Evaluation of the HillRAP Intervention in Durham County Schools 2008 - 2010 Final Report

Submitted by
Christina Christopoulos,
Katie Rosanbalm, &
David Rabiner

Center for Child and Family Policy
Duke University

May 2011
# Table of Contents

**Executive Summary** ........................................................................................................... 4  
**Key findings** ......................................................................................................................... 4  
**Summary and conclusions** ................................................................................................. 7   
**Introduction** ............................................................................................................................ 10  
  **Purpose** ................................................................................................................................. 10  
  **Background** .......................................................................................................................... 10  
**Design and Methods** ............................................................................................................. 12  
  **Design** ................................................................................................................................... 12  
  **School profiles** ...................................................................................................................... 13  
  **Student selection** ................................................................................................................. 14  
  **Teacher training** ................................................................................................................. 14  
  **Data sources** ....................................................................................................................... 15  
    - Achievement testing ........................................................................................................... 15  
    - Hill Center Database .......................................................................................................... 16  
    - Teacher form ...................................................................................................................... 17  
    - Student form ..................................................................................................................... 17  
    - Observational data ............................................................................................................ 17  
**Results** ..................................................................................................................................... 17  
**Teacher Data** ....................................................................................................................... 17  
  **Demographics** ..................................................................................................................... 17  
  **Observations: Implementation of HillRAP** ........................................................................... 17  
  **Observations: HillWrite** ...................................................................................................... 19  
**HillRAP Full Sample Student Data** ....................................................................................... 19  
  **Demographics** ..................................................................................................................... 19  
  **WJ-III achievement test scores** .......................................................................................... 21  
    - Year 1 outcomes ................................................................................................................. 21  
    - Year 2 Outcomes ................................................................................................................ 23  
    - Demographic predictors of performance on the WJ-III .................................................. 24  
    - Implementation predictors of performance on the WJ-III ............................................. 25  
**End-of-Grade Assessment Results** ....................................................................................... 26  
  **Developmental Scale Scores** .............................................................................................. 26  
  **Achievement Levels** .......................................................................................................... 28  
  **C-Scores** ............................................................................................................................. 29  
  **Growth Scores** .................................................................................................................... 30  
  **Demographic predictors of performance on the EOG** .................................................... 32  
  **Implementation predictors of performance on the EOG** .............................................. 32  
  **K-2 Assessment** ................................................................................................................ 33  
**HillWrite Student Data** ......................................................................................................... 34  
  **Demographics** ..................................................................................................................... 34  
  **WJ-III achievement test scores** .......................................................................................... 35  
**Summary and Conclusions** ................................................................................................. 37  
**References** ............................................................................................................................ 43  
**Appendix A** ............................................................................................................................ 44  
  **Fidelity ratings from 4th observation (after 1 year of HillRAP implementation)** ............. 44  

**Evaluation of HillRAP in the Durham County Schools** .......................................................... 2
Appendix B ................................................................................................................................... 47
  Year 1 demographic predictors of performance on the WJ-III........................................... 47
    Pre-intervention WJ-III standard scores ........................................................................... 47
    Grade Level ..................................................................................................................... 48
    Race .................................................................................................................................. 49
    Sex .................................................................................................................................... 50
  Year 1 implementation predictors of performance on the WJ-III........................................ 51
    Teacher characteristics .................................................................................................... 51
    Student attendance ........................................................................................................ 52
  Year 1 implementation predictors of performance on the EOG ......................................... 53
Appendix C ................................................................................................................................... 55
  Year 2 demographic predictors of performance on the WJ-III........................................... 55
    Pre-intervention WJ-III standard scores ........................................................................... 55
    Grade Level ..................................................................................................................... 56
    Race .................................................................................................................................. 58
    Sex .................................................................................................................................... 59
    IQ Score ........................................................................................................................... 59
  Year 2 implementation predictors of performance on the WJ-III........................................ 60
    Teacher characteristics .................................................................................................... 60
    Student attendance ........................................................................................................ 61
  Year 2 demographic predictors of performance on the EOG ........................................... 61
  Year 2 implementation predictors of performance on the EOG ........................................ 62
Appendix D ................................................................................................................................... 63
  Year 2 results of performance on the WJ-III for students receiving only one year of
  intervention ............................................................................................................................. 63
Appendix E ................................................................................................................................... 64
  Hill Write Comparison Study Student Data .......................................................................... 64
    Demographics ................................................................................................................... 64
Appendix F .................................................................................................................................... 67
  HillRAP Comparison Study Student Data .............................................................................. 67
    Demographics ................................................................................................................... 67
    Comparison Study Sample Year 1 Outcomes .................................................................... 69
    Comparison Study Sample Year 2 Outcomes .................................................................... 71
Executive Summary

This report presents the final findings for a two-year evaluation of the Hill Center Reading Achievement Program (HillRAP) as implemented in the Durham Public Schools from September 2008 to June 2010. The Center for Child and Family Policy at Duke University conducted this evaluation in collaboration with the Durham Public School System and the NC GlaxoSmithKline Foundation. This study is an extension of a previous evaluation of HillRAP in the Durham Public Schools that took place from 2003 to 2006, carried out by RTI International.

Key findings

The purpose of this evaluation is to examine the effectiveness of HillRAP in helping public school students in grades 1 through 8 whose reading skills are compromised. A preliminary evaluation of the HillWrite program, designed for students with writing difficulties, is also included. Specifically, we address the following questions:

1. What are the effects of HillRAP instruction on reading achievement over time?
2. How do program outcomes vary for different subgroups of children? Specifically, do the effects of the Hill Center programs vary by race/ethnicity, sex, age, grade level, English language proficiency, free/reduced lunch eligibility, exceptional children status, exceptional children classification, IQ, or history of grade retention?
3. How well does the Hill Center teacher training prepare teachers to deliver HillRAP in a public school setting? What is the level of intervention fidelity for these teachers?
4. What are the effects of attendance and model fidelity (i.e., faithful implementation of HillRAP) on reading achievement?
5. What preliminary information can we deduce about the effectiveness of the HillWrite program?

Analysis of pre- to post-intervention assessment findings for students who received HillRAP showed that:

- The HillRAP program was successful in increasing the reading achievement of students with substantial reading difficulties, as measured using a nationally-normed achievement test (Woodcock-Johnson III; WJ-III). Students showed significant improvement on Word Attack, Reading Fluency and Letter-Word Identification subtests after Year 1. Students made gains at a faster rate than their same-age peers, thus closing the gap between themselves and the average student. By the end of Year 2, students made further gains or maintained their gains in Reading Fluency and Word Attack. Students also made significant improvements in Passage Comprehension in Year 2, suggesting that once a firm foundation in basic reading skills is established, reading comprehension skills can develop.
Small gains were found on state-mandated reading achievement tests (i.e., EOG and K-2 assessments for grades 3-8 and K-2, respectively) after Year 1 of intervention, though they were not statistically significant. Given significant initial reading delays, however, grade-level tests such as the EOG are not a good measure of change for this group of students. EOGs are meant to measure performance for on-grade-level skills. Even with considerable improvement, HillRAP students may remain inadequate at reading on-grade-level material (e.g., a child starting three years behind may improve to reading one grade-level behind and still perform poorly with on-grade-level reading). For this reason, the EOG assessment is not a strong outcomes measure for these students. In acknowledgment of this concern, many low-achieving students take an alternative test known as Extend2. This measure is not standardized, however, so change over time cannot be reliably assessed.

