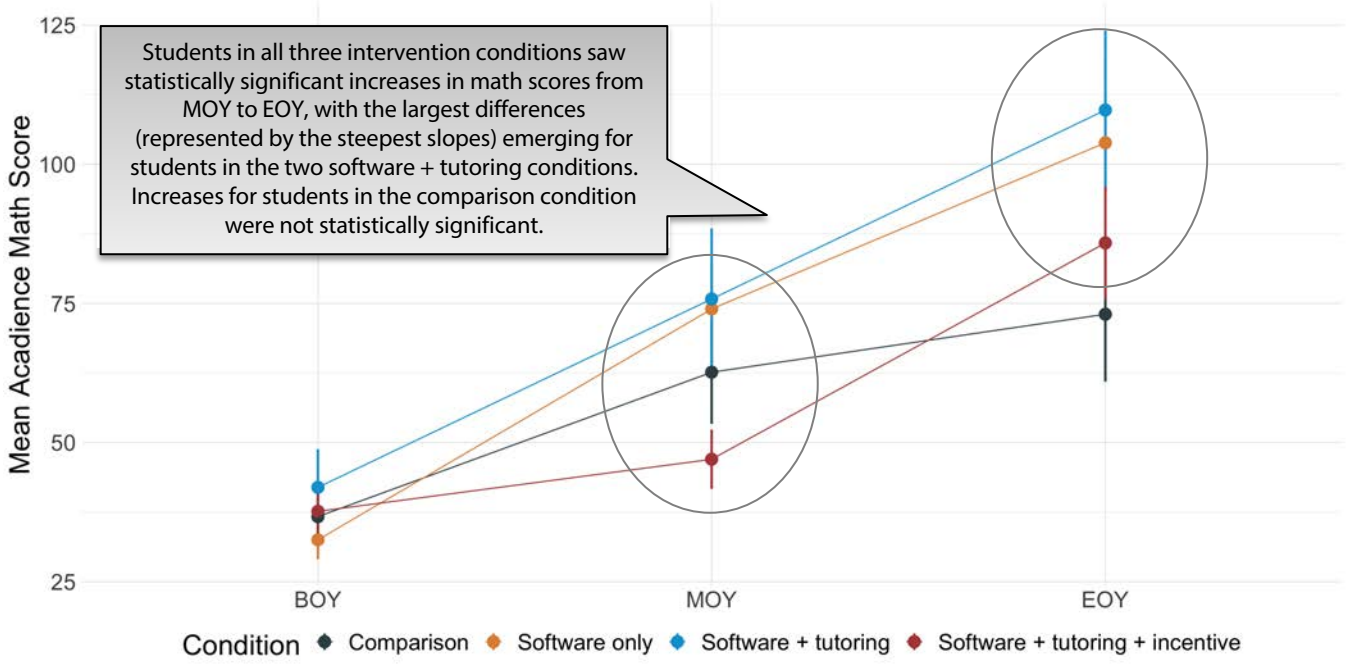


Note. Dots represent means. Bars represent the standard error of the mean. Ovals highlight statistically significant differences in means by condition.

These findings are further supported by the results of a **repeated measures ANCOVA**. For this analysis, the outcome measure was Acadience Math scores. Time (MOY, EOY) was the within-subjects factor, Condition was the between-subjects factor, and BOY math scores were included as a covariate to control for any pre-existing differences in Acadience Math scores across conditions. These analyses indicated that there was a statistically significant effect of Time, $F(1, 76) = 5.82$, $p < .05$, that was moderated by an interaction between Time and Condition, $F(3, 76) = 4.99$, $p < .01$. These findings indicate that the sample as a whole experienced increases in math scores from MOY to EOY *and* the size of the increase differed across conditions. Follow-up dependent *t*-tests indicated that students in the comparison condition did not experience statistically significant increases in Acadience Math scores from MOY to EOY, $t(18) = 2.05$, $p > .05$ However, statistically significant increases emerged for students in each of the three intervention conditions, with the largest increases emerging for students in the two tutoring conditions, *t*s > 5.40, *p*s < .001.

The full set of findings by Time and Condition are reflected graphically in Figure 4. In interpreting this figure, it is important to keep in mind that differences in mean math scores at BOY and MOY are not statistically significant. However, the *increase* in scores from MOY to EOY are significant for students in every condition except the comparison condition. The slope of the lines from MOY to EOY are reflective of the size of the increase with the steepest slopes emerging for students in the two tutoring conditions. Increases in MOY to EOY Acadience Math scores were especially strong for students in the software + tutoring + incentive condition where 31.8% of students moved from "below benchmark" at MOY to "at or above benchmark" at EOY. In comparison, 0% of students in the comparison condition moved from "below benchmark" at MOY to "at or above benchmark" at EOY.

Figure 4. Mean Acadience math scores by time and condition



Note. Dots represent means. Bars represent the standard error of the mean. Ovals highlight statistically significant differences in means by condition.

Gains in i-Ready Diagnostic scores by Condition

i-Ready Diagnostic tests were administered to students in all three intervention conditions in the fall (i.e., in October/November 2021) and, again, in the spring (i.e., in May 2022). Diagnostic tests were also administerd to these students in the winter (i.e., in January/Feburary 2022). However, these mid-year scores could not be used in analyses as program personnel noted that some students in the tutoring conditions received assistance on winter assessments from tutors. Diagnostic tests were also administered to students in the comparison condition, but only in winter and spring. These scores could also not be used in analyses as they are not comparable to the scores available for students in the three intervention conditions (where fall and spring scores were used in analyses).

