

INTERVENE

KIPP ENDEAVOR ACADEMY & OPERATION BREAKTHROUGH

by Erin Huebert



LEANLAB EDUCATION



Research

KIPP ENDEAVOR ACADEMY, OPERATION BREAKTHROUGH & INTERVENE

2020 Research Report

Erin Huebert, PhD

Vice President of Research and Programs

[LEANLAB Education](#)

SCHOOL SUMMARY AND PROBLEM OF PRACTICE

[KIPP Endeavor Academy](#) (KIPP KC) is a public charter school located in northeast Kansas City, Missouri and serves students in 5th through 8th grade. It is part of a national [KIPP](#) network of free, open-enrollment, college-preparatory schools traditionally located in educationally underserved communities. Of KIPP KC's 287 middle-school students, 92% are African-American or Hispanic students and all qualify for free lunch through the [Community Eligibility Provision](#) (KIPP Regional Report Card 2020).

A concern for teachers and administrators at KIPP KC is **progress in middle-school math**. While 77% of middle-school students are making 1+ years of academic progress in math (as measured by the 2017-18 Measures of Academic Progress (MAP) exam), 6th grade students, in particular, remain significantly behind by the end of spring semester and, thereby, enter 7th grade not at grade-level ([KIPP Regional Report Card](#)). KIPP KC decided, therefore, to focus their intervention efforts in **Mr. Justin Alt's 7th grade math class** with the goal to get students' back on grade-level math standards.

[Operation Breakthrough](#) is a before- and after-school enrichment program that serves 265 kids between Kindergarten and 8th grade. The mission of Operation Breakthrough is "to provide a safe, loving and educational environment for children in poverty and to empower their families through advocacy, emergency aid and education." Of the families served, 80% live on less than \$12,000 annually. It is located in central Kansas City and is open to all public school students.

Ms. Dajona Graves manages the tutoring classroom at Operation Breakthrough and she was looking for new and innovative approaches to helping her students advance in math. Because students arrive at Operation Breakthrough tired and not motivated to do more school work, they were particularly in need of hands-on, personal, and customized tutoring solutions that engaged students in a fun and new way.

VENTURE SUMMARY AND INTENDED EFFECT

[Intervene](#)—founded by Aaron McCloud and Mehul Shah—is a comprehensive intervention system that assesses students, analyzes data, and tutors students online to provide accessible, high-impact tutoring to schools and districts. The Intervene model combines data with tutoring to yield growth. Students take a quick assessment and an algorithm calculates what the students struggle with, so teachers don't have to figure it out. The online tutoring platform allows students to access quality tutors during or after the school day and the content is based on their unique needs. The dashboards provide recommendations for scheduling and planning tiered intervention. Finally, formative assessments give teachers and interventionists an easy way to gauge student understanding and measure progress. Taken together, the data insights and small group tutoring are believed to promote student growth.

RESEARCH GOALS

There were three research goals of this study. The first two goals focused on testing the efficacy of Intervene on two student outcomes: **math performance and social-emotional learning**. Respectively, **the first hypothesis is that students who use Intervene will see an increase in their math exam scores from the first exam to the final exam. The second hypothesis is that students will exhibit and report more positive behaviors and attitudes about math and math class after using Intervene than before.** The third goal was to gather feedback from the teacher and students on usability and implementation of Intervene in order to modify and/or enhance product features and development that met the unique needs of the classroom and its students.

METHODOLOGY

Sample

At KIPP, the sample for this study consisted of **19 students** in Mr. Alt's 7th grade Powerblock class. Powerblock is a special class that occurs for 40 minutes at the end of each day on Monday through Thursday specifically for students that are struggling in math. These were students that are performing anywhere from 1 to 4 grades below the 7th-grade math level. Of the 19 students in the sample, 35% were Latinx, 55% were Black, and 10% were white and all received free lunch through the Community Eligibility Provision. Because this sample of students was systematically different in terms of math skill-level from the other students in 7th grade, there was no available comparison group to use as a control group. Therefore, this study is purely descriptive in that it assesses pre- and post performance among just this sample of students, all of whom used Intervene.

At Operation Breakthrough, the sample consisted of **20 students** in Ms. Grave's tutoring classroom that were randomly chosen among low-performing math students that regularly attended the after-school program. They were divided into groups of four based on their age and skill level. All students in this sample were Black and all were eligible for free and reduced lunch at their respective schools. Since there was no other group at Operation Breakthrough being tutored in math, there was also no comparison group and, therefore, the results are also purely descriptive.

