Executive Summary

Evaluation of the Beable Platform

SCHOOL DISTRICT OF OSCEOLA COUNTY

2023

Authored by: David Maddock

Complete details of all analyses and literature references, can be requested from REA

Executive Summary

Evaluation of the Beable Platform

Background

In the 2022 school year, the School District of Osceola County evaluated a digital reading intervention program called Beable to determine the effect on student learning of program usage. Beable purports to be designed to identify and close the literacy gap with "speed and certainty." The results of the evaluation were mixed; there was some quantitative evidence to show that students who utilized the platform grew slightly more than students who did not utilize the platform, however quantitative user analysis found multiple emergent themes that teacher users were frustrated with the platform and found it difficult to use. While the 2022 implementation focused on high school student and teacher users, the 2023 implementation moved the platform to implementation in the middle grades. For that reason, a follow-up evaluation must be conducted to further examine the effects of the Beable platform on Osceola students.

Beable was first introduced in the Summer of 2020 and there were few analyses conducted of the platform. Beable's own internal analysis found that students who used Beable grew at a rate of five times greater than expected growth, a claim which was deeply analyzed in the 2022 evaluation. Osceola's evaluation found that Beable had an effect size of $d = 0.08$. While the 2022 evaluation compared this value against the John Hattie established baseline for the expected effects a teacher can provide a student in their growth, $d = 0.40$, a solid baseline expectation for student learning during the year, evaluations in Osceola have since trended toward using effects established by Matt Kraft (2019) at Brown University, who proposed that interventions should instead be rated as less than $d = 0.05$ as small effects, and greater than $d = 0.20$ as large effects. Under this paradigm, Beable would be considered to have moderate effects, although still measured as smaller than Achieve 3000, which measured at $d = 0.21$.

The 2022 evaluation of Beable also identified some areas of concern with the Beable platform itself, at least in relation to its usage in Osceola County, chiefly:

- Effects for students were extrapolated using a framework of twelve months of growth (rather than the ten months that students are in classroom) based from a four month sample
- The platform sometimes asked questions that were unrelated to the text a student was asked to read, and therefore unanswerable
- During adjustments, student Lexiles were not adjusted downward if a student performed lower, therefore the student and teacher only ever saw that the student was performing at their own baseline, even if the student regressed
- Beable did not report values below 0, so students lowest measured Lexile would be 0, even if they were in pre-reading Lexile ranges

Beable was provided to all students in Intensive Reading courses in Grades 6 – 8 during the 2021-2023 school year, except at Horizon Middle School where all students engaged in the platform (similar to the prior year, more on this later). Despite this fact, Horizon still did not have the most lessons completed (36,855 lessons). The school with the most lessons completed during the year was Narcoossee Middle School (43,280 lessons), followed by Horizon, and then Denn John Middle School (22,648 lessons). The schools with the lowest usage were Parkway Middle School (3,978 lessons) and Canoe Creek K-8 (2,836 lessons). New Beginnings, OCSA, and Zenith also had fewer lessons, but also disproportionately smaller target populations, and were therefore excluded from this summary. This amount of usage represents a slight decline from the year prior; while many of the schools completed between 18k and 23k lessons in 21-22, the previous analysis covered a four-month usage period rather than an eight-month period.

Finally, in the previous evaluation, the cost was analyzed as a cost-per-user effective cost rate and compared it to the actual cost of the platform. The total cost for Beable was $417,500 for the school year, and had an effective cost rate of $36.60 per student user ($47 at the middle school level and $30 at the high school level). The reason for the heightened costs was that Beable charges a schoolwide rate that includes all students, yet the program in Osceola only targeted certain student users, drastically increasing the per-student cost. If every student in the targeted schools had used the program, the cost-per-student would have been $12.93. For the updated evaluation, this figure will be recalculated and utilized in an ROI comparison.

Purpose

The purpose of this evaluation was to examine the effectiveness of the Beable Reading Intervention program in the county for the cost per student.

