WWC Intervention Report

A summary of findings from a systematic review of the evidence

ies INSTITUTE OF EDUCATION SCIENCES

Beginning Reading   **February 2018**

Achieve3000®

Intervention Description[1]

Achieve3000® is a supplemental online literacy program that provides nonfiction reading content to students in grades preK–12 and focuses on building phonemic awareness, phonics, fluency, reading comprehension, vocabulary, and writing skills.

Achieve3000® is designed to help students advance their nonfiction reading skills by providing differentiated online instruction. Teachers use the program with an entire class but the assignments are tailored to each student's reading ability level. For example, teachers assign an article and related activities to an entire class; the program then tailors the version of the article to each student by automatically increasing the difficulty of text when a student is ready for more challenging text. *Achieve3000®* provides lessons that follow a five-step routine: (1) respond to a Before Reading Poll, (2) read an article, (3) answer activity questions, (4) respond to an After Reading Poll, and (5) answer a Thought Question. Progress reports and student usage data, provided by the online tool, enable teachers to track both whole-class and individual student progress.

The program is designed for diverse student groups, including general education students, struggling readers in need of intensive tutoring, and English learners.

This intervention report presents findings from a systematic review of *Achieve3000®* conducted using the WWC Procedures and Standards Handbook (version 3.0) and the Beginning Reading review protocol (version 3.0).

Research[2]

The What Works Clearinghouse (WWC) identified one study of *Achieve3000®* that both falls within the scope of the Beginning Reading topic area and meets WWC group design standards. This study meets WWC group design standards with reservations. This study included 14,493 students in grades 2 and 3 in 32 schools in a single school district in North Carolina.[3]

According to the WWC review, the extent of evidence for *Achieve3000®* on the reading achievement outcomes of beginning readers was small for one outcome domain—reading fluency. No studies meet WWC group design standards in the three other domains, so this intervention report does not report on the effectiveness of *Achieve3000®* for those domains.[4] (See the Effectiveness Summary on p. 6 for more details of effectiveness by domain.)

Effectiveness

Achieve3000® had no discernible effects on reading fluency for beginning readers.

Table 1. Summary of findings[5]

Outcome domain	Rating of effectiveness	Improvement index (percentile points)		Number of studies	Number of students	Extent of evidence
		Average	Range			
Reading fluency	No discernible effects	+2	na	1	14,493	Small

na = not applicable

Intervention Information

Background

The developer and distributor of *Achieve3000*® is Achieve3000, Inc. The current Achieve3000 literacy product line includes *Smarty Ants*® (grades preK–2),[6] *KidBiz3000*® (grades 2–5), *TeenBiz3000*® (grades 6–8), *Empower3000*® (grades 9–12), and *Spark3000*® (adult education). In addition, *KidBizPro*®, *TeenBizPro*®, and *EmpowerPro*® deliver literacy instruction in content-area classrooms (that is, social studies and science). In this intervention report, the only study that met standards used *KidBiz3000*®. Address: 1985 Cedar Bridge Avenue, Suite 3, Lakewood, NJ 08701. Email: office@achieve3000.com. Web: http://www.achieve3000.com/ Telephone: (888) 968-6822.

Intervention details

Smarty Ants® and *KidBiz3000*® provide instruction at the grade levels applicable to the Beginning Reading review (i.e., grades K–3). *Smarty Ants*®, available in English and with Spanish support, is a foundational literacy program that provides lessons online and can be used as a core curriculum or as a supplement to whole-class instruction. Following a placement test, the adaptive content system automatically delivers instruction aligned to a student's skill level and learning pace. Instruction takes place through a series of educational games and is designed to address student competency in foundational reading skills, including alphabetics, phonemic awareness, phonics, fluency, vocabulary, and sight words (that is, commonly used words that students are taught to identify by memory instead of using decoding strategies). A teacher dashboard enables teachers to track both individual- and class-level progress.

KidBiz3000® products, which were assessed in the study that meets WWC group design standards with reservations in this review, are delivered through an online instructional format. All students begin by taking a placement test (the LevelSet™ assessment) to identify a student's reading level. Teachers can assign one nonfiction article to the whole class to read, but the program automatically selects a version of the article aligned to each student's reading ability based on the placement test results (or prior performance). Instructional materials, which include more than 15,000 nonfiction articles, focus on contemporary, real-world issues. The program monitors individual reading performance and increases the difficulty of the articles as the student's reading ability improves. *KidBiz3000*® lessons follow a five-step instruction routine:

Step 1: Respond to a Before Reading Poll. Students start each lesson by taking a poll to express their own opinions related to the subject of the article. They answer a multiple-choice question and write an explanation of why they answered the poll's question as they did.