Measurement of Outcomes

The first outcome evaluated in this study is **math performance**. At KIPP, scores from standardized unit exams were used to assess math performance. The pre- and post-exams measured the same math standards, so the overall score from the first unit exam was compared to the overall score from the final unit exam. Scores were out of 26 points.

Since there were no standardized tests or assessments of any kind at Operation Breakthrough, there was a need to design a short assessment that measured the concepts that were being taught to each group. Since there were five different groups and each was at a different skill level, five different assessments were needed. A pre- and post-assessment was created by Intervene staff and administered within Intervene itself. Scores were out of 100 points.

The second outcome evaluated in this study is *social-emotional learning (SEL)*. To assess social-emotional learning at KIPP, two data-collection efforts were used; classroom observations and student surveys. In the first approach, a research assistant conducted weekly classroom observations and used a customized rubric (Table 1) to score various social-emotional competencies. There were nine behavior indicators on the rubric that aligned to four SEL competencies, as measured in the Panorama Social-Emotional Learning User Guide (November 2015).

The four main SEL competencies were 1) student engagement, 2) perseverance/grit, 3) self-efficacy/confidence, and 4) collaboration/communication. Each student was given a score in each area that ranged from 1 (never) to 5 (almost always).

Table 1: SEL Competencies and Respective Indicators	
SEL Competency	SEL Indicator
Student Engagement	The student leans forward attentively toward device.
	The student seems to enjoy or be absorbed in interacting with the app.
	The student is interacting with the online tutor.
Perseverance/Grit	The student remains focused on the lesson.
	The student tries again if/when mistakes are made.
Growth Mindset	The student puts in a lot of effort into the lesson.
Self-Efficacy / Confidence	The student shows signs of frustration.
	The student seems confident, as expressed by comments, body language and/or facial expressions.
Collaboration and Communication	Students communicate with each other about the lesson during the lesson.
	Students get along with one another during the lesson.

In the second approach to measure SEL outcomes, pre- and post-surveys were given to students via Google Forms. Students took a survey the week before they started using Intervene and the week after they stopped using Intervene. The pre- and post-survey were identical and measured the same four SEL areas described above and used

question wording from the Panorama Social-Emotional Learning User Guide (November 2015). Some customized questions were also asked that gauged “how much students enjoyed math.” Students self-reported on a scale of 1 to 10, with higher numbers corresponding to more positive attitudes/perceptions.

At Operation Breakthrough, the same pre- and post-surveys used at KIPP were administered to all participants. As there was not a consistent time each day or week that the groups completed their tutoring at Operation Breakthrough, it was difficult to conduct consistent student observations, so classroom observations were not used.

Teacher and Student Interviews

Once a month for three months, in-person interviews were conducted with the teachers and with students at KIPP and at Operation Breakthrough to collect qualitative evidence on the impact of Intervene. Four different students were interviewed each month from each site. A standardized questionnaire was created to guide the interview and collect comparable evidence. Interviews were recorded and patterns in responses were later identified.