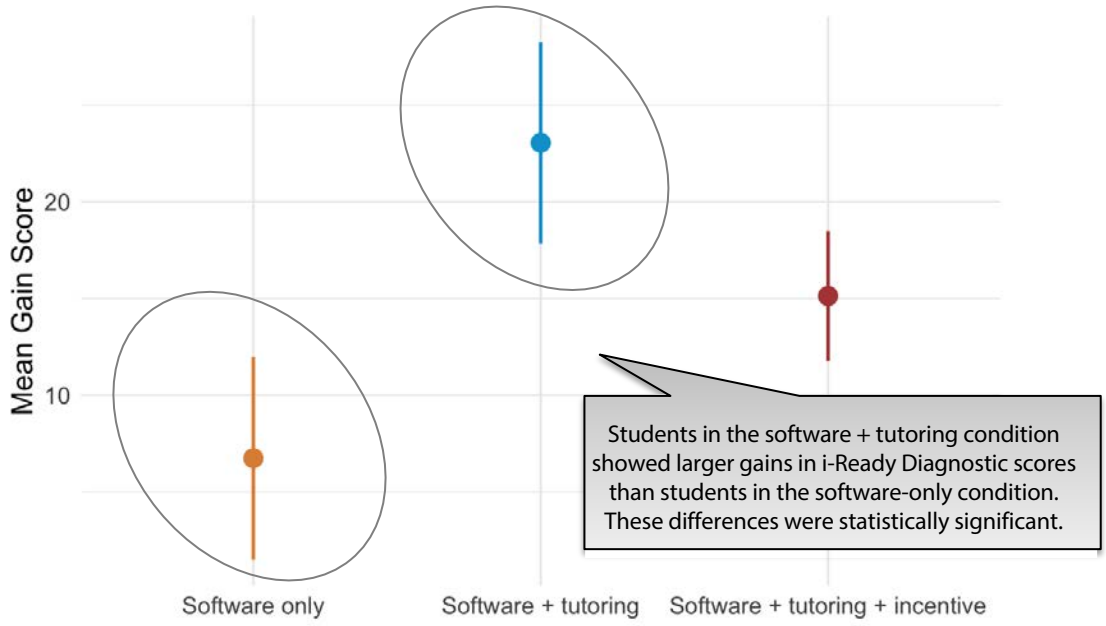
Our final set of analyses of student achievement data focused, then, on comparing fall and spring i-Ready Diagnostic test scores for students in the three intervention conditions. Given that analyses of gain scores are indistinguishable from a repeated measures ANOVA when there are only two assessments per individual, an analysis of **gain scores** was conducted.[5] Gain scores were calculated by subtracting fall scores from spring scores such that higher gain scores indicate greater increases in i-Ready Diagnostic scores from early in the school year to the end of the school year.

As show in Figure 5, there were differences in gain scores across conditions, with mean gain scores on i-Ready Diagnostic assessments ranging from 6.74 to 23.05. A preliminary one-way ANOVA indicated that there were no statistically significant differences in gain scores across conditions, $F(2, 58) = 1.99$, $p > .05$. However, because of small *n*s resulting in low power to detect differences, post-hoc Tukey tests were used to examine whether there were *any* statistically significant differences in

[5] See https://homes.ori.org/keiths/Tips/Stats_GainScores.html .

pairwise comparisons by condition. These analyses revealed one statistically significant difference: students in the software + tutoring condition showed stronger gains in i-Ready Diagnostic scores than students in the software only condition, $p < .05$.

Figure 5. Mean gain scores on i-Ready Diagnostic tests by condition



Note. Dots represent means. Bars represent the standard error of the mean. Ovals highlight statistically significant differences in means by condition.

Caveats and Next Steps

The results of analyses of Acadience Math and i-Ready Diagnostic data provide preliminary, promising evidence for the effectiveness of the One-to-One Tutoring Program in improving students' achievement outcomes. However, additional research is needed to meet more rigorous criteria for establishing program effectiveness.

In 2015, the Congress of the United States passed the *Every Student Succeeds Act* (ESSA). As part of this legislation, ESSA established four **Tiers of Evidence** for research studies on educational interventions. The current study meets standards for Tier 3: Promising Evidence as it is a correlational study with statistical controls for selection bias and demonstrates a statistically significant positive effect on one or more relevant outcomes. To meet standards for the next highest levels – Tier 2: Moderate Evidence or Tier 1: Strong Evidence – future studies would benefit from utilizing a more rigorous design. To meet Tier 2 standards, a quasi-experimental design would be required in which classroom/teacher and condition are not confounded[6] and in which controls for additional confounds are included (e.g., differences in teacher qualifications). Tier 2 standards also require that the study includes more than 350 students, and that the intervention is implemented in more than one district or school.[7] To meet Tier 1 standards, a randomized control trial (RCT) would

[6] In the current study, each condition is associated with a single teacher and classroom. As a result, it is impossible to distinguish the effect of the intervention(s) from the effect of the teacher or class. To address this "$n = 1$" confound, the What Works Clearinghouse recommends that at least two units (e.g., two teachers or classrooms) are included in each condition.

[7] See https://ies.ed.gov/ncee/edlabs/regions/midwest/pdf/blogs/RELMW-ESSA-Tiers-Video-Handout-508.pdf

be required with low levels of attrition. Tier 1 standards, like Tier 2 standards, require the participation of 350 students in at least two educational sites.

To maximize the potential of the program to meet more rigorous ESSA standards for evidence of effectiveness, it is important to gain additional insights into how the program was implemented in 2021-2022 and how tutors, teachers, and students perceived the program. To support these efforts, the UEPC administered surveys to each group of stakeholders in May 2022. The next three sections of the report provide a description of survey findings from each group.

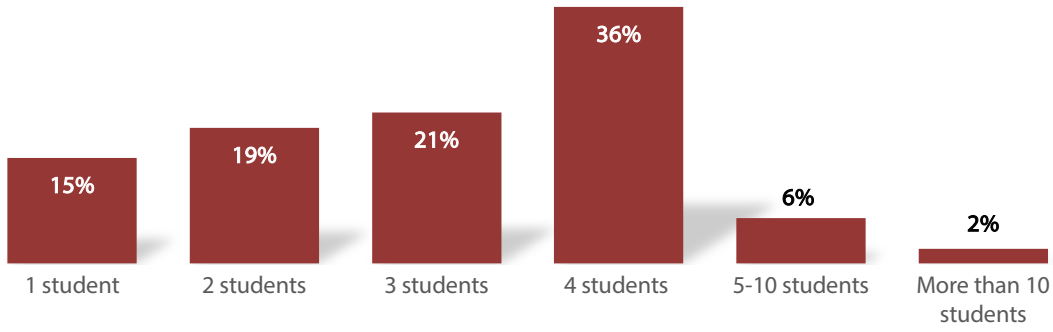
2 | Findings from Tutor Surveys

In all, 47 tutors completed the tutor survey. This represents a 63.6% response rate. The majority of respondents indicated that they were White (87.2%) and that they were employed full-time (61.7%) or retired (21.3%).