Added to this concern (and partly as a result of Extend2 usage), only 71 of the 152 HillRAP students were administered the K-2 or EOG assessments in both 07-08 and 08-09 (23 and 48 respectively). Completion of the same assessment in both years was necessary to examine pre- to post-intervention achievement. The small sample size limited statistical power and may have contributed to the lack of detectable progress on the EOG and K-2 assessment. The sample sizes were even smaller after the second intervention year, with only 33 of the 152 original HillRAP students getting the EOG assessment in both 07-08 and 09-10. With this smaller sample, significant improvement in expected EOG performance (EOG Growth Scores) was noted after the first year of HillRAP, perhaps because the students with the largest reading delays (those who could not be adequately assessed using the EOG) were assessed using the Extend2 assessment.

No student demographic variables predicted improvement on the EOG reading test, but gains on the WJ-III were significantly predicted by pre-intervention achievement scores, grade/age, race, and sex.

- Students with more substantial initial delays made the largest gains on all WJ-III subscales, suggesting that HillRAP is appropriately targeted towards students with significant reading challenges.

- Improvement on all WJ-III subscales was greater for younger students, emphasizing the need to start HillRAP as early as possible in a student’s academic career in order to maximize achievement gains.

- Overall, Latino students made the largest gains across subtests. Their reading delays may have been caused at least in part by language difficulties.
Improvement in English skills may have therefore accelerated their reading growth. More specifically, Latino students’ improvement approached significance over both African American and Caucasian students in Passage Comprehension and over African American students in Letter-Word identification for both intervention years. Caucasian students increased their performance on Reading Fluency at a faster rate than African Americans after one year of intervention. Differences in program effectiveness by race should be further assessed in future HillRAP evaluations.

- Even though it did not reach significance, the results on gender showed some interesting trends. Both males and females made gains in reading achievement on the WJ-III, but gains for males were larger across subscales. On one subscale, Passage Comprehension, females lost ground in Year 1, but regained the lost ground in Year 2. In contrast, males showed gains on Passage Comprehension after each year of intervention.

- The Hill Center teacher training appeared to prepare teachers to deliver HillRAP in a public school setting with a high level of fidelity and proficiency. Some revision of the fidelity instrument is recommended to allow more accurate assessment of implementation fidelity, however. When skills are not observed, the instrument does not currently provide a clear indication of whether the skills were not relevant to the specific unit taught (i.e., not applicable) or whether they were relevant but not taught (i.e., a break in fidelity).

- Student attendance significantly predicted performance on several WJ-III subscales as well as the EOG c-scores. The statistical significance of attendance as a predictor of both reading achievement assessments highlights the importance of students receiving the recommended level of intervention.

- Teacher demographic characteristics did not predict changes in reading achievement. Teacher proficiency, however, did. Teachers who received higher proficiency scores had students whose Passage Comprehension showed more improvement than students of teachers who were less proficient in Year 1. It may be that because reading comprehension is a higher level skill than basic word reading, it requires a more skilled and experienced teacher to deliver this aspect of HillRAP effectively. Moreover, teachers with lower proficiency scores may spend the bulk of their time on other HillRAP components and have less time to devote to comprehension instruction. This effect disappeared in Year 2, perhaps because teachers had additional practice and mentoring and were more effective in implementing HillRAP.
Preliminary examination of the small number of students who participated in the HillWrite program showed very promising results on the students’ writing achievement on the WJ-III, as well as some positive effects on reading achievement. Students in HillWrite significantly improved their Reading Fluency scores (even though this is not directly targeted in HillWrite) and showed large improvement on the Writing Samples subtest of the WJ-III. Further study of the HillWrite program with larger sample sizes is warranted to see if change of a similar magnitude is found.

**Summary and conclusions**

Generally, evaluation findings for HillRAP implementation in the Durham Public schools show that this program improved the reading proficiency of struggling readers. The effects of HillRAP on WJ-III reading achievement are stronger for younger students and for students with more significant initial achievement delays. As expected, students’ comprehension improved significantly after basic reading skills were well established. Latino students showed higher rates of improvement overall. These subgroup patterns should be monitored in future evaluations to determine whether modifications could enhance program effects for older or non-Latino students.

A comparison of the current study with the one conducted by RTI indicates that, despite the demographic differences of the two samples, the students in both studies made significant improvements in reading skills based on the Woodcock-Johnson Test of Achievement. In addition, both studies point to the importance of high attendance. Both studies also found significant improvements in developmental scale scores on the EOG, but given their limited validity as a measure of intervention effectiveness for students with reading delays, no conclusions can be drawn from them. The RTI study concluded that the impact of HillRAP is comparable across student groups, but the present study found some subgroup differences. We would recommend that this question be studied further with a larger sample.

The significant gains observed in WJ-III scores over the two-year period, with basic skill enhancement in year one building to improvement in reading comprehension in the second year, are encouraging indications that HillRAP is improving the reading skills for students who have significant delays in reading. However, without a comparison group it is impossible to know what the trajectories of these students would have been without intervention. Improvements on the EOG/K-2 assessment were not statistically significant, but these measures are unlikely to be sensitive to achievement gains for those students significantly delayed in reading. In order to draw strong conclusions about the effectiveness of the HillRAP program in Durham County, a more rigorous study including a comparison group is necessary.
Comments and future recommendations

The Durham County Public Schools selected the lowest performing schools on the 06-07 EOGs to receive the intervention for the current study. Teachers and principals within these schools selected students who they felt would most benefit from intensive reading remediation, based on classroom performance, EOG and K-2 assessment data, and EC status. Because random assignment at the school and/or individual level was not possible, a matched comparison design was adopted for the selection of a comparison group, with the goal of using this group to chart expected trajectories of reading achievement in a similar population without HillRAP intervention. This design called for reading achievement scores to be tracked over time for students who received the HillRAP intervention and for students with similar reading difficulties who received the usual reading interventions offered in the Durham Public Schools. The two groups were matched on demographic (e.g., race/ethnicity, age, sex, SES) and school variables (e.g., 07-08 EOG scores or 07-08 K-2 assessment 4th quartile reading level, grade, EC status), and the testing of both groups with the Woodcock-Johnson Achievement Test III was to be done before the intervention started as well at the end of the first and second years of intervention. Unfortunately, it was not possible to obtain consent forms for the comparison students within a timeframe that would allow them to be tested before the intervention started. Consent forms were eventually collected for the comparison students and they were tested at the end of the first and second years, but test results suggest that the two groups were not equivalent at baseline despite our best efforts to match them. This is likely because HillRAP students were selected based on teacher and administrator recommendations, whereas comparison students were selected based solely on administrative data. Teachers included qualitative classroom performance in making their selections, identifying students with obvious and ongoing reading problems. Matching based on demographics and test scores cannot account for qualitative group differences that teachers may recognize in underlying reading struggles. The results of the comparison group analyses are reported in Appendix F, but should be interpreted with extreme caution. Pre-post findings for the HillRAP participants in the current study are more reliable indicators of the changes resulting from HillRAP participation, but without a valid comparison group we do not know how these changes compare with the effects of services as usual.

Given the promising but inconclusive current results, we recommend the following:

1. Conduct a follow-up study using random assignment and a larger sample so we can draw definitive conclusions about the effectiveness of the program.

