kidaptive

Follow-Up Evaluation of the Impact of Sylvan Learning's Ace it! Program Within the Texas Alliance of Boys & Girls Clubs' Academic Innovation and Mentoring Program (Texas AIM)

An Independent Evaluation by Kidaptive, Inc.
for Sylvan Learning

November 9, 2018

Background

The Texas Academic Innovation and Mentoring Program, or Texas AIM, is a partnership between Sylvan Learning and the Texas Alliance of Boys & Girls Clubs to help Texan students succeed. One component of Texas AIM, called Ace it!, is specifically intended to support underperforming students. Through the Ace it! program, Sylvan Learning provides underperforming students with 30 hours of remedial tutoring support over the course of 10 weeks.

Two previous studies (Rockman et al, 2015; Rockman et al, 2017) have found the Ace it! program within Texas AIM to be consistently and significantly effective across subjects (Mathematics and Reading), grades (1st–8th grade), years (2012–2015), and assessments (Pearson Education's GMADE™ and GRADE™ assessments and Texas's STAAR assessment). Those studies are described briefly below; see References for full citations.

Study 1

In a preliminary report, Rockman et al (2015) found that students who were enrolled in the Ace it! program made statistically significant gains on two standardized assessments from Pearson Education, the GMADE™ and GRADE™ (testing Mathematics and Reading, respectively). Not only was this growth statistically significant, it was also rapid enough to raise Ace it! students' percentile rankings relative to a national norming sample from Pearson, indicating that the growth was higher than would be expected from students with *average* achievement levels—a strong result given that Ace it! students are by definition underperforming.

Study 2

In a more detailed investigation, Rockman et al (2017) worked with the Texas Education Agency (TEA) to obtain three years' worth of STAAR scores (i.e., pre, post, and follow-up scores) for approximately 400 Ace it! students. Using the STAAR data, researchers were able to show that Ace it! students' higher-than-expected growth on the Pearson assessments from Study 1 was mirrored by growth on the STAAR assessment that exceeded the average STAAR growth for at-risk students.

Study 3

The goal of the present study, conducted independently by Kidaptive, Inc., a learning and analytics company out of Stanford University, is to leverage the established relationship between the Pearson GMADE™ and GRADE™ scores (which are regularly administered as part of the Ace it! program and thus readily available) and STAAR scores (which must be requested from TEA and are partially masked to protect student privacy, thus providing less detailed information) to evaluate Ace it! students' academic growth relative not only to at-risk students in Texas but to typically performing students nationwide. The study also provides an evaluation of the Ace it! program's continued effectiveness through the 2015–16 and 2016–17

academic years, extending the range of Ace it! students whose outcomes have been studied to 2012–2017.

Data

Data Intake

Data Set 1: Pearson Education Group Mathematics/Reading Assessment and Diagnostic Evaluation (GMADE™/GRADE™)

Sylvan Learning provided Kidaptive with test scores from 3,186 de-identified students in grades 1–10 on two standardized assessments sold by Pearson Education: The Group Mathematics Assessment and Diagnostic Evaluation (GMADE™) and the Group Reading Assessment and Diagnostic Evaluation (GRADE™). Pretests were administered before participation in the 10-week Ace it! program; posttests were administered after completion of the program. Each student had both a pre- and a posttest score within the same school year for at least one of the two assessments (GRADE™ or GMADE™); 74 students had pre- and posttest scores within the same school year for both assessments.

Data Set 2: State of Texas Assessments of Academic Readiness (STAAR)

The Texas Alliance of Boys & Girls Clubs provided Kidaptive with test scores from 2,004 de-identified students in grades 3–8 on the State of Texas Assessments of Academic Readiness (STAAR), for both Mathematics and Reading. Each student had at least one STAAR score from a test taken in the spring of the school year in which that student participated in the Ace it! program; 264 students had an additional STAAR score from the same school year (indicating that the student retook the test).

