

# HillRAP Reading Intervention

PROMISING EVIDENCE OF PROGRAM EFFECTIVENESS

*December 2021*

Empirical Education Inc.



## Background

### ABOUT HILLRAP

The Hill Reading Achievement Program (HillRAP) is a structured literacy reading intervention building on the principles of the research-based, multisensory Orton-Gillingham approach (see Ritchey and Goeke, 2006) and is aligned with the recommendations of the National Reading Panel. HillRAP implementation includes a 4 to 1 student to teacher ratio during 45–50-minute instructional sessions, 4–5 days a week. HillRAP is intended for students with persistent reading difficulties, including those with learning disabilities and those who have failed or are at risk for failure in reading, including English language learners. It is typically provided as a Tier 3 or an Exceptional Children’s intervention.

Teachers who deliver HillRAP may pursue internationally-accredited certification by the International Multisensory Structured Language Education Council at the teacher and mentor levels. Certified instructors must successfully complete robust training, structured mentoring, coaching observations, a minimum number of required hours of instruction, phonics and sounds proficiency assessments, and a case study. [This document](#) offers a more detailed overview of the HillRAP program, professional development model, and research underpinning the design of the methodology.

### ABOUT THIS RESEARCH

At the request of Hill Learning Center, Empirical Education (Empirical) reviewed six existing studies on the effectiveness of HillRAP to assess and align the studies with the evidence tiers described in the Every Student Succeeds Act (ESSA) of 2015. Empirical used the standards and guidance documents from the U.S. Department of Education (2016), other prominent sources, and professional judgement to summarize and assess how the design and implementation of each study meets the expectations for ESSA Evidence Tier 3, as set forth by the ESSA guidelines. We found that four of the studies used research designs that met our criteria (see Appendix A for our criteria). All four of those studies show *promising evidence* (ESSA Evidence Tier 3). The remaining two studies provided indeterminate results because the context or design did not allow us to evaluate evidence of promise, according to the established criteria.

We summarize results for the four studies that demonstrated promising evidence in this report. The evaluation designs used in the four studies are on an upper tier of gain score evidence for two reasons. First, performance gains were calculated within-persons. This allows each student to serve as his or her own control, which limits bias that could arise if we were comparing outcomes for two different groups. Second, gains scores are calculated in terms of standard scores. An average child of the same age would have the same standard score at the beginning and end of the school year (based on a normative group). Therefore, a positive change indicates greater than expected growth compared to the norm used. This evidence is stronger than if gains were in terms of raw scores, which could simply indicate growth through maturation.

## Findings

The findings below are synthesized from four research studies on HillRAP in four different school districts, offering findings for students with a wide variety of background characteristics. All four studies show positive gains in reading achievement and qualify for ESSA Evidence Tier 3: Promising Evidence (see Appendix A for criteria). Gains on specific subscales were reported for the following subgroups of students: elementary, secondary, minority, traditionally underserved, receiving EC services, with learning disabilities, with lower incoming achievement, and with higher IQ.<sup>1</sup>

---

<sup>1</sup>The report uses the term “minorities.” For the definition of minority students, see *The Hill Reading Achievement Program in Carteret County Schools: Final evaluation results (2008-2011)* report (Walser et al., 2012).

## BRUNSWICK COUNTY, 2012

### Every Student Succeeds Act Evidence Level 3: Promising

Based on Empirical Education’s review of the evaluation of the HillRAP program in Brunswick County, the study meets ESSA Tier 3 evidence requirements and shows *promising evidence* based on statistically significant and positive gains on the Woodcock-Johnson III (WJ III) reading outcome measures. See Appendix A for the criteria Empirical Education used to evaluate the evidence.

### Study Sample

This study was conducted with elementary, middle and high school students who participated in HillRAP in the 2009-2010 and 2010-2011 school years. The study sample included 325 students in grades 4-8 in the first year sample and 164 elementary, middle, and high school students who continued in HillRAP for a second year.