The following evaluation questions were posed for the platform:

1) What relationship, if any, exists between the implementation of the program and student academic increases?
2) To what extent does the implementation of Beable impact student outcomes, and how does that compare to other Osceola digital reading interventions, such as Achieve Literacy?
3) What is the return on investment of the Beable platform?

Methodology

Quantitative methodologies via statistical analyses were utilized to examine the effects of each current program. Data from the Beable platform were provided by student directly from Beable. Beable's measures of starting Lexile, current Lexile, and Lexile growth were collected, as were the number of student logins to the platform, total lessons completed, and average scores on the lessons. This aligns with the manner in which Beable conducted their own Effectiveness Study, and the Osceola evaluation from 2022. Additionally for this year, data were also provided specifically on how many of the lessons were academic compared to special interest, and whether the student score met the mark of "high fidelity". It should be acknowledged that Beable has effective processes for both sharing and protecting student data, and is the gold standard in data sharing partnerships for the purposes of evaluation. Each of these data points were aligned in a row context with students' progress monitoring data from PM1, PM2, and PM3 on the Cambium FAST BEST 3 -10 ELA assessment, including scale score and student percentile. Growth in scale score were calculated between PM1 and PM2, PM2 and PM3, and PM1 and PM3. For the purposes of this study, "growth" will refer to the change in scale score between PM1 and PM3. While proficiency scores were also collected, growth is the preferred evaluation metric for this evaluation since, in most scenarios, the students selected to utilize Beable came in with lower scores and are expected to end with lower scores than students who did not use Beable. Additional data were collected from the student data warehouse relating to student demographics, prior year FSA scores, and an array of other moderator variables. Statistical tests were performed to compare differences among students. All statistical analyses were performed using SPSS 27.0.

Key Findings

Beable's Internal Validity and Effect Size

When measuring the effectiveness of an intervention platform, it is critical to see what the platform believes student growth to be internally. That is, are students showing growth *within* the platform being examined, so that we can compare those numbers to external measures and see if the platform's internal measures are valid representations of student growth, ability, and knowledge acquisition. In the prior year evaluation, it was found that there was a parabolic arc to student growth on the platform, and that students who used the platform either too little (fewer than two lessons per week) or too much (students who used the platform for five or greater lessons per week) experienced less growth than students in the middle range. For the follow-up year, a correlational analysis was conducted to determine the relationship between the metrics reported by Beable.

A statistically significant, but practically small, relationship was found between the student's Lexile score and the number of lessons completed on the platform, $r = .07$, $p < .001$. A higher relationship was found between the

average scores on a lesson and the students Lexile, $r = .48$, $p < .001$, however no relationship was found between the average score and student growth, $r = .03$, $p = .110$, indicating that the relationship is tautological; that is to say, students who score higher have higher Lexile scores to start with, which drives the relationship (the relationship is significant but extremely weak, $r = .03$, $p = .03$ when filtering the average down to only performance on "academic lessons"). The most important relationship that was uncovered is the relationship between total student logins and Lexile growth, $r = .16$, $p < .001$, indicating that there is a positive relationship between the amount of times a student gets on the platform (a similar relationship exists with lessons completed) and their Lexile score, although the overall relationship is still not strong enough to even reach the "weak" threshold of $r = .20$, which is the lowest coefficient typically used in social sciences as an indicator of an existing relationship. The strongest relationship to be found was between the number of times that students logged on to the platform and the number of lessons that they completed, $r = .60$, $p < .001$ The correlation table below shows the relationships between each of the data points provided by Beable for Osceola students.

Beable Internal Correlations	***Current Lexile***	***Lexile Growth***	***Lessons Completed***	***Logins YTD***	***Lesson Score***
Current Lexile	1	-.01	.07**	.01	.48**
Lexile Growth	-.01	1	.14**	.16**	.03
Lessons Completed	.07**	.14**	1	.60**	-.09**
Logins YTD	.01	.16**	.60**	1	-.08**
Lesson Score	.48**	.03	-.09**	-.08**	1

This indicates that usage of the Beable platform does not have a strong relationship to Lexile growth, as measured by the Beable platform.