Step 2: Read an Article. Students read a nonfiction article that discusses a contemporary issue and then review a Dig Deeper section that provides them with additional background and details about the content area. A student receives one of 12 English versions or eight Spanish versions of the article matched to his or her reading level. The program defines select vocabulary words, and an audio clip properly pronounces each word.

The program includes a Reading Connections section that provides other tools and tasks to help students comprehend what they read, such as (1) highlighting text for future reference; (2) a note field that enables students to summarize their thoughts on the article or to take notes; (3) a place for students to pose additional questions concerning the article's content; and (4) a field that encourages students to identify key themes in the article. Students are expected to complete, at a minimum, two Reading Connections per lesson.

Step 3: Answer Activity Questions. After reading an article, students respond to a series of vocabulary and reading comprehension questions (that is, around summarization, central ideas and details, and text structure and development). Based on responses to these questions, the program determines when students are ready for more complex text and then automatically adjusts the reading level of the text they receive in the next lesson. Teachers can review students' responses to each question, including whether they chose the correct answer on the first or second try.

Step 4: Respond to an After Reading Poll. Students return to the poll question (Step 1) to again express their opinions, factoring in any new information they might have acquired from the article they read. This step aims to teach students the importance of evidence, and provides an opportunity to share and reflect on their learning.

Step 5: Answer a Thought Question. In the final step of the lesson, students provide a written response to a question based on the article they read in Step 2 that includes examples, reasons, and evidence to support their responses.

In addition, each lesson includes a Stretch Article that is written at a level higher than the student's instructional level. A student can complete this article and accompanying Stretch Activity as homework or immediately after the lesson if he or she finishes early. The Stretch Article can also be embedded into the lesson; students can use the new information they have acquired to revise their answers to the Thought Question.

Professional learning for classroom teachers is available onsite, live online, and via online videos. Sessions include hands-on practice for teachers to master implementation strategies, monitor student data, and create an action plan for each student.

Recommended use is 80 lessons over the course of the school year. *KidBiz3000*® has program versions tailored to the standards of each state. The program is accessible on multiple devices and platforms, including Apple, Android, and Chromebook products. Apps enable students to access lessons with or without Internet connectivity from school or from home.

Cost

Achieve3000, Inc. offers program packages that include software access for the academic school year. As of January 2017, the unit price of $14,675 covered up to 12 teacher licenses, 250 student licenses, 250 parent/guardian licenses, and 2 days of professional development. The company also offers a per-student pricing option of $42 with a 100-student minimum. Professional development is required with the per-student option and sold separately at $2,300 per day. For more information about program options and pricing, contact Achieve3000, Inc. at office@achieve3000.com.

Research Summary

The WWC identified two eligible studies that investigated the effects of *Achieve3000®* on the reading achievement of beginning readers. Nine additional studies were identified but do not meet WWC eligibility criteria (see the Glossary of Terms in this document for a definition of this term and other commonly used research terms) for review in this topic area. Citations for all 11 studies are in the References section, which begins on p. 7.

Table 2. Scope of reviewed research

Grade	2–3
Delivery method	Whole class
Program type	Supplement

The WWC reviewed two eligible studies against group design standards. One study is a randomized controlled trial that meets WWC group design standards with reservations. This report summarizes the study. The remaining study does not meet WWC group design standards.

Summary of studies meeting WWC group design standards without reservations

No studies of *Achieve3000®* meet WWC group design standards without reservations.

Summary of studies meeting WWC group design standards with reservations

Hill and Lenard (2016) conducted a cluster, or group-based, randomized controlled trial examining the effects of *KidBiz3000®* (the elementary school version of *Achieve3000®*) on students in grades 2–5 in 32 elementary schools in North Carolina. The study matched schools based on End-of-Grade reading composite scores from spring 2013. Within each pair of matched schools, the authors randomly assigned one school to receive the intervention and the other to receive business-as-usual literacy instruction. Because the study was a cluster randomized controlled trial that analyzed outcomes for students who enrolled in the school after school random assignment, the integrity of the study's random assignment was jeopardized. However, the authors demonstrated baseline equivalence of the intervention and comparison groups in the analytic sample.[7] The study took place over two school years: 2013–14 and 2014–15. The WWC based its effectiveness rating on findings from students in grades 2 and 3 in these two school years: the 2013–14 sample included 7,197 students and the 2014–15 sample included 7,296 students. The findings for students in grades 4 and 5 were not reviewed for this intervention report because these are not eligible for review under the Beginning Reading topic area protocol.

