

## The Impact of Achieve3000 on Elementary Literacy Outcomes: Randomized Control Trial Evidence, 2013-14 to 2014-15

*Darryl V. Hill and Matthew A. Lenard*

This evaluation reports the two-year results from a randomized controlled trial (RCT) of Achieve3000 in the Wake County Public School System (WCPSS). Achieve3000 is an early literacy program that differentiates non-fiction reading passages based on individual students' Lexile scores. The driving force behind implementing Achieve3000 in particular and focusing on literacy in general was North Carolina's 2012 Read to Achieve legislation, which mandated that students failing to meet proficiency standards in grade 3 reading must attend summer camp and pass the state's Read to Achieve test before being promoted to grade 4. WCPSS Academics staff were eager to launch a program that could help increase reading proficiency among students receiving core instruction. After a pilot in 12 schools in 2012-13, Achieve3000 was randomly assigned to 16 elementary schools for the 2013-14 (hereafter 2014) school year, which included approximately 8,000 students in grades 2 through 5. The program continued for a second year in 2014-15 (hereafter 2015) and was available to roughly 18,000 students over the course of the two-year implementation. Program staff decided to continue implementation into a third year (2015-16). In order to measure implementation fidelity, we benchmarked district usage against activity-completion categories established by the vendor. To measure impact, we estimated how well students in schools using Achieve3000 performed on various Lexile measures compared to their counterpart schools, which did not have access to Achieve3000.

### *Abstract*

In 2013-14, the Wake County Public School System (WCPSS) launched Achieve3000 as a randomized controlled trial in 16 elementary schools. Achieve3000 is an early literacy program that differentiates non-fiction reading passages based on individual students' Lexile scores. Two-year results show that Achieve3000 did not have a significant impact on student outcomes. However, both intent-to-treat and treatment-on-treated estimates show that in 2015, the second year of implementation, students in the treatment group outperformed their control-group counterparts by 0.13 standard deviation units (SD) on the year-end Achieve3000 LevelSet Lexile test. This effect size is consistent with mean empirical effect sizes reported by Lipsey et al. (2012). Yet in neither the pooled nor annual results did Achieve3000 significantly impact student performance on additional Lexile outcomes (EOG or DIBELS ORF). Both implementation and impact results for Achieve3000 suggest that the ability of this particular technology-based literacy solution to improve student performance beyond that of a control group fell short of vendor-defined and empirical expectations.

### Table of Contents

Achieve3000 Background	3
Program Launch	4
Descriptive Information	7
Implementation Findings	9
Analytic Strategy	14
Student Achievement Impacts	15
Conclusions and Recommendations	18
References and Appendices	21

Specifically, in evaluating whether Achieve3000 was an effective resource for increasing student achievement over the two-year study period, we examined the following research questions:

- Was Achieve3000 implemented with fidelity across the 16 treatment schools that were offered the program in 2014 and 2015?
- Did students who received the offer of Achieve3000 outperform students who did not (measured with intent-to-treat impacts)?
- Did students who used Achieve3000 outperform students who did not use the program (measured with treatment-on-treated impacts)?
- Did the performance of students who used Achieve3000 differ by grade or by subgroup?

The findings from this evaluation revealed the following:

**Implementation of Achieve3000 improved slightly in 2015, but was weak in both years.**

Benchmark usage data provided in Achieve3000's 2012 and 2015 National Lexile Reports suggested that roughly 25% of students, at best, were able to complete at least 80 activities—the goal agreed to by the district and vendor. In those reports, this level was correlated with higher Lexile levels compared with lower activity ranges (1-39 and 40-79). In 2014, fewer than 10% of WCPSS students met this threshold and approximately 10% met it in 2015. These activity completion rates fell far short of expectations. Implementation was relatively consistent across grade levels, although grades 3 and 4 had slightly higher percentages of students meeting the 80-activity compared with their counterparts in grades 2 and 5.

- *Recommendation:* Staff should maintain modest expectations in the face of aggressive program usage goals. The district should establish an overarching implementation fidelity framework and, at minimum, each large program should have an associated implementation team that monitors usage and progress at least monthly. District leadership members, such as assistant and area superintendents, should be aware of efforts to monitor usage especially when their schools are included in a treatment group.

**Achievement goals failed to meet district, vendor, or empirical standards.** District staff selected Achieve3000 with the hope that, at minimum, it would contribute to student achievement gains when compared to the control group. In the two-year combined results, this did not occur for any of the three Lexile outcomes of interest (Achieve3000 Lexile, EOG Lexile, and DIBELS ORF Lexile), nor in 11 of the 12 different ways staff estimated impacts—by year, outcome, and model (two years, three outcomes, and two models for 12 total specifications). The lone exception was the positive and significant impact on the vendor's LevelSet Lexile score in 2015, a result that was in-line with empirical estimates (0.13 *SD*). This amounted to 31 Lexile points, which fell short of Lexile gains promoted by the vendor and expected by district staff.

- *Recommendation:* Staff should develop a roster of evidence-based programs that can be deployed when new core and supplemental resources are needed. At the time of adoption, Achieve3000 did not have existing experimental evidence (see Table 1 for definitions) of effectiveness to support its use in the district. Indeed, few classroom-based technology products currently do, which is why staff appropriately launched the program through the use of a randomized controlled trial (RCT). Now that causal impact results are available, staff should consider alternative programs in the event that additional years of outcomes data (i.e., 2016) fail to meet expectations.

**Achieve3000 positively impacted select subgroups.** While Achieve3000 did not have an impact on limited English proficient (LEP) students, students with disabilities (SWD) and academically and intellectually gifted (AIG) students appeared to benefit to a small degree from the program compared with their non-identified peers. These impacts were small compared with empirical average effects.

- *Recommendation:* SWD and AIG program staff should investigate why these subgroups may have benefited from Achieve3000. But they should exercise caution when making programming decisions on the basis of these effects which, while statistically significant, were small in magnitude.

**Table 1**  
**Nature of the Data Provided and Valid Uses**

Research Design	Conclusions that Can be Drawn
<input checked="" type="checkbox"/> Experimental	We can conclude that the program or policy caused changes in outcomes because the research design used random assignment.
<input type="checkbox"/> Quasi-Experimental	We can reasonably conclude that the program or policy caused changes in outcomes because an appropriate comparison strategy was used.
<input type="checkbox"/> Descriptive <input type="checkbox"/> Quantitative <input type="checkbox"/> Qualitative	These designs provide outcome data for the program or policy, but differences cannot be attributed directly to it due to lack of a comparative control group.

Sources: List, Sadoff, & Wagner (2011) and What Works Clearinghouse (2014).

## Achieve3000 Background

In 2012, the North Carolina General Assembly passed House Bill 950, which enacted the statewide *Read to Achieve* (R2A) program.<sup>1</sup> The act's goal is to ensure that students become proficient in reading by the end of grade 3 or else be promoted to grade 4 pending the successful completion of a summer reading camp. In order to prepare for the 2013-14 school year in the era of R2A, district staff bolstered efforts to implement programs and structures in an attempt to help students clear this proficiency hurdle. A large part of this effort was devoted to strengthening implementation of the district's early literacy framework called The Daily Five, through which students participate in five different types of literacy tasks: (1) reading to self; (2) reading to someone else; (3) listening to reading; (4) working on phonics or vocabulary skills; or (5) working on writing (Rhea, 2012). Achieve3000 emerged as a core instructional resource that could be used across any or all of these five literacy tasks.

The Achieve3000 suite includes six separate computer-based applications, four of which — KidBiz3000, TeenBiz3000, Empower3000, and Spark3000 — the company claims are inspired by the work of R.C. Anderson on prior knowledge, Carol Ann Tomlinson on differentiation, Michael Kamil on the role of technology, and Linda Duncan on vocabulary development (Achieve3000, 2011). Achieve3000's *theory of action* is that students will become college and career ready if they are able to read non-fiction texts at Lexile levels that exceed 1,350. To help students reach that level, the company's applications administer an assessment that establishes a

<sup>1</sup> Complete bill text: <http://www.ncleg.net/gascripts/BillLookUp/BillLookUp.pl?BillID=H950&Session=2011>

baseline Lexile level. From there, students are exposed to non-fiction adaptive reading passages that are aligned to their Lexile levels and adjust on the basis of end-of-lesson assessments. This way, reading passages are custom tailored to each student's Lexile level so that students do not spend valuable instructional time reading passages that are neither too easy nor too difficult (Achieve3000, 2011).

