SREE Spring 2017 Conference

Title:
"The Impact of Achieve3000 on Elementary Literacy Outcomes: Final Results from a Three-Year Randomized Trial"

Authors:
Darryl V. Hill, Wake County Public School System
Matthew A. Lenard, Wake County Public School System
Lindsay C. Page, University of Pittsburgh

Abstract:
School districts across the United States have increasingly adopted information and communications technology (ICT) in an attempt to improve student achievement across a wide range of educational settings (Knezek, 2008). Spending on educational hardware is expected to grow from $13 billion in 2013 to $19 billion in 2018, an annualized increase of eight percent, while districts spend an estimated $8 billion annually on software (Nagel, 2014; Richards & Stebbins, 2012). A growing body of research has emerged attempting to measure the causal impact of ICT-based resources on various student achievement outcomes (Barrow, Markman, & Rouse, 2008; Campuzano, Dynarski, Agodini, & Rall, 2009; Given, Wasserman, Chari, Beattie, & Eden, 2008; James-Burdumy et al., 2009; Roschelle et al., 2007; Rouse & Krueger, 2004; R. S. Savage, Abrami, Hipps, & Deault, 2009; R. Savage et al., 2013; Wijekumar, Meyer, & Lei, 2012). This body of work, as well as more recent meta-analytic reviews of software- and hardware-based interventions (Takacs, Swart, & Bus, 2015; Van der Kleij, Feskens, & Eggen, 2015; Zheng, Warschauer, Lin, & Chang, 2016), has produced largely mixed results and suggests that a closer inspection of ICT impacts is warranted across a range of geographies, grade levels, and subjects.

This paper examines the impacts of Achieve3000, an early literacy program that differentiates non-fiction reading passages based on individual students' Lexile scores. To estimate the causal impact of Achieve3000, we randomly assigned the program in 2013-14 in grades 2-5 among 32 elementary schools in the Wake County Public School System (WCPSS). WCPSS has the 15th-largest student enrollment in the U.S. and the largest in North Carolina. The district's motivation for using Achieve3000 stemmed from 2012 legislation that enacted the statewide Read to Achieve (R2A) program, which mandated that students not proficient in reading by the end of grade 3 successfully complete summer school before promotion. To prepare for the 2013-14 school year in the era of R2A, district staff identified new programs—including Achieve3000—to help students clear the grade 3 proficiency hurdle.

Our analytic sample of 32 representative schools includes those that expressed a high level of interest in the program. These schools were sorted on the percentage of students proficient in reading based on 2011-12 End-of-Grade (EOG) data and Achieve3000 was randomly assigned to one school within each ranked pair (see Table 1 for balance statistics). Schools expressing interest in Achieve3000 committed to reserving at least two half-hour blocks each week for students to work toward completing 80 activities in a year ("full implementation"). Treatment schools identified an "Achieve3000 Leader"—the school's point of contact for coordinating

implementation and attending implementation team meetings at the school and district levels. In addition, the district central office fielded an implementation team consisting of roughly ten staff members across a range of departments as well as two representatives from the company who provided support to treatment schools and monthly usage reports to the district team.

To estimate the impact of Achieve3000 on students, we fit administrative and program usage data to a cluster two-level model with random effects. First, we estimate the causal impact of Achieve3000 on summative performance on their year-end Achieve3000 Lexile score, EOG Lexile score, and DIBELS Oral Reading Fluency (ORF) Lexile conversion, in each case controlling for prior achievement and various student-level covariates. Finally, we fit multilevel models for change to estimate the longitudinal impact of Achieve3000 on Lexile scores for individual students with repeated data points.

Results suggest that Achieve3000 had mixed impacts across the range of outcomes and models. On the vendor's own Lexile test, intent-to-treat effects were slightly negative in 2014 (σ = -.05, p < .05) but positive in 2015 (σ = .13, p < .01), 2016 (σ = .20, p < .01), and in the three-year pooled sample (σ = 0.09, p < .01). Treatment-on-treated impacts were similar (Tables 2-3; Figures 1-2). However, on the additional two Lexile measures, there was largely no effect (Figures 3-6), suggesting that while impacts on the vendor's test were both substantively and statistically meaningful, results may be associated with the embedded nature of the assessment. Longitudinal results showed small levels of growth over time for the treatment group compared with the control group (0.008 < σ < .02 per assessment wave; p < .01).