Beable's Lexile Measure

Beable measures a student's initial Lexile via an assessment within their platform. In the previous evaluation, an extensive analysis was conducted to determine the validity of the measurement construct within Beable. It was determined that the Lexile in Beable held a strong correlation to the Lexile in NWEA ($r = .691$, $p < .001$). It was also noted that two major issues existed with the Lexile measure in the platform. The first was that no students had negative growth (that is to say, a student never measured below their initial Lexile on subsequent assessments, even when skill atrophy was detected on other assessments for the same student), and the second was that 58% of students had a Lexile growth score of 0 on the platform, despite consistent usage. In the current year, both of these issues persisted; no students had a detected negative change in Lexile, and 1,244 students (out of 4,465 active users) had a Lexile change of 0 despite platform usage. This indicates that 27.9% of students who use the platform never see growth as measured by the platform.

While the prior year analysis used NWEA data to determine validity, Osceola middle school students did not take the NWEA assessment in the 2023 school year. Instead, they completed the Florida Assessment of Student Thinking on the Benchmarks for Excellent for Student Thinking Progress Monitoring Assessment (FAST BEST PM, herein referred to as "PM") three times during the year, Fall (PM1), Winter (PM2), and Spring (PM3). Students were provided scale scores for their performance, which were utilized to conduct similar analyses to the prior year.

First, the relationship between Beable's Lexile and the scale score from PM3 were compared for association. A correlational analysis was conducted to determine the similarity between the two constructs. Beable's Lexile score was rather strongly correlated with the PM3 scale score ($r = .631$, $p < .001$), at nearly the same level it was correlated to the NWEA Lexile, indicating that the Beable assessment at least had similar construct measurement to the state assessment (students who scored well on one were likely to score well on the other, and the inverse). There

was also an extremely weak but statistically significant relationship between Lexile growth on Beable and scale score growth on the progress monitoring (r = .06, p < .001). A full table of correlations between the two platforms can be seen below.

Beable External Correlations	***Current Lexile***	***Lexile Growth***	***Lessons Completed***	***Logins YTD***	***Lesson Score***
PM1 Scale Score	.60**	-.12**	.04**	-.14**	.43**
PM3 Scale Score	.63**	-.06**	-.02	-.10**	.48**
PM1 to PM3 Growth	.10**	.06**	.05**	.03	.11**

This indicates that the assessment of student Lexile in Beable is a valid measure of student performance. However, it also indicates that there is not a very strong relationship between growth on Beable and growth on the progress monitoring assessment, indicating that one of the two growth measures may have some validity errors.

On the other side of the coin from the associative analysis is the differential analysis. These tests are utilized to determine how specific groups of students compare, and allow for analysis to include students who did not use Beable for comparison. To begin with, a *t*-test was conducted to determine the difference in progress monitoring scores based on whether or not the student utilized Beable. A comparison based on student scale scores is not useful (t(16314) = -53.34, p < .001, d = -0.96) since the only students who use the platform are in an intensive reading course, which means this is much more a measure of the difference between students in IR courses versus Research courses rather than a measure of the success of the platform. What is instead important to examine is the growth that comes from the platform. Since many of the students in Beable are scoring lower already and have greater room for growth than students not using Beable within the platform (and they receive twice the dosage of reading instruction), the logic should follow that students who use Beable in their Intensive Reading course would experience greater growth when compared to students who did not use the program. This was not the case. Students who used Beable grew less than students who did not use Beable, (t(10747) = -7.54, p < .001, d = -0.15). What's more is that disaggregating this data down to point-in-time analysis shows that this gap did not exist to this extent earlier in the school year, with growth between PM1 and PM2 showing far less difference between the groups when compared to growth for the whole year (t(10525) = -2.531, p < .001, d = -0.05). Students who were using Beable grew less than students who did not use Beable during the 2023 school year.