RESULTS

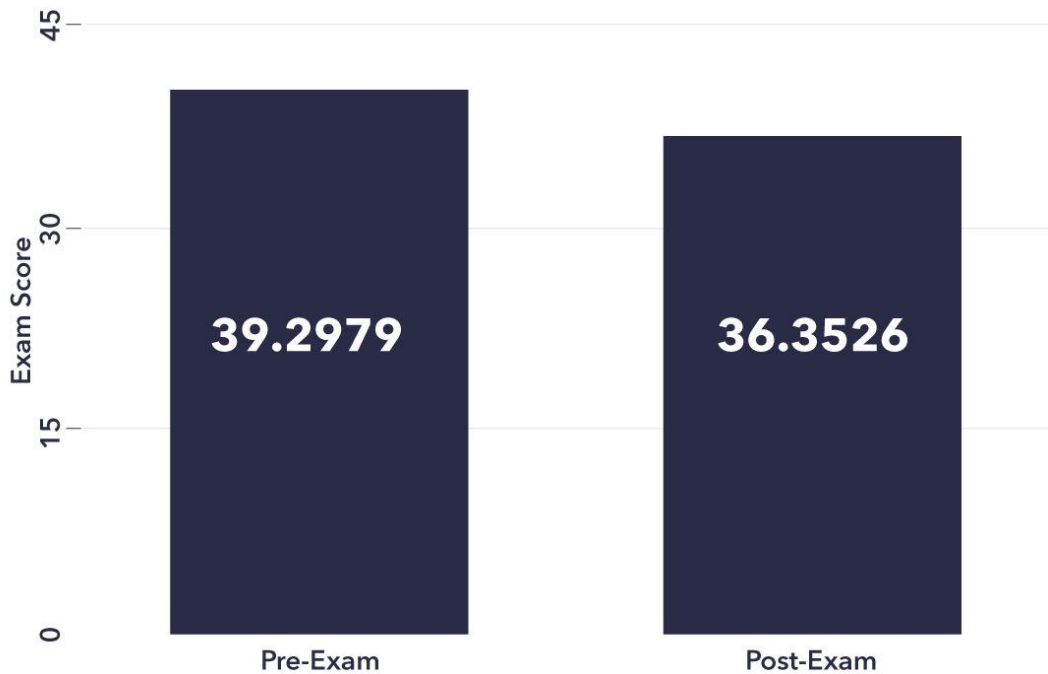
In this section, the quantitative results are presented in regards to the two research outcomes: math performance and social-emotional learning. The evidence presented here is strictly descriptive. The results illustrate pre- and post-changes in student exam scores, observed behavior, and survey responses among only the students who used Intervene. There was no control group and no other confounding variables were controlled for, so causation is neither tested nor implied in this study.

MATH PERFORMANCE

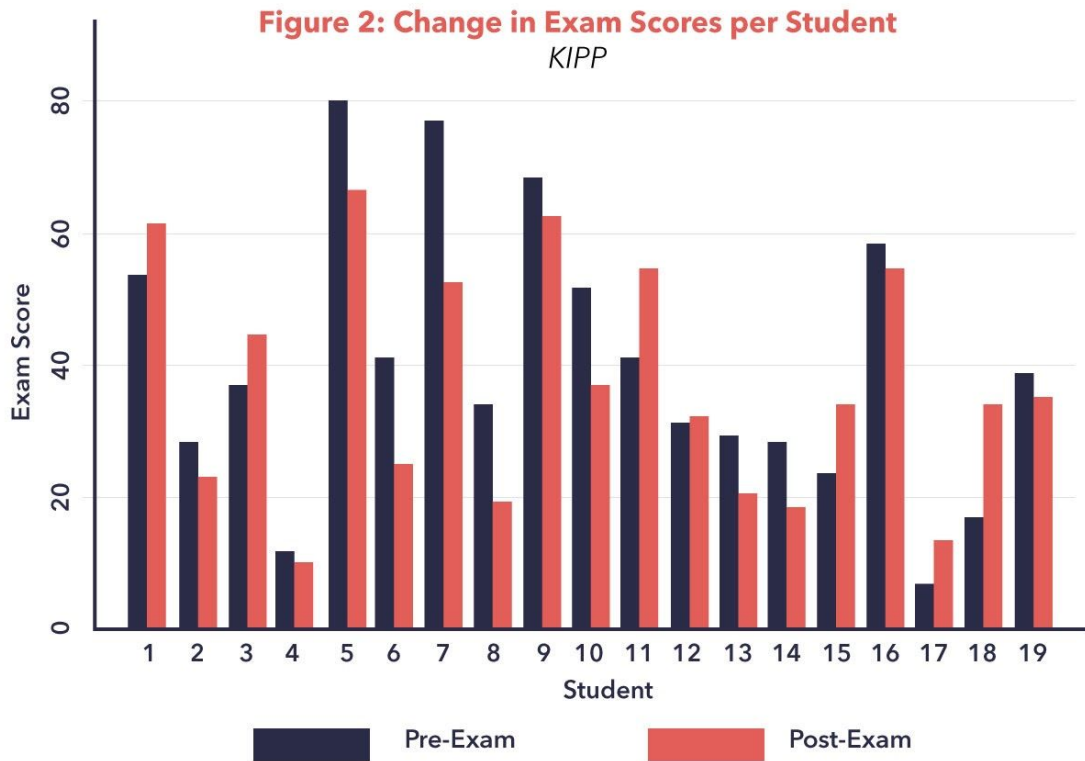
Overall, there is little support for the first hypothesis that students who used Intervene will experience an increase in math exam scores from the first to the second exam.