Over the three-year study period, implementation improved dramatically: in year 3, nearly a quarter of students met the 80+ activities goal (Figure 7), nearly identical to the rate reported in the vendor's national benchmark study (Achieve3000, 2015). While an impact of .20 *sd* in 2016 appears notably large, it translates into only 52 Lexile points, which is far fewer than the 94 points expected for students who met conservative levels of program fidelity in the national benchmark study. Moreover, students did not experience any significant gains on the EOG Lexile equivalent, and growth on the DIBELS Lexile equivalents was decidedly mixed. Our next step in this work is to apply a principal stratification framework in order to determine whether certain activity threshold levels are driving the results that we do see on the vendor's Lexile test. In the end, however, impacts for this popular and fast-growing program suggest that it fell short of delivering results promised by the vendor and suggested by empirical meta-analyses of program effects.

Appendix – Bibliography, Tables, and Figures

Achieve3000. (2015). National Lexile Study, 2014-2015.

Barrow, L., Markman, L. & Rouse, C. E. (2008). *Technology's edge: The educational benefits of computer-aided instruction*.

Campuzano, L., Dynarski, M., Agodini, R. & Rall, K. (2009). Effectiveness of Reading and Mathematics Software Products: Findings From Two Student Cohorts. NCEE 2009-4041. *National Center for Education Evaluation and Regional Assistance*.

Given, B. K., Wasserman, J. D., Chari, S. A., Beattie, K. & Eden, G. F. (2008). A randomized, controlled study of computer-based intervention in middle school struggling readers. *Brain and Language*, *106*(2), 83–97.

James-Burdumy, S., Mansfield, W., Deke, J., Carey, N., Lugo-Gil, J., Hershey, A., … others. (2009). Effectiveness of Selected Supplemental Reading Comprehension Interventions: Impacts on a First Cohort of Fifth-Grade Students. NCEE 2009-4032. *National Center for Education Evaluation and Regional Assistance*.

Knezek, D. (2008). Technology and Student Achievement— The Indelible Link.

Nagel, D. (2014). Spending on instructional tech to reach $19 billion within 5 years. *THE Journal*.

Richards, J. & Stebbins, L. (2012). Education Division of The Software & Information Industry. *SIIA*, 1–12.

Roschelle, J., Tatar, D., Shechtman, N., Hegedus, S., Hopkins, B., Knudsen, J. & Stroter, A. (2007). Can a technology-enhanced curriculum improve student learning of important mathematics. *Results from 7th Grade, Year, 1*.

Rouse, C. E. & Krueger, A. B. (2004). Putting computerized instruction to the test: a randomized evaluation of a "scientifically based" reading program. *Economics of Education Review*, *23*(4), 323–338.

Savage, R., Abrami, P. C., Piquette, N., Wood, E., Deleveaux, G., Sanghera-Sidhu, S. & Burgos, G. (2013). A (Pan-Canadian) cluster randomized control effectiveness trial of the ABRACADABRA web-based literacy program. *Journal of Educational Psychology*, *105*(2), 310.

Savage, R. S., Abrami, P., Hipps, G. & Deault, L. (2009). A randomized controlled trial study of the ABRACADABRA reading intervention program in grade 1. *Journal of Educational Psychology*, *101*(3), 590.

Takacs, Z. K., Swart, E. K. & Bus, A. G. (2015). Benefits and Pitfalls of Multimedia and Interactive Features in Technology-Enhanced Storybooks A Meta-Analysis. *Review of Educational Research*, 0034654314566989.

Van der Kleij, F. M., Feskens, R. C. & Eggen, T. J. (2015). Effects of Feedback in a Computer-Based Learning Environment on Students' Learning Outcomes A Meta-Analysis. *Review of Educational Research*, *85*(4), 475–511.

Wijekumar, K. K., Meyer, B. J. & Lei, P. (2012). Large-scale randomized controlled trial with 4th graders using intelligent tutoring of the structure strategy to improve nonfiction reading comprehension. *Educational Technology Research and Development*, *60*(6), 987–1013.