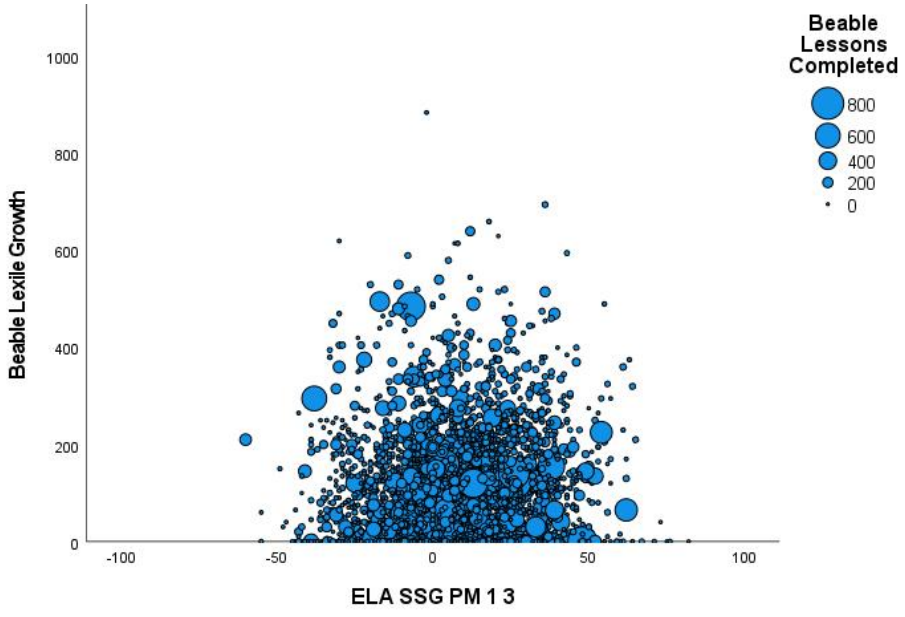
Beable also provides a measure of whether or not a student met the threshold of being a "high-fidelity" user, which was determined the previous year to have an impact on student outcomes. Since the potential exists that the difference in growth was related to how effective the student was at using the Beable platform, and analysis was conducted comparing the growth of students who used Beable with high-fidelity against all other students. The results were slightly more positive, t(3819) = 1.726, p = .042, d = 0.06, which aligns with previous findings that the implementation of the platform does have an impact on student success, but the effect size is quite small at d = 0.06. It is also possible that the data was skewed by students who were selected for the platform, although the fact that Horizon, where all students used the platform, saw an absolute zero-point difference between high-fidelity and non-high-fidelity users would discredit the idea of skewness being introduced by moderator variables, especially given that the students the *t*-test for all students found that students in intensive reading performed about one standard deviation below students not in intensive reading. If the student-level variable were moderating the difference, it would be expected that the difference between the groups of fidelity would have *greater* differences than the whole model, not *lesser* differences. It is also of value to note that Horizon utilized the platform with all students in the prior year, and, exactly the same as this year, saw a difference of d = 0.00 between struggling and non-struggling readers, regardless of their platform usage. The following table shows the mean growth for each group by school.

This indicates that students who spend time on Beable grow slower than students who spend time in other interventions or learning environments. This effect is somewhat lessened if a student uses the program appropriately, but Beable still yields little growth beyond standard expectations for a student.

School	*Non-Beable User PM Growth*	*Beable User PM Growth*	*High-Fidelity Beable User PM Growth*
BLK8	11.0	8.2	9.7
CCK8	10.4	8.1	10.6
CK8S	13.9	11.9	11.4
DJMS	4.8	5.7	5.5
DISC	11.0	9.6	10.3
HRMS	12.0	10.3	10.2
HZMS	-	7.5	7.5
KMMS	7.1	2.4	2.7
NBEC	5.7	9.5*	1.5*
NCMS	11.2	7.8	8.3
NPMS	11.7	9.0	8.8
OCSA	12.1	20.8*	-
PWMS	7.4	4.5	6.4
SCMS	8.4	7.6	6.7
WSK8	10.1	9.1	9.3
ZENP	6.6	-2.6	-2.5
* indicates a group smaller than $n = 15$			

To further illustrate the relationship, a scatterplot is presented showing the relationship between growth on Beable and growth on the FAST Progress Monitoring Assessment. In an ideal model, all dots would be to the right of the 0 on the x-axis (showing growth from PM1 to PM3), and their distance from 0 would show a similar distance from 0 on the y-axis. This would indicate a solid relationship, and a trend in growth. Instead, there are a large amount of students who are to the left of the 0 on the x-axis (meaning they scored better on PM1 than PM3), a substantial amount of whom had great growth on Beable. Multiple points can be found where students experienced 200-600 Lexile growth in Beable, yet declined in PM scale score throughout the year.