Data Cleaning

Data Set 1

From the original set of Pearson assessment data for 3,186 de-identified students, Kidaptive removed data for the following reasons:

- Duplicate rows of data
- Data from retaken assessments (to be conservative, Kidaptive evaluates only first successful attempts at each assessment)
- Data with uninterpretable scores (e.g., impossible differences between raw scores and scaled scores, with no way to determine which is correct)
- Data with clear data-entry mistakes on grade level (i.e., students enrolled in Ace it! for both school years under evaluation but whose records showed a single grade for both

years—despite confirmation from Texas that no students in the sample repeated a grade—with no way to determine the students' actual grades in each of the two years)
- Data from students in grades above 6th (because those grades lack sufficient data to meaningfully evaluate outcomes)

After removing 194 students' data as part of this data cleaning, Kidaptive had Pearson assessment data for 2,992 students, distributed as follows across subjects, grades, and years (note: tables sum to 3,086 because 60 students participated in both subjects in the same school year and another 34 students participated in one subject in each of the two school years):

2015–16	1st grade	2nd grade	3rd grade	4th grade	5th grade	6th grade
GMADE™	35	96	142	139	67	17
GRADE™	89	179	218	211	178	52

2016–17	1st grade	2nd grade	3rd grade	4th grade	5th grade	6th grade
GMADE™	38	111	175	199	103	39
GRADE™	82	185	268	265	157	41

Data Set 2

From the original set of STAAR assessment data for 2,004 de-identified students, Kidaptive removed data for the following reasons:

- Data from retaken assessments (to be conservative, Kidaptive evaluates only first successful attempts at each assessment)
- Data with uninterpretable scores (e.g., scores of "NA" or a STAAR-supplied score code indicating an abnormal score)
- Data from students who took their tests in Spanish (because scores from Spanish-language versions of assessments are scaled differently from the corresponding English-language versions, making direct comparisons impossible)
- Data from students in grades above 6th (because those grades lack sufficient data to meaningfully evaluate outcomes)

After removing 78 students' data as part of this data cleaning, Kidaptive had STAAR assessment data for 1,926 students, distributed as follows across subjects, grades, and years (note: tables sum to 2,051 because 125 students were present in both of the two school years):

2015–16	3rd grade	4th grade	5th grade	6th grade
Math/Reading	351	289	220	51

2016–17	3rd grade	4th grade	5th grade	6th grade
Math/Reading	413	405	253	69

Data Merging

Before submitting data sets to Kidaptive, Sylvan Learning and the Texas Alliance of Boys & Girls Clubs collaborated to provide an arbitrary numeric identifier for every student by matching personally identifiable information wherever possible. This matching effort succeeded for some but not all students. Each of the two data sets (i.e., the Pearson-assessment data set from Sylvan Learning and the STAAR-assessment data set from the Texas Alliance of Boys & Girls Clubs) was given an additional data column with the new arbitrary numeric identifiers, and all personally identifiable information was removed from both data sets before their submission to Kidaptive.

After cleaning both data sets, Kidaptive merged the remaining student data where possible by joining student records that shared the same arbitrary numeric identifier. This merging of Pearson assessment data for 2,992 students and STAAR assessment data for 1,926 students only succeeded for 1,276 students. Kidaptive confirmed with Sylvan Learning and the Texas Alliance of Boys & Girls Clubs that the remaining students could not be matched across data sets due to missing or inconsistent identifying information. Kidaptive could not assume that the missing information was missing completely at random, a requirement known as MCAR in the statistical methodology literature. In the absence of the MCAR assumption, missingness might be correlated with outcomes. To give one example, students whose parents were separated or divorced might be more likely to have had inconsistent surnames, and parental relationship instability is plausibly related to academic achievement. Therefore, Kidaptive did not analyze relationships between Pearson assessment data and STAAR assessment data.

Analysis and Findings

Kidaptive had two main goals in this evaluation: first, to establish that the Ace it! students were in fact underperforming according to the state of Texas, and second, to determine whether participation in the Ace it! program was associated with higher-than-expected rates of academic growth, as found in Studies 1 and 2.

STAAR Analysis and Findings

To achieve the first goal, which is an evaluation of *achievement*, Kidaptive compared Ace it! students' STAAR scores against statewide means for at-risk students. As shown in Figure 1, Ace it! students' mean scores were lower than the statewide means for at-risk students in nearly all subjects, grades, and years, confirming that the Ace it! students were indeed underperforming relative to their peers.