Effectiveness Summary

The WWC review of *Achieve3000*® for the Beginning Reading topic area includes outcomes in four domains: alphabetics, reading fluency, comprehension, and general reading achievement. The one study of *Achieve3000*® that meets WWC group design standards reported findings in one of the four domains: reading fluency. The following findings present the authors' estimates and WWC-calculated estimates of the size and statistical significance of the effects of *Achieve3000*® on beginning readers. Additional comparisons are available as supplemental findings in Appendix D. The supplemental findings do not factor into the intervention's rating of effectiveness. For a more detailed description of the rating of effectiveness and extent of evidence criteria, see the WWC Rating Criteria on p. 15.

Summary of effectiveness for the reading fluency domain
Table 3. Rating of effectiveness and extent of evidence for the reading fluency domain

Rating of effectiveness	Criteria met
No discernible effects *No affirmative evidence of effects.*	In the one study that reported findings, the estimated impact of the intervention on outcomes in the *reading fluency* domain was neither statistically significant nor substantively important.
Extent of evidence	**Criteria met**
Small	One study that included 14,493 students in 32 schools reported evidence of effectiveness in the *reading fluency* domain.

One study that meets WWC group design standards with reservations reported findings in the reading fluency domain.

Hill and Lenard (2016) reported findings on the Lexile® score of the Dynamic Indicators of Basic Early Literacy Skills (DIBELS) Oral Reading Fluency Test for students in grades 2 and 3 in the 2013–14 and 2014–15 school years. The authors reported, and the WWC confirmed, no statistically significant findings on the DIBELS in the 2014–15 school year. The authors also reported a statistically significant positive finding on the DIBELS in the 2013–14 school year, but this finding was not statistically significant after correcting for multiple comparisons. The WWC characterizes this study finding as an indeterminate effect.

Thus, for the reading fluency domain, one study reported an indeterminate effect. This results in a rating of no discernible effects, with a small extent of evidence.

References

Studies that meet WWC group design standards without reservations

None.

Study that meets WWC group design standards with reservations

Hill, D. V., & Lenard, M. A. (2016). *The impact of Achieve3000 on elementary literacy outcomes: Randomized control trial evidence, 2013-14 to 2014-15* (DRA Report No. 16.02). Cary, NC: Wake County Public School System, Data and Accountability Department.

Additional source:

Hill, D. V., Lenard, M. A., & Page, L. C. (2016, March). *The impact of Achieve3000 on elementary literacy outcomes: Evidence from a two-year randomized control trial.* Paper presented at the Society for Research on Educational Effectiveness (SREE) Spring Conference, Washington, DC. Retrieved from https://eric.ed.gov/?id=ED567483

Study that does not meet WWC group design standards

Shannon, L., & Grant, B-J. (2015). *A final report for the evaluation of the Achieve3000 programs.* Charlottesville, VA: Magnolia Consulting. Retrieved from https://eric.ed.gov/?id=ED563449 The study does not meet WWC group design standards because the analysis does not provide a credible measure of the effectiveness of the intervention.

Additional source:

Magnolia Consulting. (2017). *An addendum to the final report for the evaluation of the Achieve3000 programs: Impact analyses for grades 6 and 9.* Charlottesville, VA: Author.

Studies that are ineligible for review using the Beginning Reading Evidence Review Protocol

Achieve3000. (2009). *National study: Lexile® analysis reveals dramatic non-fiction reading skill increases, 2009.* Lakewood, NJ: Author. The study is ineligible for review because it does not use an eligible design.

Achieve3000. (2011). *National Lexile study: 2009-2010.* Lakewood, NJ: Author. The study is ineligible for review because it does not use an eligible design.

Achieve3000. (2016). *National Lexile Study: 2015-2016.* Lakewood, NJ: Author. The study is ineligible for review because it does not use an eligible design.

Banister-Doughty, S., Hollingsworth, G., & Tittle, F. (2015). *Effective literacy interventions at two middle Tennessee high schools* (Doctoral dissertation). Available from ProQuest Dissertations and Theses database. (UMI No. 3733242) The study is ineligible for review because it does not use an eligible design.

Borman, G. D., Park, S. J., & Min, S. (2015). *The district-wide effectiveness of the Achieve3000 program: A quasi-experimental study.* Madison: Measured Decisions, Inc. and the University of Wisconsin-Madison. Retrieved from https://eric.ed.gov/?id=ED558845 The study is ineligible for review because it does not use a sample aligned with the protocol.

Tracey, D. H., & Young, J. W. (2004). *Evaluation of KidBiz3000: Bayonne study final report.* Lakewood, NJ: Achieve3000. The study is ineligible for review because it does not use a sample aligned with the protocol.

Additional source:

Tracey, D. H., & Young, J. W. (2005). *Bayonne, NJ schools 2003-2004.* Lakewood, NJ: Achieve3000.

Urdegar, S. M. (2014). *Achieve 3000: An analysis of usage and impact, 2013– 2014. Technical Note. Volume 3, Number 1.* Miami, FL: Miami-Dade County Public Schools, Office of Assessment, Research, and Data Analytics. Retrieved from https://eric.ed.gov/?id=ED561897 The study is ineligible for review because it does not use a sample aligned with the protocol.