In Figure 1, the class average exam scores from the first to the second unit exams at KIPP are shown. On the first exam (pre-exam), the class averaged 39.29%. On the second exam (post-exam), the class averaged 36.35%. Contrary to expectations, the class-average actually decreased slightly, although the decrease is not statistically significant (based on a paired t-test).

Figure 1: Change in Class-Average Exam Scores
KIPP



In Figure 2, the scores on the pre- and post-exam for each of the 19 students at KIPP are presented, showing that 7 students increased their scores, while 12 students' scores decreased. There is extreme variation among the students in the extent of how much change occurred from the first to the second exam, which further validates a lack of a positive empirical association between Intervene and exam scores.



In Figure 3, the class average exam scores from the first to the second exam at Operation Breakthrough are shown. Recall that the exam at Operation Breakthrough was a customized exam created by Intervene staff and administered within Intervene itself. It is not, therefore, a standardized test like the one taken at KIPP. And each group of students took a different exam in line with their skill level. Scores among the 20 students are not, therefore, directly comparable. On the first exam (pre-exam) that students took at Operation Breakthrough, the class averaged 48.86%. On the second exam (post-exam), the class averaged 61.14%. The increase is statistically significant at the $p < .10$ level (based on a paired t-test) but refer to the discussion section below for causation on this finding.

Figure 3: Change in Class-Average Exam Scores
Operation Breakthrough

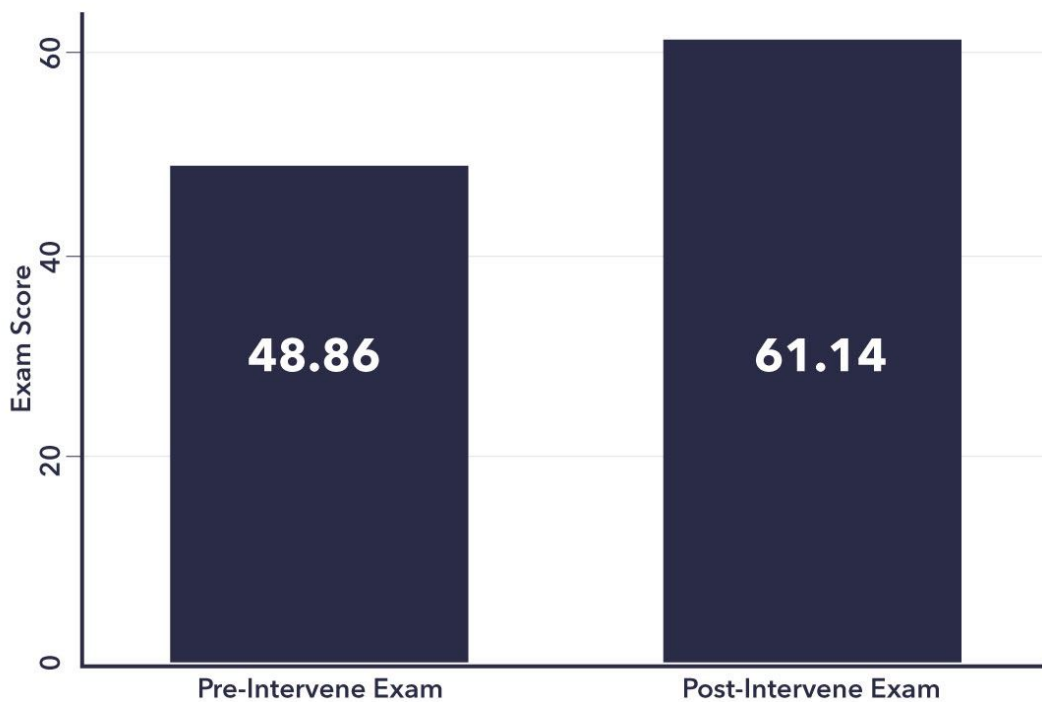
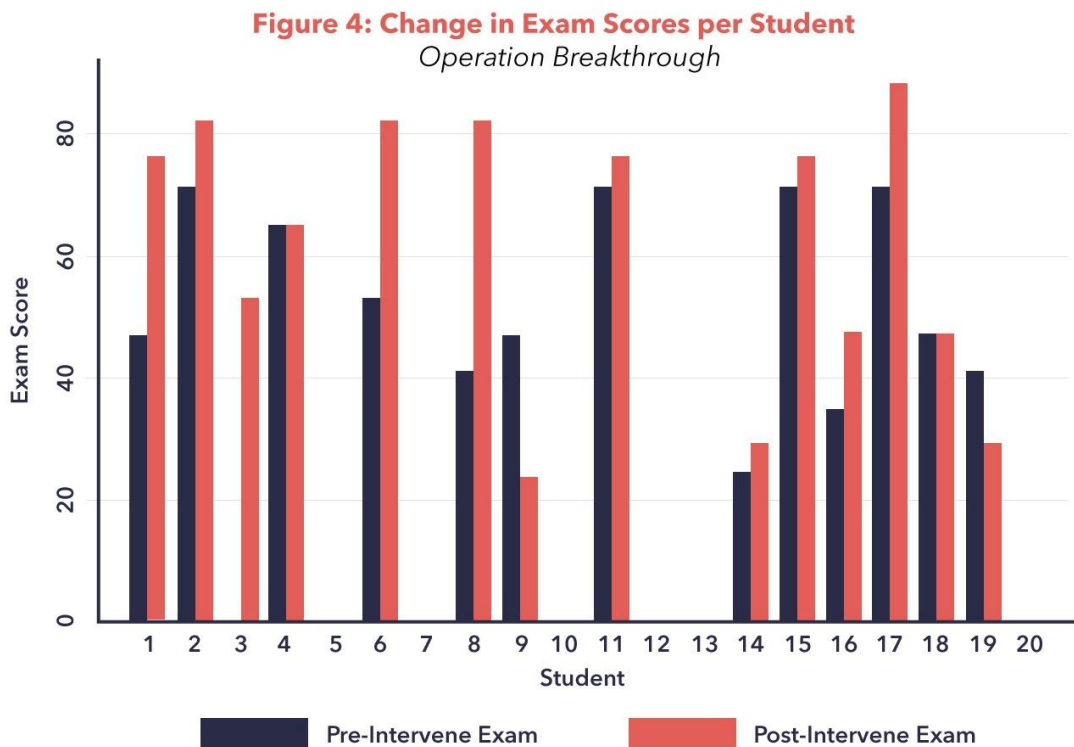