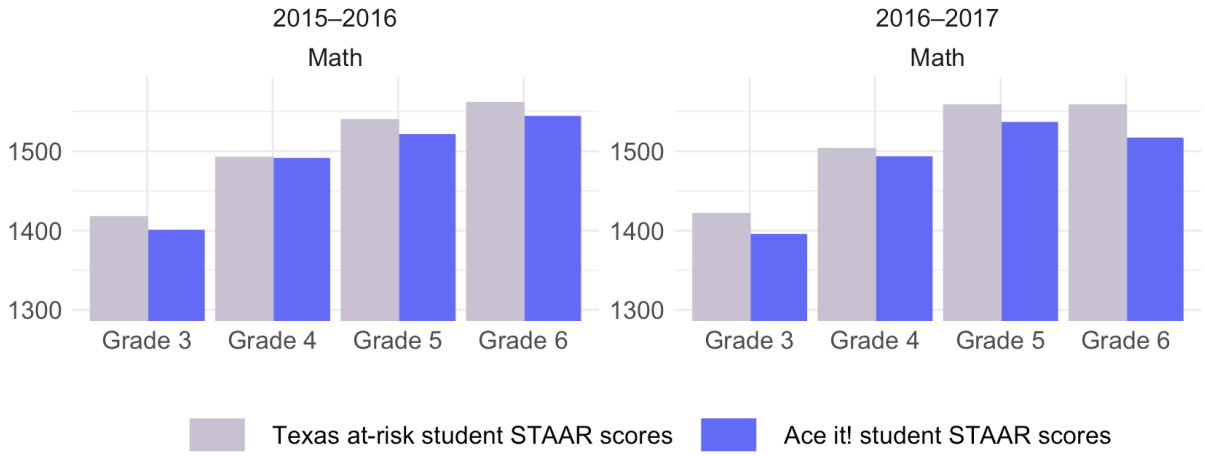

***Figure 1*. Performance of Ace it! students vs. Texas at-risk students in Mathematics.**

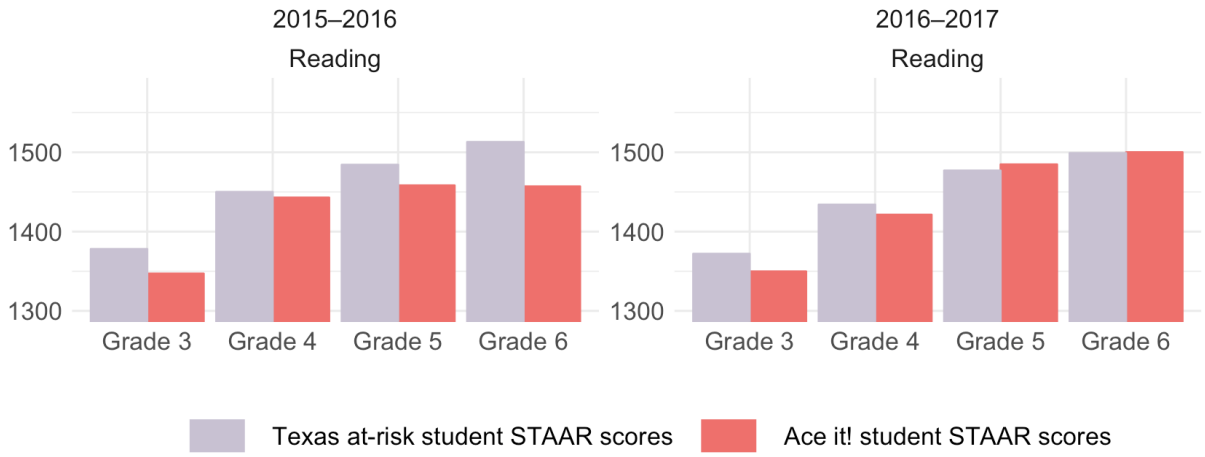

***Figure 2*. Performance of Ace it! students vs. Texas at-risk students in Reading.**

GMADE™/GRADE™ Analysis and Findings

To achieve the second goal, which was an evaluation of *growth*, Kidaptive analyzed the changes in students' Pearson-assessment scores from before and after their participation in the Ace it! program.