Wright, R. (2010). *Impact of a web-based reading program on sixth-grade English language learners* (Doctoral dissertation). Retrieved from https://eric.ed.gov/?id=ED546286 The study is ineligible for review because it does not use a sample aligned with the protocol.

Young, J. W. (2005). *Evaluation study: Chicago public schools 2003-2004.* Lakewood, NJ: Achieve3000. The study is ineligible for review because it does not use a sample aligned with the protocol.

Appendix A: Research details for Hill and Lenard (2016)

Hill, D. V., & Lenard, M. A. (2016). *The impact of Achieve3000 on elementary literacy outcomes: Randomized control trial evidence, 2013–14 to 2014–15* (DRA Report No. 16.02). Cary, NC: Wake County Public School System, Data and Accountability Department.

Additional source:
Hill, D. V., Lenard, M. A., & Page, L. C. (2016, March). *The impact of Achieve3000 on elementary literacy outcomes: Evidence from a two-year randomized control trial.* Paper presented at the Society for Research on Educational Effectiveness (SREE) Spring Conference, Washington, DC. Retrieved from https://eric.ed.gov/?id=ED567483

Table A. Summary of findings   **Meets WWC group design standards with reservations**

| Outcome domain | Sample size | Study findings | |
		Average improvement index *(percentile points)*	Statistically significant
Reading fluency	32 schools/ 14,493 students	+2	No

Setting

The study was conducted in the Wake County Public School System (WCPSS) in Raleigh, North Carolina. As a countywide district, WCPSS has schools representing suburban, urban, and rural areas.

Study sample

The authors used a cluster randomized controlled trial design to study the effects of *KidBiz3000®* on the reading achievement of students in grades 2-5 (only analyses for grades 2-3 were eligible for review under the Beginning Reading Protocol). The study took place over two school years (from 2013–14 to 2014–15) in 32 elementary schools. In summer of 2013, the authors matched pairs of schools based on their average 2013 End-of-Grade (EOG) reading composite scores, and then from within each matched pair, randomly assigned one school to the intervention group and one school to the comparison group. In both study years, the same 16 *KidBiz3000®* schools and 16 comparison schools participated in the study.

The WWC considers random assignment to be jeopardized because the analytic sample included students who enrolled in study schools after random assignment. The 2-year combined analysis sample included 14,493 students: 7,540 students were in the *Achieve3000®* group, and 6,953 students were in the comparison group. The reported sample sizes count some students more than once because some second-grade students in 2013–14 may also appear as third-grade students in 2014–15.

No demographic data were available on the study sample in grades 2–3; however, in the 32 participating schools, the student population was 51% White, 26% African American, and 19% Hispanic. Moreover, 12% of students had disabilities, 9% of students in study schools were English learners, and 7% of students were academically and intellectually gifted (AIG). Approximately one-third of the district's students were certified for free or reduced-price lunch.

Intervention group

KidBiz3000® was implemented in 30-minute sessions, two times per week over a school year. On initial use, students took a 30-minute test to measure their baseline reading achievement. For each lesson, students followed *Achieve3000®*'s five-step literacy routine. Across both years of the study, about 8% of intervention students completed at least 80 lessons (i.e., the developer's recommended dosage), about 21% used 40–79 lessons, about 50% used 1–39 lessons, and 22% of students completed no activities. The *Achieve3000®* intervention was used to supplement a standard core reading curriculum; however, the authors did not identify which core curriculum was used.

Comparison group

The comparison condition was business-as-usual reading instruction. Classrooms in comparison schools did not receive a supplemental curriculum.

Outcomes and measurement

Outcomes were measured in spring 2014 and 2015, and the pretests were administered at the beginning of the school year, in the fall of 2013 and 2014, respectively. All findings reported in the study reflect the impact of the intervention after 1 year of student exposure; in particular, while some students whose outcomes were analyzed in the second year of the study had received the intervention in both years, outcomes from spring 2015 were analyzed using a fall 2014 pretest (which was administered 1 year after the intervention had begun). The authors used the Dynamic Indicators of Basic Early Literacy Skills (DIBELS) Oral Reading Fluency assessment for students in grades 2–3. The DIBELS outcome was reviewed in the reading fluency domain. For a more detailed description of these outcome measures, see Appendix B.

The authors also measured outcomes using the state EOG assessment and the Achieve3000 LevelSet Lexile® assessment. Because the authors presented findings that were combined across grades 2–5 or grades 4–5 for these measures, they are reviewed under the Adolescent Literacy topic area and are reported on in the *Achieve3000®* Adolescent Literacy intervention report.