Figure 1. Scatterplot of Beable Lexile Growth Versus Progress Monitoring Scale Score Growth



Beable Stakeholder Feedback

In the previous evaluation, one particular point of relevant analysis was that 100% of respondents to the qualitative analysis had a negative sentiment towards usage of the platform. Due to the strength of this finding, the same analysis was conducted for 2023. Whereas in 2022 feedback was only collected from reading coaches, in the 2023 feedback was solicited from all adult users of the platform.

A net promoter score is a measure of how likely users are to recommend a platform to their friends and colleagues, and is an industry standard measure for sentiment towards a platform, a useful heuristic for whether or not a platform's usage is likely to grow or shrink in the future based on the amount of promoters and detractors that exist within the target population. In Osceola County, Beable was measured with a net promoter score of -67 NPS (n = 31), with 73% of participants falling in the detractor range. Only two respondents fell in the promoter range for the platform, which is one of the lowest ever measured NPS in Osceola evaluations.

This indicates that there is a general negative sentiment towards Beable and it is possible that usage rates would decline in the future unless strict compliance and monitoring measures were put in place.

Compared to the prior year, which only used open ended items and qualitative theming, a series of Likert-type agreement statements were provided to further examine teacher sentiment. None of the items had a percent agreement higher than 50%, which indicates that a majority of respondents disagreed with all of the positively worded statements. The most agreement was found on the statement "Beable aids the students in developing their reading skills" (48% agreed), and most disagreement was found with the statement "I would rather use Beable than other programs I have used in the past" (13% agreed). The following table shows each of the statements, the percent of respondents who either agreed or strongly agreed with the statement, and the weighted average score (a 3.0 represents antipathy towards the statement, a 1.0 strong general disagreement, and a 5.0 strong general agreement).

Agreement Statements	*Percent Agreement*	*Weighted Average*
Teacher focused		
The initial training I received prepared me to use Beable	43%	3.00
I continue to get the support and training I need to use Beable	42%	2.90
Students are using Beable with fidelity	23%	2.63
I would rather use Beable than other programs I have used in the past	13%	2.03
Student focused		
Students' reading levels have increased from using Beable	39%	3.00
Beable aids the students in developing reading skills	48%	3.29
Beable meets the needs of different levels of students	39%	3.03
Students are engaged and on-task while using Beable	35%	2.65
Students enjoy using Beable	19%	2.23
Platform focused		
Beable is a quality tool	29%	2.81
Beable is easy to use	38%	3.06
The data reports on Beable are useful	32%	2.68

For the purposes of evaluative fidelity, the above questions are the same as those used on evaluations of other reading interventions, such as HD Word, Lexia, and Achieve3000. Open ended items, similarly to the 2022 survey, were provided after the agreement statements. In the previous evaluation, the emergent themes were that the reports within the platform were difficult to use, that changes were needed, that the platform was ineffective, and that teachers were unhappy with Beable. A similar qualitative process was followed for the current year, and the following themes emerged from the responses (some responses broken up to partials for theming):