Estimating "Expected" Growth

One complicating factor for this analysis is that Pearson's Growth Scale Value (GSV) scores, the scaled scores that are reported for GMADE™ and GRADE™ assessments, are designed to map student growth to a single scale over time; therefore, a student taking the test later in the school year would be expected to score higher than that same student taking the test earlier in the school year as a result of natural growth and typical schooling experiences.

Because of this property of GSV scores, Kidaptive obtained national mean scores for the GMADE™ and GRADE™ assessments for all grades under investigation at two time points: fall and spring of the school year. Using these two time points, Kidaptive was able to construct a trajectory of "expected" growth against which to compare Ace it! students' actual growth. (See Appendix for details of this estimation process.) As shown in Figures 3–6, this comparison demonstrates that **Ace it! students started from lower levels of achievement but outperformed national rates of academic growth in both subjects (Mathematics and Reading), in all grades (1–6), and in both years under investigation (2015–16 and 2016–17)**.

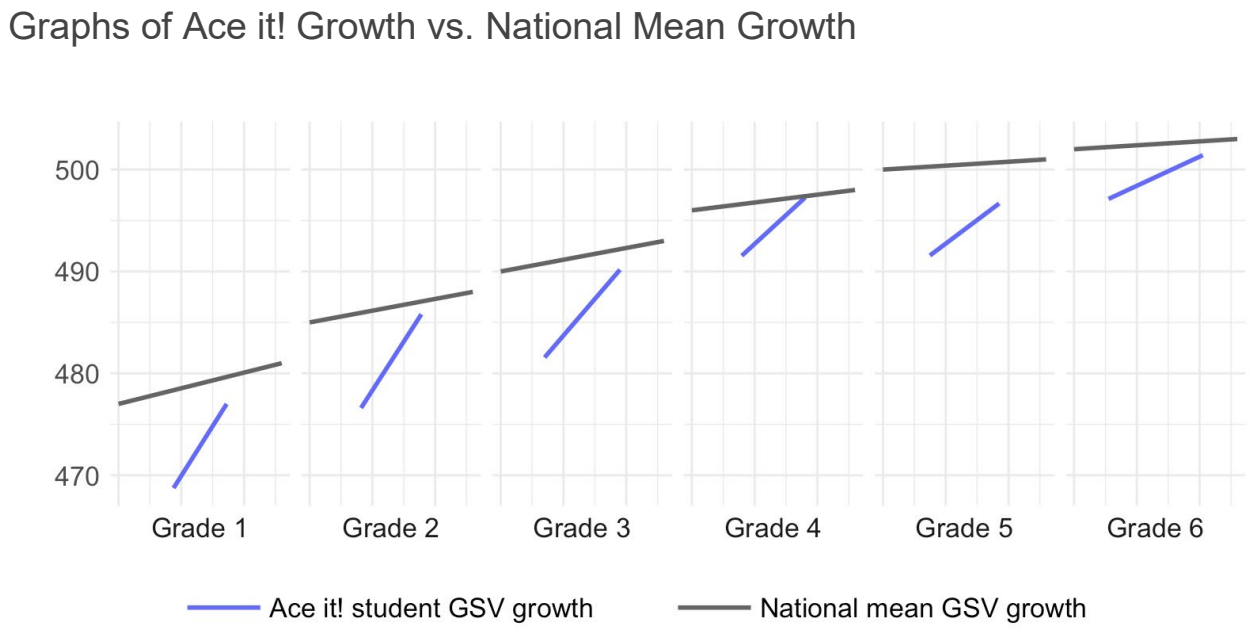

***Figure 3*. 2015–16 Growth of Ace it! students vs. national means in Mathematics.**

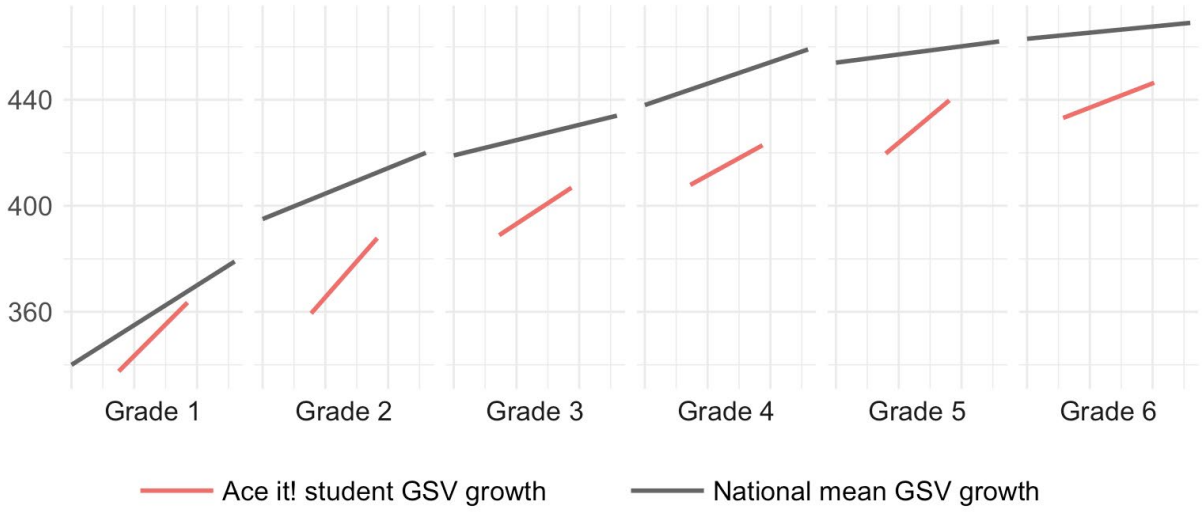