The authors present treatment on the treated (TOT) estimates of *Achieve3000®* impacts on the DIBELS outcome for each sample. These findings do not meet WWC complier average causal effect (CACE) guidance since the study is a cluster RCT that includes joiners (i.e., students who joined the sample after randomization took place).

The study also presented supplemental findings for a subgroup of academically and intellectually gifted (AIG) students. These supplemental findings are reported in Appendix D and do not factor into the intervention's rating of effectiveness.

The authors also conducted subgroup analyses by special education status (students with disabilities) and English learners. These subgroup analyses are not eligible for review under the Beginning Reading review protocol.

Support for implementation

The study included professional development to train teachers, consisting of two 2.5-hour large-group training sessions and one 1-hour small-group session. Teachers were able to obtain follow-up help if needed.

Appendix B: Outcome measures for the reading fluency domain

Comprehension
Reading fluency
Dynamic Indicators of Basic Early Literacy Skills (DIBELS) Oral Reading Fluency (ORF) subtest   The DIBELS ORF subtest is a standardized, individually-administered assessment that measures students' reading accuracy and reading rate. Reading rates are measured as the number of words read correctly per minute. MetaMetrics, Inc. 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 (as cited in Hill & Lenard, 2016).

Appendix C: Findings included in the rating for the reading fluency domain

Outcome measure	Study sample	Sample size	Mean (standard deviation)		WWC calculations			p-value
			Intervention group	Comparison group	Mean difference	Effect size	Improvement index	
Hill and Lenard (2016)[a]								
DIBELS Oral Reading Fluency subtest	2014 sample	32 schools/ 7,197 students	674.88 (300.98)	656.43 (308.97)	18.45	0.06	+2	.04
DIBELS Oral Reading Fluency subtest	2015 sample	32 schools/ 7,296 students	663.89 (297.52)	650.85 (300.51)	13.05	0.04	+2	> .05
Domain average for reading fluency (Hill & Lenard, 2016)						0.05	+2	**Not statistically significant**
Domain average for reading fluency across all studies						0.05	+2	**na**

Table Notes: For mean difference, effect size, and improvement index values reported in the table, a positive number favors the intervention group and a negative number favors the comparison group. The effect size is a standardized measure of the effect of an intervention on outcomes, representing the average change expected for all individuals who are given the intervention (measured in standard deviations of the outcome measure). The improvement index is an alternate presentation of the effect size, reflecting the change in an average individual's percentile rank that can be expected if the individual is given the intervention. The WWC-computed average effect size is a simple average rounded to two decimal places; the average improvement index is calculated from the average effect size. The statistical significance of the study's domain average was determined by the WWC. Some statistics may not sum as expected due to rounding. na = not applicable.

[a] For Hill and Lenard (2016), a correction for multiple comparisons was needed and resulted in a WWC-computed critical p-value of .025 for the DIBELS 2014 sample; the specific p-value for this contrast was reported as < .05 in the original study, but an exact p-value of .044 was obtained via an author query. As a result, the WWC does not find this result to be statistically significant after the correction for multiple comparisons. The p-value for the DIBELS 2015 sample was reported in the original study. Findings from 2014 and 2015 are presented separately since these samples partially overlap (i.e., second-grade students in the 2014 sample appear as third-grade students in the 2015 sample), and because the 2015 sample used a different point of baseline measurement. Findings from both years (2014 and 2015) reflect 1-year impacts for students. The adjusted group means, unadjusted standard deviations, and sample sizes were obtained through an author query. This study is characterized as having an indeterminate effect because the mean effect is neither statistically significant nor substantively important. For more information, please refer to the WWC Procedures and Standards Handbook (version 3.0), p. 26.

Appendix D: Description of supplemental findings for the reading fluency domain

Outcome measure	Study sample	Sample size	Mean (standard deviation)		WWC calculations			p-value
			Intervention group	Comparison group	Mean difference	Effect size	Improvement index	
Hill and Lenard (2016)[a]								
DIBELS Oral Reading Fluency subtest	2014 AIG students	32 schools/ 331 students	1,049.40 (192.30)	1,066.30 (nr)	−16.90	nr	nr	.73
DIBELS Oral Reading Fluency subtest	2015 AIG students	32 schools/ 173 students	1,063.60 (180.27)	1,107.20 (nr)	−43.60	nr	nr	.73

Table Notes: The supplemental findings presented in this table are additional findings from studies in this report that meet WWC design standards with or without reservations, but do not factor into the determination of the intervention rating. For mean difference, effect size, and improvement index values reported in the table, a positive number favors the intervention group and a negative number favors the comparison group. The effect size is a standardized measure of the effect of an intervention on outcomes, representing the average change expected for all individuals who are given the intervention (measured in standard deviations of the outcome measure). The improvement index is an alternate presentation of the effect size, reflecting the change in an average individual's percentile rank that can be expected if the individual is given the intervention. Some statistics may not sum as expected due to rounding. nr = not reported. AIG = academically and intellectually gifted.