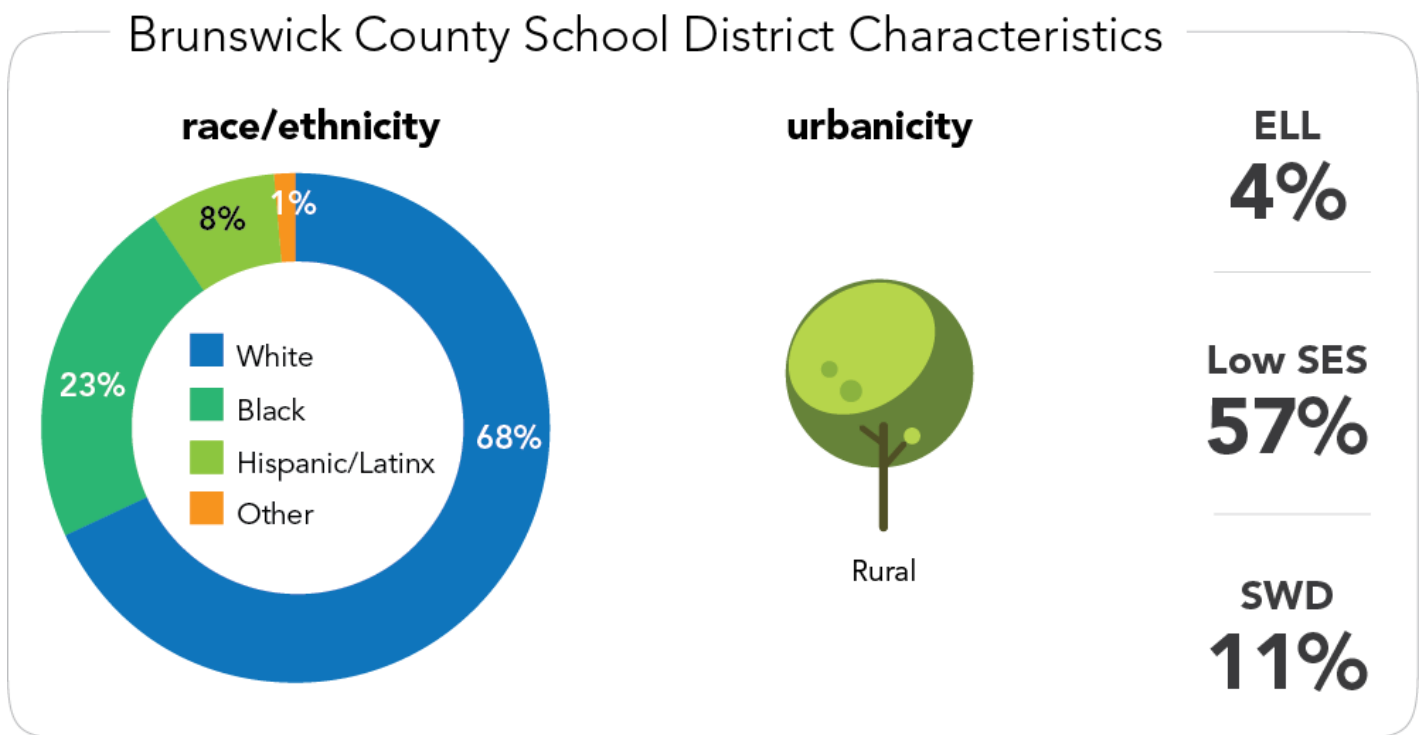


FIGURE 1. DEMOGRAPHICS OF BRUNSWICK COUNTY

Note. ELL stands for English language learners. SES stands for socio-economic status. SWD stands for students with disabilities. Data obtained from the 2009-2010 National Center for Education Statistics Common Core of Data.

### Study Design, Procedures, and Selection Criteria

Students were selected for HillRAP instruction based on criteria specified in the Hill Reading Achievement Program Student Selection Checklist. Students attended Brunswick County Schools and were enrolled in HillRAP for two years. Impacts were assessed using four reading subscales of the established (WJ III) Tests of Achievement (Letter-Word Identification, Reading Fluency, Passage Comprehension and Word Attack). Students were assessed at the pretests, posttest and one-year follow-up time points.

**Main Findings**

Use of HillRAP is associated with improvement in the reading achievement of struggling readers. HillRAP students demonstrated statistically significant and greater than expected growth on the four subscales of WJ III reading assessment that were used in the study. (See Appendix B for a description of WJ III). Specific subgroups of students who received HillRAP instruction also

*Overall Means, Standard Deviations (SD), and Change Scores for First Year Sample and Two Year Longitudinal Sample Age-Referenced Standard Scores on Woodcock Johnson Tests of Achievement III Reading Tests*

Test	First Year Sample (n = 325) <sup>a</sup>			Two Year Longitudinal Sample (n = 164) <sup>b</sup>			
	Fall 2009 Mean	Spring 2010 Mean	2009– 2010 Change	Fall 2009 Mean	Spring 2010 Mean	Spring 2011 Mean	2009– 2011 Change
Letter-Word Identification	87.39 (12.39)	89.07 (12.34)	1.67**	85.46 (12.76)	87.24 (12.68)	87.34 (13.02)	1.88**
Reading Fluency	84.74 (11.24)	87.80 (11.78)	3.06**	83.67 (11.35)	86.46 (11.81)	86.88 (11.71)	3.21**
Passage Comprehension	81.70 (11.25)	84.92 (11.64)	3.22**	79.55 (11.72)	83.15 (11.73)	83.50 (11.71)	3.95**
Word Attack	88.89 (10.24)	92.09 (9.70)	3.20**	87.17 (10.46)	90.95 (9.11)	92.48 (10.60)	5.31**

<sup>a</sup>n size for First Year Sample is 324 for Reading Fluency and Word Attack tests

\*p < .05. \*\*p < .01

FIGURE 2. RESULTS AFTER 1 YEAR AND 2 YEARS OF USAGE

Note. This is Table 5 in *The Hill Center Regional Educational Model: Evaluation Results of the Hill Reading Achievement Program in Brunswick County Schools* report (Walser et al., 2012).

showed significant growth. These subgroups included Black<sup>2</sup> and Hispanic<sup>3</sup> students and students identified to receive special education services. Conclusions from analysis of 2-year gains are limited by a sample reduction of 49.6%, compared to the first year sample, which may affect the external validity of the result.

<sup>2</sup>For Black students receiving HillRAP instruction, the study found statistically significant positive growth on three WJ III measures. However, after one year, gains on the Letter-Word Identification subscale were not statistically significant, and after two years, gains on the Letter-Word Identification and Reading Fluency subscales, were not statistically significant.

<sup>3</sup>For Hispanic students receiving HillRAP instruction, the study found statistically significant positive growth on all four WJ III measures. However, after one year, gains on the Passage Comprehension subscale were not statistically significant and after two years, gains on the Letter-Word Identification subscale were not statistically significant for Hispanic students.