Implementation

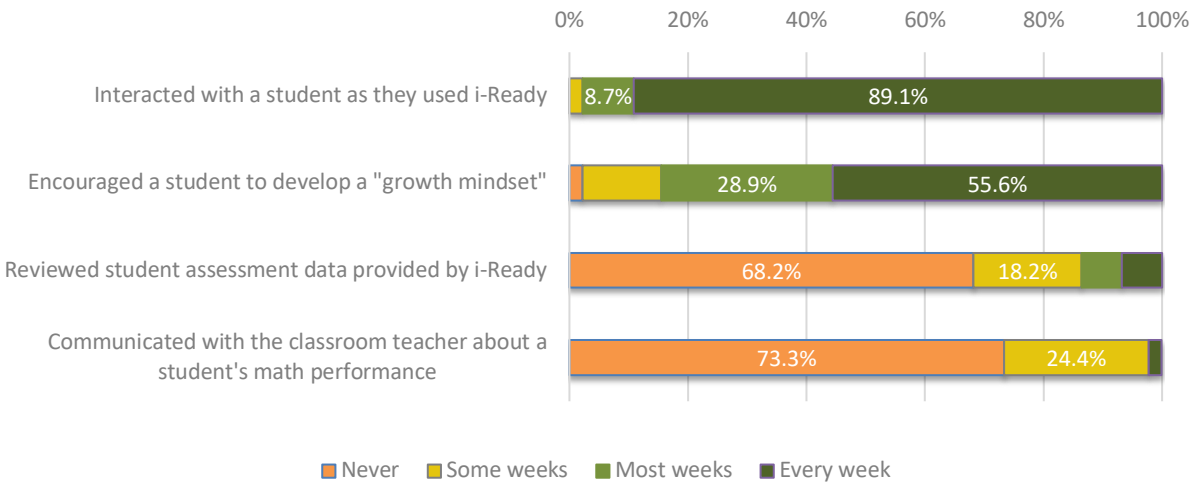
Tutors were asked to report on the number of hours they tutored per week and the number of different students they tutored from January 2022 to May 2022. Most tutors (n = 42, 89.4%) reported spending one to two hours tutoring per week on average. A small number of tutors reported tutoring for less than one hour per week (n = 4, 8.5%) or more than two hours per week (n = 1, 2.1%). The number of different students with whom tutors worked varied considerably from tutor to tutor. As shown in Figure 6, although most tutors (91.5%) reported working with four of fewer students during the tutoring initiative, 8.5% worked with five or more students.

Figure 6. Number of different students with whom tutors worked from January 2022 to May 2022



Tutors were also asked to indicate how often they engaged in four specific activities: 1) interacted with a student as they used i-Ready, 2) encouraged a student to develop a "growth mindset," 3) reviewed student assessment data provided by i-Ready, and 4) communicated with the classroom teacher about a student's performance. As shown in Figure 7, consistent with program objectives, most tutors reported that they interacted with students as they used i-Ready and encouraged students to develop a growth mindset "most weeks" or "every week." In contrast, most tutors indicated that they reviewed student assessment data or communicated with the classroom teacher about students' math performance "never" or only "some weeks."

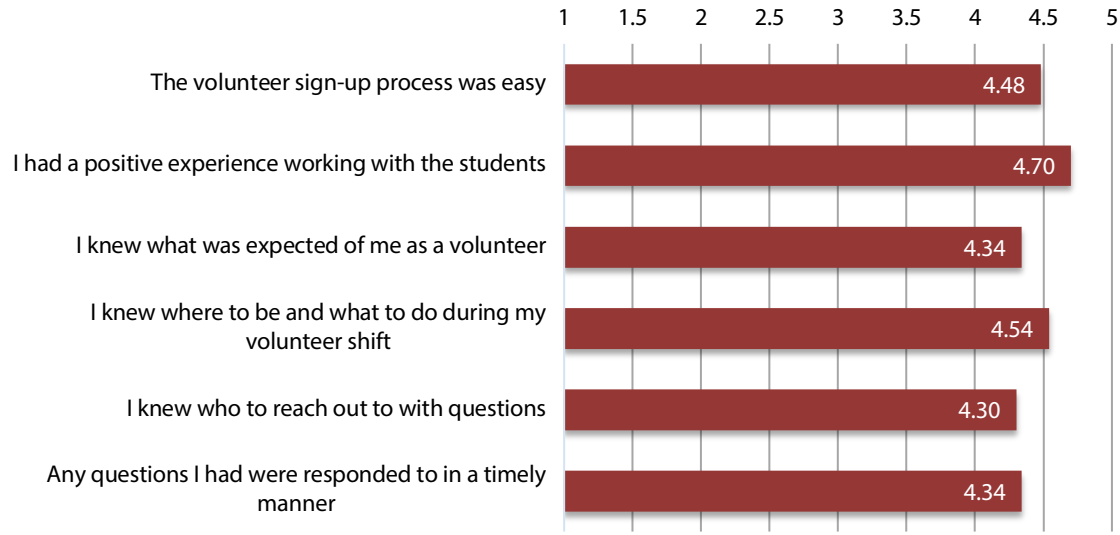
Figure 7. Self-reported frequency of engagement in four specific tutoring-related activities



Tutor Attitudes Toward the Tutoring Program

Tutors were asked to rate six items designed to assess tutor attitudes toward the tutoring program on a five-point scale ranging from 1 ("strongly disagree") to 5 ("strongly agree"). As shown in Figure 8, ratings were all above the midpoint of the scale indicating strong, positive attitudes among tutors about working with students and the level of volunteer support.