Random assignment is of particular importance in future evaluations of HillRAP in order to ensure clear results. Through random assignment of students to HillRAP or a comparison group, evaluators could create two groups of students who are similar on every aspect other than intervention type. As a result, if HillRAP students’ reading achievement improved...
significantly over that of the comparison group, evaluators could definitively conclude that their improvement was attributable to HillRAP and nothing else. Random assignment is the only way to ensure group equivalence and therefore unbiased findings. Given the promising early results of HillRAP, rigorous evaluation is warranted to further examine program effectiveness.

2. Do not reach conclusions about the effectiveness of the HillRAP program based on EOG scores.

The lack of statistically significant improvements in EOG scores should not be considered a measure of the HillRAP effectiveness. The End-of-Grade test is a criterion-referenced test based on the North Carolina Standard Course of Study for the curriculum at each grade level. As a result the items on the test are focused on skills and content specific to each grade. Furthermore, the text of the EOG is written at a difficulty level that a majority of students at that grade level could read and comprehend. The children who participated in HillRAP were selected because of their delays in reading, and many were performing two or more years below their actual grade level. Consequently, it does not make sense to expect them to perform acceptably on material written at their grade level even after a year of intensive intervention. The lack of statistically significant improvements on the EOG after one or even two years of intervention is not surprising, though EOG scores are moving marginally closer to those of their grade peers. An additional concern is that only a small number of students took the EOG test on consecutive years. Because the content of EOG test is grade-specific and overly challenging for students with significant delays, many instead take the Extend2, an alternative nonstandardized assessment. Such dramatic loss of data could have certainly contributed to the current results, but also speaks to the inappropriateness of the EOG to adequately measure achievement for these students.
Evaluation of the HillRAP Intervention in Durham County Schools: Final Report

Introduction

This report presents the final findings of a two-year evaluation of the Hill Center Reading Achievement Program (HillRAP) as implemented in the Durham Public Schools from September 2008 to June 2010. Some pilot data on the Hill Center Writing Achievement Program (HillWrite) are also presented. The Center for Child and Family Policy at Duke University conducted this evaluation in collaboration with the Durham Public School System and the NC GlaxoSmithKline Foundation.

Purpose

The purpose of this evaluation is to examine the effectiveness of HillRAP in helping public school students whose reading skills are compromised. A preliminary evaluation of the HillWrite program, designed for students with writing difficulties, is also included. Specifically, we address the following questions:

1. What are the effects of HillRAP instruction on reading achievement over time?
2. How do program outcomes vary for different subgroups of children? Specifically, do the effects of the Hill Center programs vary by race/ethnicity, sex, age, grade level, English language proficiency, free/reduced lunch eligibility, exceptional children status, exceptional children classification, IQ, or history of grade retention?
3. How well does the Hill Center teacher training prepare teachers to deliver HillRAP in a public school setting? What is the level of intervention fidelity for these teachers?
4. What are the effects of attendance and model fidelity (i.e., faithful implementation of HillRAP) on reading achievement?
5. What preliminary information can we deduce about the effectiveness of the HillWrite program?

Background

The Hill Learning Development Center was established in 1977 to provide an intensive remediation program for students with specific learning disabilities and/or attention deficit disorder. To date The Hill Center has served over 8,000 students, trained more than 10,000 teachers, and successfully partnered with hundreds of public and private schools on a national and international basis. Since its inception, The Hill Center has been committed to research to provide evidence on the effectiveness of its programs.
The “Hill Methodology” is a comprehensive system of instruction that has been developed and refined over many years in the model school. The teaching model is based on the Orton-Gillingham approach and has evolved as a result of the experience of the Center’s certified learning disabilities specialists. In order to make the program more widely available to students in public schools, in 2003 a private foundation provided funds to develop the Hill Reading Achievement Program (HillRAP). This adapted version of the Hill Methodology in reading is more prescriptive, easier for teachers to implement, and more cost-effective to deliver than traditional Hill Methodology delivered at The Hill Center.

HillRAP was first implemented in the Durham Public Schools in a pre-post design with encouraging results through funds provided by NC GlaxoSmithKline Foundation and evaluated by Research Triangle International (RTI). Eight elementary schools actively used HillRAP during the study period, which lasted from 2003 to 2006. The sample, which ranged across all elementary school grades, was majority African American (55%), with white (30%) and Hispanic (15%) minorities. Sixty-one percent of students were eligible for free/reduced lunch, 14% spoke English as a second language, 79% were identified as Exceptional Children and 52% had IQs below 85. Woodcock-Johnson Achievement Test III (WJ-III) and the North Carolina End-Of-Grade test (EOG) were the achievement tests used. The study found significant improvements in three of the four Woodcock-Johnson subtests (Reading Comprehension, Reading Fluency and Word Attack) after one year of HillRAP and a further significant improvement on Reading Fluency in the second year. Moreover, End-Of-Grade Reading scores (developmental scale score and achievement levels) improved after one and two years of intervention (though again, these EOG scores are not standardized). Students received an average of 56 hours of instruction in the first year and 60 in the second. Number of instructional hours was a significant predictor achievement of three of the four Woodcock-Johnson subtests used. Finally, teachers’ level of fidelity to the HillRAP model did not correlate with student achievement scores, possibly because the majority of teachers were proficient or very proficient in implementing HillRAP.

In 2008 NC GlaxoSmithKline Foundation provided the Hill Center with further funds to continue the evaluation of the HillRAP program in the Durham Public Schools. The current evaluation assesses the results of the second wave of evaluation and also includes a pilot study of 11 students who received HillWrite.

The Hill Center Reading Achievement Program (HillRAP) includes the five essential components of a successful reading program as put forth in the National Reading Panel Report of 2000 (National Reading Panel, 2000). Phonological awareness, phonics, fluency, vocabulary, and comprehension are the core of the daily instructional program that is intended to be implemented in 45- to 60-minute sessions five days a week. The core components of the HillWrite daily instructional program are spelling, handwriting, copying/dictation, and the writing process.

Evaluation of HillRAP in the Durham County Schools
For both programs students work in small groups of four, and each student has an individualized curriculum to provide instruction in areas where there are demonstrated skill deficits in reading or writing. Small units of information are presented sequentially and practiced daily until a set criterion is met for three to five consecutive days and overlearning is achieved. Mastered skills are reviewed weekly to ensure retention. Classes are designed to maximize opportunities for oral and written student responses and success experiences. Student responses are graphed and charted daily by the teachers and students in order to document mastery before advancing to a higher-level skill. Student-teacher interaction focuses on praise and positive reinforcement for correct answers or approximations of the correct response.

**Design and Methods**

**Design**

The most scientific approach to testing the effectiveness of an intervention is to create two equivalent groups by randomly assigning schools or students to the intervention or comparison groups. Any changes in the intervention group can then be attributed directly to the intervention and not to other factors, such as higher motivation of the intervention schools/students, etc. Understandably, the Durham County Public Schools selected the lowest performing schools on the 06-07 EOGs to receive the intervention, which meant that the schools/students were not randomly assigned. Teachers and principals within these schools selected students who they felt would most benefit from intensive reading remediation, based on classroom performance, EOG and K-2 assessment data, and EC status. As a result, the evaluation had to find a different way to create two groups that were as equivalent as possible. A matched comparison design was the next best choice. A group of comparison students was selected from non-intervention schools who were matched to the students receiving the Hill Center programs on certain demographic and school variables (e.g., 07-08 EOG scores or 07-08 K-2 assessment 4th quartile reading level, grade). This design attempts to equate HillRAP and comparison groups using available data, but does not ensure the equivalence of the two groups. The study called for testing of both intervention and comparison students with the WJ-III before the intervention started (baseline) as well as at the end of the first and second years of the intervention. By having all students tested before and after the intervention and having matched the two groups as closely as possible, the evaluation could determine whether any achievement changes found after the intervention were more likely to be the result of HillRAP and not of normal development in reading. Upon selection of both the intervention and comparison students, the school district obtained parental consent for participation in the evaluation. Unfortunately, it was not possible to obtain consent forms for the comparison students within a timeframe that would allow them to be tested before the intervention started. As a result the comparison students were not administered the WJ-III achievement test at the beginning of the year and changes in the comparison students’ achievement scores over the first year could not be tested. However, consent forms were
obtained later in the year and W-J III testing was completed for both groups at the end of the first and second years.

