A Practitioner’s Report for the Evaluation of the Achieve3000 Programs

November 3, 2015
Findings at a Glance

- Overall, Achieve3000 users’ learning gains were statistically significant and substantively important based on the What Works Clearinghouse (WWC) threshold of 0.25. This was true for all areas assessed: the GMRT-4 Vocabulary, Reading Comprehension, and Total Reading tests, as well as the LevelSet Lexile reading assessment.

- Across grades, Achieve3000 had a statistically significant positive impact on GMRT-4 Reading Comprehension and Total Reading when compared to study schools’ standard English Language Arts (ELA) curricula. However, the effect sizes did not meet the WWC threshold of 0.25 for being substantively important.

- Within-grades, Achieve3000 had positive impacts on sixth-grade reading that were not statistically significant but approached the WWC threshold of 0.25 for being substantively important. Furthermore, the program had substantively important positive impacts on ninth-grade GMRT-4 Vocabulary, Reading Comprehension, and Total Reading.

- Teachers generally reported that Achieve3000 was effective, and most teachers noted that they would use the program again next year.

Executive Summary

Achieve3000, the publisher of differentiated online literacy programs, understands the importance of demonstrating the efficacy of its products through independent evaluation. Therefore, it contracted with Magnolia Consulting, LLC, an independent evaluation consulting firm, to conduct this randomized control trial study of its Achieve3000 programs. Magnolia Consulting conducted this study with third-, sixth-, and ninth-grade teachers and students during the 2014/15 school year. This study sample of 46 teachers and 1,012 students came from 16 schools across four districts located in three different regions of the United States: the West South region, the East North Central region and the Pacific region. Two districts were classified as “Suburb: Large” and two districts were classified as “City: Large” (U.S. Department of Education, National Center for Education Statistics, 2015).

Study Design & Methods

To conduct this study, Magnolia Consulting evaluators used a randomized control trial in which teachers were randomly assigned to participate in either the treatment or comparison condition. Treatment teachers implemented the Achieve3000 program with their students, and comparison teachers implemented their usual ELA materials but not the Achieve3000 program. This design enabled evaluators to estimate an unbiased impact of Achieve3000 on student learning in reading.

This efficacy study used multiple student and teacher measures. Student measures included the Gates MacGinitie Reading Test, fourth edition (GMRT-4) and Achieve3000’s LevelSet. The GMRT-4 is a group-administered, norm-referenced assessment that yields scores for Vocabulary, Reading Comprehension, and Total Reading. Teachers in both study conditions administered this assessment to their students at the beginning of the school year as a pretest and at the end of the school year as a posttest. The LevelSet is an online assessment that uses
the Lexile Framework® to assess students’ Lexile reading level. Teachers in the treatment condition administered it to their students at the beginning and end of the school year. In addition to student measures, the study used multiple teacher measures: weekly treatment teacher online implementation logs, a spring comparison-teacher survey, and spring classroom observations of treatment and comparison teachers.

**Achieve3000 Program Implementation and Comparison Program Implementation**

**KEY FINDING:**
Overall, treatment teachers implemented the Achieve3000 program with moderate fidelity.

The study’s minimum implementation requirements, developed jointly by Magnolia Consulting and Achieve3000, asked teachers to implement the Achieve3000 program for at least 90 minutes per week. The analyses of implementation data from weekly implementation logs, observations and student usage data, revealed an average implementation score of 71% (out of a possible total of 100%) across treatment teachers. This score reflects real-world implementation variation due to competing district and state requirements, assessments, holidays, weather delays, technology issues, and other issues.

**Treatment Group Student Learning Results**

**KEY FINDING:**
As a group, students who used Achieve3000 during the 2014/15 school year demonstrated statistically significant and substantively important gains on the GMRT-4 Vocabulary, Reading Comprehension, and Total tests. They also demonstrated statistically significant and substantively important gains in LevelSet Lexile points.

Statistical analyses examining GMRT-4 gains showed that on average, treatment students who used Achieve3000 demonstrated statistically significant and substantively important learning gains in all areas assessed (see Table 1). More specifically, Achieve3000 users’ GMRT-4 Vocabulary, Reading Comprehension, and Total Reading test gains corresponded to effect sizes of 0.43, 0.47, and 0.48. Their average LevelSet Lexile gain corresponded to an average gain of 110 points.