Zheng, B., Warschauer, M., Lin, C.-H. & Chang, C. (2016). Learning in One-to-One Laptop Environments A Meta-Analysis and Research Synthesis. *Review of Educational Research*, 0034654316628645.

Table 1. Pre-Intervention Balance between Treatment and Control Groups

Variable	Treatment Group	Control Group	Difference (T - C)	p-value
Male	0.512	0.513	-0.001	0.895
Black	0.205	0.313	-0.108	0.035*
Hispanic	0.199	0.170	0.029	0.335
LEP	0.107	0.081	0.026	0.168
SWD	0.118	0.124	-0.006	0.584
SES	0.363	0.374	-0.011	0.828
AIG: Reading & Math	0.070	0.077	-0.007	0.670
LevelSet Lexile	409.232	395.331	13.901	0.690
EOG Lexile	858.817	866.013	-7.196	0.718
DIBELS ORF Lexile	421.134	414.379	6.755	0.798

Note:
T-C: Treatment group mean minus control group mean.
LevelSet, EOG, and DIBELS ORF Lexiles expressed as raw scores.
Student-level means calculated using mixed-effects regression with robust standard errors.
* $p<0.05$ ** $p<0.01$ *** $p<0.001$

Table 2. Achieve3000 ITT Impacts on Achieve3000 Lexile Test, Grades 2-5, 2014-2016

	(1) 2014	(2) 2015	(3) 2016	(4) Pooled
Achieve3000	-0.042*	0.133***	0.191***	0.091***
	(0.024)	(0.033)	(0.040)	(0.025)
Prior Lexile	0.769***	0.866***	0.806***	0.816***
	(0.006)	(0.004)	(0.004)	(0.003)
Male	-0.006	0.016***	0.010	0.007*
	(0.009)	(0.006)	(0.007)	(0.004)
Black	-0.119***	-0.084***	-0.078***	-0.092***
	(0.013)	(0.009)	(0.010)	(0.006)
Hispanic	-0.045***	-0.042***	-0.046***	-0.046***
	(0.015)	(0.010)	(0.011)	(0.007)
LEP	-0.165***	-0.068***	-0.100***	-0.107***
	(0.019)	(0.012)	(0.014)	(0.009)
SWD	-0.245***	-0.122***	-0.172***	-0.178***
	(0.015)	(0.011)	(0.012)	(0.007)
SES	-0.126***	-0.069***	-0.109***	-0.100***
	(0.012)	(0.008)	(0.009)	(0.005)
AIG: Reading & Math	0.318***	0.158***	0.110***	0.205***
	(0.017)	(0.013)	(0.016)	(0.009)
Constant	-0.584	-1.087**	-0.659	-0.747*
	(0.453)	(0.541)	(0.666)	(0.415)
School-level controls	Y	Y	Y	Y
sd of residuals (within)	0.043***	0.072***	0.089***	0.056***
	(0.008)	(0.010)	(0.012)	(0.007)
sd of residuals (overall)	0.443***	0.344***	0.381***	0.395***
	(0.003)	(0.002)	(0.002)	(0.001)
Observations	9732	12851	12486	35069
R^2				

Standard errors in parentheses
* $p < .10$, ** $p < .05$, *** $p < .01$

Table 3. Achieve3000 TOT Impacts on Achieve3000 Lexile Test, Grades 2-5, 2014-2016

	(1) 2014	(2) 2015	(3) 2016	(4) Pooled
Achieve3000	-0.053	0.137***	0.193***	0.096***
	(0.045)	(0.042)	(0.040)	(0.030)
Prior Lexile	0.768***	0.866***	0.806***	0.816***
	(0.006)	(0.004)	(0.004)	(0.003)
Male	-0.006	0.016***	0.010	0.007*
	(0.009)	(0.006)	(0.007)	(0.004)
Black	-0.119***	-0.084***	-0.078***	-0.092***
	(0.013)	(0.009)	(0.010)	(0.006)
Hispanic	-0.045***	-0.041***	-0.046***	-0.046***
	(0.015)	(0.010)	(0.011)	(0.007)
LEP	-0.167***	-0.067***	-0.100***	-0.106***
	(0.019)	(0.012)	(0.014)	(0.009)
SWD	-0.246***	-0.121***	-0.171***	-0.177***
	(0.015)	(0.011)	(0.012)	(0.007)
SES	-0.125***	-0.069***	-0.109***	-0.101***
	(0.012)	(0.008)	(0.009)	(0.005)
AIG: Reading & Math	0.318***	0.157***	0.111***	0.205***
	(0.017)	(0.013)	(0.016)	(0.009)
School-level controls	Y	Y	Y	Y
Constant	-0.452	-1.080	-0.641	-0.768
	(0.727)	(0.676)	(0.658)	(0.474)
Observations	9732	12851	12486	35069
R^2				