In Year 2, changes in the WJ-III were examined for those students with a second year of HillRAP relative to the matched comparison students. Unfortunately, the numbers of students with End-Of-Grade scores or K-2 assessment scores across the years of the study were quite small. Given the lack of random assignment and baseline data on WJ-III reading performance, evident differences in reading performance at the end of Year 1 (first comparable time point), and the possible presence of additional qualitative differences in the two groups, results from this comparison with usable data were unreliable and inconclusive. These data were therefore not included in the body of this report or in our conclusions. For completeness, however, full findings are presented in Appendix F. The main body of the present report includes the results of the full HillRAP sample before and after one and two years of intervention, as well as pilot results for HillWrite.

**School profiles**

Thirteen elementary and three middle schools were targeted to receive the intervention. The table below provides descriptive information about these schools.

<table>
<thead>
<tr>
<th>Table 1. School Demographics</th>
</tr>
</thead>
<tbody>
<tr>
<td>HillRAP schools</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>Average school size 08-09</td>
</tr>
<tr>
<td>(number of students)</td>
</tr>
<tr>
<td>Percent race/ethnicity 08-09</td>
</tr>
<tr>
<td>White</td>
</tr>
<tr>
<td>Black/African American</td>
</tr>
<tr>
<td>Hispanic/Latino</td>
</tr>
<tr>
<td>Other</td>
</tr>
<tr>
<td>Percent eligible for free/reduced lunch 08-09</td>
</tr>
<tr>
<td>Average percent who passed the reading EOG in 07-08 (the year before program implementation)</td>
</tr>
</tbody>
</table>
**Student selection**

Principals and teachers in the intervention schools selected 201 students for participation in HillRAP and HillWrite programs based on the following criteria: previously failed EOGs, risk for failing EOGs, risk for grade retention, LD identification, EC identification with an IQ < 85, and/or an inability to read despite having completed other reading interventions. Students with one or more of these criteria were eligible for intervention. Once eligible students were identified, school personnel contacted families to inform them about the HillRAP/HillWrite program and evaluation and to obtain informed consent for participation. Consent was obtained for 176 students. Table 2 presents the distribution of inclusion reasons for consented students (students may have been selected for multiple reasons). It is important to note that the schools in this evaluation are very low-performing as evidenced by the low average 07-08 EOG scores shown in Table 1. Forty-seven percent of selected students were identified LD, 22% had an IQ below 85 and 2% were identified as autistic. These percentages highlight the severity of challenges the selected students were facing and provide a framework within which to consider the results of the study.

**Table 2. Selection criteria for consented students**

<table>
<thead>
<tr>
<th>Criteria</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Previously failed End-of-Grade Test</td>
<td>38</td>
</tr>
<tr>
<td>At risk for failing End-of-Grade Test</td>
<td>59</td>
</tr>
<tr>
<td>At risk for failing grade</td>
<td>43</td>
</tr>
<tr>
<td>LD identified</td>
<td>47</td>
</tr>
<tr>
<td>EC identified with IQ &lt;85</td>
<td>22</td>
</tr>
<tr>
<td>Failing other reading interventions</td>
<td>26</td>
</tr>
<tr>
<td>Autism classification</td>
<td>2</td>
</tr>
</tbody>
</table>

**Teacher training**

Fifteen Exceptional Children’s teachers from the Durham Public Schools participated in the training. All teachers received both the HillRAP and HillWrite training. Teachers were expected to receive 45 hours of course work, which consisted of a 5-day workshop and two half-day follow-up workshops. The 5-day workshop included a 1-day training in Phonics Breaking the Code, a 2-day HillRAP workshop, and a 2-day HillWrite workshop. Teachers had the option to participate in the 5-day training in June or July of 2008. During training, teachers focused on the fundamentals of phonological awareness, phonics, spelling, paragraph-level writing, grammar, and handwriting as well as intensive instruction on the Hill Center’s precision teaching approach. In addition they were trained in the Hill Center’s Reading and Writing Assessment instruments, which help create a profile of strengths and difficulties and allow the teacher to start instruction at the most appropriate level for each student.
During school year 08-09, each teacher received at least four classroom observations from a Hill Center Master Mentor, an instructor certified to train HillRAP teachers (Master Mentors are required to have at least 2 years and 900 hours of experience teaching with the HillRAP method, 90 hours of coursework, and a supervised practicum in training HillRAP teachers). A fifth observation was completed in school year 09-10 to make sure fidelity was maintained over the summer break. At each visit the mentor scored the quality and fidelity with which the teacher adhered to specified aspects of the Hill Center Methodology. At the completion of each observation, the mentor gave the teacher an overall proficiency rating from 1 (=not proficient) to 3 (=proficient) and discussed the session with the teacher, providing positive feedback as well as suggestions about what could be added or done differently. Training also included a Phonics Test and a Speech Sound Test that teachers are required to pass with 80% accuracy to ensure consistency in teaching the English language across all teachers.

During 08-09, teachers participated in two half-day workshops where the focus was on discussing and expanding upon methodology, fine tuning teacher skills and expanding on the vocabulary and reading comprehension components.

Data sources

Achievement testing

All consented students were tested with the Woodcock-Johnson III NU Tests of Achievement (WJ-III) before (i.e., September) and after (i.e., April/May) they received HillRAP in Year 1 and after they received HillRAP in Year 2. The testing was coordinated by the evaluators, the school district, and the teachers and was carried out by graduate students in clinical psychology experienced in the administration of the instrument. The WJ-III is a nationally normed, widely used achievement test. Four subtests were used to measure the students’ achievement in reading:

- Letter-Word Identification measures the ability to identify letters and words.
- Word Attack measures the ability to pronounce nonsense words based on previous knowledge of letter sounds.
- Reading Fluency measures the ability to quickly read simple sentences silently and respond to True/False questions about them.
- Passage Comprehension measures the ability to understand what you read by filling in the appropriate blank word in a sentence.

Two additional subtests were used for students in HillWrite:

- Writing Fluency measures the ability to formulate and write simple sentences quickly.
- Writing Samples measures the ability to write higher quality sentences with increasing passage length, level of vocabulary, grammatical complexity and level of concept abstraction.
Grade-based computerized scoring was used for the WJ-III, which yields raw scores, standard scores and percentile rank scores.