CARTERET COUNTY, 2012

**Every Student Succeeds Act Evidence Level 3: Promising**

Based on Empirical Education’s review of the evaluation of the HillRAP program in Carteret County, the study meets ESSA Tier 3 evidence requirements and shows *promising evidence* based on statistically significant and positive gains on the Woodcock-Johnson III (WJ III) reading outcome measures. See Appendix A for the criteria Empirical Education used to evaluate the evidence.

**Study Sample**

This study was conducted with elementary, middle, and high school students who participated in HillRAP in the 2008-2009, 2009-2010, and 2010-2011 school years. The study sample included 194 elementary, middle, and high school students attending Carteret County schools. Most students were at the elementary level (78.35%), White (75.13%) and identified to receive special education services (77.72%).

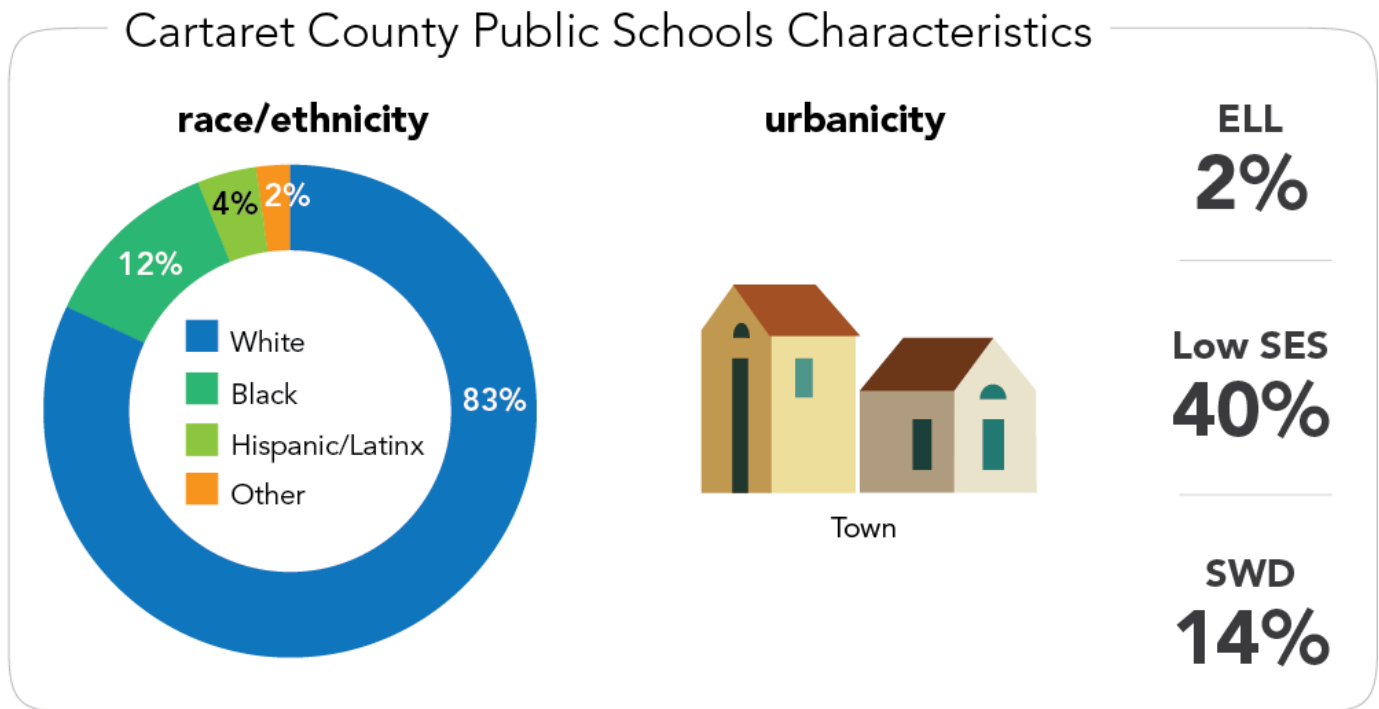


FIGURE 3. DEMOGRAPHICS OF CARTERET COUNTY

Note. ELL stands for English language learners. SES stands for socio-economic status. SWD stands for students with disabilities. Data obtained from the 2009-2010 National Center for Education Statistics Common Core of Data.

**Study Design, Procedures, and Selection Criteria**

School and district leadership identified teachers to receive HillRAP training. The HillRAP teachers selected students for HillRAP instruction based on their understanding of the program and the students it is intended to benefit. Most students were enrolled in HillRAP for one year. About a third of the students received two years of HillRAP instruction. Impacts were assessed using four reading subscales of the established WJ III Tests of Achievement (Letter-Word Identification,

Reading Fluency, Passage Comprehension, and Word Attack). Students were assessed at the pretest, posttest and one-year follow-up time points.