Emergent Themes and Relevant Response Segments
Theme: The platform is difficult to use
Beable has been very difficult to work with. The grading system is inconvenient. There have been several instances where the correct answer was marked as incorrect. The program needs a lot of work
Not a fan of how to view student scores
The reports are so hard to find and access.
I have no idea what skills a student is working on at a given time.
Students have trouble reading and understanding their growth data
Student data is a bit tough to monitor as you need to go into each student individually. I would prefer to see statistics on background/word study/core reading in one report vs having to go into each individual profile. It is very time consuming.
There is no English Language Learners scaffolding and there needs to be a mandatory video explaining Beable to students when they first log in.
Theme: Beable is not accurate
The Lexile levels on Beable are NOT accurate. 100L is the level of Kindergarten, and often my 8th graders test at this level, even when they are trying.
Because the program is so new- there have been many instances where the student will ask for help with a question, and it doesn't even match to the reading they had
Some of the questions are not specific when asking about foundational skills (for example, the question might say what type of syllable is in this word- but it has multiple syllables, and each answer choice could technically be right).
I have too often witnessed students answering questions correctly and Beable marking them wrong. This brings a students confidence down.
Some of the obstacles I encountered were, Beable reading levels not moving up even though they were really trying, reading levels are not accurate.
I also feel that their Lexile Levels are not an accurate representation of their independent reading levels.
Theme: Beable is frustrating
Students complain about not getting points for doing more than an allowable amount of articles--so they stop after a point
Some kind of report needs to be created to show what standards the kids are mastering. It is very time consuming and frustrating with how things are currently set up in the program.
There should be lower level questions for beginning second language learners. Some of the questions are too confusing for all students.
This was my first year using Beable and it was disappointing. I found it to be much less teacher and student friendly then Achieve 3000 or IReady.
Some of the students' comments are that it is boring, it is too long.
Theme: The problems lie with the student
Beable is a good tool but the students do not want to read.
Like any program we use in school, some students use Beable with fidelty, others do not. For the students who are engaged in their education, Beable provides instant feedback on their progress. They can check their profile at any time and it allows teachers to have good conversations with students about their progress.
Struggling readers are easily frustrated with this program. Many of the questions require multiple answers or sequencing of events. When struggling readers are performing well below grade level, these lessons and question styles do not build their confidence and reading skills… They will usually give up if they are working independently.
Beable can be a good program however, the students do not like Beable.

One additional point that was mentioned by the stakeholders in the prior year was that sometimes students receive questions where the text being read does not seem to relate to the question being asked during the Reading Challenges. Pictures of these issues were presented in an appendix. It would appear that this issue was not totally resolved, as some respondents mentioned it happening again during the 2022-2023 school year on the platform as well.

Finally, it must be noted that during Osceola's evaluation of the Achieve3000 platform, the analysis of the emergent themes for Achieve were considerably more difficult since many of the respondents used the open-ended response space to comment on their feelings towards Beable rather than discuss Achieve.

Beable Return on Investment

The final question in the evaluation was related to the return on investment that Beable provides. Since 2022, the School District of Osceola County has sought to quantify the educational value of a platform on not only effect size (academic growth) but also the return on investment that it provides. Since schools are limited by resources in many ways – fiscal, human resource, time – it is critical to align around specific usages of platforms that provide not just growth, but the *most growth per dollar* possible. Using the methodological practices provided by the ROI Institute, we must first determine the *true cost* of Beable. This is not simply the amount of contract paid to Beable, but also the human resource cost of training, providing access to devices and internet, and the cost of having a certified teacher in the room with the student.

For digital interventions, the ROI goal is 0%. When a student is academically behind, there is a cost to "catch them up" to the Tier 1 core that is usually paid in terms of additional intervention time, academic tutoring, and eventually credit recovery programs. In the case of students who are behind in reading, entire FTE positions are dedicated to helping these students return to the core via Intensive Reading classes and teachers. If a program achieves a 0% ROI, this means that the cost of the program is at least as cost-effective as putting the student in any other intervention, but there are also hidden benefits that are not enumerated in the ROI: the teacher gets to spend more time with students with the greatest need, and suffers less stress from having to carefully plan every moment of intervention time. For these reasons, a 0% ROI can be considered a worthy investment for a digital intervention (the expectation is higher for a human-resource based intervention). Of course, a greater ROI is more desirable. For comparison purposes, other digital interventions - using this exact calculation - such as Dreambox (math digital intervention) and Edgenuity (credit recovery) had measured ROIs of 196% and 22% respectively.