Achieve3000 outlined its goals for elementary school student performance in a national study it released in 2012 and updated in 2015 (Achieve3000, 2012). The initial analysis of more than 90,000 elementary school students in 1,291 schools concluded that students using the elementary school suite, KidBiz3000, experienced Lexile growth of 124 points, which was 46 more Lexile points over the course of an academic year compared with "normal" growth of 78 Lexile points. "Normal" growth is defined by MetaMetrics, the Durham, NC-based developer of the LevelSet Lexile assessment, which Achieve3000 uses to establish baseline Lexile levels (Williamson, 2006). Thus, students using Achieve3000 should conceivably outperform their peers by a growth factor of roughly 1.5 on Lexile levels. In its updated 2015 benchmark study of nearly 300,000 elementary school students, Achieve3000 focused on the relationship between Lexile growth and program use. Dividing "use" into three categories — 1-39 activities completed (low use), 40-79 (moderate), and 80 or above (high use)—the company concluded that students completing at least 80 activities would expect to see Lexile growth exceed "normal" growth by 72 points (Achieve3000, 2015). Taken together, the results from these benchmark studies suggest that, on average, students using Achieve3000 should expect to outgrow their peers by 46 points and, in cases where implementation fidelity is strong, by 72 points.

Staff from the district and the vendor jointly decided that in order to reach the annual goal of 80 completed activities, students would utilize KidBiz3000 twice weekly for 30 minutes. Upon initial use at the beginning of each school year, students would take a 30-minute assessment to obtain a baseline Lexile score. For each activity, students would follow a five-step procedure: (1) take a poll and respond to it through the KidBiz3000 email application; (2) read a non-fiction article aligned with their current Lexile level; (3) complete a series of multiple choice questions; (4) vote in a post-reading poll; and (5) answer a "Thought Question."

### Program Launch

Achieve3000 was first piloted in WCPSS during the 2012-13 (hereafter 2013) school year based on voluntary adoption by principals in 12 schools across the three school levels, including seven elementary schools, two middle schools, and three high schools. WCPSS Data, Research and Accountability (DRA) department staff conducted a preliminary examination of usage data, which suggested the program was fairly well-implemented across the 12-school sample. Following the results of these analyses, and after receiving positive feedback from pilot school principals and teachers, the Senior Director for Elementary Programs (SDEP) decided to expand the implementation of Achieve3000 during the 2013-14 school year (hereafter 2014).

In contrast with the voluntary adoption of Acheive3000 in 2013, district leadership decided to implement the program using random assignment in 2014. The decision to randomly assign Achieve3000 to schools was motivated by the district's Enhancing Data Use (EDU) framework, which requires instructional programs and policies to be supported by rigorous evidence or,

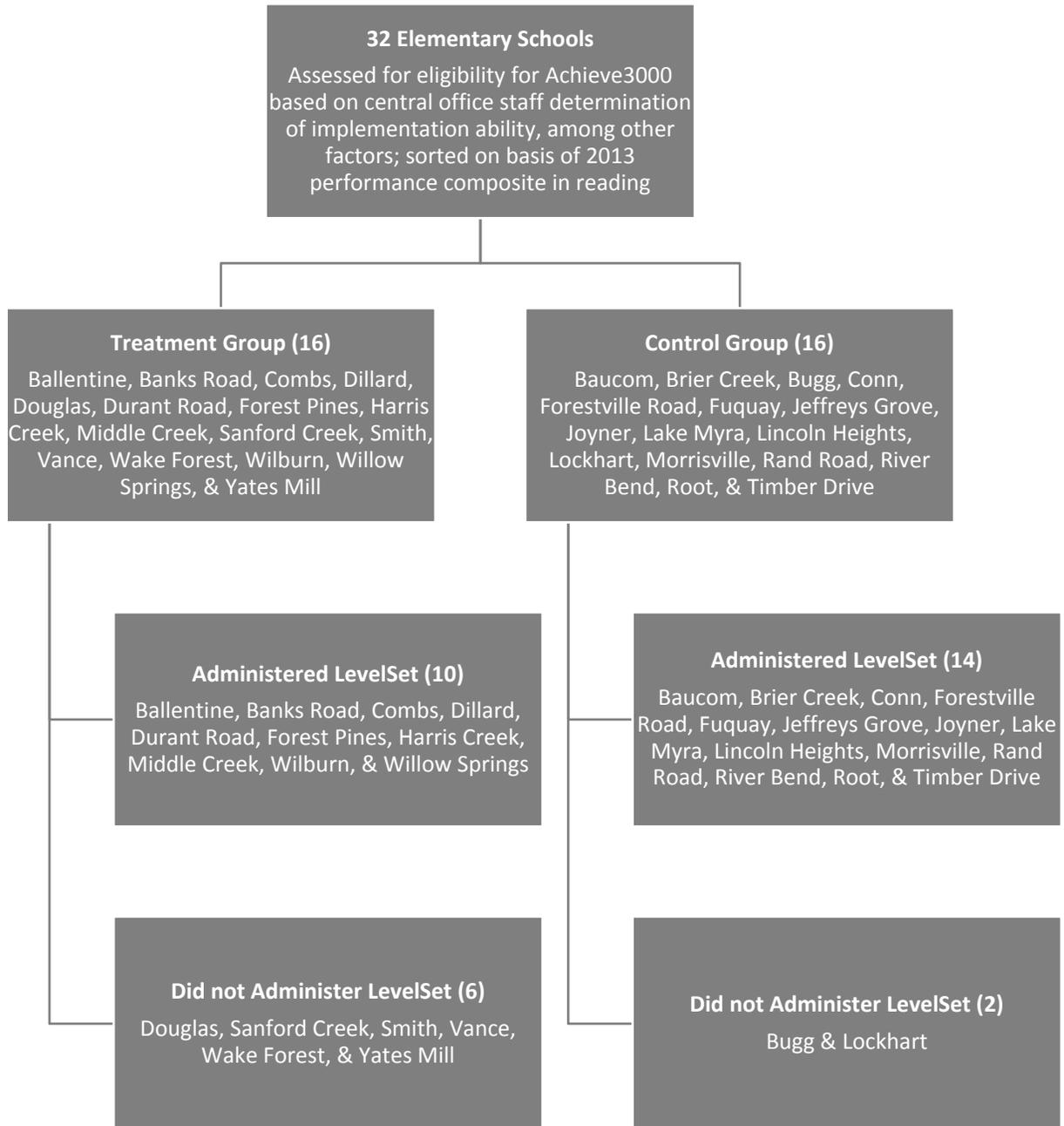
where existing evidence is nonexistent or weak, implemented in a way that generates rigorous evidence. Since neither independent researchers nor the vendor had produced rigorous, experimental evidence supporting the effectiveness of Achieve3000, the program was implemented through the use of a randomized control trial (RCT).

Using the strategy of random assignment allowed program staff to assign a program to a “treatment” group of schools that would ultimately be compared to a “control” group of similar schools. The major benefit of an RCT is that it creates two groups of schools that are similar on the basis of observed and unobserved characteristics. While a group of schools that volunteers for a program can be ultimately compared to a comparison group, there is no guarantee that unobservable factors—e.g., parental involvement, student motivation, and school leadership—will not influence the results. Randomization accounts for both observable *and* unobservable factors such that any difference in outcomes between the treatment and control groups can be attributed to the program itself and not to any other factor.

The process of randomly assigning Achieve3000 to schools was relatively straightforward. Motivated by a new statewide and district focus on elementary literacy, Academics Department staff decided to limit implementation to elementary schools. In order to evaluate impacts, staff determined that at least 12 schools had to receive the program. In order to generate a population of schools eligible for the program, Academics and DRA evaluation staff began with the district’s then 105 elementary schools and removed from consideration the 12 Achieve3000 2013 pilot schools, as well as schools that used SuccessMaker, another commercial product with similar anticipated effects on elementary literacy outcomes. The SDEP then notified principals at the remaining schools about the opportunity to implement Achieve3000 in 2014. Approximately 40 principals replied that they were interested. After discovering a few unanticipated fidelity challenges, the list was further reduced to 32, from which 16 schools were randomly assigned to receive Acheive3000; the remaining 16 schools were assigned to the control group (Figure 1). Schools were first sorted pairwise on the basis of their 2013 reading performance composites, and from within each pair, one school was selected to belong to the treatment group.

Figure 1 shows that 24 of 36 schools administered the LevelSet Lexile test—required to begin using the program—in fall 2013. This included 10 of 16 treatment schools and 14 of 16 control schools. Statistical techniques described below permitted staff to account for the differential rates of program adoption in the treatment and control groups.

**Figure 1**  
**Random Assignment of Schools to Treatment and Control Conditions**



Sources: WCPSS administrative records and Achieve3000 LevelSet Lexile assessment results.

In establishing whether Achieve3000 was an effective resource for increasing student achievement, we examined the following research questions:

- Was Achieve3000 implemented with fidelity across the 16 treatment schools that were offered the program in 2014 and 2015?
- Did students who received the offer of Achieve3000 outperform students who did not (measured with intent-to-treat impacts)?
- Did students who used Achieve3000 outperform students who did not use the program (measured with treatment-on-treated impacts)?
- Did the performance of students who used Achieve3000 differ by grade or by subgroup?