## Main Findings

Use of HillRAP is associated with improvement in the reading achievement of struggling readers. HillRAP students demonstrated statistically significant and greater than expected growth on the four subscales of the WJ III reading assessment that were used in the study after one year of implementation. Positive statistically significant gains were also observed for all four subscales for one of the two study cohorts after two years. (See Appendix B for a description of WJ III). Specific subgroups of students who received HillRAP instruction also showed significant growth. These subgroups included students receiving special education services and minority students.<sup>4</sup>

### *Means, Standard Deviations (SD), and Change Scores for Combined Sample Age-Referenced Standard Scores on Woodcock Johnson Tests of Achievement III Reading Tests (n = 194)*

<b>Test</b>	<b>Pretest Mean</b>	<b>Posttest Mean</b>	<b>One Year Change</b>
Letter-Word Identification	78.43 (18.66)	81.50 (18.52)	3.07**
Reading Fluency	74.78 (21.27)	80.24 (18.16)	5.46**
Passage Comprehension	71.97 (19.12)	78.17 (18.19)	6.20**
Word Attack	83.82 (17.00)	89.64 (12.04)	5.82**

\*p < .05. \*\*p < .01

FIGURE 4. CHANGE IN MEAN RESULTS AFTER 1 YEAR OF USAGE

Note. This figure is Table 9 in *The Hill Reading Achievement Program in Carteret County Schools: Final evaluation results (2008-2011)* report (Walser et al., 2012).

<sup>4</sup>For minority students receiving HillRAP instruction, the study found statistically significant positive growth on three WJ III measures. However, after one year the Reading Fluency positive growth was not statistically significant. For the definition of minority students, see *The Hill Reading Achievement Program in Carteret County Schools: Final evaluation results (2008-2011)* report (Walser et al., 2012).

DAVIE COUNTY MIDDLE SCHOOLS, 2011

**Every Student Succeeds Act Evidence Level 3: Promising**

Based on Empirical Education’s review of the evaluation of the HillRAP program in Davie County Middle Schools, the study meets ESSA Tier 3 evidence requirements and shows *promising evidence* based on statistically significant and positive gains on the Woodcock-Johnson III (WJ III) reading outcome measures. See Appendix A for the criteria Empirical Education used to evaluate the evidence.

**Study Sample**

This study was conducted with middle school students who participated in HillRAP in the 2008-2009 and/or 2009-2010 school years. The study sample included 142 middle school students in grades 6-8 in Davie County Public Schools.

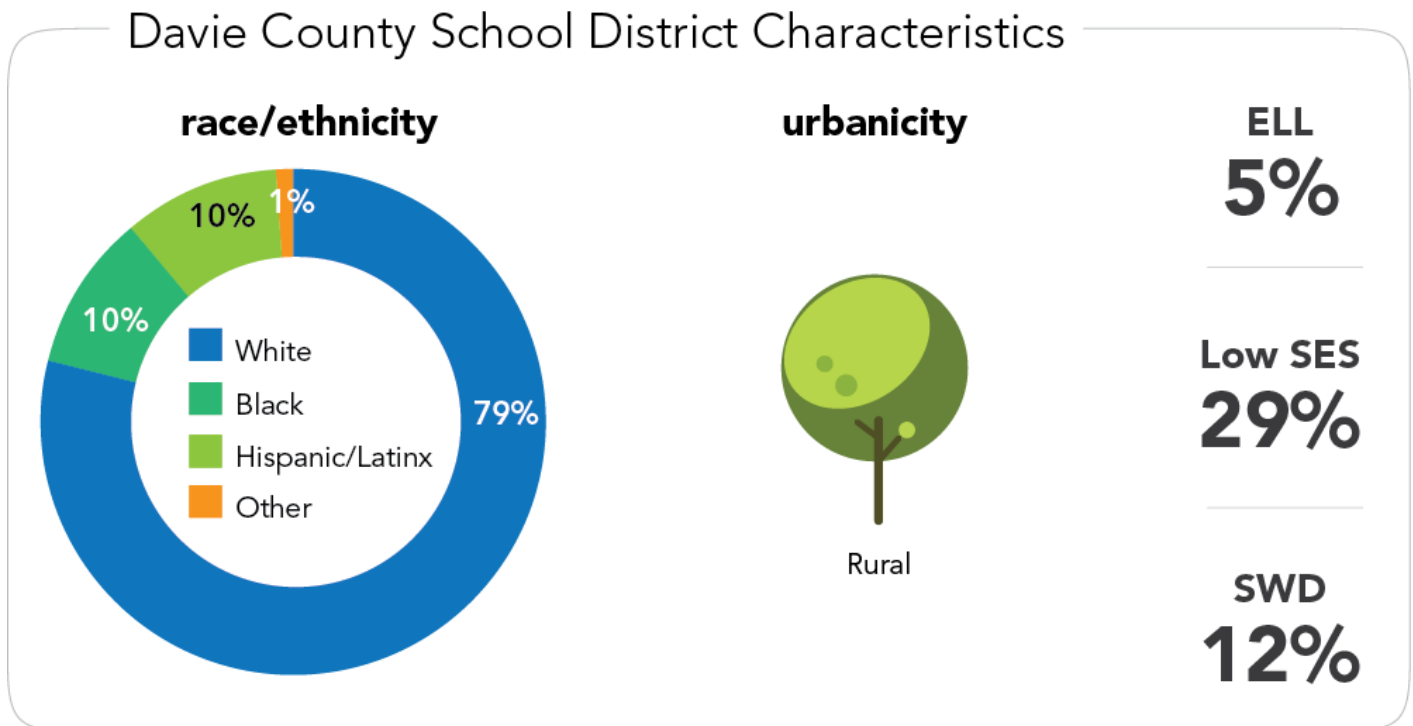


FIGURE 5. DEMOGRAPHICS OF DAVIE COUNTY

Note. ELL stands for English language learners. SES stands for socio-economic status. SWD stands for students with disabilities. Data obtained from the 2009-2010 National Center for Education Statistics Common Core of Data.