The Beable contract for the 2023-2024 school year cost $169,587. This is a considerable decrease in price from the prior year cost for all middle schools ($203,000). Considerably fewer students used the platform in 2023, where only 4,537 students logged in versus the prior year count of 11,409 students, which resulted in an effective cost of $37.38 per student, up slightly from the pervious year value of $36.60 (lower than the middle school only effective rate of $47 in the prior year). In Exhibit A of the Beable contract, the stated cost of the Beable platform is $30 per student, indicating that some students were paid for who did not use the platform, and drove up the price slightly (not as vast as the prior year cost differentials of $36 vs $13). This contract price also included three professional developments at 45 minutes per session, and one full day of PD on-site.

Beable is operated during the Intensive Reading period, which is scheduled into the students' day and requires a teacher to at least be present in the room with them to ensure that they log on to the platform, answer questions as needed, and provide supervisions for student safety. This requires us to factor in the teacher as a cost of operating Beable as a platform in their classroom; that is to say, were students to only utilize Beable after school, it would likely have far fewer costs in the ROI calculation, although the expected usage and effects would also likely be commensurately different. Prior Osceola program evaluations have used an average teacher pay rate of $35 per hour, which is the Osceola average, however, since all teacher users of the program were known explicitly for this study, pay per employee was utilized for a finer calculation (the average teacher pay for teacher of IR courses that use Beable was $34.26 per hour, slightly lower than the district average, likely because 26 of the teachers in the study were first year teachers, which brought down the average). Across the 64 employees who used Beable, the individual salaries were collected, and a benefits rate of 19.79% was added, which represents inescapable benefit costs of employing the teacher. This led to a human capital cost of $2,992,585 in salaries, and $592,233 in benefits, for a total human resource cost of $3,584,818 to be loaded into the cost of providing Beable for the students. While this may seem like a large value to add the cost in the ROI model, it is essential to note that 1) higher *true costs* work to strengthen the value found from the ROI calculation and 2) this exact model has been used multiple times in

Osceola and yielded positive ROIs as high as 201% return. This is because a truly valuable intervention will potentially save the district *years* of costs in remediation if a student is caught up before their senior year.

The monthly cost of internet for last year was $336 at each site, for a total cost of $53,760 for all sixteen schools combined. This brings an all-in cost for providing the intervention to a maximum cost of $3,808,165, not counting unavoidable sunk costs such as facilities usage and technology. To reiterate, the majority of this cost is burdened by the usage of teacher time to provide Beable, which does go a ways to explain the emotions teachers presented in relation to platform via their survey responses.

In SY2023, the School District of Osceola County received a base per-pupil expenditure of $8,629 (FLDOE, 2022). For simplicity, weighted FTE will not be used during the analysis to increase simplicity in the ROI measure (it would take a hierarchical model of effects to weight student growth by cost to educate), although this may *slightly* increase the ROI. Middle school students in Osceola County have class periods that average out to 49 minutes (this is slightly different for schools that have "block scheduling", where the block ranges from 90 to 106 minutes). Since a student usually has seven periods a day, for 182 school days, it can be determined that one hour of learning at the high school level is worth approximately $6.77 (for comparison, the DreamBox evaluation found that in elementary schools, one hour of learning was work $7.29 at its base level, a rate that is slightly different after accounting for time spent in transition and longer lunches at middle schools). This metric means that one student, learning for one hour, is worth approximately $6.77 of a teacher's time (students with greater need garner greater dollars to meet their needs), and can be used in calculating the costs avoided in interventions. For example, a student who was 100 hours behind, by this measure, would cost $677 in teachers' time to remediate back to Tier 1. For the purposes of this calculation, students spend approximately 149 hours a year in intensive reading courses (49 minutes per period), so a student who was "a year behind" would cost $1,009 in employee wages to remediate. For this reason, it becomes important to determine how far behind a student actually is in this formula.