***Figure 4*. 2015–16 Growth of Ace it! students vs. national means in Reading.**

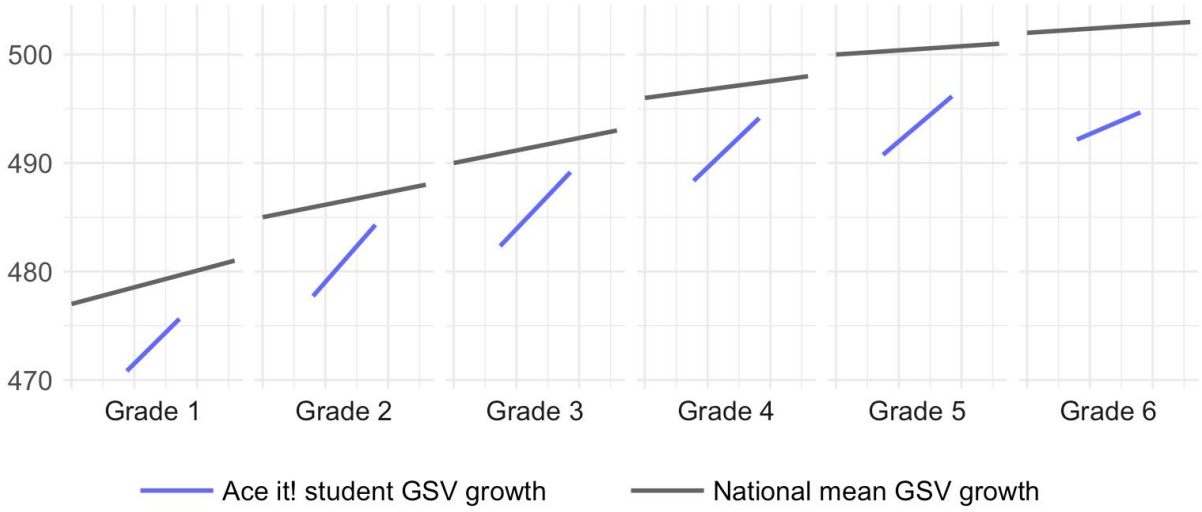

***Figure 5*. 2016–17 Growth of Ace it! students vs. national means in Mathematics.**

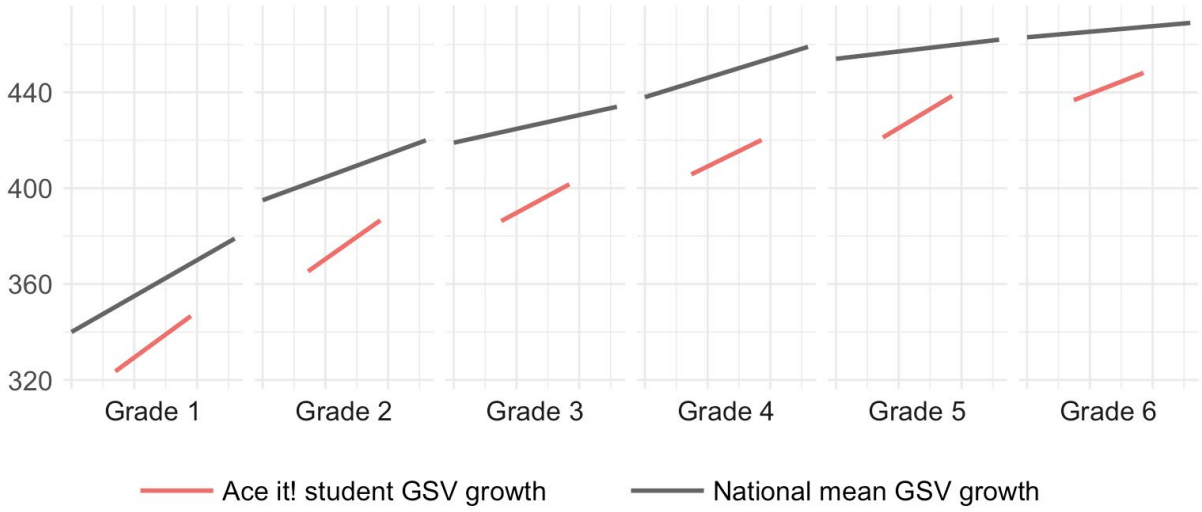

***Figure 6*. 2016–17 Growth of Ace it! students vs. national means in Reading.**

Statistical inference testing (one-tailed paired *t*-tests) confirmed that **Ace it! students' rates of academic growth were significantly higher than those reflected in the GSV national means in every subject (Mathematics and Reading), grade (1st–6th), and school year (2015–16 and 2016–17) tested by Kidaptive**, with all *p*-values < .05 even after correction for family-wise error rate using the Holm-Bonferroni method. As with the findings of Study 1, this outperformance is especially noteworthy given that Ace it! students are by definition underperforming, whereas the national mean scores obtained from Pearson include students at all levels of performance.

Conclusions

Kidaptive's analysis of STAAR assessment data revealed that in the 2015–16 and 2016–17 school years the Ace it! program continued to serve students who need extra support, and Kidaptive's analysis of Pearson assessment data revealed that in those school years students achieved higher-than-expected academic growth while enrolled in the Ace it! program. Taken together, these findings suggest that **the Ace it! program continues to be an effective intervention to support underperforming students in Texas**.

References

Rockman et al. (2015). Academic impact of the Ace it! component of TEXAS AIM: Preliminary Report. Retrieved November 9, 2018, from: http://sylvanresearchinstitute.com/pdf/TXAIMPreliminaryReport4.7.15%5B2%5D.pdf.

Rockman et al. (2017). The impact of the Ace it! component of the Texas Alliance of Boys & Girls Clubs' Academic Innovation & Mentoring Program. Retrieved November 9, 2018, from: http://www.sylvaneducationresearch.com/library/public/PDF-Folders/Case-Studies/TXAIMIntegratedReport-Final-Feb-2017-v-2.pdf.

Appendix

Estimation of "Expected" Growth From GSV National Means

Pearson's technical manuals for the GMADE™ and GRADE™ assessments indicate that standardization for the fall of the school year took place in September, October, and November of 2002 and that standardization for the spring of the school year took place in March, April, and May of 2003. Kidaptive established that the mean time between those two periods is 26 weeks (182 days), then subtracted the fall mean score for each grade from the spring mean score for that grade and divided the result by 26 to yield an estimated weekly growth for that grade.