**Study Design, Procedures, and Selection Criteria**

School principals selected students for HillRAP instruction based on the student’s achievement on previous state-wide testing, special education status, risk for grade retention and/or ability to read despite having completed other reading interventions. Cohort 1 included 92 students who enrolled in HillRAP in 2008-2009; 33 of them received two years of HillRAP instruction. Cohort 2 included 50 students who were enrolled in HillRAP for one year (2009-10). Impacts were assessed within-subject and using four reading subscales of the established WJ III Tests of Achievement (Letter-Word

Identification, Reading Fluency, Passage Comprehension, and Word Attack). Students were assessed at the pretest, posttest and one-year follow-up time points.

**Main Findings**

Use of HillRAP is associated with improvement in the reading achievement of struggling readers in middle school. After one year, Cohort 1 and 2 HillRAP students demonstrated statistically significant growth on Reading Fluency subscale of the WJ III reading tests. After one year, Cohort 1 HillRAP students also demonstrated statistically significant growth on Letter Word Identification. (Passage Comprehension and Word Attack were not statistically significant.) Students who received HillRAP instruction for two years demonstrated statistically significant growth on three of four reading tests (all but Passage Comprehension). Two subgroups of students who received HillRAP instruction also showed significant growth: students with lower pre-intervention achievement and students with above-average IQs. Conclusions from analysis of 2-year gains are limited by a sample reduction of 64.1%, which may affect the external validity of the result.

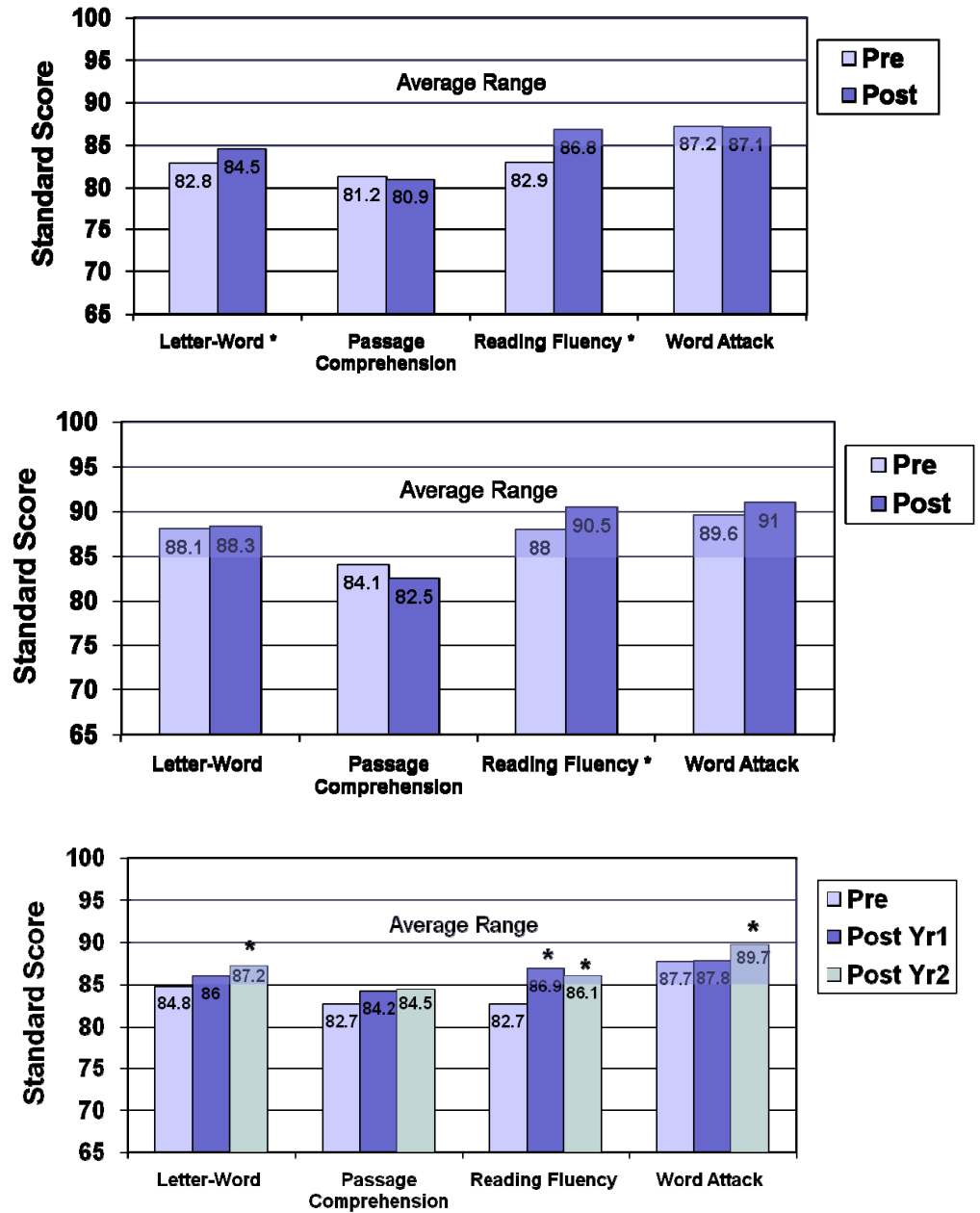


FIGURE 6. CHANGES IN STANDARD SCORES FROM PRETEST TO POSTTEST (TOP: COHORT 1, YEAR 1; MIDDLE: COHORT 2, YEAR 1; BOTTOM: COHORT 1, PRE TO POST YEAR 2)

Note. This figure is from the Evaluation of the HillRAP Intervention in Davie County Middle Schools, 2008-2010 Report (Christopoulos et al., 2011). Asterisks (\*) represents a statistically significant change from pretest ( $p < .05$ ).