Standard errors in parentheses
* $p < .10$, ** $p < .05$, *** $p < .01$

Figure 1: ITT Impact on LevelSet Lexile

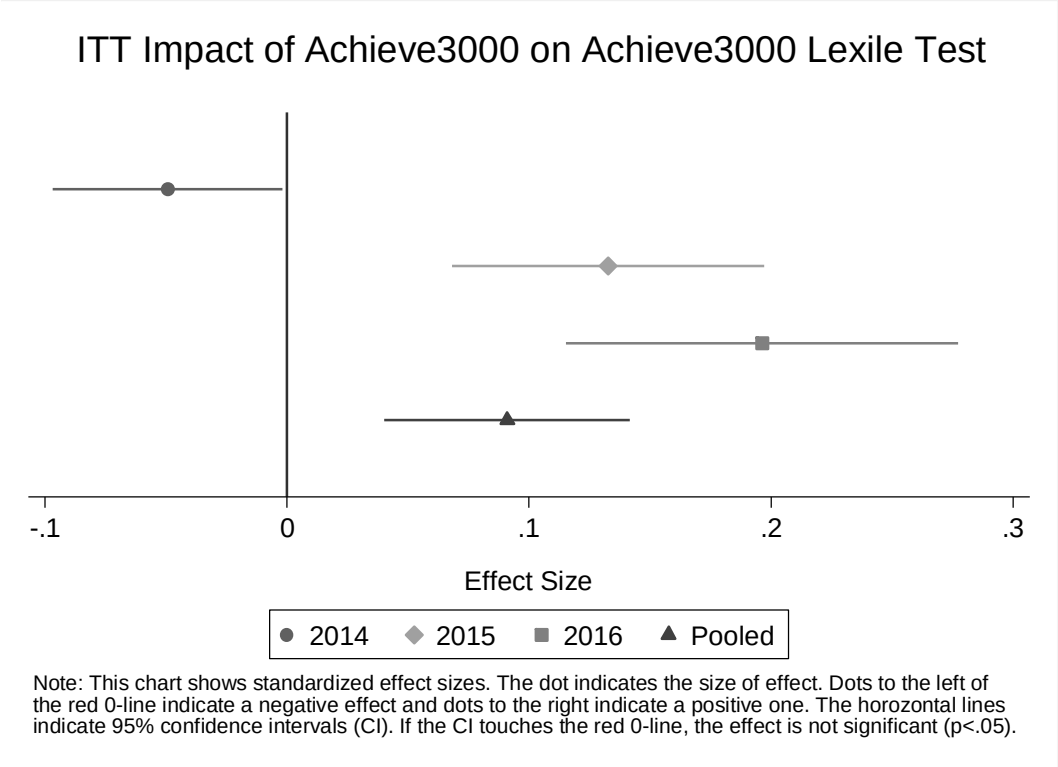


Figure 2: TOT Impact on LevelSet Lexile