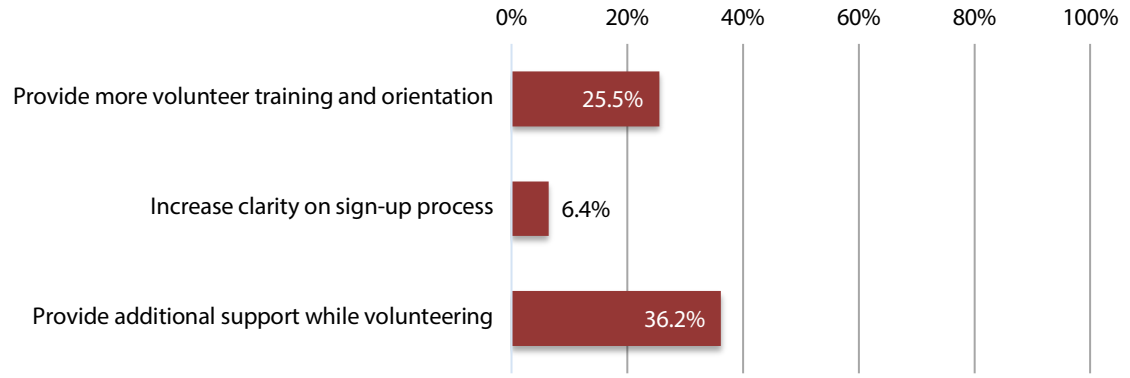
Figure 8. Tutor ratings of items tapping tutor attitudes toward tutoring program



These findings are consistent with tutors' general perceptions of the program. When asked to rate their "overall volunteer experience," 95.5% of tutors indicated that the experience was "good" or "awesome." When asked how likely they would be to participate as a tutor in the future, 73.6% indicated that they would be "very likely" or "extremely likely" to do so. When asked how likely they would be to recommend serving as a tutor to a colleague, friend, or family member, 81.4% indicated that they would be "very likely" or "extremely likely" to do so.

Still, some responses from tutors indicated areas for improvement. As shown in Figure 9, more than 1/3 of respondents indicated that the program might be improved by providing additional support to tutors during their time volunteering and 1/4 indicated that additional volunteer training and orientation would improve the program

Figure 9. Percentage of tutors who indicated that the program might be improved through additional training, clarity, or support



In response to open-ended questions inviting tutors to indicate how the program might be improved to increase impact on students, tutors offered myriad suggestions. Six themes emerged.

1. Tutors indicated that they would benefit from additional training on effective mathematics tutoring practices. For example:

   "[It would be helpful to have] more communication with tutors re: growing as a tutor, best practices in tutoring pairs, etc."

   "Some strategies for tutors to know how to get students to interact would be helpful."

   "I don't have any background in the math they're doing, so I was learning along with them. Some prep for how they're being taught to do math would be great!"

2. Tutors indicated that they would benefit from additional opportunities to access and share information on student performance, progress, and activities. For example:

   "I would have loved to see diagnostic or progress information on the students I worked with to know how to better support."

   "I do not know how to look at their progress. Maybe it could be emailed to us …"

   "[It would be helpful] to communicate with teachers about tutor observations … or [to have] teachers giving tutors information on where students might need additional support."

   "I was never quite sure how this program fit into the rest of what they were doing in Mathematics. Did they work on the same program everyday by themselves?"

3. Tutors indicated that they would benefit from more consistent tutor-student pairings. For example:

> "The more consistent you can be with the students assigned to volunteers, the better. I would assume it would increase student performance, but also I think it would increase volunteer retention."
>
> "I think it would be best to match up student/tutor and keep that the same throughout the timeframe for tutoring."
>
> "It is good to assign one student to one tutor most of the time... but changing it up now and then is good too so that tutors can see the range of student achievement."

4. Tutors indicated that in-person or hybrid sessions might be more effective. For example:

> "My student seemed to be driven to distraction by the other students moving around and talking. I believe that, at least for my student, it would have been better for me to be physically in the classroom with the student than being on zoom."
>
> "In person opportunities might create a better impact. It was also difficult to often have the student focus as they were closely surrounded by many others in the classroom."
>
> "The virtual world limited the personal interaction with the students. It was difficult to talk to them when competing with i-Ready in the background."
>
> "I think my student struggles with focusing for long periods of time. With it being virtual, I felt like it was easier for him to get distracted and took a while to get refocused."
>
> "While I think zoom worked well for most, some need one-on-one in person. I also wonder if we were assigned to a student and had the opportunity to start with a session in person, we might develop a better relationship."
>
> "Some topics like fractions would do better if there was i-Ready usage mixed with working with the student in person to visually understand fractions."

5. Tutors indicated that technology challenges sometimes made tutoring difficult. For example:

> "Having the technical difficulties the first few weeks discouraged some tutors and they left the program."
>
> "We had tech issues."
>
> "Decrease the problems technology. Easier said than done."

6. Tutors expressed concerns that the software was not an effective tool for providing an engaging, effective personalized learning experience for some students and expressed interest in finding solutions (e.g., by overriding the software program to assign more appropriate content). For example:

> "Some of my students seemed stuck in a loop and not advancing, despite what appears to me as mastery of the concept."

"i-Ready is not adaptable for students who only speak Spanish. Incorporating Spanish translation would be really helpful. I'm not sure how much value I added to the students. Sometimes I felt like I was mostly helping them stay on task and not doing a lot to build relationships or actually tutor them."

"I was paired with 13 different students. I was able to see the huge deviation of student use of the program. Students who had a lower diagnostic score and were stuck in the lower lessons were bored easily and unmotivated. The animations took a while to load and seemed geared for a younger audience. There was lag in entering answers and with click/drag/drop activities. There is also a lot more repetition in the lower modules. They might get 100% on a quiz, only to spend a couple more modules working on the same thing. This may be fine for grades K-2, but for these 3rd graders working through these modules, I think it was having the opposite of the intended impact."

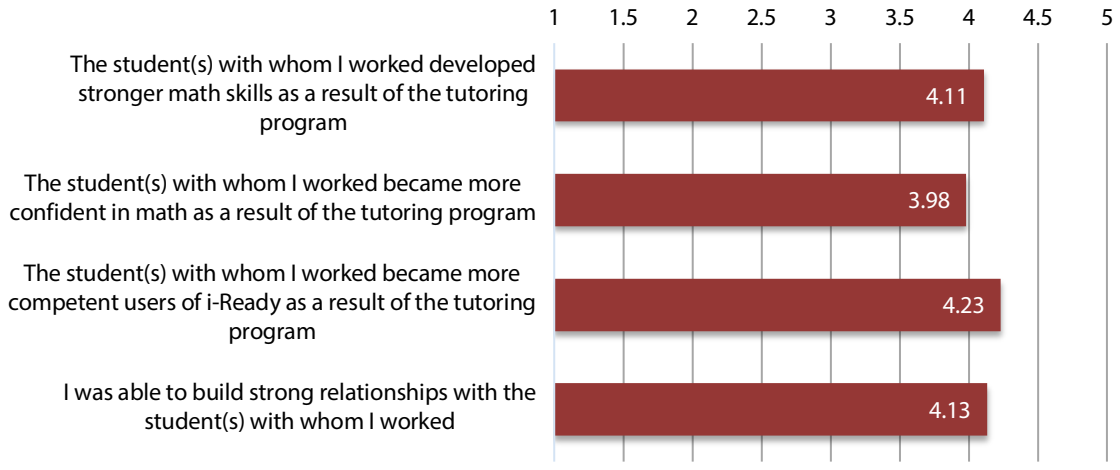
"At the end of the school year, there were 3rd graders still working on subtracting small amounts on a number line (e.g., 13 - 4). Today my student was working on this module while also completing his multiplication tables on paper, for an ice cream sundae activity later in the day. He was very unmotivated and bored with i-Ready, but clearly doing math outside of i-Ready that was more advanced than where he was working on in the modules."