**Table 1. Treatment Students’ GMRT-4 and LevelSet Pretest to Posttest**

<table>
<thead>
<tr>
<th>Outcome Variable</th>
<th>Coefficient</th>
<th>Standard Error</th>
<th>t-value</th>
<th>Approx. df</th>
<th>p-value</th>
<th>Effect Size</th>
</tr>
</thead>
<tbody>
<tr>
<td>GMRT-4 Vocabulary</td>
<td>19.82</td>
<td>3.32</td>
<td>5.97</td>
<td>22</td>
<td>&lt;.001*</td>
<td>0.43**</td>
</tr>
<tr>
<td>GMRT-4 Reading Comprehension</td>
<td>21.24</td>
<td>4.10</td>
<td>5.18</td>
<td>22</td>
<td>&lt;.001*</td>
<td>0.47**</td>
</tr>
<tr>
<td>GMRT-4 Total Reading</td>
<td>20.24</td>
<td>3.11</td>
<td>6.51</td>
<td>22</td>
<td>&lt;.001*</td>
<td>0.48**</td>
</tr>
<tr>
<td>LevelSet Lexile level</td>
<td>109.62</td>
<td>15.96</td>
<td>6.87</td>
<td>22</td>
<td>&lt;.001*</td>
<td>0.33**</td>
</tr>
</tbody>
</table>

* Statistically significant at the 0.05 level.
** Substantively important based on the WWC Standards.
Relationships between Treatment Teachers’ Implementation Fidelity of Achieve3000 and Student Learning Gains

The degree to which treatment teachers implemented Achieve3000 with fidelity varied. Within the range of implementation for this study, there were positive relationships between implementation fidelity and learning gains, but they were not statistically significant for the GMRT-4 Vocabulary, Reading Comprehension, or Total Reading Tests. However, the relationship between implementation fidelity and Lexile point gains was statistically significant, with implementation fidelity increases of 10% associated with average gains of 31 Lexile points.

Student Learning Results Comparing Treatment and Comparison Groups

**KEY FINDING:**
Achieve3000 had a statistically significant positive impact on GMRT-4 Reading Comprehension and Total Reading performance. It did not have a statistically significant impact on GMRT-4 Vocabulary performance.

As shown in Table 2, below, statistical analyses comparing reading performance of Achieve3000 users and non-users showed that Achieve3000 had a statistically significant positive impact on students’ GMRT-4 Reading Comprehension and Total Reading performance. The respective effect sizes of 0.22 and 0.20 approached the WWC standards for substantively important effects. Achieve3000 did not have a statistically significant or substantively important impact on students’ GMRT-4 Vocabulary performance.

<table>
<thead>
<tr>
<th>Outcome Variable</th>
<th>Coefficient</th>
<th>Standard Error</th>
<th>t-value</th>
<th>Approx. df</th>
<th>p-value</th>
<th>Effect Size</th>
<th>Improvement Index</th>
</tr>
</thead>
<tbody>
<tr>
<td>Vocabulary</td>
<td>5.04</td>
<td>3.57</td>
<td>1.41</td>
<td>31</td>
<td>.17</td>
<td>0.12</td>
<td>0.05</td>
</tr>
<tr>
<td>Reading Comprehension</td>
<td>9.49</td>
<td>4.55</td>
<td>2.09</td>
<td>31</td>
<td>.045*</td>
<td>0.22</td>
<td>0.09</td>
</tr>
<tr>
<td>Total Reading</td>
<td>7.76</td>
<td>3.44</td>
<td>2.26</td>
<td>31</td>
<td>.03*</td>
<td>0.20</td>
<td>0.08</td>
</tr>
</tbody>
</table>

* Statistically significant at the 0.05 level.

Exploratory Analyses Comparing Achieve3000 Impacts by Grade

**KEY FINDING:**
Findings suggest that the impact of Achieve3000 on reading performance varied by grade, with the greatest impacts on ninth-grade reading.

Exploratory analyses conducted to examine the impacts of Achieve3000 separately for each grade revealed differences by grade. On average, third-grade students who participated in Achieve3000 performed similarly to comparison-group students who used their schools’ typical literacy programs (with effect sizes of -0.02, 0.02, and 0.06 for Vocabulary, Reading Comprehension, and Total Reading). For sixth-grade students, there were no statistically
significant differences in posttest GMRT-4 scores by study condition, but the effect sizes of 0.21, 0.22, and 0.22 for Vocabulary, Reading Comprehension, and Total Reading favored Achieve3000 users and approached the WWC standards for substantively important effects. Finally, although there were also no statistically significant differences in average ninth-grade treatment and comparison-group posttest reading performance, the effect sizes of 0.28, 0.51, and 0.44 for Vocabulary, Reading Comprehension, and Total Reading favored Achieve3000 users and exceeded WWC standards for substantively important effects. Because these subgroup analyses had smaller sample sizes and less statistical power than main analyses to detect effects, readers should use caution when interpreting the statistical significant of findings.

**Exploratory Analyses Comparing Achieve3000 Impacts for ELL Students**

**KEY FINDING:** Findings suggest that ELL students who used Achieve3000 performed similarly on the GMRT-4 as ELL students who used their schools’ typical literacy programs.