### Descriptive Information

The 32 school leaders who volunteered to join the experimental sample represented a diverse array of elementary schools in terms of prior achievement, demographics, and geography. Because our primary outcomes of interest were academic in nature, we paired schools on the basis of their 2013 elementary school End-of-Grade (EOG) reading composite score. Within each pair, one school was randomized to the treatment condition — receiving access to Achieve3000 — and one was randomized to the control condition through the use of a computer-based random number generator. Data for this study came from the district’s administrative and testing records, Amplify, Inc.’s *mClass* reporting system, and Achieve3000’s activity completion and LevelSet Lexile pre- and post-test assessment results.

Before presenting implementation and impact findings, descriptive information on baseline balance appears below. Essential to a successful RCT is baseline equivalence between treatment and control groups prior to program implementation. If the two groups differ too widely on observable characteristics, adjustments must be made during the analysis stage in order to account for these differences. According to What Works Clearinghouse’s (WWC) *Rules and Procedures Handbook* (Clearinghouse, 2014), no adjustment is needed where differences between the two groups are less than 0.05 standard deviation units (*SD*).

Table 2 compares mean characteristics across a host of variables, as well the degree of balance between the treatment and control groups at the school-level, which was the level of randomization. Column 3 in Table 2 demonstrates that randomization at the school level resulted in a successful balancing of all school-type characteristics, and all but one demographic characteristic. This one instance of imbalance occurred for black students, in which the control group had 11% more Black students than the treatment group ( $p < .05$ ). We adjusted for this difference in our analysis stage using WWC guidance. Table 3 compares these same means and balance at the student-level, which closely resemble corresponding school-level balance.

**Table 2**  
**School-Level Descriptive Statistics and Balance across Control and Treatment Groups**

	Control	Treatment	C – T	SE	p-value
<b>Student characteristics</b>					
Male	0.512	0.511	0.002	0.007	0.804
Black	0.314	0.206	0.108**	0.053	0.050
Hispanic/Latino	0.171	0.203	-0.032	0.031	0.313
LEP	0.080	0.106	-0.027	0.019	0.165
SWD	0.118	0.112	0.006	0.010	0.536
Economically Disadvantaged	0.370	0.356	0.014	0.052	0.796
AIG: Reading & Math	0.067	0.064	0.003	0.015	0.831
Magnet School	0.250	0.188	0.063	0.151	0.681
Year-Round Calendar	0.438	0.563	-0.125	0.181	0.495
Title I	0.625	0.688	-0.063	0.173	0.721
<b>Baseline achievement</b>					
Achieve3000 Lexile Pretest	-0.083	0.006	-0.089	0.103	0.391
EOG Lexile Pretest	-0.012	-0.028	0.016	0.095	0.869
DIBELS Lexile Pretest	-0.038	-0.016	-0.022	0.099	0.826
N	16	16			
Proportion	0.500	0.500			

\*  $p < .10$ ; \*\*  $p < .05$ ; \*\*\*  $p < .01$

Note: C – T: Control mean minus Treatment mean; SE: standard errors; Lexile scores expressed as standardized values.

**Table 3**  
**Student-Level Descriptive Statistics and Balance across Control and Treatment Groups**

	Control	Treatment	C - T	SE	p-value
<b>Student characteristics</b>					
Male	0.512	0.511	0.001	0.007	0.856
Black	0.314	0.207	0.108**	0.052	0.038
Hispanic/Latino	0.171	0.203	-0.032	0.031	0.300
LEP	0.080	0.106	-0.026	0.018	0.152
SWD	0.118	0.117	0.001	0.010	0.563
AIG: Reading & Math	0.067	0.064	0.003	0.015	0.823
<b>Baseline achievement</b>					
Achieve3000 Lexile Pretest	-0.082	0.006	-0.089	0.101	0.378
EOG Lexile Pretest	-0.012	-0.027	0.016	0.094	0.868
DIBELS Lexile Pretest	-0.038	-0.016	-0.022	0.097	0.823
N	16,619	18,013			
Proportion	0.480	0.520			

\*  $p < .10$ ; \*\*  $p < .05$ ; \*\*\*  $p < .01$

Note: C – T: Control mean minus Treatment mean; SE: robust standard errors; Lexile scores expressed as standardized values; student-level means calculated using two-level random effects regression with robust standard errors.

The primary outcomes of interest are the Achieve3000 LevelSet Lexile test, EOG Lexile, and DIBELS Oral Reading Fluency (ORF) Lexile equivalent. Tables 2 and 3 report the mean baseline standardized scores of these outcome variables for treatment and control groups at the school- and student-levels. Because Achieve3000 Lexile scores were missing for roughly one-third of the sample — too many to constitute a reliable outcome measure (see Table 4, below) — we used additional outcome measures, including Lexile equivalents for EOG and the DIBELS ORF assessment. The North Carolina Department of Public Instruction (NC-DPI) contracts with MetaMetrics, Inc., the Lexile framework developer, to create an EOG-Lexile crosswalk (MetaMetrics, 2009). The company also provides conversion formulas in order to create a crosswalk between DIBELS ORF raw scores and a corresponding Lexile score for students in grades 2-3. See Appendix A for more information about the conversion and correlations between measures.

### Implementation Findings

Before the RCT began, principals who expressed interest in potentially receiving Achieve3000 through random assignment were asked to commit to the following five guidelines in order for their schools to be considered to receive the program:

- Require all 2nd through 5th grade students to use the program for at least 30 minutes, twice weekly;
- Ensure that students followed Achieve3000’s “five-step literacy routine” (see "Achieve3000 Background");
- Utilize the writing component and the math problem associated with each article;
- Integrate Achieve3000 with core instruction and intervention only as appropriate; and
- Implement the program within The Daily Five structure in a way that meets the needs of teachers and students.

To monitor implementation throughout the two-year study period, the Academics staff led the creation an implementation team of roughly a dozen individuals, which included staff from Academics, Student Support Services, the former Office of School Performance, and two members of Achieve3000’s North Carolina team. The district implementation team met monthly to discuss LevelSet Lexile assessment administration, activity completion rates, school-based initiatives designed to increase program usage, and communication strategies, among other issues. In addition, each of the 16 treatment schools was asked to name an “Achieve3000 Leader” who would become the point of contact for the district implementation team and the vendor’s North Carolina staff, who visited schools regularly to monitor implementation. This evaluation measures implementation fidelity in two ways.

First, students in both treatment and control groups were required to take Achieve3000’s LevelSet Lexile pre-test so that the two groups could be compared on the major outcome of interest. Students in the treatment group were required to complete the pre-test in order to begin using Achieve3000. As Figure 1 (above) shows, 10 of 16 treatment schools and 14 of 16 control schools administered this assessment in 2014. Table 4 shows the student-level counts and rates of pre- and post-test completion in each of the two study years. From 2014 to 2015, a higher percentage of students in both groups completed pre- and post-test LevelSet Lexile assessments. However, while the percentage of treatment group students completing the pre-test increased

from 60% in 2014 to 79% in 2015, this still reflected that roughly a fifth of students in treatment schools did not use the program.

**Table 4**  
***Achieve3000 LevelSet Lexile Assessment Completion***

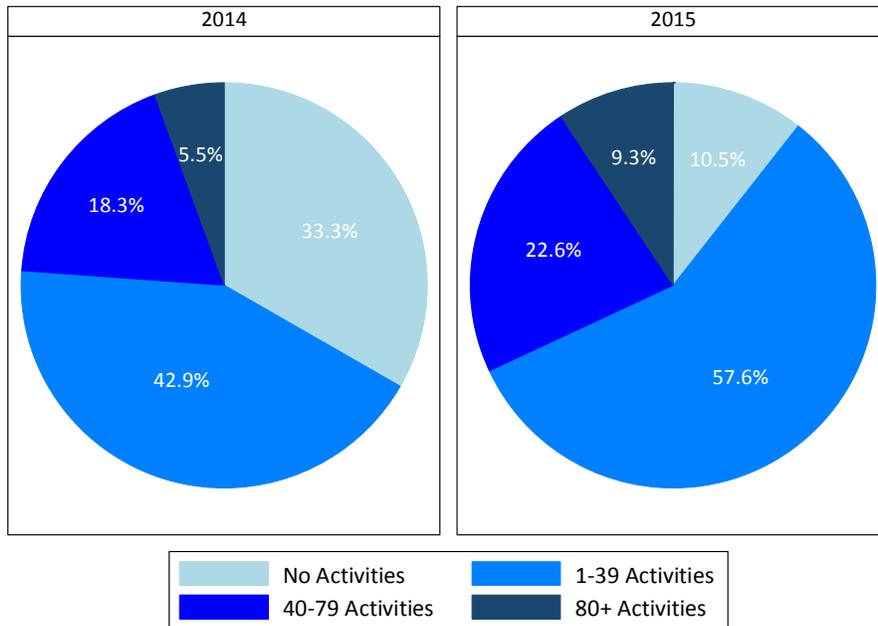
	<b>2014</b>	<b>2014</b>	<b>2014</b>	<b>2015</b>	<b>2015</b>	<b>2015</b>	
	<b>Control</b>	<b>Treatment</b>	<b>Total</b>	<b>Control</b>	<b>Treatment</b>	<b>Total</b>	<b>Total</b>
Sample	7,962	8,713	16,675	8,657	9,300	17,957	34,632
Took Pre-Test	5,524	5,193	10,717	6,286	7,311	13,597	23,314
<i>% of Sample</i>	<i>69%</i>	<i>60%</i>	<i>64%</i>	<i>73%</i>	<i>79%</i>	<i>76%</i>	<i>67%</i>
Took Post-Test	4,837	5,024	9,861	5,950	7,140	13,090	22,951
<i>% of Sample</i>	<i>61%</i>	<i>58%</i>	<i>59%</i>	<i>69%</i>	<i>77%</i>	<i>73%</i>	<i>66%</i>

Source: Achieve3000 LevelSet Lexile assessment results

The second indicator used to determine implementation fidelity was the number of Achieve3000 activities completed in a given year. Recall that Achieve3000 defines “high use” as completion of 80 or more activities in a given year. Accordingly, WCPSS set this as its goal for individual student use. In Achieve3000’s 2012 and 2015 benchmark studies, roughly a quarter of students who completed at least one activity in a given year went on to meet the 80-activity goal.