"My student was very good at math, and it seemed he was outpacing the i-Ready program. Maybe a little more advanced sessions, particularly early on, would have made the time more valuable."

Tutor Perceptions of Student Outcomes

Tutors were asked to rate four items designed to assess tutor perceptions of student outcomes on a five-point scale ranging from 1 ("strongly disagree") to 5 ("strongly agree"). As shown in Figure 10, ratings were all above the midpoint of the scale indicating that tutors perceived that they were able to form strong relationships with students, and that the tutoring program had a strong positive impact on students' math skills, confidence in math, and competence in using the i-Ready program.

Figure 10. Tutor ratings of items tapping tutor perceptions of student outcomes



In response to open-ended questions inviting tutors to indicate what aspects of the program had the most impact on students, tutors, as a group, suggested that three program elements were especially impactful.

1. Tutors indicated that consistent tutor-student pairings were critical to program success insofar as consistency allowed tutors to track student progress and build strong relationships. For example:

   "I have tutored for United Way before, in math, online, at an elementary school. During that time, I was assigned different students each week, and so was unable to form a connection very well with a student. This time, I was assigned the same student each session, which worked much better. He was kind and expressed his gratitude to me for working with him. I was able to help him with his math skills due to the relationship we built."

   "Having that 1:1 with the same student allowed me to connect with my student on a more personal level, creating a friendship that allowed commonality to drive our sessions. Because of that 1:1, I saw my student being impacted not only by the math skills she learned but by the relationship we have."

   "Meeting with the same student week after week had a noticeable impact. The weeks I had a different student were a bigger struggle."

   "I think being with the same student the whole time was so great. We got to know each other really well and it was nice to see her get more confident."

   "Consistency in working with the same student was very helpful for both volunteer and student. Most weeks, I had the same student and was able to build a relationship and understand where he needed help."

2. Tutors indicated that the pairing of a software intervention (i.e., the i-Ready program) and a tutoring intervention was critical to program success. For example:

   "Some students probably would do great with i-Ready on their own, but I think having someone there to help redirect them when they get distracted/unmotivated is the thing that really helps. They also get someone to celebrate their wins or encourage them when they're struggling."

   "i-Ready is engaging. Holds students' attention. Having a tutor seems helpful, mostly to keep them concentrating and to bring them back from distractions."

   "I do think having a live person connect with the student, even if virtually, did help the student feel more engaged with i-Ready."

3. Tutors indicated that a positive, growth-mindset-oriented, relationship-centered approach was critical to program success. For example:

   "[It was important to] develop a relationship with the student on a weekly basis. [I was] able to share my experience, strength and hope with the student regarding my difficulties with math and how I overcame them. Also, communicating to the student how important math is to learn, that making an honest effort is all that you can do and not to be so focused on the result."

“I liked having a check-in question at the beginning of each tutoring session so I could get to know my student and they could get to know me. This helped us develop a relationship.”

“Building rapport with students with whom I worked created a sense of belonging for students … Tutoring time became an opportunity to chat and share about their school day.”

“I like the 1:1 time to watch the student's progress through materials, but I especially enjoyed cheering them on and being available to hear the bits of information about themselves they liked to share with me (e.g., birthday celebration, magic trick, how their day is going, what they're looking forward to in the rest of the day).”

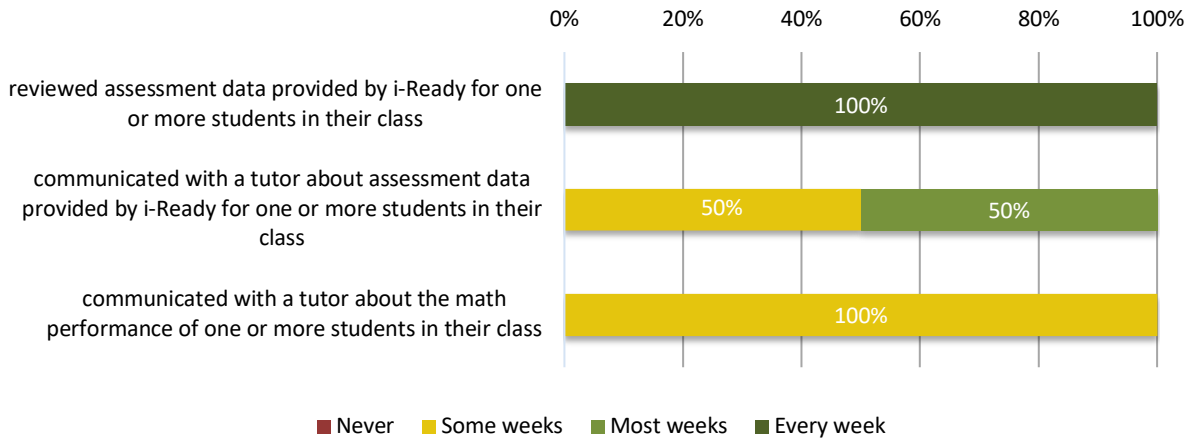
3 | Findings from Teacher Surveys

Both of the teachers whose students were paired with tutors completed the teacher survey. These data were collected anonymously (e.g., teachers were not asked to provide their names).