DURHAM PUBLIC SCHOOLS, 2007

**Every Student Succeeds Act Evidence Level 3: Promising**

Based on Empirical Education’s review of the evaluation of the HillRAP program in Durham Public Schools, the study meets ESSA Tier 3 evidence requirements and shows *promising evidence* based on statistically significant and positive gains on the Woodcock-Johnson III (WJ III) reading outcome measures. See Appendix A for the criteria Empirical Education used to evaluate the evidence.

**Study Sample**

This study was conducted with elementary school students who participated in HillRAP in the 2003-2004, 2004-2005, and/or 2005-2006 school years. The study sample included 137 students attending nine elementary schools in Durham Public Schools across the 3-year study period. The average school enrollment was 577 students (range from 200-747 students). For students enrolled in the HillRAP program, 78.8% were identified to receive special education services, 13.9% were English language learners and 61.3% were eligible for free or reduced-price lunch, The percentage of African American students was 54.7%, Hispanic students was 15.3% and White students was 29.9%. Approximately 85% of students were in the second, third or fourth grade at the time they entered the study; just over half of these students were third graders.

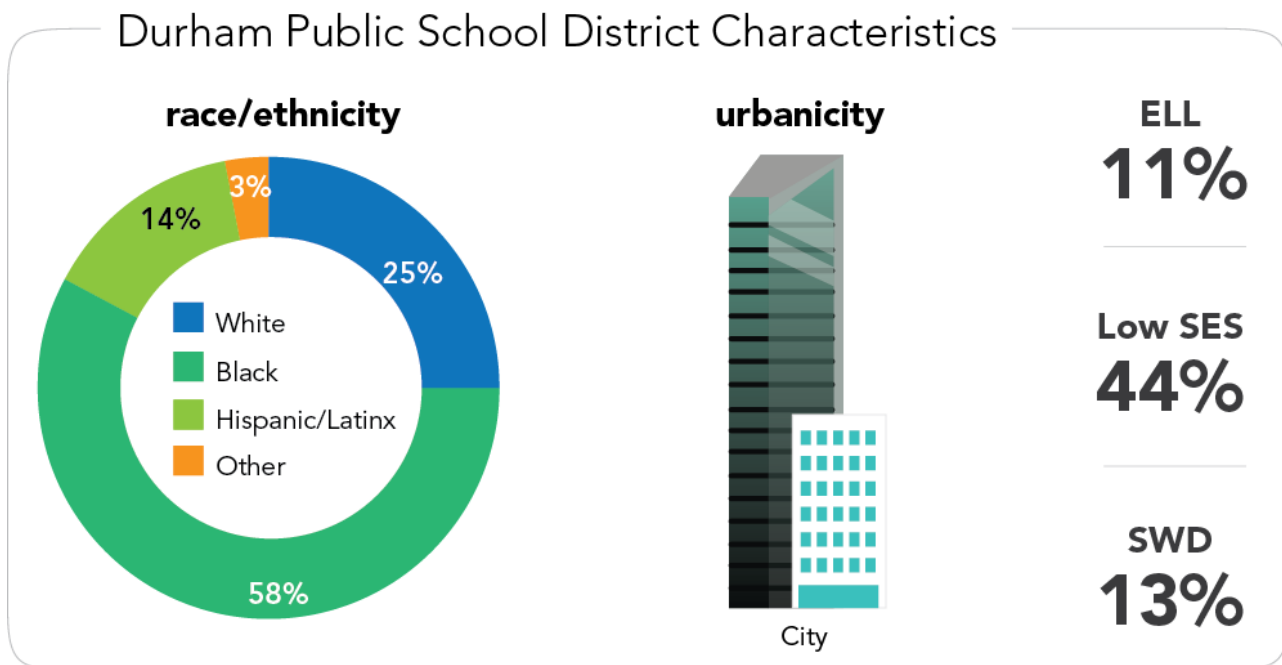


FIGURE 7. DEMOGRAPHICS OF DURHAM COUNTY

Note. ELL stands for English language learners. SES stands for socio-economic status. SWD stands for students with disabilities. Data obtained from the 2009-2010 National Center for Education Statistics Common Core of Data.

### Study Design, Procedures, and Selection Criteria

School and district leadership identified nine elementary schools to implement the program. Over the 3-year study period, 23 special education teachers received HillRAP training. The HillRAP teachers selected students with learning disabilities or students struggling with reading who might benefit from HillRAP instruction. Nearly half of the students (n = 61) were enrolled in HillRAP for one year. Another 58 students received two years of HillRAP instruction, and 18 students enrolled for 3 years. Impacts were assessed using four reading subscales of the established WJ III Tests of Achievement (Letter-Word Identification, Reading Fluency, Passage Comprehension, and Word Attack). Students were assessed at the pretest, posttest, and one-year follow-up time points.