[a] For Hill and Lenard (2016), the WWC did not need to make corrections for clustering, multiple comparisons, or to adjust for baseline differences. The *p*-values presented here were obtained through an author query. The adjusted group means, unadjusted standard deviations, and sample sizes were also obtained through an author query.

Endnotes

[1] The descriptive information for this intervention comes from publicly available sources: the program's website http://www.achieve3000.com (accessed April 3, 2017), Achieve3000 Lessons and Resources, 2010 Price Sheet, Shannon and Grant (2015), and the EdSurge product review (https://www.edsurge.com/product-reviews/achieve3000). The What Works Clearinghouse (WWC) requests distributors review the intervention description sections for accuracy from their perspective. The WWC provided the distributor with the intervention description in April 2017, and the WWC incorporated feedback from the distributor. Further verification of the accuracy of the descriptive information for this intervention is beyond the scope of this review. The WWC published a separate intervention report under the Adolescent Literacy topic area, which covers the combined grades 2–5 samples: https://ies.ed.gov/ncee/wwc/InterventionReport/691.

[2] The literature search reflects documents publicly available by February 2017. Reviews of the studies in this report used the standards from the WWC Procedures and Standards Handbook (version 3.0) and the Beginning Reading review protocol (version 3.0). The evidence presented in this report is based on available research. Findings and conclusions may change as new research becomes available.

[3] Per the Beginning Reading topic area protocol, the current intervention report includes findings for students in grades 2 and 3. Hill and Lenard (2016) also reported findings for additional outcome measures using a combined sample of students in grades 2–5 and a sample of students in grades 4 and 5. These findings are not eligible for review under the Beginning Reading topic area protocol but are reviewed in the Adolescent Literacy *Achieve3000*® intervention report. Furthermore, the reported sample size of 14,493 students overestimates the number of unique students in the eligible study sample of students in grades 2 and 3. Hill and Lenard (2016) included students who attended study schools in two consecutive school years (2013–14 and 2014–15). Some second grade students in the 2013–14 sample appear as third-grade students in the 2014–15 sample, and are therefore counted twice in the overall sample size presented in this report.

[4] Please see the Beginning Reading review protocol (version 3.0) for a list of all the outcome domains.

[5] For criteria used to determine the rating of effectiveness and extent of evidence, see the WWC Rating Criteria on p. 15. These improvement index numbers show the average and range of individual-level improvement indices for all findings across the studies.

[6] Achieve3000, Inc. acquired *Smarty Ants*® from Smarty Ants, Inc. in August 2015. The program provides game-based instruction to students in prekindergarten through second grade, but does not use the *Achieve3000*® five-step model.

[7] The WWC Reviewer Guidance, for use with the WWC Procedures and Standards Handbook (version 3.0), indicates that if the authors of a cluster randomized controlled trial study characterize the intervention as having effects on student scores (rather than only on cluster-level scores), and some students enter clusters after random assignment, then the study must demonstrate equivalence of the analytic intervention and comparison groups at baseline.

Recommended Citation

U.S. Department of Education, Institute of Education Sciences, What Works Clearinghouse. (2018, February). *Beginning Reading intervention report: Achieve3000*®*.* Retrieved from https://whatworks.ed.gov

WWC Rating Criteria

Criteria used to determine the rating of a study

Study rating	Criteria
Meets WWC group design standards without reservations	A study that provides strong evidence for an intervention's effectiveness, such as a well-implemented RCT.
Meets WWC group design standards with reservations	A study that provides weaker evidence for an intervention's effectiveness, such as a QED or an RCT with high attrition that has established equivalence of the analytic samples.

Criteria used to determine the rating of effectiveness for an intervention

Rating of effectiveness	Criteria
Positive effects	Two or more studies show statistically significant positive effects, at least one of which met WWC group design standards for a strong design, AND No studies show statistically significant or substantively important negative effects.
Potentially positive effects	At least one study shows a statistically significant or substantively important positive effect, AND No studies show a statistically significant or substantively important negative effect AND fewer or the same number of studies show indeterminate effects than show statistically significant or substantively important positive effects.
Mixed effects	At least one study shows a statistically significant or substantively important positive effect AND at least one study shows a statistically significant or substantively important negative effect, but no more such studies than the number showing a statistically significant or substantively important positive effect, OR At least one study shows a statistically significant or substantively important effect AND more studies show an indeterminate effect than show a statistically significant or substantively important effect.
Potentially negative effects	One study shows a statistically significant or substantively important negative effect and no studies show a statistically significant or substantively important positive effect, OR Two or more studies show statistically significant or substantively important negative effects, at least one study shows a statistically significant or substantively important positive effect, and more studies show statistically significant or substantively important negative effects than show statistically significant or substantively important positive effects.
Negative effects	Two or more studies show statistically significant negative effects, at least one of which met WWC group design standards for a strong design, AND No studies show statistically significant or substantively important positive effects.
No discernible effects	None of the studies shows a statistically significant or substantively important effect, either positive or negative.