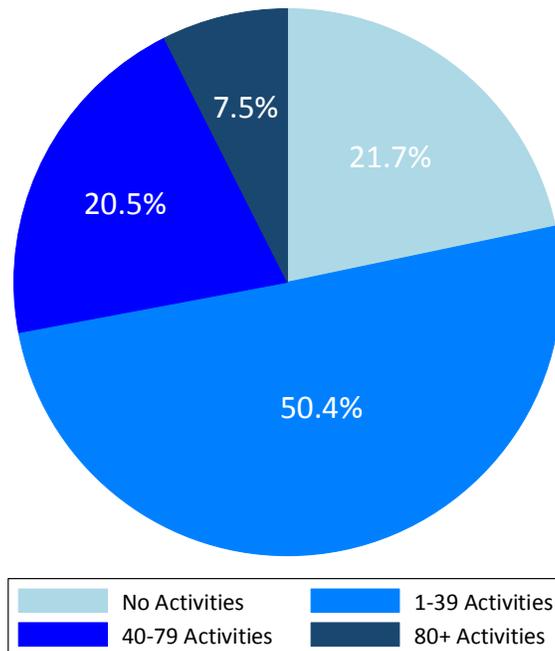
To paint a more complete picture of activity completion in WCPSS, we calculated rates for each of Achieve3000’s pre-established categories in addition to the percentage of students who did not complete any activities. In 2014, 5.5% of students completed 80 or more activities, which rose to 9.3% in 2015 (Figure 2). Consistent with Achieve3000’s completion rate calculation, removing from the sample students who did not complete a single activity, the percentage of WCPSS students meeting the 80-activity goal became 8.3% in 2014 and 10.3% in 2015. While a modest increase, these annual rates—as well as the combined rates (Figure 3)—remained far lower than the benchmark rates of roughly a quarter reported in Achieve3000’s national benchmark studies.

**Figure 2**  
**Achieve3000 Activity Completion Frequency, by Year**



Note: "2014" refers to the 2013-14 school year and "2015" refers to the 2014-15 school year.  
Sources: WCPSS and Achieve3000 administrative records

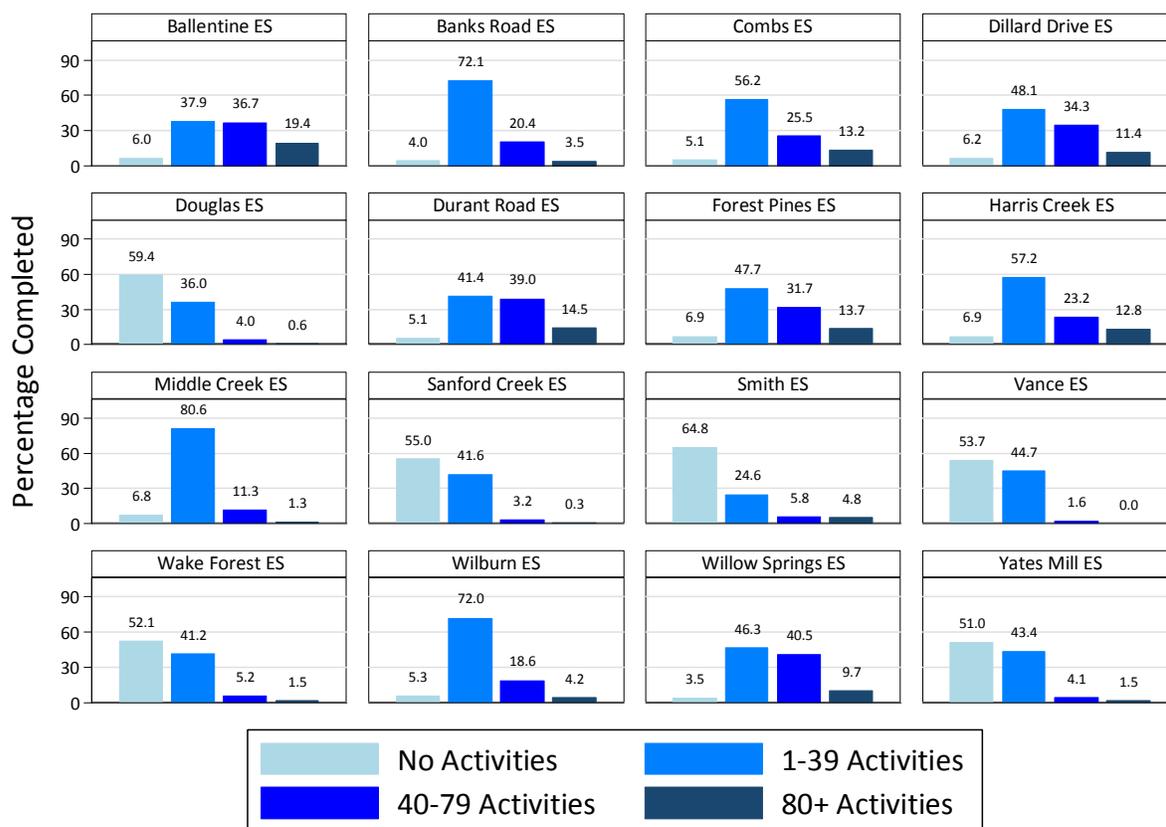
**Figure 3**  
**Achieve3000 Activity Completion Frequency, Pooled**



Sources: WCPSS and Achieve3000 administrative records

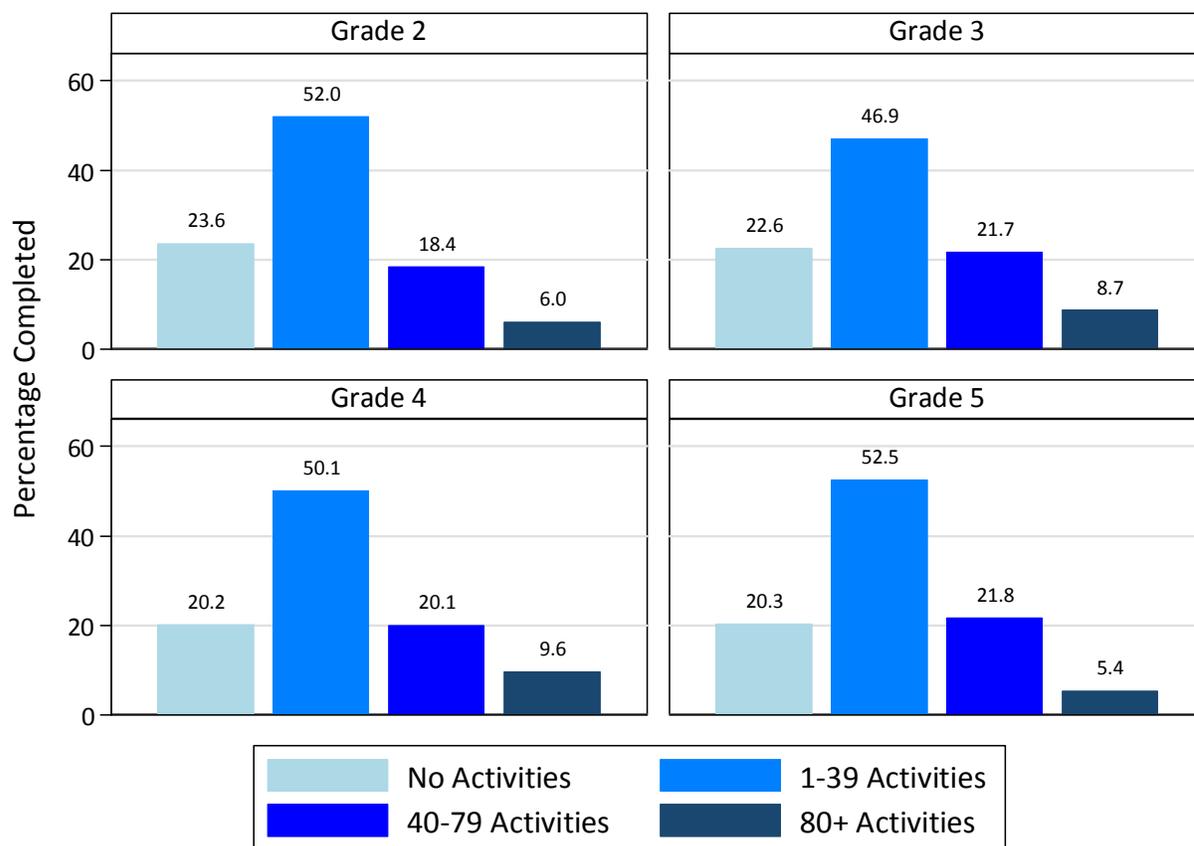
To gain a deeper understanding of what drove activity completion rates, we reviewed school- and grade-level rates. Figure 4 shows that no single school eclipsed the Achieve3000 benchmark 80-activity completion rate, though Ballentine ES came close, and in five additional schools, more than 10% of students met the 80-activity goal. Across grade levels, a slightly higher percentage of students in grades 3 and 4 met the 80-activity goal compared with their counterparts in grades 2 and 5 (Figure 5).