Figure 4 illustrates the variation in change in exam scores among the students at Operation Breakthrough and also highlights a significant amount of missing data. While four students dramatically improved their scores, others improved a little, and still others did not improve at all. Of importance is the fact that seven students (35% of the total) did not take the exam. For these reasons, the results presented in Figures 3 and 4 are, unfortunately, full of bias and measurement error, which are discussed in greater detail below.



Social-Emotional Learning

There is also little support for the second hypothesis that students who used Intervene will become more engaged in the classroom and express more positive attitudes towards math and math class.

First, the *classroom observations* at KIPP yielded no statistically significant findings. While many students did exhibit positive behavior growth in one or more of the SEL competencies listed in Table 1, the increases were not large enough to reach statistical significance. This data collection had, however, some major challenges that will be discussed below. Classroom observations were not used at Operation Breakthrough.

Second, the results from the *pre- and post-survey* also do not yield any statistically significant growth. Figure 5 illustrates change in student responses from the pre- to the post-survey at KIPP on some key SEL indicators. While there was some growth on the question of “how well [students] like math” and “how interested in math class” students were, the growth was not statistically significant. The response rate on the post-survey at Operation Breakthrough was, unfortunately, too low to be able to calculate reliable average change.

Figure 5: Change in Class-Average Engagement Indicators



PRODUCT MODIFICATIONS

The third goal of this research study was to gather feedback from teachers and students about product usability and implementation requirements. A number of suggestions and modifications were made. In response to teacher feedback that it was hard for the teacher to track student progress or to see what students were doing during live sessions, Intervene created a new interface that is more transparent for the teacher. The new interface is easier to navigate and allows teachers more direct access to troubleshooting set up and technical problems. Before, teachers would have to call Intervene for technical assistance which was timely. Intervene also created a customer service line so that a support technician was immediately available for assistance.

The research process also helped Intervene understand the conditions of implementation that led to more successful classroom engagement. First, when students had the same tutor each session, they were able to develop a relationship with the tutor that led to more positive engagement and experience for those students. For the students that did not have the same, consistent tutor throughout the semester, they were generally not as engaged and reported less positive experience.