The school district provided the evaluators with the End-of-Grade test scores from years 06-07, 07-08, 08-09 and 09-10 for all of the students participating in the HillRAP evaluation. The North Carolina End-of-Grade test (EOG) is the achievement test of choice of the North Carolina Department of Public Instruction (NCDPI Division of Accountability Services, 2007). The two components of the test—Reading Comprehension and Mathematics—are aligned to the North Carolina Standard Course of Study for the curriculum at each grade level. They are multiple-choice tests administered in grades 3 through 8 in the last three weeks of the school year. Developmental scale scores, standardized scores (i.e., C-scores) and growth scores are reported for the EOGs. Achievement levels are calculated based on the developmental scale scores. An achievement level of III or IV is considered on grade level or above. An alternative test, the North Carolina Extend2 End-of-Grade test, is administered to students with current Individual Education Plans (IEPs) who do not have a significant cognitive disability and whose instruction content follows the grade-level North Carolina Standard Course of Study, but whose achievement is measured against modified academic achievement standards. The Extend2 is not standardized in a way that allows comparison of scores across years, so this test was not used in the current evaluation; students tested using the Extend2 could only be included in the WJ-III portion of the evaluation. This practice results in a drastic reduction of the number of students with EOG scores that can be used for analyses (as seen in the End-of-Grade Assessment Results section below), and it considerably affects the confidence with which conclusions can be made.

For students in grades K through 2, the North Carolina Department of Public Instruction requires that students are tested through the NC K-2 Literacy Assessment to assess their reading and writing achievement. For the purposes of the study, the fourth quarter text reading level was used as the indicator of reading achievement. (This only applied to HillRAP students as all HillWrite students were in third grade or above.)

Because the K-2 assessment data are unnormed, they cannot be compared with EOG data in longitudinal analyses. Only students who completed the same assessment both pre- and post-intervention (i.e., K-2 pre and post or EOG pre and post) could be included in the analysis of change over time.

**Hill Center Database**

To accompany the HillRAP intervention, The Hill Center developed a web-based database to facilitate data collection, monitoring, and skill tracking of student progress. This database is used by the teachers to list currently enrolled students, their daily progress, and their attendance in HillRAP sessions. For the purposes of this report, the Hill Center Database was used solely to extract attendance information. HillWrite students were not entered in the database.
Teacher form

Each teacher was asked to provide the following demographic information: sex, race/ethnicity, EC teacher status, highest level of education and number of years teaching.

Student form

After parental consents were collected, teachers of both intervention and comparison students provided the following information on each participating student: reason for selection (intervention students only), previous achievement and IQ testing dates and scores, student status (i.e., Limited English Proficiency [LEP], English as a Second Language [ESL], Exceptional Children’s status [EC], etc.), and if EC, type of exceptionality (entered as dichotomous yes/no variables for each exceptionality category).

Observational data

Master Mentors conducted classroom observations four times for each teacher during the 08-09 school year, and one more observation was conducted in 09-10. Mentors rated HillRAP or HillWrite implementation quality and fidelity using a HillRAP or HillWrite observation form with a scale of 1 (not proficient) to 3 (proficient) or 0 (not observed). The HillRAP form includes ratings for specific activities and teaching behaviors in the following HillRAP components: drill activities, phonological awareness, word attack, fluency, vocabulary, and comprehension. Teachers were also rated on classroom management. A total of 60 sub-components in these areas were included. The HillWrite form includes ratings for the following components: drill activities, oral/written spelling, copying, dictation, handwriting, grammar/mechanics, composition, and classroom management. A total of 56 sub-components were included.

Results

Teacher Data

Demographics

A total of 15 teachers across the 16 Durham County elementary and middle schools implemented the HillRAP intervention. These teachers are predominantly female (93.3%). Two-thirds are Caucasian, 26.7% African American, and 6.7% multi-racial. Just under half of the teachers (46.7%) have a Master’s degree, and the remainder have a Bachelor’s degree. They have an average of 12.6 years teaching experience (range = 1.5 to 30 years) and are all EC teachers.

Observations: Implementation of HillRAP

Based on observation data recorded by a Hill Center Master Mentor, teachers improved their implementation of HillRAP slightly between the first observation in October and the fourth
observation in April, but were highly proficient at both observations. A summary of proficiency scores in each of the core components is provided in Table 2. Proficiency was rated 1 (not proficient), 2 (somewhat proficient), or 3 (proficient).

### Table 3. Teacher Proficiency in HillRAP Implementation

<table>
<thead>
<tr>
<th>HillRAP Component</th>
<th>1&lt;sup&gt;st&lt;/sup&gt; Observation</th>
<th>4&lt;sup&gt;th&lt;/sup&gt; Observation</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>% with each rating</td>
<td>% with each rating</td>
</tr>
<tr>
<td>Drill Activities</td>
<td>3 = 100%</td>
<td>3 = 100%</td>
</tr>
<tr>
<td>Phonological Awareness</td>
<td>2 = 7.1%</td>
<td>3 = 100%</td>
</tr>
<tr>
<td></td>
<td>3 = 92.9%</td>
<td></td>
</tr>
<tr>
<td>Word Attack</td>
<td>2 = 15.4%</td>
<td>3 = 100%</td>
</tr>
<tr>
<td></td>
<td>3 = 84.6%</td>
<td></td>
</tr>
<tr>
<td>Fluency</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Timed Reading from a List</td>
<td>3 = 100%</td>
<td>3 = 100%</td>
</tr>
<tr>
<td>Timed Reading from a Text</td>
<td>1 = 12.5%</td>
<td>2 = 11.1%</td>
</tr>
<tr>
<td></td>
<td>3 = 87.5%</td>
<td>3 = 88.9%</td>
</tr>
<tr>
<td>Vocabulary</td>
<td>2 = 10.0%</td>
<td>3 = 100%</td>
</tr>
<tr>
<td></td>
<td>3 = 90.0%</td>
<td></td>
</tr>
<tr>
<td>Comprehension</td>
<td>3 = 100%</td>
<td>2 = 7.1%</td>
</tr>
<tr>
<td></td>
<td></td>
<td>3 = 92.9%</td>
</tr>
<tr>
<td>Classroom Management</td>
<td>2 = 15.4%</td>
<td>3 = 100%</td>
</tr>
<tr>
<td></td>
<td>3 = 84.6%</td>
<td></td>
</tr>
<tr>
<td><strong>Overall Proficiency</strong></td>
<td>2 = 15.4%</td>
<td>3 = 100%</td>
</tr>
<tr>
<td></td>
<td>3 = 84.6%</td>
<td></td>
</tr>
</tbody>
</table>

In addition to the overall component ratings represented in Table 3, the Master Mentors also rated a number of specific teaching skills within each component area. Ratings indicated a very high level of overall proficiency. A few specific skills in Reading Comprehension and Vocabulary were rated as not observed in a large proportion of the observations, but based on the current fidelity measure it is impossible to ascertain whether they were not observed because they were not applicable to the given lesson or student, or whether their omission was truly a
A break in fidelity (and as such, requires remediation). A revision of the fidelity measure is recommended to better assess program delivery in these areas. The new measure should focus on the critical components of the HillRAP intervention that are necessary for the program to achieve intended effects, and should also include a rating option for “Not Applicable”. A detailed list of all the skills within each area and the proportion of time they were observed during the 4th observation is included in Appendix A.

**Observations: HillWrite**

Across all four observations, all components and sub-components on the HillWrite observation form were given ratings of either 3 or not observed. This indicates one or both of the following: (1) teachers were strongly proficient in HillWrite implementation from very early in the school year, and/or (2) the fidelity measure would benefit from revision to more clearly delineate levels of skill proficiency in each teaching area (so that variability and improvement over time can be documented). As with the HillRAP observation measure, it is impossible to ascertain whether a given skill was not observed because it was not applicable, or whether it was truly a break in fidelity (and therefore requires remediation).