**Figure 4**  
**Achieve3000 Activity Completion Frequency, by School**



Sources: WCPSS and Achieve3000 administrative records

**Figure 5**  
**Achieve3000 Activity Completion Frequency, by Grade**



Sources: WCPSS and Achieve3000 administrative records

In conclusion, implementation in each of the two study years fell short of expectations. This was true both for LevelSet Lexile pre-test administration and activity completion. Several explanations were offered by school-based staff, implementation team members, and other district staff. For example:

- Late implementation in fall 2013 impacted year-round schools, which constituted nine of 16 schools in the treatment sample. While this was not a problem for traditional schools, staff from year-round schools noted that by the time they received access to Achieve3000 in October, students had already completed three months of instructional time and many teachers found it hard to seamlessly incorporate Achieve3000 into their instruction.
- A technology refresh in fall 2014 resulted in many treatment schools losing hardware resources in anticipation of receiving new resources. Since the program requires the use of an electronic device, this led to implementation challenges.
- Principal turnover at a handful of schools led to a resetting of priorities, which may have contributed to decreased usage under new leadership.

## Analytic Strategy

Education data are typically “clustered,” whereby students are grouped in classrooms, classrooms are grouped in schools, and schools are grouped in districts. To disentangle the contributions of these different groupings to a program’s impact on student achievement, researchers use statistical models that incorporate information for additional groups such as classrooms and schools. In this evaluation, the impact of Achieve3000 on student outcomes is estimated by grouping students within schools. Here, the major outcomes of interest are end-of-year Achieve3000 LevelSet Lexile scores, EOG Lexile scores, and DIBELS ORF Lexile conversions. In each case, corresponding prior-year test scores are included to account for the explanatory power of prior achievement.

To measure the impact of Achieve3000 on students grouped within schools, two types of estimation strategies are used: Intent-to-treat (ITT) estimation and treatment-on-treated (TOT) estimation. ITT estimation answers the question: *What is the effect of being offered Achieve3000 on a Lexile outcome?* TOT estimation answers the question: *What is the impact of using Achieve3000 on a Lexile outcome?* Both types of treatment effects are relevant to district policy decisions. ITT impacts allow district staff and leadership to understand what would likely happen if Achieve3000 were offered to *all* elementary schools — not simply 16 elementary schools. TOT impacts help district staff and leadership understand the effects for students who not only have the offer to use the program but actually use the program. This is important because we know that roughly a quarter of students who had the opportunity to use Achieve3000 did not use it, and therefore we want to learn about Achieve3000’s impact on the three-quarters who took the LevelSet Lexile pre-test and proceeded to use the program.

The ITT model is specified as:

$$OUTCOME_{ij} = \beta_0 + \beta_1(A3KOFFER)_{ij} + Z_{ij} + \varepsilon_{ij} + u_i$$

Where *OUTCOME* represents our test score of interest (end-of-year score for each literacy measure) for student *j* grouped in school *i*;  $\beta_0$  represents the constant;  $\beta_1$  represents the coefficient of *A3KOFFER*, which represents the *offer* of Achieve3000, for student *j* grouped in school *i*; *Z* represents a matrix of student- and school-level control variables;  $\varepsilon_{ij}$  represents the random effect of student *j* in school *i*; and  $u_i$  represents the random effect of school *i* (these are known as error terms).

To estimate the treatment-on-treated impact of *using* Achieve3000, we employ a two-stage least squares regression model with an instrumental variable for actual program usage in the first stage. The first stage of this model is specified as:

$$A3KUSE_{ij} = \beta_0 + \beta_1(A3KOFFER)_{ij} + Z_{ijk} + \varepsilon_{ij} + u_i$$

Where *A3KUSE* represents our dependent variable of interest (Achieve3000 use) for student *j* grouped in school *i*;  $\beta_1$  represents the coefficient of our *A3KOFFER* predictor (the offer of Achieve3000) for student *j* grouped in school *i*. The second stage of this model is specified as:

$$OUTCOME_{ij} = \beta_0 + \beta_1(\widehat{A3KUSE})_{ij} + Z_{ij} + \varepsilon_{ij} + u_i$$

Where the fitted value  $\widehat{A3KUSE}$  represents an estimate of the exogenous portion of the offer variable. The instrument for use is a conservative measure of implementation, which is completion of a single Achieve3000 activity during the entire school year. Figure 2 shows that in 2014 and 2015, 67% and 90% of students, respectively, completed at least one activity; the TOT estimation here demonstrates the impact of Achieve3000 on those students. Finally, since students and teachers in the second year of the RCT have potentially had greater access to and familiarity with the program, both ITT and TOT models include a variable to control for “cohort effects.” These effects only appear in the two-year, pooled sample.

### Student Achievement Impacts

This section includes both ITT and TOT impacts of Achieve3000 on each of three outcome measures for 2014, 2015, and the combined two-year sample. The estimate for “Achieve3000 impact” in Table 6—for 2014, 2015, and pooled—represents the difference in achievement between students in the treatment group compared with students in the control group.

All impact estimates are expressed in standard deviation units (*SD*) in order to enable comparisons across grade levels and years. When a test is standardized for all students in a sample, the mean score is set at 0 and the standard deviation is 1. For an example, see Table 6. The ITT estimate in the first row shows a negative value for 2014, which means students in the treatment group scored lower than their peers in the control group. In 2015 of this same row, the value is positive, which means the treatment group outperformed the control group on this particular measure. Finally, in the “pooled” column, while the estimate is slightly positive, it is not significantly different from zero. Thus, there was no impact for the treatment group on this measure over the two-year period.

To further put standard deviation units in context, Table 5 reproduces data from Lipsey et al. (2012) showing mean and median effect sizes across interventions that resemble Achieve3000 by intervention type and target recipients. Achieve3000 in WCPSS was randomized at the “whole school” level and most closely resembles a “curriculum or broad instructional program.” Thus, compared with similar RCTs, the expected effect size, measured in standard deviation units, would be somewhere between 0.08 and 0.14, inclusive (the range of medians and means in these two categories).

**Table 5**  
***Achievement effect sizes from RCTs by intervention Type and Recipient***

<b>Type of Intervention</b>	<b>Number of effect sizes</b>	<b>Median</b>	<b>Mean</b>
Instructional format	52	0.13	0.21
Teaching technique	117	0.27	0.35
Instructional component or skill training	401	0.27	0.36
Curriculum or broad instructional program	227	0.08	0.13
Whole school program	32	0.17	0.11
<b>Target recipients</b>			
Individual students	252	0.29	0.40
Small group	322	0.22	0.26
Classroom	176	0.08	0.18
Whole school	35	0.14	0.10
Mixed	44	0.24	0.30

Source: Lipsey, et al. (2012). "Translating the Statistical Representation of the Effects of Education Interventions into More Readily Interpretable Forms." Reproduced from Table 10, page 36.

### *Achieve3000 Impacts*

Table 6 displays the full results for the school years 2013-14 and 2014-15, as well as the full two-year sample. Intent-to-treat (ITT) estimates represent the effect of the school-level offer to receive Achieve3000. These estimates are listed for Achieve3000's impact on the LevelSet Lexile, EOG Lexile, and DIBELS ORF Lexile, respectively. We used a group of control variables (race/ethnicity, LEP status, SWD status, sex, type of school attended) in the analyses, but we omit the results here for simplicity. Table 6 shows that ITT estimates for the Achieve3000 Lexile post-test were negative ( $-0.05$  *SD*) and statistically significant ( $p < 0.05$ ) in 2014, positive ( $0.13$  *SD*) and statistically significant ( $p < 0.01$ ) in 2015, and positive but not statistically significant in the pooled sample. The offer of Achieve3000 did not impact EOG Lexile scores in either year or in the pooled sample. On the DIBELS ORF Lexile, Table 6 shows that Achieve3000 had a positive ( $0.06$  *SD*) and significant ( $p < 0.05$ ) impact in 2014, but not in 2015 or in the pooled sample. These results suggest that only Achieve Lexile score impacts were in line with empirical estimates provided by Lipsey et al. (2012) and reported above in Table 5. The 2015 impact of  $0.13$  *SD* translates into a difference of 31 Lexile points.