Implementation

Consistent with reports from tutors, both teachers reported that the average student in their classroom spent one to two hours per week working with a tutor from January 2022 to May 2022. Teachers were also asked to indicate how often they engaged in three specific activities: 1) reviewed assessment data provided by i-Ready for one or more students in their class, 2) communicated with a tutor about assessment data provided by i-Ready for one or more students in their class, and 3) communicated with a tutor about the math performance of one or more students in their class. As shown in Figure 11, both teachers reported that they reviewed assessment data provided by i-Ready "every week." While one teacher indicated that the communicated with a tutor about assessment data provided by i-Ready "some weeks," the other teacher indicated doing so "most weeks." Both teachers indicated that they communicated with a tutor about students' math performance more generally only "some weeks."

Figure 11. Self-reported frequency of engagement in three specific tutoring-related activities



In response to an open-ended question inviting teachers to explain any of their ratings, both teachers indicated that they would have been more likely to view communicating with tutors as a useful pedagogical strategy if there were more consistency in which tutors were paired with which students. Specifically:

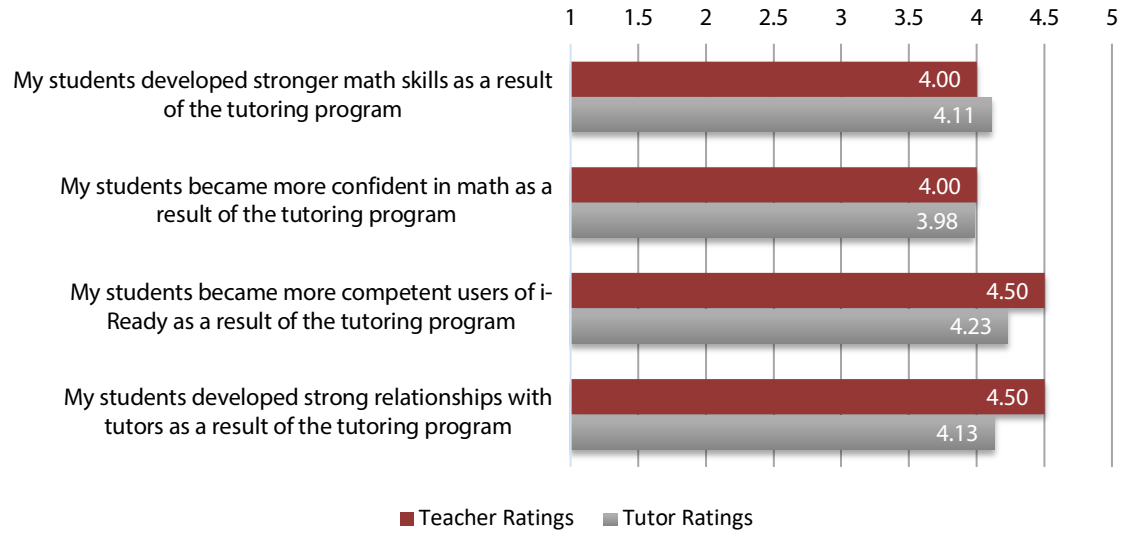
> "Not all the tutors were consistent. It would have been helpful to have a way to communicate to tutors regarding individual needs. Although I would chat with tutors from time to time, it was mostly about technological issues and not students' needs."

> "Sometimes the tutors weren't consistent enough to share data with them to make a difference."

Teacher Perceptions of Student Outcomes

Teachers were asked to rate four items designed to assess their perceptions of student outcomes on a five-point scale ranging from 1 ("strongly disagree") to 5 ("strongly agree"). As shown in Figure 12, ratings were all above the midpoint of the scale indicating that teachers perceived that students were able to form strong relationships with tutors, and that the tutoring program had a strong positive impact on students' math skills, confidence in math, and competence in using the i-Ready program. Tutor ratings (see Figure 10, above) of similarly worded items are included here for comparison.

Figure 12. Teacher (and tutor) ratings of items tapping teacher (and tutor) perceptions of student outcomes



In response to an open-ended question inviting teachers to explain any of their ratings, one teacher noted that "some tutors were very engaging and went out of their ways to connect with a student."

Teacher Attitudes Toward the Tutoring Program

Teachers were asked to rate how valuable the tutoring program was for their students and for them on a scale ranging from "not valuable" to "highly valuable". Both teachers indicated that the program was "moderately valuable" for students. One teacher indicated that the program was also "moderately valuable" for them as a teacher while the second teacher indicated that the program the value of the program was "neutral" for them as a teacher. One teacher noted that "having an adult to speak to when students ran into a stumbling block was helpful."

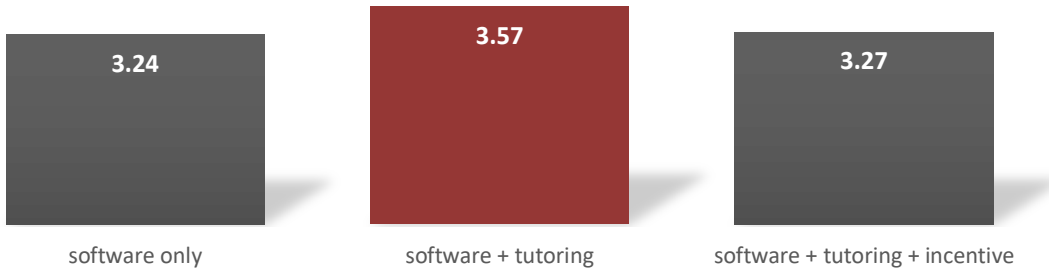
4 | Findings from Student Surveys

In all, 61students completed the student survey. Respondents were distributed nearly evenly across the three intervention conditions, with 20 or 21 students responding to the survey from each classroom.

Student Attitudes Toward Math

Attitudes toward math were assessed with six items (e.g., "I like math" and "I know I can do well in math if I try hard"). Students were asked to respond to each item on a four-point scale ranging from 1 ("NO!") to 4 ("YES!"). The items were combined to form a single scale with higher numbers indicating more positive attitudes about math. As shown in Figure 13, students in the i-Ready + tutoring condition had math attitudes that were more positive than students in both the i-Ready + tutoring + incentive and i-Ready only conditions. These differences were statistically significant, *p*s < .05. The difference in math attitudes between students in the i-Ready + tutoring + incentive and i-Ready only conditions were not statistically significant, $p > .05$.

Figure 13. Mean ratings of items tapping attitudes toward math by condition



Note. Statistically significant differences in math attitudes are represented by different color bars.