**HillRAP Full Sample Student Data**

**Demographics**

Of the 176 consented students, 152 HillRAP students attended at least 20 days of intervention and were tested with the WJ-III both before and after the HillRAP intervention in Year 1. One hundred and eleven of these students attended at least 20 days of HillRAP in Year 2 and were again tested with the WJ-III at the end of the school year. Twenty-nine students completed Year 1 of HillRAP and remained in the Durham Public School system in Year 2, but did not receive the intervention in Year 2 (most moved to non-intervention schools). Table 4 describes the demographic characteristics of the full HillRAP sample and the sample who received intervention only for one year.

At the end of Year 1, HillRAP students had an average age of 10.6 (SD = 1.9, range = 7.6 to 14.5) and had attended an average of 77.7 hours of HillRAP (SD = 13.9, range = 16.5 to 109.5 hours). The subset for whom IQ scores were available (n = 105) had an average IQ of 81.7 (SD = 13.0, range = 44 to 113), which is in the low average range (an IQ score between 85 and 115 is considered average).

At the end of year 2, students with a second year of intervention averaged 11.3 years of age (SD = 1.9, range = 8.6 to 15.5). During Year 2, students attended an average of 77.5 hours of HillRAP (SD = 20.6, range = 35 to 117 hours). The 72 students for whom IQ scores were available had an average IQ of 83.4 (SD = 12.2, range = 59 to 113).
Students in this study were generally more disadvantaged than those in the previous Durham HillRAP evaluation conducted by RTI. Students in the RTI study were all in elementary school, and the sample included a much higher percentage of Caucasian students (30% as compared to 6.6% in the present study) and a much lower percentage of Hispanic students (15% vs. 29.6%). In addition, the RTI sample included students of higher socioeconomic means and fewer children of Exceptional status.

Table 4. Demographic Characteristics of HillRAP Study Sample

<table>
<thead>
<tr>
<th></th>
<th>Yr 1 Full HillRAP sample (n=152)</th>
<th>Yr 2 Full HillRAP sample (n=111)</th>
<th>Intervention Students w/ no Yr 2 HillRAP (n=29)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age</td>
<td>Mean (SD)</td>
<td>Mean (SD)</td>
<td>Mean (SD)</td>
</tr>
<tr>
<td></td>
<td>10.6 (1.9)</td>
<td>11.3 (1.9)</td>
<td>12.4 (2.0)</td>
</tr>
<tr>
<td>Attendance (hours)</td>
<td>77.7 (13.9)</td>
<td>77.5 (20.6)</td>
<td>--</td>
</tr>
<tr>
<td>Full Scale IQ</td>
<td>81.7 (13.0)</td>
<td>83.4 (12.2)</td>
<td>77.9 (14.3)</td>
</tr>
<tr>
<td>Sex</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>63.8%</td>
<td>62.2%</td>
<td>65.5%</td>
</tr>
<tr>
<td>Female</td>
<td>36.2%</td>
<td>37.8%</td>
<td>34.5%</td>
</tr>
<tr>
<td>Race</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>African American</td>
<td>59.2%</td>
<td>59.5%</td>
<td>62.1%</td>
</tr>
<tr>
<td>Latino</td>
<td>29.6%</td>
<td>29.7%</td>
<td>27.6%</td>
</tr>
<tr>
<td>Caucasian</td>
<td>6.6%</td>
<td>5.4%</td>
<td>10.3%</td>
</tr>
<tr>
<td>Multiracial</td>
<td>4.0%</td>
<td>4.5%</td>
<td>0.0%</td>
</tr>
<tr>
<td>Asian</td>
<td>0.7%</td>
<td>0.9%</td>
<td>0.0%</td>
</tr>
<tr>
<td>Grade</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1st grade</td>
<td>0.7%</td>
<td>0.0%</td>
<td>0.0%</td>
</tr>
<tr>
<td>2nd grade</td>
<td>20.4%</td>
<td>0.9%</td>
<td>0.0%</td>
</tr>
<tr>
<td>3rd grade</td>
<td>26.3%</td>
<td>20.9%</td>
<td>20.7%</td>
</tr>
<tr>
<td>4th grade</td>
<td>17.1%</td>
<td>31.8%</td>
<td>10.3%</td>
</tr>
<tr>
<td>5th grade</td>
<td>14.5%</td>
<td>20.0%</td>
<td>3.5%</td>
</tr>
<tr>
<td>6th grade</td>
<td>13.2%</td>
<td>5.5%</td>
<td>34.5%</td>
</tr>
<tr>
<td>7th grade</td>
<td>7.9%</td>
<td>11.8%</td>
<td>20.7%</td>
</tr>
<tr>
<td>8th grade</td>
<td>0.0%</td>
<td>9.1%</td>
<td>10.3%</td>
</tr>
</tbody>
</table>
The WJ-III was used to measure changes in reading achievement from pre- to post-intervention. The grade-based WJ-III computerized scoring program yields raw and standard scores. A standard score is a transformation of the raw score based on national norms for students in a given grade. By using standard scores, we can draw conclusions about individual performance as compared to the performance of the group. Students who have the same standard score before and after receiving HillRAP (i.e., post standard score – pre standard score = 0) remain at the same level relative to their peers. In other words, they have learned at the average rate for their grade level. On the other hand, students whose standard scores increase after they receive a year of HillRAP (i.e., post standard score – pre standard score > 0) have learned at a rate faster than would be expected for same-grade students with comparable baseline achievement. As a result, they are closing the gap between themselves and the average student. The average standard score for the WJ-III is 100, with a standard deviation of 15; scores between 85 and 115 are considered to be in the average range (Mather and Woodcock, 2001).

One hundred fifty-two students attended at least 20 sessions of HillRAP in Year 1 and completed the WJ-III reading subtests at both the beginning (September) and end (April/May) of
the 08-09 school year. At the beginning of the school year, prior to the start of the HillRAP intervention, these students performed an average of 1.8 to 2.3 years below grade level on each of the reading subtests (elementary school students were an average of 1.4 to 1.9 years behind, middle school students were an average of 3.2 to 3.9 years behind, suggesting that middle school students would need longer remediation to reach the level of their grade mates).

Figure 1 shows the mean standard scores on all four WJ-III subtests before and after HillRAP. Students who received HillRAP improved significantly on three of the four WJ-III subtests after one year of intervention: Letter-Word Identification, Reading Fluency, and Word Attack. Passage Comprehension showed no significant change over time. Overall, these results are consistent with the previous evaluation conducted by RTI in which WJ-III scores improved on three of the four subtests (i.e., Reading Fluency, Word Attack and Passage Comprehension).

Figure 1. Changes in Woodcock-Johnson Standard Scores Pre- to Post-Year 1 (n = 152)

![Bar chart showing changes in standard scores before and after HillRAP.](image)

**Subtest**

*statistically significant (p < .001)

Individual change over time showed considerable variability. These results are summarized in Appendix B.
**Year 2 Outcomes**

One hundred eleven students completed at least 20 sessions of HillRAP in Year 2 and were assessed a final time at the end (April/May) of the 09-10 school year. An additional 29 students completed only Year 1 of HillRAP, but were still attending school in DPS during 09-10, and were thus able to complete a final assessment at the end of Year 2. These data give us a tentative look at how well students maintain improvements if they only receive one year of intervention. The results of these 29 students are presented in Appendix D.