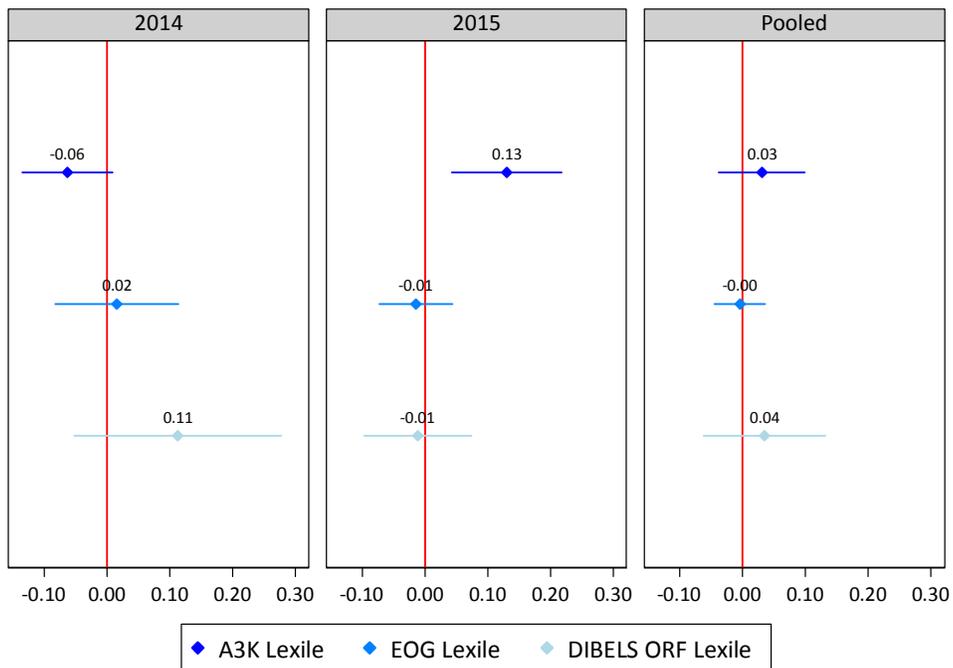
Treatment-on-treated (TOT) estimates represent the effect of Achieve3000 use on student outcomes. Figure 6 displays these same results graphically. We selected a conservative measure of at least one activity completed as a proxy for use. Table 6 shows that TOT estimates for the Achieve3000 Lexile post-test were similar to the ITT estimates reported above for both 2014 ( $-0.06$  *SD*;  $p < 0.10$ ), 2015 ( $0.13$  *SD*;  $p < 0.01$ ), and in the pooled sample. The program did not have TOT impacts on the EOG Lexile scores in either year or in the pooled sample. In contrast to 2014 ITT impacts on the DIBELS Lexile, 2014 TOT impacts, while positive in direction, were not statistically significant. Appendix B contains more detailed tables for each of the six different assessment outcomes.

**Table 6**  
**Achieve3000 Impacts on Elementary Literacy Outcomes**

	Estimate	2014	2015	Pooled
Achieve3000 LevelSet Lexile	ITT	-0.054** (0.022)	0.129*** (0.033)	0.029 (0.024)
	TOT	-0.064* (0.037)	0.131*** (0.045)	0.031 (0.035)
End-of-Grade Lexile	ITT	0.009 (0.029)	-0.016 (0.023)	-0.004 (0.020)
	TOT	0.015 (0.051)	-0.015 (0.030)	-0.005 (0.021)
DIBELS Oral Reading Fluency Lexile	ITT	0.063** (0.031)	-0.013 (0.032)	0.025 (0.029)
	TOT	0.114 (0.085)	-0.012 (0.045)	-0.035 (0.051)

Note: "ITT" is intent-to-treat and "TOT" is treatment-on-treated. Standard errors are in parenthesis.  
\* p<0.1; \*\* p<0.05; \*\*\* p<0.01

**Figure 6**  
**Achieve3000 Treatment-on-Treated Impacts on Elementary Literacy Outcomes**



Note: Diamonds represent Achieve3000 impacts on the three Lexile outcomes. For example, The "A3K Lexile" value of 0.13 in 2015 is the graphical representation of "0.129\*\*\*" in Table 6. Horizontal lines represent 95% confidence intervals which indicate that the impact is not significantly different from zero ( $p < .05$ ) if they intersect the red "0.00" line.

The main claim from the Achieve3000 vendors is that the program results in roughly 1.5 times Lexile growth for students using the program compared with normal growth, which is represented by our control group. Table 6 demonstrates that over the two-year study period, neither the offer nor the use of Achieve3000 contributed to that goal being met. This is not to say that students using Achieve3000 did not achieve Lexile gains—they simply achieved Lexile gains that were statistically the same as their peers in the control group. At best, ITT and TOT impacts of Achieve3000 on the LevelSet Lexile test of 0.13 *SD*, or 31 points more than the control group, represents two-thirds of the growth suggested in Achieve3000’s 2012 national benchmark study (31 points ÷ 46 points).

### *Subgroup Analyses*

Since Achieve3000 promotes its products as “differentiated, instructional solutions designed to reach a school’s entire student population (MetaMetrics, 2015),” we examined select subgroups effects, including those for Limited English Proficient (LEP) students, students with disabilities (SWD), and Academically and Intellectually Gifted (AIG) students. Given the proposed ability of KidBiz3000 to differentiate content, the program would presumably provide LEP students and SWD students alike with the opportunity to benefit from Achieve3000’s leveled articles that adjust in real time. In addition, AIG students with access to the program could presumably advance to Lexile levels higher than a traditional book might allow, and thus provide differential benefits to these students as well.

For LEP students, there were no significant ITT impacts in 2014 or 2015 in the pooled sample, nor on any of the three outcome measures. Achieve3000 did, however, positively impact SWD performance to a small and statistically significant degree on the Achieve3000 Lexile test in 2015 (0.08 *SD*;  $p < .01$ ) and in the pooled sample (0.05 *SD*;  $p < 0.01$ ). AIG students also experienced small benefits from the program on the 2015 Achieve3000 Lexile test (0.05 *SD*;  $p < 0.05$ ) and the 2014 EOG Lexile (0.08 *SD*;  $p < .10$ ). However, there was a small negative impact on the 2015 EOG Lexile test (-0.07 *SD*;  $p < .10$ ).

## Conclusions and Recommendations

The district’s two-year randomized controlled trial (RCT) of Achieve3000 represents an important milestone in its effort, through the Enhancing Data Use framework, to identify evidence-based instructional strategies. It also represented a model of program implementation to inform future efforts, as the district has since launched two additional RCTs. Such efforts represent the “gold standard” in educational research and provide a wide range of stakeholders with causal impact estimates of program effectiveness.

Overall, this RCT of Achieve3000 is a significant step toward increasing the district’s understanding of the role of technology in instruction in a number of ways. First, while dozens of programs have been shown to positively impact student performance (Kim & Quinn, 2013; Slavin, Lake, Chambers, Cheung, & Davis, 2009; Slavin, Lake, Davis, & Madden, 2011), few districts have examined the impact of strictly technology-based initiatives. This is especially true of increasingly visible programs like Achieve3000, which has existed for more than a decade,

serves more than 1 million U.S. students, and is consistently ranked by Inc. Magazine as one of the fastest growing private education companies in the United States (Inc. Magazine, 2015). WCPSS, as well as its peer districts, should continue to evaluate the effectiveness of such programs in experimental settings.

Second, the findings herein suggest that district staff should maintain modest expectations for the promise of education technology — both in terms of implementation and effectiveness. These findings suggest that for the most part, Achieve3000, a program designed as a component of core instruction, produced gains that were not substantively different than those in the control group (except in the 2015 impact on Achieve’s LevelSet Lexile outcome). Findings like these suggest that staff should remain realistic and vigilant when faced with aggressive performance targets promised by vendors.

Third, implementation of technology programs is hard work and requires staff resources. The district’s implementation team of roughly a dozen staff members attempted to maintain outreach to schools where implementation was weak. At the vendor level, one staff member was assigned to exclusively monitor implementation across the treatment schools in both years. Yet these efforts may not have been sufficient in the first year. In addition, while implementation ramped up in 2015, the percentage of students taking pre- and post-tests as well as those meeting the 80-activity goal fell far short of expectations. This tells us that school leaders who express interest in a particular product might not actually need the product. Based on conversations with district staff, we learned that many schools experience varying degrees of resource bloat but nonetheless welcome the offer of new programs. Thus, it is not entirely surprising that implementation fidelity could suffer in schools managing a multitude of programs and interventions across grades, subjects, and subgroups.