Second, students reported that they liked the small group tutoring and collaboration with their group, and two aspects of implementation made collaboration and engagement better. The first was to disengage the mute feature of the program. Usually the tutor would have complete control of the mute feature and would have the students muted in order to control the background noise. As a result, tutors would not realize when students were asking questions, which frustrated students. Instead, students used headphones that blocked background noise and, as a result, tutors didn't have to use the mute function, which helped students remain engaged and ask more questions. The other thing that helped student engagement and collaboration was controlling when students were able to write on the screen. While students were anxious to write on the screen, it was a distraction when they had unlimited access to do so. For successful implementation, teachers either needed to closely monitor what and when students are writing on the screens or the tutor needed to have the ability to limit when a student had access to writing.

Finally, the feedback from Ms. Graves at Operation Breakthrough is unique in that it is an after-school program. Given that students came to the program after school and were tired, they did not look forward to doing more school work, so she recommended that Intervene create special programming that is more game-based. She believed a program that is more fun and different from the typical classroom setting would help engagement with Intervene at other after-school programs.

DISCUSSION AND NEXT STEPS

While the efficacy results of this study do not support the hypotheses linking Intervene with math performance and social emotional learning, the study was limited in many ways and additional research is nonetheless warranted. First, the sample size in this study was very low. With only 19 students in one group and 20 students in the other, it was a very small group to infer generalizability to a wider context. More importantly, this study did not compare changes in exam scores or attitudes of a control group. As stated earlier, there simply was not a viable comparison group at KIPP or Operation Breakthrough. We cannot say in this study that any of the results we see are due to Intervene without having a control group and/or controlling for other possible explanations for change. Future research should, therefore, have a much larger sample size and include a proper control group and collect data on confounding variables to control for in a statistical analysis. A more robust design may yield very different efficacy results, so we caution readers from drawing definitive conclusions regarding the null findings in this study.

Second, the consistency in attendance at Operation Breakthrough severely limited the results in this study. We were only able to get about 20% of the sample (4 students) to complete the post-survey, which is far too low of a response rate to even report average change. Additionally, in a given week, not one student attended all three tutoring sessions and sometimes students only attended one session a week. Beyond this, some students would get picked up by their parents early and then they wouldn't even have the chance to complete the session when they were there. As a result, it was extremely hard to evaluate the impact or even the conditions of successful implementation with such inconsistency. Furthermore, the math assessment at Operation Breakthrough was not a standardized exam created by an external entity. Since Intervene created the assessment itself based on its own content, it lends itself to bias. Future studies should use an externally validated assessment for pre- and post-math comparisons that are also directly comparable. It is also recommended that future research either be confined to a typical classroom setting or at an after school program where attendance is consistent.

Third, the pre- and post-survey results at KIPP on SEL competencies yielded little change, which is likely due to the short timeline (ten weeks) between surveys. Changes in SEL outcomes likely take much longer to notice significant changes. In future research, different measures of SEL that can be measured and assessed on a weekly basis may be helpful, such as student self-reports on exit tickets or discipline referrals, etc. Objective measures of SEL would also be helpful to avoid potential reliability bias with survey data. The classroom observation data was an attempt to validate student reports on surveys by observing and recording actual behavior aligned to SEL outcomes. Since the observations were, however, only conducted once a week, a couple of limitations arose. First, several students had significant missing data, as they happened to be absent on the observation day. Second, observing students just once a week is not necessarily representative of their typical behavior, as some students may exhibit an “off” day on the days of observation leading to validity bias. This further stresses the importance of new and creative approaches to measure SEL outcomes.

Despite these limitations, there is sufficient preliminary anecdotal evidence to warrant a larger usability study of Intervene. As Intervene is creating new features, platforms, and service offerings, a robust usability study would help decipher the conditions of successful implementation of their existing features and platforms as well as identify the need and demand for additional features or services.

CONCLUSION

While KIPP KC chose not to continue working with Intervene, they were grateful for the opportunity to co-design a new solution with Intervene and Intervene in return proved to be a very responsive partner. Operation Breakthrough was interested primarily in the assessment feature of Intervene that determines which math skill(s) a student needs particular attention in. They decided to pursue additional support from Intervene in that manner. While the efficacy evidence in this study is only descriptive and limited by sample size, reliability bias, and validity bias, there is sufficient positive anecdotal evidence from students that lends support for a larger, more robust usability study.

REFERENCES

KIPP Regional Report Card. Accessed on March 9, 2020.

<https://www.kipp.org/wp-content/uploads/report-card-results/2018/school/kipp-endavor-academy.pdf>

Gehlbach, Hunter. November 2015. *User Guide: Social-Emotional Learning*. Panorama Education.