Figure 2 shows the mean standard scores on all four WJ-III subtests at pretest, Year 1, and Year 2. At the end of Year 2, students improved significantly from baseline on Passage Comprehension. They also maintained significant improvements on Reading Fluency and Word Attack. Letter-Word standard scores declined slightly in Year 2, thus were no longer significantly higher than baseline. These results are consistent with the earlier RTI study that also found significant improvements in three of the four WJ-III subtests (i.e., Passage Comprehension, Reading Fluency and Word Attack). Interestingly, the RTI study found an increase in Passage Comprehension after the first year of intervention, whereas the present study found significant improvements in Passage Comprehension only after two years.

**Figure 2. Changes in Woodcock-Johnson Standard Scores Pre- to Post-Year 2 (n=111)**

*statistically significant change from Pretest*
Demographic predictors of performance on the WJ-III

To examine whether HillRAP is equally effective across all student subgroups we looked at the following student characteristics: pre-intervention achievement scores, grade/age, sex, free/reduced lunch eligibility status, IQ score, IQ/achievement discrepancy, grade retention history, limited English proficiency status/English as a second language (LEP and ESL), and specific learning disability status. The student characteristics that were statistically significant predictors of improvement in academic achievement are described below. For a more detailed discussion of analyses and the demographic predictors of performance on the WJ-III, including graphs of findings, see Appendices B and C.

Students with larger delays in reading at baseline showed the largest improvements on all four subtests of the Woodcock-Johnson in both Year 1 and Year 2, suggesting that HillRAP is appropriately targeted towards students with significant reading challenges. Students with lower initial standard scores demonstrated gains 3 to 13 points greater than did students with relatively higher pre-intervention achievement scores.

Across Years 1 and 2, students in lower grades made larger gains on all of the four Woodcock-Johnson subtests, emphasizing the need to start HillRAP as early as possible in a student’s academic career in order to maximize achievement gains. Indeed, middle school students showed no gains on any subtest after two years of intervention. This may be because older students are more concerned about the stigma of being pulled out for HillRAP or are more frustrated with their long-standing reading problems, and as a result do not put forth the same effort as the younger students. Reading difficulties may also be more entrenched by middle school, and thus harder to remediate. In addition, older students are entering puberty and are in a different developmental phase, with more emphasis on peer interaction than on academics.

Race was a significant predictor of improvement on the WJ-III across both years, though the specific subtests showing these effects varied. In Year 1, Caucasian students improved significantly more than did African American students in Reading Fluency, whereas Latino students showed marginally better improvement than all other groups in Letter-Word identification and Passage Comprehension. In Year 2, Latino students improved significantly more than other racial groups in Passage Comprehension and marginally more in Letter Word. The reading delays for Latino students may have been caused at least in part by language difficulties. Improvement in English skills may have therefore accelerated their reading growth.

In general, both males and females improved across subtests, but males made marginally more gains in Letter-Word Identification and in Passage Comprehension in Year 1 and in Word Attack in Year 2.
Full-scale IQ score emerged as an important predictor of improvement over time in Passage Comprehension in Year 2. Those with lower IQs showed the largest improvement in Passage Comprehension.

**Implementation predictors of performance on the WJ-III**

One of the evaluation’s research goals was to examine the effect of faithful implementation and attendance on students’ reading achievement. The following variables were selected for that purpose: teachers’ years of teaching experience, level of education (Bachelor’s or Master’s), the average Overall Proficiency score teachers received during their observations, and hours of attendance (for a more detailed discussion see Appendices B and C).

Teacher education level and years of teaching experience were not significant predictors of improvement on any subtest in either year of the intervention. Average Overall Proficiency scores, however, were marginally predictive of Passage Comprehension gains in Year 1, suggesting that teachers who implemented the program with more proficiency and greater fidelity had students whose Passage Comprehension improved more after one year of HillRAP. It may be that reading comprehension is a higher level skill that requires a more skilled and experienced teacher to teach effectively. Moreover, teachers with lower proficiency scores may spend the bulk of their time on other HillRAP components and have less time to devote to comprehension instruction. Proficiency was no longer a predictor of any subscale after Year 2. It may be that teacher differences in implementation evened out by the second year of HillRAP instruction. Similar results were found in the RTI study: teachers’ overall proficiency scores did not have a significant impact on student achievement after two years of intervention, possibly because by the end of the second year all teachers were highly proficient.

Students with higher attendance performed significantly better over time on all WJ-III subtests, suggesting that high attendance is crucial for reading improvement. Specifically, higher attendance significantly enhanced Letter-Word and Passage Comprehension performance in Year 1, with marginal effects on Reading Fluency. In Year 2, attendance significantly predicted improvement in Word Attack and marginally predicted improvement in Letter-Word and Reading Fluency. The RTI study found similar results, emphasizing the importance of strong student attendance. The Hill Center recommends that students receive at least 108 hours of HillRAP (the equivalent of 144 class periods) in one school year to maximize outcomes. Importantly, students in this study attended an average of 77.7 (SD = 13.9) hours of HillRAP during Year 1 and 77.5 (SD = 20.6) hours in Year 2. Only three students (2%) attended 108 hours or more in Year 1, and seven students (6%) attended 108 or more hours in Year 2. With better attendance, it is likely that students would have made even larger gains in reading achievement.
To illustrate the effects of attendance on reading gains, Figure 3 depicts findings from Year 1 of the intervention. Attendance is divided into four groups with equal numbers of children in each. For more basic reading skills, such as Letter-Word identification, attendance thresholds are modest; students attending at least 67 ½ hours (i.e., the top 75% for hours attended) all made equivalent gains. More complex reading comprehension required much higher attendance, however, as gains were seen only for students in the top quartile (those who attended more than 88 hours).

Figure 3. Hours of Attendance as a Predictor of Improvement

*statistically significant difference (p < .05)  ^trend difference (p < .10)

End-of-Grade Assessment Results

The Department of Public Instruction uses the End-of-Grade (EOG) assessment to measure performance over time in grades 3 through 8. EOG reading performance is reported using several different scoring metrics. These include developmental scale scores, achievement levels, standardized “c-scores” and “growth scores.” Results on each of these metrics will be described below.

Developmental Scale Scores

Developmental scale scores are calculated from the EOG raw scores to allow for the comparison of a student’s achievement from one grade to the next. These scores are expected to increase every year with development, like height for example. There is no clear benchmark for the amount of increase to expect, however. As a result, if a child participating in a reading
intervention has a higher score after the intervention, one cannot say how much of the score increase is attributable to the intervention and how much was because of normal development. For this reason, developmental scale scores are not appropriate to use in examining an intervention’s efficacy, but because they are widely used, their analyses were included in this report.

Only 48 of the 152 students who received HillRAP the first year had EOG scores in both 07-08 (prior to the start of HillRAP) and 08-09 (after one year of intervention). The remainder were either in grades K–2 during one or both of these school years, completed the alternative Extend2 assessment, or were not in the Durham County school system during one of these two school years. As expected, students’ developmental scale scores significantly increased after one year of HillRAP, moving from an average of 326 to 334.