Finally, implementation may take more time for some programs than for others. The literacy program *Success for All*, for example, began to impact additional student outcomes to a greater magnitude after the second and third years of implementation (Borman et al., 2005a, 2005b, 2007). If Achieve3000 does turn out to be an effective program, staff may have to be patient to realize its potential effects—a choice that may be costly in terms of time, money, and student achievement given the program’s impacts thus far. District leadership will be making decisions about the future of Achieve3000 in the coming months, taking into account the results of this study as well as an internal review of Achieve3000 usage during the current (2015-16) school year.

Conclusions and associated recommendations follow below:

**Implementation of Achieve3000 improved slightly in 2015, but was weak in both years.**

Benchmark usage data provided in Achieve3000’s 2012 and 2015 National Lexile Reports suggested that roughly 25% of students, at best, were able to complete at least 80 activities—the goal agreed to by the district and vendor. In those reports, this level was correlated with higher Lexile levels compared with lower activity ranges (1-39 and 40-79). In 2014, fewer than 10% of WCPSS students met this threshold and approximately 10% met it in 2015. These activity completion rates fell far short of expectations. Implementation was relatively consistent across

grade levels, although grades 3 and 4 had slightly higher percentages of students meeting the 80-activity compared with their counterparts in grades 2 and 5.

- *Recommendation:* Staff should maintain modest expectations in the face of aggressive program usage goals. The district should establish an overarching implementation fidelity framework and, at minimum, each large program should have an associated implementation team that monitors usage and progress at least monthly. District leadership members, such as assistant and area superintendents, should be aware of efforts to monitor usage especially when their schools are included in a treatment group.

**Achievement goals failed to meet district, vendor, or empirical standards.** District staff selected Achieve3000 with the hope that, at minimum, it would contribute to student achievement gains compared with the control group. In the two-year combined results, this did not occur for any of the three outcomes of interest. Nor did it occur in 11 of the 12 different ways staff estimated impacts—by year, outcome and model (two years, three outcomes, and two models for 12 total specifications). The lone exception was the positive and significant impact on the vendor’s LevelSet Lexile score in 2015, a result that was in-line with empirical estimates (0.13 *SD*). This amounted to 31 Lexile points, which fell short of Lexile gains promised by the vendor and expected by district staff.

- *Recommendation:* Staff should develop a roster of evidence-based programs that can be deployed when new core and supplemental resources are needed. Achieve3000 did not have existing experimental evidence of effectiveness to support its use in the district. Indeed, few products do, which is why staff appropriately launched the program as an RCT. Now that causal impact results are available, staff must consider alternative programs in the event that additional years of outcomes data (i.e., 2016) fail to meet expectations.

**Achieve3000 positively impacted select subgroups.** While Achieve3000 did not have an impact on LEP students, SWD and AIG students appeared to benefit to a small degree from the program compared with their non-identified peers. These impacts were small compared with empirical average effects.

- *Recommendation:* SWD and AIG program staff should investigate why these subgroups may have benefited from Achieve3000. But they should exercise caution when making programming decisions on the basis of these effects since, while statistically significant, were quite small in magnitude.

## References

- Achieve3000. (2011). Research to practice: How Achieve3000 differentiated literacy solutions use research to prepare students to thrive in the 21st century. Retrieved from [http://doc.achieve3000.com/intranet/A3K12\\_00102\\_ResearchToPractice.pdf](http://doc.achieve3000.com/intranet/A3K12_00102_ResearchToPractice.pdf).
- Achieve3000. (2012). National Lexile Study: 2010-11 Lexile Study. *Lakewood, NJ: Achieve3000*.
- Achieve3000. (2015). National Lexile Study, 2014-2015. *Lakewood, NJ: Achieve3000*.
- Borman, G. D., Slavin, R. E., Cheung, A. C., Chamberlain, A. M., Madden, N. A. & Chambers, B. (2005a). The national randomized field trial of Success for All: Second-year outcomes. *American Educational Research Journal*, 42(4), 673–696.
- Borman, G. D., Slavin, R. E., Cheung, A. C., Chamberlain, A. M., Madden, N. A. & Chambers, B. (2007). Final reading outcomes of the national randomized field trial of Success for All. *American Educational Research Journal*, 44(3), 701–731.
- Borman, G. D., Slavin, R. E., Cheung, A., Chamberlain, A. M., Madden, N. A. & Chambers, B. (2005b). Success for All: First-year results from the national randomized field trial. *Educational Evaluation and Policy Analysis*, 27(1), 1–22.
- Kim, J. S. & Quinn, D. M. (2013). The effects of summer reading on low-income children’s literacy achievement from kindergarten to grade 8: A meta-analysis of classroom and home interventions. *Review of Educational Research*, 83(3), 386–431.
- Lipsey, M. W., Puzio, K., Yun, C., Hebert, M. A., Steinka-Fry, K., Cole, M. W., Roberts, M., Anthony, K.S., & Busick, M. D. (2012). Translating the statistical representation of the effects of education interventions into more readily interpretable forms. *U.S. Department of Education (NCSE 2013-3000)*.
- List, J. A., Sadoff, S. & Wagner, M. (2011). So you want to run an experiment, now what? Some simple rules of thumb for optimal experimental design. *Experimental Economics*, 14(4), 439–457.
- Inc. Magazine. (2015). The 2015 Inc. 5000: Meet the fastest-growing private companies in America. Retrieved from <http://www.inc.com/inc5000>.
- MetaMetrics. (2009). Linking DIBELS 6th edition Oral Reading Fluency with the Lexile framework for reading. *Durham, NC: MetaMetrics, Inc.*
- MetaMetrics. (2015). Achieve3000 LevelSet (version 2): Development and Technical Guide.

- Rhea, A. (2012). A preliminary examination of The Daily Five initiative. *Cary, NC: Wake County Public School System.*
- Slavin, R. E., Lake, C., Chambers, B., Cheung, A. & Davis, S. (2009). Effective reading programs for the elementary grades: A best-evidence synthesis. *Review of Educational Research, 79*(4) 1391-1466.
- Slavin, R. E., Lake, C., Davis, S. & Madden, N. A. (2011). Effective programs for struggling readers: A best-evidence synthesis. *Educational Research Review, 6*(1), 1–26.
- What Works Clearinghouse. (2014). WWC procedures and standards handbook, version 3.0. *Washington, DC: Retrieved January, 1, 2014.*
- Williamson, G. L. (2006). What is expected growth? A white paper from MetaMetrics, Inc. *Durham, NC: MetaMetrics, Inc.*

## Appendix A

In 2007, MetaMetrics, creator of the Lexile framework, partnered with Dynamic Measurement Group and Wireless Generation to develop a conversion formula to link Lexile and DIBELS Oral Reading Fluency (ORF) scores (MetaMetrics, 2009). The following two formulas achieve this conversion for students in grades 2 and 3:

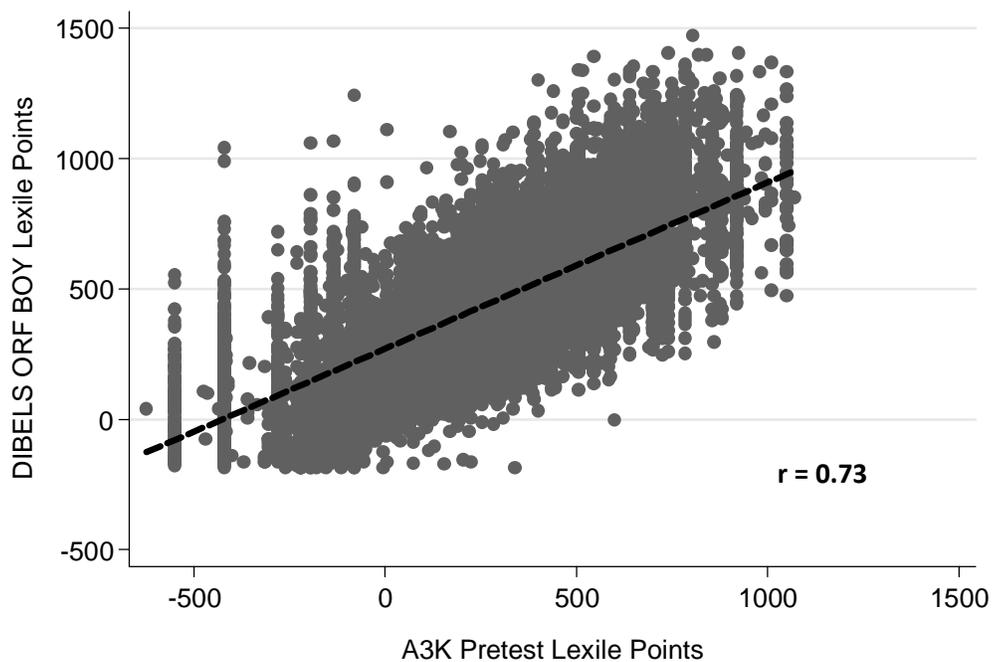
$$\text{Grade 2: Lexile measure} = 7.31829214450681 * \text{ORF} + -185.479047114992$$

$$\text{Grade 3: Lexile measure} = 7.29760592369798 * \text{ORF} + -170.258972906792$$

Figure A1 applies these formulas in order to generate corresponding Lexile measures for DIBELS ORF and demonstrates the strong correlation between the two measures ( $r = 0.73$ ). Figure A2 demonstrate that pre- and post-intervention Lexile conversions drawn from prior End-of-Grade (EOG) tests also provided strong a correlation ( $r = 0.77$ ). The relationship between these measures—EOG and DIBELS ORF Lexiles—and Lexile LevelSet scores supports their use as outcomes in addition to LevelSet scores.