Criteria used to determine the extent of evidence for an intervention

Extent of evidence	Criteria
Medium to large	The domain includes more than one study, AND The domain includes more than one school, AND The domain findings are based on a total sample size of at least 350 students, OR, assuming 25 students in a class, a total of at least 14 classrooms across studies.
Small	The domain includes only one study, OR The domain includes only one school, OR The domain findings are based on a total sample size of fewer than 350 students, AND, assuming 25 students in a class, a total of fewer than 14 classrooms across studies.

Glossary of Terms

Attrition Attrition occurs when an outcome variable is not available for all subjects initially assigned to the intervention and comparison groups. If a randomized controlled trial (RCT) or regression discontinuity design (RDD) study has high levels of attrition, the validity of the study results can be called into question. An RCT with high attrition cannot receive the highest rating of *Meets WWC Group Design Standards without Reservations*, but can receive a rating of *Meets WWC Group Design Standards with Reservations* if it establishes baseline equivalence of the analytic sample. Similarly, the highest rating an RDD with high attrition can receive is *Meets WWC RDD Standards with Reservations*.

For single-case design research, attrition occurs when an individual fails to complete all required phases or data points in an experiment, or when the case is a group and individuals leave the group. If a single-case design does not meet minimum requirements for phases and data points within phases, the study cannot receive the highest rating of *Meets WWC Pilot Single-Case Design Standards without Reservations*.

Baseline A point in time before the intervention was implemented in group design research and in regression discontinuity design studies. When a study is required to satisfy the baseline equivalence requirement, it must be done with characteristics of the analytic sample at baseline. In a single-case design experiment, the baseline condition is a period during which participants are not receiving the intervention.

Clustering adjustment An adjustment to the statistical significance of a finding when the units of assignment and analysis differ. When random assignment is carried out at the cluster level, outcomes for individual units within the same clusters may be correlated. When the analysis is conducted at the individual level rather than the cluster level, there is a mismatch between the unit of assignment and the unit of analysis, and this correlation must be accounted for when assessing the statistical significance of an impact estimate. If the correlation is not accounted for in a mismatched analysis, the study may be too likely to report statistically significant findings. To fairly assess an intervention's effects, in cases where study authors have not corrected for the clustering, the WWC applies an adjustment for clustering when reporting statistical significance.

Confounding factor A confounding factor is a component of a study that is completely aligned with one of the study conditions, making it impossible to separate how much of the observed effect was due to the intervention and how much was due to the factor.

Design The method by which intervention and comparison groups are assigned (group design and regression discontinuity design) or the method by which an outcome measure is assessed repeatedly within and across different phases that are defined by the presence or absence of an intervention (single-case design). Designs eligible for WWC review are randomized controlled trials, quasi-experimental designs, regression discontinuity designs, and single-case designs.

Effect size The effect size is a measure of the magnitude of an effect. The WWC uses a standardized measure to facilitate comparisons across studies and outcomes.

Eligibility A study is eligible for review and inclusion in this report if it falls within the scope of the review protocol and uses either an experimental or matched comparison group design.

Equivalence A demonstration that the analytic sample groups are similar on observed characteristics defined in the review area protocol.

Glossary of Terms

Extent of evidence An indication of how much evidence from group design studies supports the findings in an intervention report. The extent of evidence categorization for intervention reports focuses on the number and sizes of studies of the intervention in order to give an indication of how broadly findings may be applied to different settings. There are two extent of evidence categories: small and medium to large.

- **small:** includes only one study, or one school, or findings based on a total sample size of less than 350 students and 14 classrooms (assuming 25 students in a class)
- **medium to large:** includes more than one study, more than one school, and findings based on a total sample of at least 350 students or 14 classrooms

Gain scores The result of subtracting the pretest from the posttest for each individual in the sample. Some studies analyze gain scores instead of the unadjusted outcome measure as a method of accounting for the baseline measure when estimating the effect of an intervention. The WWC reviews and reports findings from analyses of gain scores, but gain scores do not satisfy the WWC's requirement for a statistical adjustment under the baseline equivalence requirement. This means that a study that must satisfy the baseline equivalence requirement and has baseline differences between 0.05 and 0.25 standard deviations *Does Not Meet WWC Group Design Standards* if the study's only adjustment for the baseline measure was in the construction of the gain score.