Student Attitudes Toward Math Tutors

Attitudes toward math tutors were assessed with four items (e.g., "I get along well with my tutor(s)" and "I would be excited to learn from my tutor(s) again next year"). Students were asked to respond to each item on a four-point scale ranging from 1 ("NO!") to 4 ("YES!"). The items were combined to form a single scale with higher numbers indicating more positive attitudes toward math tutors. As shown in Figure 14, students in the i-Ready + tutoring + incentive and students in the i-Ready + tutoring condition had similarly positive attitudes toward their tutors. The difference was not statistically significant, $p > .05$.

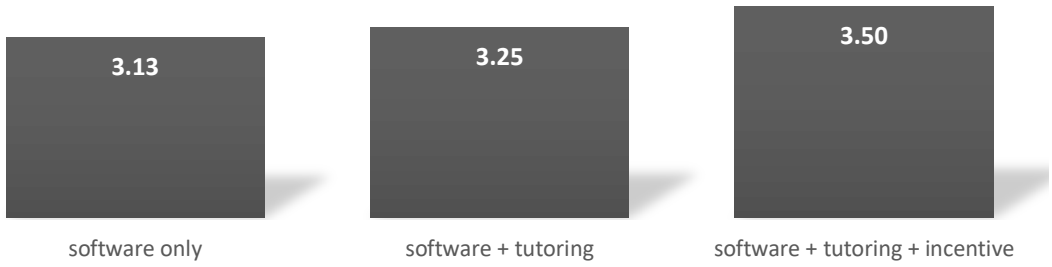
Figure 14. Mean ratings of items tapping attitudes toward math tutors by condition

Note. Differences are not statistically significant.

Student Attitudes Toward i-Ready

Attitudes toward i-Ready were assessed with six items (e.g., "I like using i-Ready" and "I would be excited about using i-Ready again next year"). Students were asked to respond to each item on a four-point scale ranging from 1 ("NO!") to 4 ("YES!"). The items were combined to form a single scale with higher numbers indicating more positive attitudes about i-Ready. As shown in Figure 15, students' attitudes toward i-Ready did not vary significantly across conditions, *p*s > .05. However, the pattern of results was in the hoped-for direction with students in the i-Ready + tutoring condition and i-Ready + tutoring + incentive condition reporting more positive attitudes than students in the i-Ready only condition.

Figure 15. Mean ratings of items tapping attitudes toward i-Ready by condition



Note. Differences are not statistically significant.

Conclusions and Recommendations

Conclusions

Findings from the current study provide preliminary, **promising evidence** for the effectiveness of the One-to-One Tutoring Program for improving student achievement outcomes in mathematics. Analyses of Acadience Math data indicate that students in both the software + tutoring and software + tutoring + incentive conditions showed larger gains in math scores than students in the comparison condition. These differences were statistically significant, meaning that they were unlikely to have occurred by chance. Likewise, analyses of i-Ready Diagnostic data indicate that students in the software + tutoring condition showed larger gains in math scores than students in the comparison condition. This difference is also statistically significant.

Program Recommendations

Based on its review of program materials, achievement data, and survey responses, the UEPC offers the following short list of recommendations for sustaining and strengthening the One-to-One Tutoring Program. Importantly, these recommendations are aligned with recommendation provided by the National Student Support Accelerator (2021). The Accelerator is a program of the Annenberg Institute at Brown University that has developed a number of evidence-based tools – including a Toolkit for Tutoring Programs – designed to assist practitioners in creating and implementing effective, high-impact tutoring programs (see https://studentsupportaccelerator.org/tutoring). These recommendations should be evaluated by program personnel and school staff in light of program goals and local constraints, including constraints associated with recruitment and return-on-investment.

1. **Increase training and support opportunities.** As a group, tutors indicated a desire for more training including training focused on effective practices for tutoring students in mathematics using i-Ready and building strong relationships with students. The National Student Support Accelerator (2021) recommends that in blended tutoring environments (where live tutoring is blended with the use of personalized learning software) tutors should be trained on relevant features of the software including how to access and use data from the software to inform their tutoring. The Accelerator notes that, for all program types, tutors benefit from both in-service training *and* ongoing support focused on creating low-stress, high trust environments in which tutors express interest in students' lives outside the classroom and in which students can experience authentic enjoyment of learning mathematics. Tutor comments indicate broad support for these principles. Pairing more and less experienced tutors and creating learning communities can be effective strategies for providing additional training and ongoing support when time and other resources may be limited (Garringer, Kupersmidt, Rhodes, Stelter, & Tai, 2015; Kupersmidt, Stelter, Garringer, & Bourgoin, 2018; Robinson et al., 2021).

2. **Increase opportunities for communication and coordination between tutors and teachers.** Both tutors and teachers expressed a desire for greater communication and coordination between tutors and teachers. The National Student Support Accelerator (2021) notes that the benefits of ongoing communication include giving teachers opportunities to provide tutors the information they need to adjust their instruction to meet student needs, giving tutors opportunities to provide updates to teachers on student progress, and ensuring that tutors and teachers have a shared understanding of when math software activities may be misaligned with student skill levels so that activities can be adjusted appropriately.

3. **Increase access to and actionable use of data.** As a group, tutors expressed uncertainty about how to access software-provided data on student progress and interest in **using** this data to inform their instruction, communicate with teachers, and celebrate student progress. The National Student Support Accelerator (2021) strongly recommends that tutoring programs intentionally set aside time to collect, review, and reflect on data for program improvement. In addition to the types of summative data that were collected for the current study (e.g., assessment data and end-of-program surveys), the Accelerator recommends that tutors collect and share regular "session assessment" data to determine the degree to which students have mastered session content and to tailor subsequent tutoring sessions. The Accelerator also recommends that program personnel or an external evaluator collect and share data from "pulse" surveys that might be administered weekly, monthly, or quarterly to identify strengths and challenges in real time.

Recommendations for Research and Evaluation

The current study provides preliminary, **promising** evidence for the effectiveness of the One-to-One Tutoring Program for improving student outcomes in mathematics. To meet ESSA standards for **moderate evidence** or **strong evidence**, future studies would benefit from utilizing more rigorous quasi-experimental or experimental research designs with a larger sample of students (i.e., 350+) receiving the tutoring intervention at two or more educational sites.