To determine the benefit from using Beable, the platform's internal measure of Lexile was utilized. In past ROI evaluation from Osceola that have used a platform's internal metrics to determine growth, the ROI has come out as *higher* than when calculated using external measures such as NWEA data or FSA data. This is important to note as the ROI here may accordingly be reported as more beneficial that the true effect. Student Lexile's were mapped out to a student grade level, with "on grade level" set at Metametric's published 50%ile score, early grade-level based on the 40%ile (calculated using a percent-under-curve formula from the reported percentiles), and late grade-level based on the 60%ile (students under or above the forty-to-sixty range moved into different school years). This created a growth measure reported in thirds, similar to the "buckets growth" model used by the Florida Department of Education. Critically, only the students who were reported as having zero Lexile change were given a growth of zero under the model.

Out of the 4,537 students who logged in and were given a reported Lexile, 1,244 students experienced zero growth. A further 1,876 students experienced less than one years' growth, 1,262 students had one year of growth, 144 students had two years' growth, and eleven students had greater than two years' growth. Based on Beable's Lexile measure and Metametric's 50%ile grade level measure, when compared to students' actual grades, the students on Beable were approximately a combined 24,999 years behind expected proficiency. Combined across all students, Beable helped students close 2,741 years' worth of gaps. Based on the Osceola ROI formula, this is a benefit of $2,765,669 worth of avoided intervention course costs. With the total cost of $3,808,165 and following the ROI equation of benefit-minus-cost-divided-by-cost, we arrive at an ROI of -27%. This means that offering Intensive Reading with Beable is less cost-effective than simply offering traditional Intensive Reading courses, which aligns with the effect size finding as well.

This suggests that the cost of Beable does not provide a good return on the investment into the platform. Compared to the ROI of other platforms, such as Lexia, Beable is on the lower end of returns.

Conclusion

Based on the quantitative results, it appears that students who use Beable grow at a slower rate than students who do not use it. This is a worrying finding given that the Intensive Reading courses have traditionally held higher growth scores than the general courses, however there is a strong possibility that this difference may be due to either teacher effects of facets of the new FAST assessment. What is much more tangible of an outcome is that there was a net promoter score of -67, showing a sharp, negative perception from the teacher surveys. This poses a question as to the long-term success of Beable implementation in Osceola since it is likely that, unless strict implementation

measures are put into place, users of the platform have the potential to atrophy. This negative perception has not changed from the prior year, and, as was stated in 2022, is likely associated with the quantity of students that report little growth when using the platform (an issue continually cited in the feedback).

There is a wealth of evidence to support the fact that Beable works better for non-struggling students than for students who struggle with reading. The highest growth observed in Osceola was for students with starting Lexiles over 800; there was little correlation between Beable growth and standardized assessment growth yet the Lexile measure was strongly correlated with the standardized scale score; in traditionally higher performing schools, such as OCSA, HRMS, and CK8S, greater growth was observed; teacher feedback indicates that struggling students have great difficulty interacting with the platform and need personalized assistance to find success; and in the school where all students used the platform, the students receiving interventions grew less than students not receiving interventions. This makes a strong case that Beable is likely more effective as a Tier 1 support than a Tier 2 or 3 intervention, and explains the overall low performance of the platform. Indeed, the only students targeted for platform usage were students receiving Tier 2 and 3 interventions.

Beable still has a lower overall cost than similar programs, although the fact remains that since so few students use the platform in comparison to the amount paid, the effective cost makes it one of the most expensive digital interventions utilized by the district. As with last year, the platform still shows a slight, positive effect overall, but it was measured a bit lower than the year before (d = 0.08). Compared to the effect size of the Achieve 3000 intervention (measured at d = 0.42 within Osceola, and at d = 0.21 for students overall), Beable has a lower effect. Beable was also found to have a return on investment of -27%, as measured by their own internal growth measures. This is lower than other reading interventions platforms have been measured at in Osceola County in the past.

For the reasons of negative teacher perception, poor general user feedback, a small effect size on students, and a measured negative return on the investment in the platform, the continued usage of the Beable platform is not recommended.