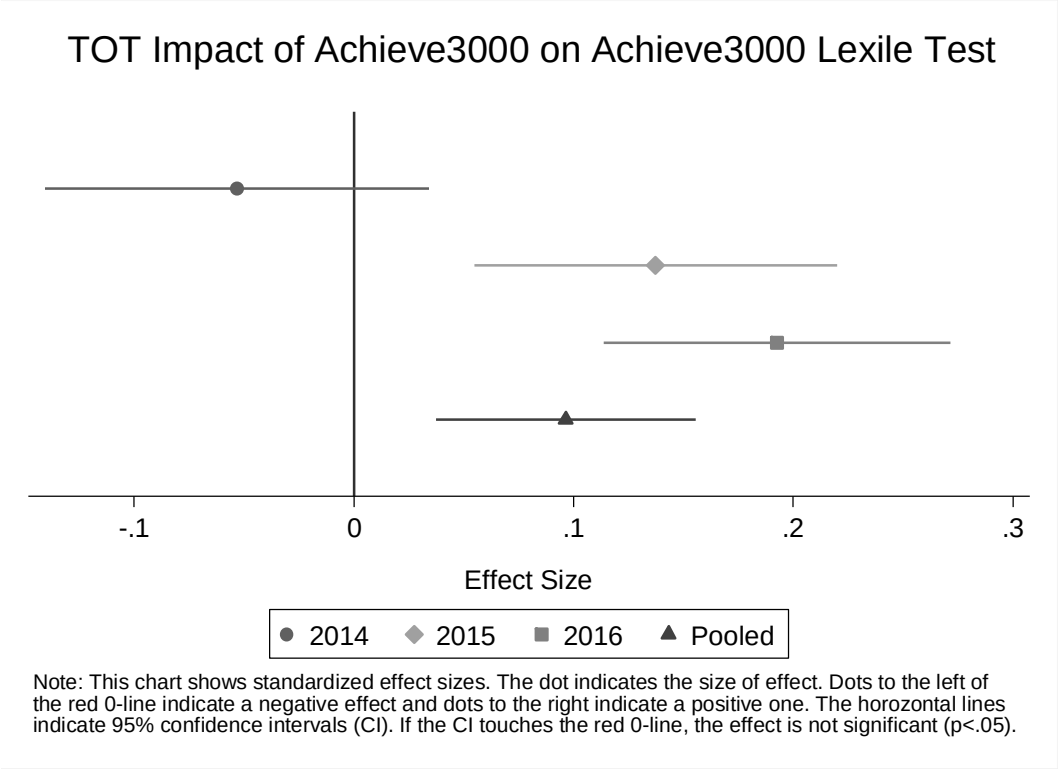
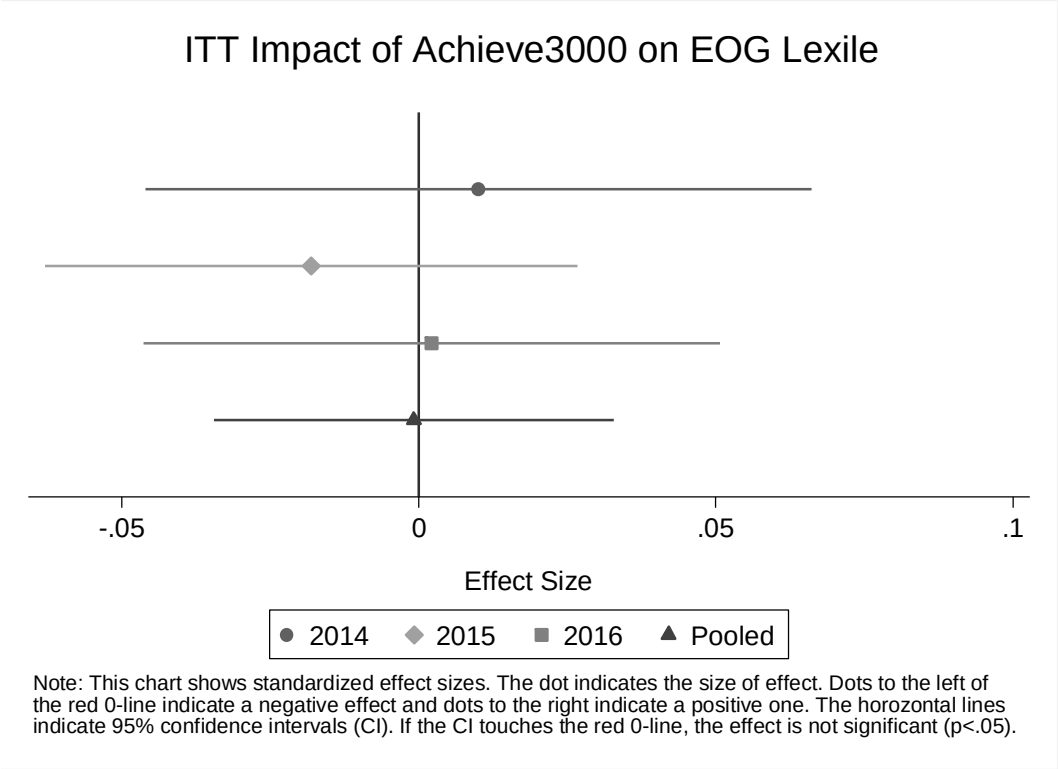
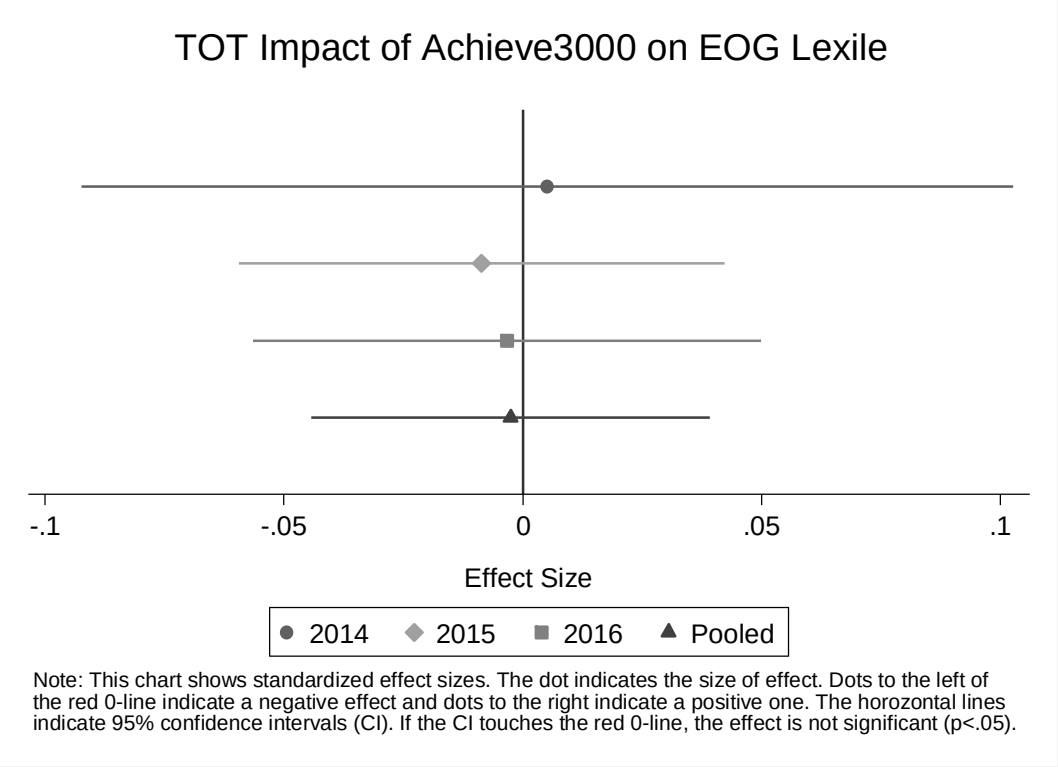