Future work will also be important in beginning to identify the elements of the One-to-One Tutoring Program that are most critical for achieving positive outcomes. Findings from tutor and teacher surveys indicate that key strengths of the program include the 1:1 tutor-student ratio, consistent tutor-student pairings, the combination of a software intervention with a tutoring intervention, and the focus on creating a positive, growth-mindset-oriented, relationship-centered approach for learning. More work is needed to determine how critical incentives are to program success. Results from the current study suggest that incentives may not be necessary for improving student achievement outcomes and may, in fact, result in less positive math attitudes among students.[8] Evaluation efforts that use standardized measures and metrics to look at the effectiveness of programs that take different approaches to implementing tutoring and software interventions across the state of Utah would be especially informative.

Future evaluations should also collect data that can be used to assess the degree to which the program is being implemented with fidelity. This would include data on the timing, frequency, duration and focus of tutor training opportunities as well as data on the frequency, duration, and focus of tutoring sessions and the frequency and focus of support opportunities (e.g., "check-in" meetings) between tutors and teachers or tutors and program staff. Likewise, additional data should be collected to determine whether additional intended outcomes for teachers, tutors, and students are being achieved. For tutors, these outcomes may include increased self-efficacy. For students, these outcomes may include increased feelings of belongingness.

[8] This finding should be evaluated in light of an ongoing debate about the value of educational incentives (see Cerasoli, Nicklin, & Ford, 2014).

References

Altermatt, T. W., Altermatt, E. R., Rorrer, A. K., & Moore, B. (2022). *Math Personalized Learning Software: Examining Usage and Associations with Achievement in Utah During the Covid-19 Pandemic*. Salt Lake City, UT: Utah Education Policy Center.

Cerasoli, C. P., Nicklin, J. M., & Ford, M. T. (2014). Intrinsic motivation and extrinsic incentives jointly predict performance: A 40-year meta-analysis. *Psychological Bulletin, 140*(4), 980–1008.

Cheung, A. C., K, & Slavin, R. E. (2013). The effectiveness of educational technology applications for enhancing mathematics achievement in K-12 classrooms: A meta-analysis. *Educational Research Review, 9*, 88-113.

Garringer, M., Kupersmidt, J., Rhodes, J., Stelter, R., & Tai, T. (2015). *Elements of effective practice for mentoring (4th Edition).* Boston, MA: MENTOR: The National Mentoring Partnership.

Kupersmidt, J., Stelter, R., Garringer, M., & Bourgoin, J. (2018). *STEM mentoring: Supplement to the Elements of effective practice for mentoring.* Boston, MA: MENTOR: The National Mentoring Partnership.

Hillmayr, D., Ziernwald, L., Reinhold, F., Hofer, S. I., & Reiss, K. M. (2020). The potential of digital tools to enhance mathematics and science learning in secondary schools: A context-specific meta-analysis. *Computational Education, 153*.

Koehler, M. & Mishra, P. (2009). What is Technological Pedagogical Content Knowledge (TPACK)? *Contemporary Issues in Technology and Teacher Education, 9*(1), 60-70.

Ma, W., Adesope, O. O., Nesbit, J. C., & Liu, Q. (2014). Intelligent tutoring systems and learning outcomes: A meta-analysis. *Journal of Educational Psychology, 106*(4), 901–918.

National Student Support Accelerator (2021). Toolkit for Tutoring Programs. Retrieved from National Student Support Accelerator: https://doi.org/10.26300/5n7h-mh59

Nickow, A., Oreopoulos, P., & Quan, V. (2020). *The Impressive Effects of Tutoring on PreK-12 Learning: A Systematic Review and Meta-Analysis of the Experimental Evidence* (Working Paper No. 27476; Working Paper Series). National Bureau of Economic Research.

Owens, R., Rorrer, A.K., Ni, Y., Onuma, F.J., Pecsok, M. & Moore, B. (2020). L*ongitudinal Evaluation of the Math Personalized Learning Software Grant Program.* University of Utah, Utah Education Policy Center. Published, 07/2020.

Pane, J, F., Steiner, E. D., Baird, M. D., & Hamilton, L. S. (2015). *Continued Progress: Promising Evidence on Personalized Learning*, Santa Monica, Calif.: RAND Corporation, RR-1365-BMGF, 2015. As of June 23, 2022: https://www.rand.org/pubs/research_reports/RR1365.html

Pane, J, F., Steiner, E. D., Baird, M. D., Hamilton, L. S., & Pane, J. D. (2017). *Informing Progress: Insights on Personalized Learning Implementation and Effects*, Santa Monica, Calif.: RAND Corporation, RR-2042-BMGF, 2017. https://www.rand.org/pubs/research_reports/RR2042.html

Pellegrini, M., Neitzel, A., Lake, C., & Slavin, R. (2021). Effective programs in elementary mathematics: A best-evidence synthesis. *AERA Open*, 7 (1), 1-29.

REL Mid-Atlantic (2017). *Research Review: Impacts and Implementation of Blended Learning.* https://ies.ed.gov/ncee/edlabs/infographics/pdf/REL_MA_What_is_Blended_Learning.pd

Robinson, C. D., Kraft, M. A., Loeb, S., & Schueler, B. E. (2021). *Accelerating Student Learning with High-Dosage Tutoring* (Design Principles Series, p. 10). Ed Research for Recovery.

Sarker, M. N. I., Wu, M., Cao, Q., Alam, G. M., & Li, D. (2019). Leveraging Digital Technology for Better Learning and Education: A Systematic Literature Review. *International Journal of Education Information Technology*, *9*(7), 453-461.

Steenbergen-Hu, S., & Cooper, H. (2013). A meta-analysis of the effectiveness of intelligent tutoring systems on K–12 students' mathematical learning. *Journal of Educational Psychology, 105*(4), 970–987.

Su, Y., Rorrer, A., Owens, R., Pecsok, M., Moore, B., & Ni, Y. (2020). *K-12 Mathematics Personalized Learning Program 2018-19 Student Outcomes, Evaluation Addendum for Year 3.* Salt Lake City, UT: Utah Education Policy Center.

Yeager, D. S., & Dweck, C. S. (2020). What can be learned from growth mindset controversies?. *The American psychologist*, *75*(9), 1269–1284. https://doi.org/10.1037/amp0000794

Young, J. R., Gorumek, F., & Hamilton, C. (2018). Technology effectiveness in the mathematics classroom: a systematic review of meta-analytic research. *Journal of Computers in Education, 5*, 133-148.