**Table 4-1. Change in WJ-III Standard Test Scores from Time 1 to Time 2**

	Time 1 Mean	Time 2 Mean	Change in Standard Score	Number of Students	Growth per Instructional Hour
<b>W-JIII test data</b>					
Passage Comprehension	80.34	83.54	3.20***	149	<b>0.06***</b>
Reading Fluency <sup>1</sup>	80.61	82.89	2.29***	124	<b>0.04***</b>
Letter-Word Identification	81.25	81.39	0.14	149	<b>0.00</b>
Word Attack	81.36	86.04	4.67***	147	<b>0.08***</b>

Note: Results presented here do not include IQ score in the model because including IQ reduces the N by 17 percent.

<sup>1</sup>The number of students given the Reading Fluency test was likely lower than the other WJ-III tests because a higher level of reading ability is required to take this test compared with other tests.

\*\*\* p < .001.

**Table 4-2. Change in WJ-III Standard Test Scores from Time 2 to Time 3**

	Time 2 Mean	Time 3 Mean	Change in Standard Score	Number of Students	Growth per Instructional Hour
<b>WJ-III test data</b>					
Passage Comprehension	83.54	83.23	-0.30	134	<b>-.01</b>
Reading Fluency	82.89	84.97	2.07*	124	<b>.04*</b>
Letter-Word Identification	81.39	81.84	.45	134	<b>.01</b>
Word Attack	86.04	87.24	1.21	134	<b>.02</b>

Note: Results presented here do not include IQ score in the model because including IQ reduces the N by 17 percent.

\* p < .05.

FIGURE 8. STANDARD SCORE INCREASE

Note. This figure is Tables 4-1 and 4-2 from *The Hill Center Reading Achievement Program in Durham Public Schools* report (Downing et al., 2007).

### Main Findings

Use of HillRAP is associated with improvement in the reading achievement of struggling readers. HillRAP students demonstrated statistically significant and greater than expected growth after one year on three of the four<sup>5</sup> WJ III reading tests. Students who received HillRAP instruction for two years demonstrated greater growth on the Reading Fluency WJ III subtest than those who received it for one year. Specific subgroups of students who received HillRAP instruction also showed significant growth.

<sup>5</sup> After one year, the Letter-Word Identification positive growth was not statistically significant.

## Conclusion

Based on Empirical Education’s review of the evaluation of the HillRAP program conducted by independent evaluators in four studies, the program meets ESSA Tier 3 evidence requirements and shows *promising evidence* based on statistically significant and positive gains. The benefits were demonstrated most clearly after one year. Reductions in samples in the second year of several of the studies offer less conclusive results about added benefits beyond one year.

Two strengths of the evaluation designs that were used to assess gains are that (1) performance gains were calculated within-person, which allows individuals to serve as their own controls and thereby limits bias arising from individuals selecting into treatment or comparison groups, and (2) the use of standard scores as outcomes—an average child of the same age would have the same standard score at the beginning and end of the school year (based on a normative group). Therefore, a positive change indicates greater than expected growth, and the measure sets a much higher bar for demonstrating growth than would a basic measure of pre-to-post change in raw scores.

We recommend efforts to conduct conceptual replications to update the evidence base concerning the promise of the HillRAP reading intervention. Notably, since the date that the studies were conducted, the program has changed in some respects, including through the addition of technology-based components. A relevant question is whether additional achievement gains may be expected with the integration of these enhancements. We also recommend use of impact evaluation designs that would allow HillRAP to decisively demonstrate not just evidence of promise, but potentially the causal impacts of the program on student reading outcomes that satisfy higher tiers of evidence as set out by the U.S. Department of Education.

## References

- Christopoulos, C., Rosanbalm, K., & Rabiner, D. (2011). *Evaluation of the HillRAP intervention in Davie County Middle Schools, 2008-2010* (Final report). Center for Child and Family Policy, Duke University. <https://www.hillcenter.org/wp-content/uploads/2018/09/DavieCoMiddleSchoolHillRAP-Duke-2011.pdf>
- Downing, J., Williams, J., Lasater, B., & Bell, K. (2007). *The Hill Center Reading Achievement Program in Durham Public Schools: Final report*. RTI International. <https://www.hillcenter.org/wp-content/uploads/2018/09/RTIRAPFinalReport-2007.pdf>
- Ritchey, K. D., & Goeke, J. L. (2006). Orton-Gillingham and Orton-Gillingham based reading instruction: A review of the literature. *Journal of Special Education, 40*(3), 171-183.
- Schrank, F. A., McGrew, K. S., & Woodcock, R. W. (2001). *Technical Abstract* (Woodcock-Johnson III Assessment Service Bulletin No. 2). Riverside Publishing.
- U.S. Department of Education (2016). *Non-Regulatory Guidance: Using Evidence to Strengthen Education Investments*. <https://www2.ed.gov/policy/elsec/leg/essa/guidanceeuseinvestment.pdf>
- U.S. Department of Education's Office of Elementary and Secondary Education (OESE) (2020). *Selecting Evidence-Based Practices for Tiers 1, 2, and 3: Navigating Clearinghouses and Databases*. <https://oese.ed.gov/resources/oese-technical-assistance-centers/state-support-network/resources/selecting-evidence-based-practices-tiers-1-2-3-navigating-clearinghouses-databases/>
- Walser, T., Bell, C., Parker, M., & Tart, M. (2012). *The Hill Reading Achievement Program in Carteret County Schools: Final evaluation results (2008-2011)*. Watson School of Education, University of North Carolina Wilmington. <https://www.hillcenter.org/wp-content/uploads/2018/09/carteretfinal.pdf>
- Walser, T., Grace, E., Bell, C., Parker, M., & Tart, M. (2012). *The Hill Center regional educational model: Evaluation results of the Hill Reading Achievement Program in Brunswick County Schools*. Watson School of Education, University of North Carolina Wilmington. <https://www.hillcenter.org/wp-content/uploads/2018/09/robertsonfinal.pdf>
- Wendling, B. J., Schrank, F. A., & Schmitt, A. J. (2007). *Educational Interventions Related to the Woodcock-Johnson III Tests of Achievement* (Assessment Service Bulletin No. 8). Riverside Publishing.