Figure 3: ITT Impact on EOG Lexile



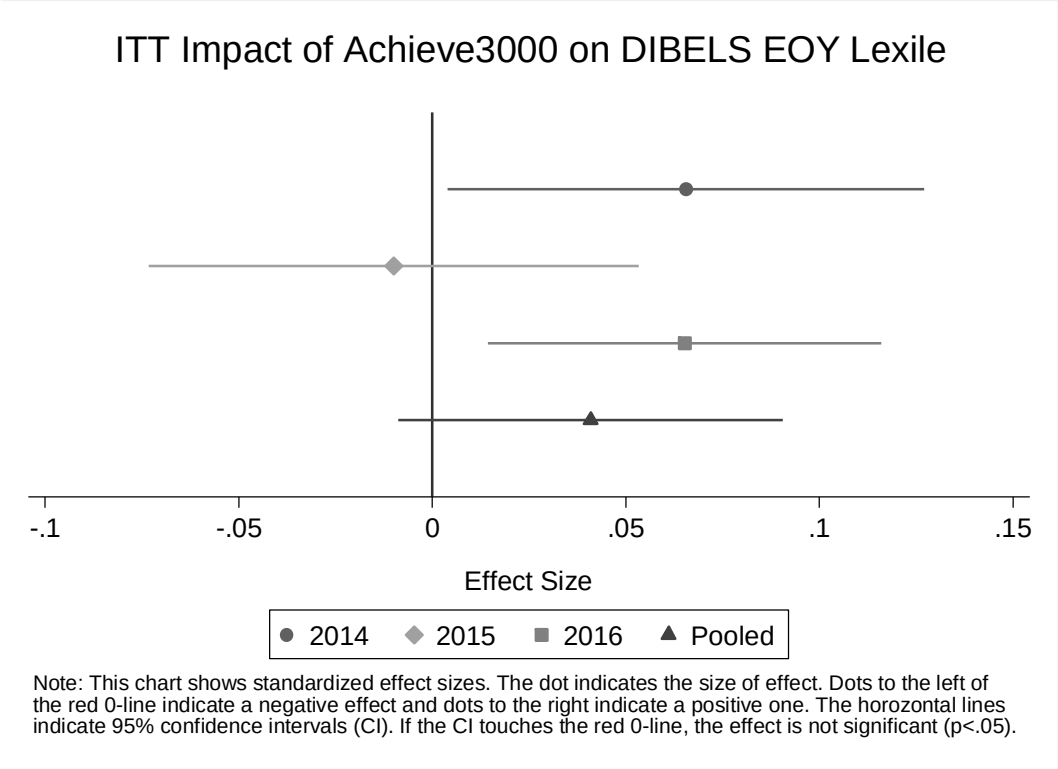
Note: This chart shows standardized effect sizes. The dot indicates the size of effect. Dots to the left of the red 0-line indicate a negative effect and dots to the right indicate a positive one. The horozontal lines indicate 95% confidence intervals (CI). If the CI touches the red 0-line, the effect is not significant (p<.05).

Figure 4: TOT Impact on EOG Lexile



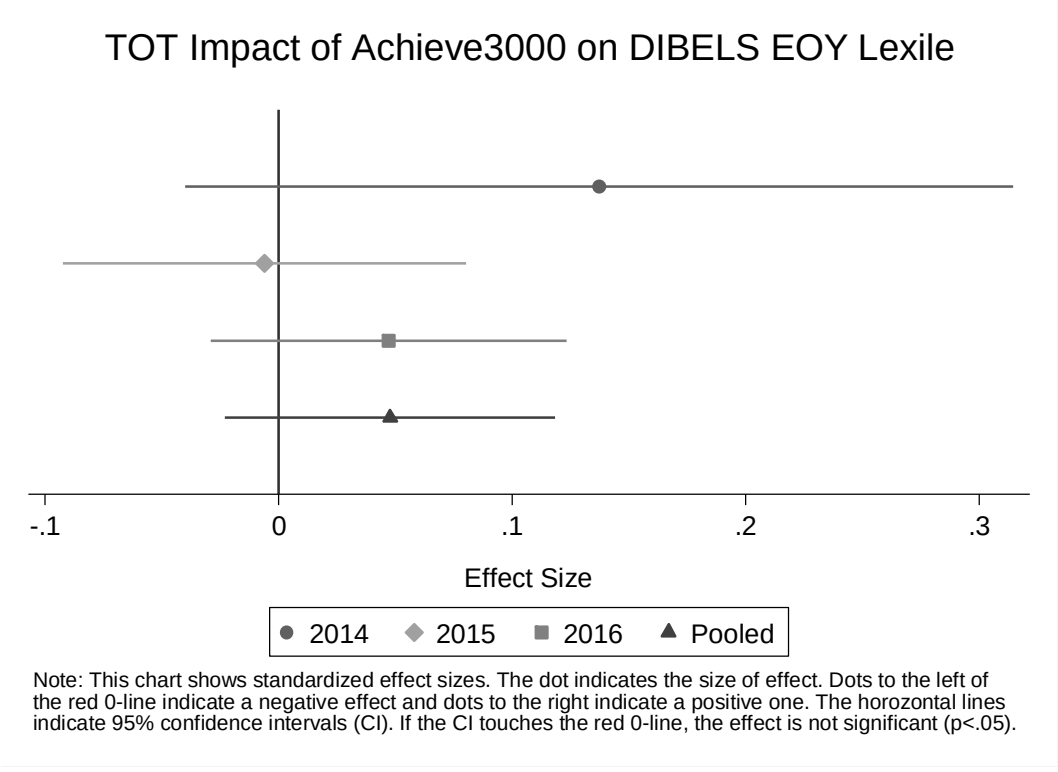
Note: This chart shows standardized effect sizes. The dot indicates the size of effect. Dots to the left of the red 0-line indicate a negative effect and dots to the right indicate a positive one. The horozontal lines indicate 95% confidence intervals (CI). If the CI touches the red 0-line, the effect is not significant (p<.05).