Exploratory analyses showed that among ELL students, Achieve3000 did not have a statistically significant or substantively important impact on Reading Vocabulary, Reading Comprehension, or Total Reading. These findings suggest that on average, ELL students who used Achieve3000 performed similarly to comparison-group ELL students who used their schools’ typical literacy programs. Readers should interpret these findings with caution because of the small sample sizes for this ELL subgroup.

**Comparisons of Teacher Program Perceptions across Study Conditions**

**KEY FINDING:** Treatment teachers described the Achieve3000 program as having higher student engagement, having the appropriate amount of materials to cover, and more applicable pacing than comparison programs.

Treatment teachers generally reported that the Achieve3000 program components were useful and comprehensive, and they described many benefits to the program. However, some treatment teachers also expressed frustration with various aspects of the program, such as the repetitive routine and time it took away from core instruction. Comparisons of perceptions across conditions showed that treatment teachers often described Achieve3000 as having higher student engagement, having the appropriate amount of materials to cover, and more suitable pacing than comparison programs. According to study teachers, Achieve3000 more adequately or very adequately supported below-level, on-level and advanced-level readers, English Language Learners and special education students than comparison programs. For student skills, Achieve3000 more effectively supported building academic vocabulary, comprehending complex text, and critically evaluating informational text than comparison programs, while comparison programs more effectively supported reading fluency.
Study Limitations

This rigorous study had some limitations worth noting. First, findings only generalize to schools that met the participation requirements for this study, which included specific technology access and infrastructure as well as multiple ELA teachers at participating grades. Because two teachers who were uncomfortable with the program’s technology requirements dropped out, it is unclear how teachers who are less skilled with technology might implement the program, and if that would have impacted study findings. It is possible that if teachers had received additional training earlier in the study, their implementation fidelity might have increased, which could have impacted the study findings. Finally, readers should use caution when interpreting within-grade subgroup analyses, as the relatively smaller sample sizes limited the statistical power of the analyses to detect effects.

Summary and Conclusions

Results from this 2014/15 evaluation of Achieve3000 showed that on average, treatment teachers implemented the program with moderate fidelity. Students who used Achieve3000 demonstrated statistically significant and substantively important gains on the GMRT-4 Vocabulary, Reading Comprehension, and Total Reading tests, as well as in their LevelSet Lexile levels. Comparisons of students who used Achieve3000 with students who used the schools’ standard ELA programs showed that overall, Achieve3000 had a statistically significant but not substantively important impact on GMRT-4 Reading Comprehension and Total Reading. Within-grade analyses showed that Achieve3000 had impacts on sixth-grade reading that were not statistically significant but approached the WWC threshold for being substantively important, and it had substantively important impacts on ninth-grade GMRT-4 Vocabulary, Reading Comprehension, and Total Reading. Finally, ELL subgroup analyses revealed no statistically significant differences or substantively important effect sizes regarding the impact of Achieve3000 on ELL students’ GMRT-4 Reading Vocabulary, Reading Comprehension, or Total Reading. Overall, treatment teachers expressed many positive perceptions regarding Achieve3000. Although several teachers shared suggestions for improvement, most teachers noted that that they would continue using the program.
An Efficacy Study of the Achieve3000 Programs

WHO WAS IN THE STUDY?

16 46 1012

During the 2014/15 school year, 16 schools in four suburban and city districts across the U.S (the West South region, the East North Central region and the Pacific region) implemented Achieve3000. The study included 46 teachers and 1012 students.

HOW DID THEY IMPLEMENT THE PROGRAMS?

Treatment Teachers implemented the Achieve3000 program for 2 days per week, averaged 90 minutes on the program, and at least one lesson.

Comparison Teachers implemented their existing ELA programs.

WHAT DID THEY SAY?

“I enjoyed that each article was pushed out at their own level. I thought the articles were interesting and so did the students. I loved the growth reports and the Sunday report that was sent. Overall one of the best programs I have worked with!”
Teacher Quote

“Treat teachers disliked the monotony of the program routine, the amount of time the program took away from their core curriculum, the brevity of the training, program navigation, and technology issues.

“What were the Impacts on Student Achievement?

ACHIEVE 3000 STUDENTS
Statistically significant and substantively important learning gains on the GMRT-4 and LevelSet assessments

ACHIEVE 3000 VS. COMPARISON
Statistically significant positive impact on GMRT-4 Reading Comprehension and Total Reading when compared to study schools’ standard ELA curricula.

ACHIEVE 3000 BY GRADE
Substantively important positive impacts on ninth-grade GMRT-4 Vocabulary, Reading Comprehension, and Total Reading.