## Appendix A. Every Student Succeeds Act Tier 3 Defined

To achieve Tier 3 evidence, a program must meet the promising evidence requirement.

In 2016, the U.S. Department of Education defined promising evidence as follows.

To be supported by promising evidence, there must be at least one well-designed and well-implemented correlational study with statistical controls for selection bias on the intervention. The Department considers a correlational study to be “well-designed and well-implemented” if it uses sampling and/or analytic methods to reduce or account for differences between the intervention group and a comparison group. Additionally, to provide promising evidence, the study should:

1) Show a statistically significant and positive (i.e., favorable) effect of the intervention on a student outcome or other relevant outcome; and

2) Not be overridden by statistically significant and negative (i.e., unfavorable) evidence on that intervention from findings in studies that meet WWC Evidence Standards with or without reservations or are the equivalent quality for making causal inferences

To form our assessment of the requirements for a study to achieve Tier 3 evidence, we note first that the definition above provides a certain latitude of interpretation concerning designs that give evidence of promise.

To establish our benchmarks for Tier 3 evidence, Empirical Education consulted with multiple sources—including the Department of Education and Evidence for ESSA—and we adopted the following criteria to evaluate the evidence.

1. Currently we consider for review only student outcomes that would be acceptable for review under WWC.
2. We base our rating on the largest sample for which results are reported (i.e., studies will often also report results disaggregated by subgroups.) If results are reported only by subgroup our conclusions will depend on the proportion of findings that show positive and statistically significant gains.
3. We provide separate ratings for each distinct measurement instrument (i.e., we do not aggregate results across different assessments unless results are provided in that form in the report, and only if we judge the aggregation to be technically sound.)
4. To evaluate pre-to-post gains, we require a “within-person” design in which pretests and posttests are obtained from exactly the same sample of students. (By allowing each person to effectively serve as his or her own control, this design controls for selection bias consistent with requirements for meeting Tier 3 evidence. The design does not however rule out effects of maturation or practice effects in the results.)
5. It must be clear in reporting that pretest and posttest scores are measured on the same scale. For example, if a pretest score is a specific weighted average of subscale scores and the posttest score is a different weighted average (e.g., if a subscale is completely removed from the posttest average) then the pre-post analysis is invalidated. We allow for assessment of gains on several metrics. For example, raw scores on the same instrument may be measured at pre and post; alternatively, age-referenced standard scores may be reported, which indicate expected growth relative to a specific norm group, with zero gains indicating on-track growth relative to the norm group. It is normally easier to show gains using raw scores than standard scores. Both are acceptable; however, we will report use of standard scores as potentially yielding stronger evidence. If both age-referenced standard scores and raw scores are reported, and if the former does not show positive gains, we will assess promise in terms of gains on the raw scores (i.e., using more lenient criteria).

6. Both pretest and posttest scores on which gains are calculated must satisfy minimum reliability standards as specified by WWC. Developer-created assessments are allowed.
7. We flag if multiple comparison adjustments are not applied in analysis and when reporting  $p$ -values; however, if they are not applied, it does not affect the rating.
8. We flag if an adjustment for clustering of observations within classes, teachers or schools is not applied; however, if the adjustment for clustering is not applied, it does not affect the rating.
9. We assume a  $p$  value of .050 or less is adequate to demonstrate statistical significance.
10. Program implementation has to be conducted in the United States.
11. Program implementation has to have occurred in previous 20 years.
12. The sample size of participants on which gains are calculated should be more than 20.
13. We reserve the option to change the above criteria and include additional criteria to evaluate if studies meet Tier 3 evidence, especially as we encounter additional study designs, analyses, and contexts.

## Appendix B. Woodcock-Johnson III Measure

The WJ III Tests of Achievement (Woodcock, McGrew, & Mather, 2007) include 22 tests for measuring skills in reading, mathematics, and writing. The WJ III provides norm-referenced measures of academic abilities. The present studies use four reading subscales of the established Woodcock-Johnson III Tests of Achievement (Letter-Word Identification, Reading Fluency, Passage Comprehension, and Word Attack). Each of these tests measures one or more narrow, or specific, psychometrically defined abilities. The Letter-Word Identification test measures reading decoding. The Reading Fluency test measures reading speed and semantic processing speed. The Passage Comprehension test measures reading comprehension and cloze ability. The Word Attack test measures reading decoding and phonetic coding. The reliabilities for each WJ III test of achievement used in this study were all greater than 0.80 and can be found in the following table (Schrank, McGrew, & Woodcock, 2001).

TABLE 1. MEDIAN TEST RELIABILITY STATISTICS FOR WJ III TESTS OF ACHIEVEMENT

Test	Median internal consistency reliabilities
Letter-Word Identification	0.94
Reading Fluency	0.90
Passage Comprehension	0.88
Word Attack	0.87

Note. The internal consistency reliabilities were calculated using the split-half procedure (corrected for length of the published test using the Spearman-Brown correction formula).