**Figure A1**

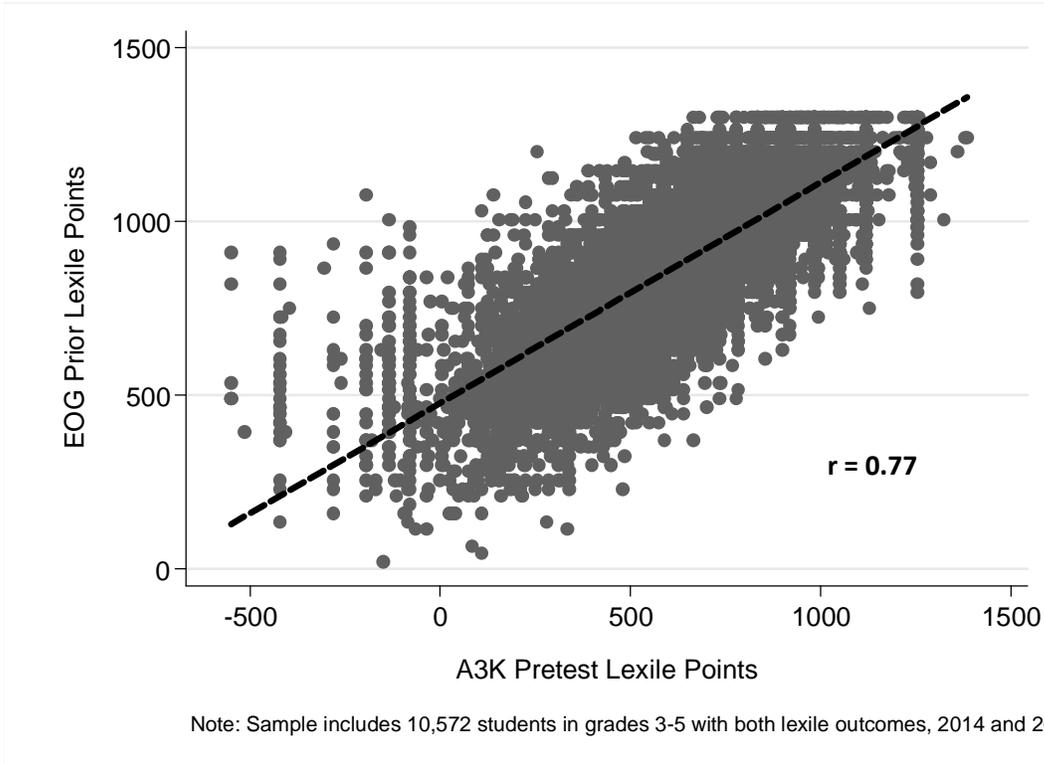
***Correlation between Achieve3000 LevelSet Lexile Pre-Test and BOY DIBELS ORF Lexile Points***



Note: Sample includes 12,180 students in grades 2-3 with both lexile outcomes, 2014 and 2015.

Figure A2

Correlation between Achieve3000 LevelSet Lexile Pre-Test and BOY EOG Lexile Points



## Appendix B

**Table B1**  
***Achieve3000 ITT Impact on Achieve3000 Lexile***

	<b>2014</b>	<b>2015</b>	<b>Pooled</b>
Achieve3000 impact	-0.054** (0.022)	0.129*** (0.033)	0.029 (0.024)
Prior Achieve3000 Lexile	0.769*** (0.006)	0.866*** (0.004)	0.821*** (0.003)
Cohort effect	— —	— —	0.027*** (0.006)
Constant	-0.201 (0.410)	-0.784 (0.496)	-0.535 (0.353)
School-level variance contribution	0.038*** (0.007)	0.072*** (0.009)	0.050*** (0.007)
Student-level variance contribution	0.443*** (0.003)	0.344*** (0.002)	0.397*** (0.002)
<i>N</i>	9,732	12,851	22,583

\*  $p < 0.1$ ; \*\*  $p < 0.05$ ; \*\*\*  $p < 0.01$

**Table B2**  
***Achieve3000 ITT Impact on EOG Lexile***

	<b>2014</b>	<b>2015</b>	<b>Pooled</b>
Achieve3000 impact	0.009 (0.029)	-0.016 (0.023)	-0.004 (0.020)
Prior EOG Lexile	0.670*** (0.009)	0.617*** (0.009)	0.643*** (0.006)
Cohort effect	— —	— —	0.072*** (0.010)
Constant	0.410 (0.429)	0.303 (0.347)	0.330 (0.305)
School-level variance contribution	0.049*** (0.010)	0.034*** (0.010)	0.035*** (0.007)
Student-level variance contribution	0.554*** (0.005)	0.519*** (0.005)	0.538*** (0.003)
<i>N</i>	6,235	6,307	12,542

\*  $p < 0.1$ ; \*\*  $p < 0.05$ ; \*\*\*  $p < 0.01$

**Table B3**  
*Achieve3000 ITT Impact on DIBELS Lexile*

	<b>2014</b>	<b>2015</b>	<b>Pooled</b>
Achieve3000 impact	0.063** (0.031)	-0.013 (0.032)	0.025 (0.029)
DIBELS Lexile Pretest	0.842*** (0.006)	0.829*** (0.006)	0.834*** (0.004)
Cohort effect	— —	— —	0.003 (0.007)
Constant	0.234 (0.470)	0.014 (0.481)	0.116 (0.431)
School-level variance contribution	0.063*** (0.010)	0.065*** (0.010)	0.061*** (0.009)
Student-level variance contribution	0.412*** (0.003)	0.421*** (0.003)	0.418*** (0.002)
<i>N</i>	7,196	7,296	14,492

\*  $p < 0.1$ ; \*\*  $p < 0.05$ ; \*\*\*  $p < 0.01$

**Table B4**  
*Achieve3000 TOT Impact on Achieve3000 Lexile*

	<b>2014</b>	<b>2015</b>	<b>Pooled</b>
Achieve3000	-0.064* (0.037)	0.131*** (0.045)	0.031 (0.035)
Prior Achieve3000 Lexile	0.769*** (0.006)	0.866*** (0.004)	0.821*** (0.003)
Cohort	— —	— —	0.029*** (0.006)
Constant	-0.072 (0.592)	-0.775 (0.646)	-0.554 (0.490)
<i>N</i>	9,732	12,851	22,583

\*  $p < 0.1$ ; \*\*  $p < 0.05$ ; \*\*\*  $p < 0.01$

**Table B5**  
***Achieve3000 TOT Impact on EOG Lexile***

	<b>2014</b>	<b>2015</b>	<b>Pooled</b>
Achieve3000	0.015 (0.051)	-0.015 (0.030)	-0.005 (0.021)
Prior EOG Lexile	0.670*** (0.009)	0.615*** (0.009)	0.643*** (0.006)
Cohort	—	—	0.072*** (0.010)
Constant	0.384 (0.497)	0.376 (0.454)	0.350 (0.261)
<i>N</i>	6,217	6,288	12,505

\*  $p < 0.1$ ; \*\*  $p < 0.05$ ; \*\*\*  $p < 0.01$

**Table B6**  
***Achieve3000 TOT Impact on DIBELS Lexile***

	<b>2014</b>	<b>2015</b>	<b>Pooled</b>
Achieve3000	0.114 (0.085)	-0.012 (0.045)	0.035 (0.051)
Prior DIBELS Lexile	0.841*** (0.006)	0.825*** (0.006)	0.832*** (0.004)
Cohort	—	—	-0.000 (0.009)
Constant	0.108 (0.753)	0.053 (0.630)	0.119 (0.579)
<i>N</i>	7,173	7,200	14,373

\*  $p < 0.1$ ; \*\*  $p < 0.05$ ; \*\*\*  $p < 0.01$

## **Acknowledgements**

### **Authors**

Darryl V. Hill  
Office of the Chief of Staff and Strategic Planning

Matthew A. Lenard  
Data, Research and Accountability

### **Acknowledgements**

The authors would like to acknowledge the contributions and collaboration of Cathy Moore (Deputy Superintendent for Academic Advancement), James Overman (Area Superintendent for Elementary Support), Sherri Miller (Director of K-12 Literacy), Lindsay Page (University of Pittsburgh), WCPSS's Achieve3000 implementation team members, school-based Achieve3000 Leaders, and staff from Achieve3000's North Carolina implementation team.