Group design A study design in which outcomes for a group receiving an intervention are compared to those for a group not receiving the intervention. Comparison group designs eligible for WWC review are randomized controlled trials and quasi-experimental designs.

Improvement index Along a percentile distribution of individuals, the improvement index represents the gain or loss of the average individual due to the intervention. As the average individual starts at the 50th percentile, the measure ranges from –50 to +50.

Intervention An educational program, product, practice, or policy aimed at improving student outcomes.

Intervention report A summary of the findings of the highest-quality research on a given program, product, practice, or policy in education. The WWC searches for all research studies on an intervention, reviews each against design standards, and summarizes the findings of those that meet WWC design standards.

Multiple comparison adjustment An adjustment to the statistical significance of results to account for multiple comparisons in a group design study. The WWC uses the Benjamini-Hochberg (BH) correction to adjust the statistical significance of results within an outcome domain when study authors perform multiple hypothesis tests without adjusting the *p*-value. The BH correction is used in three types of situations: studies that tested multiple outcome measures in the same outcome domain with a single comparison group; studies that tested a given outcome measure with multiple comparison groups; and studies that tested multiple outcome measures in the same outcome domain with multiple comparison groups. Because repeated tests of highly correlated constructs will lead to a greater likelihood of mistakenly concluding that the impact was different from zero, in all three situations, the WWC uses the BH correction to reduce the possibility of making this error. The WWC makes separate adjustments for primary and secondary findings.

Please see the WWC Procedures and Standards Handbook (version 3.0) for additional details.

Outcome domain	A group of closely-related outcomes. A domain is the organizing construct for a set of related outcomes through which studies claim effectiveness.
Quasi-experimental design (QED)	A quasi-experimental design (QED) is a research design in which study participants are assigned to intervention and comparison groups through a process that is not random.
Randomized controlled trial (RCT)	A randomized controlled trial (RCT) is an experiment in which eligible study participants are randomly assigned to intervention and comparison groups.
Rating of effectiveness	For group design research, the WWC rates the effectiveness of an intervention in each domain based on the quality of the research design and the magnitude, statistical significance, and consistency in findings. For single-case design research, the WWC rates the effectiveness of an intervention in each domain based on the quality of the research design and the consistency of demonstrated effects. The criteria for the ratings of effectiveness are given in the WWC Rating Criteria on p. 15.
Regression discontinuity design (RDD)	A design in which groups are created using a continuous scoring rule. For example, students may be assigned to a summer school program if they score below a preset point on a standardized test, or schools may be awarded a grant based on their score on an application. A regression line or curve is estimated for the intervention group and similarly for the comparison group, and an effect occurs if there is a discontinuity in the two regression lines at the cutoff.
Single-case design	A research approach in which an outcome variable is measured repeatedly within and across different conditions that are defined by the presence or absence of an intervention.
Standard deviation	The standard deviation of a measure shows how much variation exists across observations in the sample. A low standard deviation indicates that the observations in the sample tend to be very close to the mean; a high standard deviation indicates that the observations in the sample tend to be spread out over a large range of values.
Statistical significance	Statistical significance is the probability that the difference between groups is a result of chance rather than a real difference between the groups. The WWC labels a finding statistically significant if the likelihood that the difference is due to chance is less than 5% ($p < .05$).
Study rating	The result of the WWC assessment of a study. The rating is based on the strength of the evidence of the effectiveness of the educational intervention. Studies are given a rating of *Meets WWC Design Standards without Reservations, Meets WWC Design Standards with Reservations,* or *Does Not Meet WWC Design Standards,* based on the assessment of the study against the appropriate design standards. The WWC has design standards for group design, single-case design, and regression discontinuity design studies.
Substantively important	A substantively important finding is one that has an effect size of 0.25 or greater, regardless of statistical significance.
Systematic review	A review of existing literature on a topic that is identified and reviewed using explicit methods. A WWC systematic review has five steps: 1) developing a review protocol; 2) searching the literature; 3) reviewing studies, including screening studies for eligibility, reviewing the methodological quality of each study, and reporting on high quality studies and their findings; 4) combining findings within and across studies; and, 5) summarizing the review.

Please see the WWC Procedures and Standards Handbook (version 3.0) for additional details.

Intervention Report   Practice Guide   Quick Review   Single Study Review
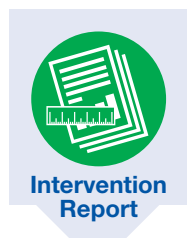
An **intervention report** summarizes the findings of high-quality research on a given program, practice, or policy in education. The WWC searches for all research studies on an intervention, reviews each against evidence standards, and summarizes the findings of those that meet standards.

This intervention report was prepared for the WWC by Mathematica Policy Research under contract ED-IES-13-C-0010.