Figure 5: ITT Impact on DIBELS ORF Lexile



Note: This chart shows standardized effect sizes. The dot indicates the size of effect. Dots to the left of the red 0-line indicate a negative effect and dots to the right indicate a positive one. The horozontal lines indicate 95% confidence intervals (CI). If the CI touches the red 0-line, the effect is not significant (p<.05).

Figure 6: TOT Impact on DIBELS ORF Lexile



Note: This chart shows standardized effect sizes. The dot indicates the size of effect. Dots to the left of the red 0-line indicate a negative effect and dots to the right indicate a positive one. The horozontal lines indicate 95% confidence intervals (CI). If the CI touches the red 0-line, the effect is not significant (p<.05).

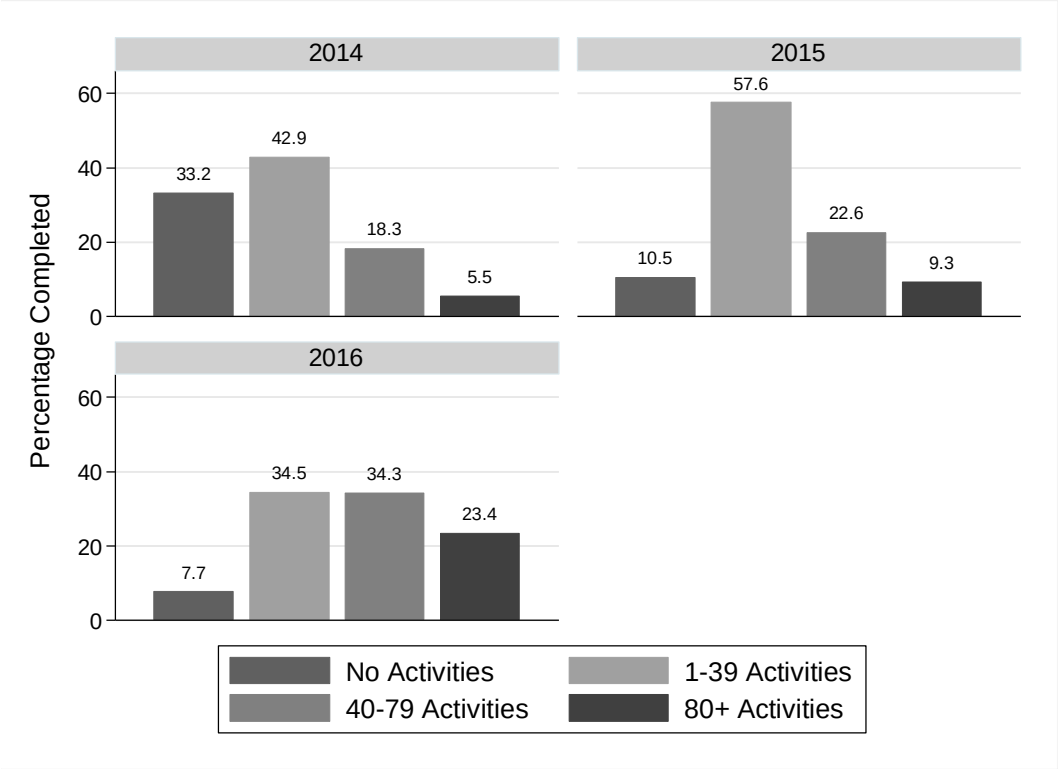



Figure 7: Activity Completion