The number of HillRAP students who took the EOG test again after the second year of program implementation was small: only 33 students from the HillRAP sample completed the reading EOG assessment in 07-08 (prior to the start of the HillRAP intervention), 08-09 (after one year of intervention), and 09-10 (after two years of intervention). Many of the students with previous EOG scores switched to the alternative Extend2 assessment in 09-10, further limiting the generalizability of these findings. Those still taking the EOG had the same mean scores in 07-08 and 08-09 as the larger sample from Year 1, however, suggesting this subset may be representative of the larger sample with EOG data. Again, as expected, developmental scale scores increased significantly during the second year of HillRAP. This improvement is depicted in Figure 4.
Achievement Levels

Achievement levels provide a wider view of student performance and are based on developmental scale scores. Levels I and II correspond to below grade-level achievement, whereas levels III and IV correspond to on- or above-grade-level achievement, respectively. Because achievement levels are based on scale scores, they are also not the best metric to use when examining the efficacy of an intervention. Once again, because of their wide use they are included here. As seen in Figure 5, the proportion of students who earned Level 2 or Level 3 scores increased slightly after each year of intervention, though the majority of students continued to perform at a Level 1.

*statistically significant difference (p < .0001)
C-Scores

In contrast to the scores presented above, c-scores are standardized scores that allow one to examine how students perform in comparison with other students and to evaluate changes in scores over time in a reliable way. A c-score is calculated by subtracting the state-level mean scale score for the student’s grade from the student’s scale score and then dividing by the state-level standard deviation:

\[
c\text{-score} = \frac{\text{Student score} - \text{State-level mean for the norming year}}{\text{State-level standard deviation for the norming year}}
\]

Thus, a c-score describes a student’s reading performance relative to all same-grade students in the state. A score of 0 means that the student performed at the same level as the average student in the state, whereas a negative score means the student performed worse than the average student in the state.

HillRAP students showed improvements in EOG reading c-scores between the 07-08 (pre-HillRAP) and 08-09 school years (post-HillRAP), increasing an average of 0.12 points (SD = 0.6) (we could not examine scores prior to 07-08 because so few students in this sample had earlier scores). Students with EOG c-scores in 09-10 showed additional improvements after the second year of HillRAP. EOG c-score improvements were not statistically significant, however. These findings suggest that even though the HillRAP program appeared to help students improve their EOG scores relative to their grade-mates, the improvement was too small to be considered a reliable change (i.e., it could have been due to chance alone). This may be the result of the small
number of students who took the EOGs in both years as well as the fact that they started with considerable delays in reading (thus even with significant improvement may not perform well on a standardized test written at their grade level). Figure 6 depicts changes to EOG c-scores for students who took EOGs for all three years of the analysis.

**Figure 6. Change in EOG c-scores from Baseline (07-08; n = 33)**

![Chart showing changes in EOG c-scores from 07-08 to 09-10.](chart)

**Growth Scores**

The Department of Public Instruction also calculates growth scores to assess overall student changes in performance over time. A growth score compares a student’s current EOG performance with their predicted performance based on previous years’ testing:

\[
\text{Growth score} = \text{Current c-score} - \text{Predicted c-score}
\]

Predicted scores are calculated as the average c-score from the past two years, multiplied by an adjustment coefficient that models expected change in c-scores over time based on expected regression to the mean:

\[
\text{Predicted c-score} = \text{average of past 2 c-scores} \times 0.92
\]

If the current year’s c-score is higher than predicted, the growth score will be positive. This suggests that the student has made more gains than expected over the current school year. Importantly, though, for students with significant reading delays, even those who make greater gains than they usually do can show a negative growth score. The growth score adjustment assumes that students will move closer to the average c-score of 0 each year, regardless of
previous performance. For students with random fluctuation in performance, this is an accurate assumption and aids in monitoring true change. For those with actual, significant delays in reading, however, this assumption may not hold true—students who are considerably behind may indeed be expected to fall farther behind, showing a downward trajectory of growth scores across the years. Growth scores that remain negative but are less extreme may indicate an improvement in reading trajectory for these students.

Based only on the 24 students with growth score data in both 07-08 and 08-09, the average growth score improved by about .15 following participation in Year 1 of HillRAP, though this change is not significant. Seventeen students had enough information available to examine growth score changes through 09-10. For this small subsample, growth scores improved significantly between 07-08 and 08-09, moving to nearly 0 and remained at that level in 09-10 (see Figure 7 below). This suggests that prior to intervention, students were on a downward trajectory, with EOG scores decreasing each year. Following HillRAP intervention, these students were able to maintain EOG performance from one year to the next.

**Figure 7. Changes in EOG Reading Growth Scores (n = 17)**

![Figure 7](image)

* statistically significant (p<.05)

Taken together, the EOG results show that students significantly improved their reading skills across both years (based on the developmental scale scores and achievement levels), but their improvement was only minimally (and non-significantly) greater than that of their grade-mates. Because c-scores and growth scores are standardized, they are more valid indicators of achievement than developmental scores and levels of achievement. The dearth of large, significant results on c-scores and growth scores should not be interpreted as HillRAP’s failure.
to increase achievement, however. EOG tests are aligned to curricula of certain grades and measure what an average student in that grade knows. Students selected to participate in the current study were performing below their grade level. As a result it is not reasonable to expect them to perform at the level of the average student in that grade. The fact that their c-scores and growth scores on the average are improving suggests that they are making gains, but these gains may not reach the level of the average student in that grade within one or even two years of intervention. Moreover, because a high number of study students take the Extend2 EOG, which is not standardized, their scores cannot be included in the analyses and the sample becomes very small. Conclusions based on small samples can only be tentative and cannot be generalized to the population in general.

**Demographic predictors of performance on the EOG**

Considerable variability was also found for EOG scores (see Appendices B and C for a more detailed discussion). To examine why some students improved while others did not, we explored possible predictors of improved EOG scores. The following variables were examined: student sex, race, and English language proficiency; free/reduced lunch eligibility; exceptional children status and classification; and history of grade retention. No significant demographic predictors of EOG scores emerged.

**Implementation predictors of performance on the EOG**

To examine the effect of faithful implementation and attendance on students’ reading achievement on EOGs, the following variables were used as predictors: teachers’ years of teaching experience, level of education (Bachelor’s or Master’s), the average Overall Proficiency score teachers received during their observations, and hours of attendance (for a more detailed discussion see Appendix B). These analyses were only conducted for Year 1 data; Year 2 sample sizes were too small to yield valid results.
Teacher education level and years of teaching experience were not significant predictors of improvement on the EOGs. In addition, Overall Proficiency scores were not predictive of achievement on the EOGs. HillRAP attendance, on the other hand, predicted improvement on the EOG reading c-scores between 07-08 and 08-09. Figure 8 divides attendance into four groups with equal numbers of children in each to demonstrate this effect. This finding shows very clearly the importance of high attendance to produce effects that can be observed based on EOG scores. The more hours students attended, the higher their improvement in EOG scores. In fact, those with the lowest attendance (i.e., < 78 hours) showed a worsening performance as compared to their grade mates, widening the achievement gap.

Figure 8. Attendance as a Predictor of Change in EOG Reading C-scores

![Bar chart showing change in EOG reading c-scores for different attendance groups.]

**K-2 Assessment**

For students in grades kindergarten through 2, the Department of Public Instruction uses the K-2 assessment to measure change in early academic skills over time. For the purposes of this report, only the reading subtest was examined. In this measure, teachers assess each student’s level of reading from 0 (no reading) to 44 (reading books at a 5th grade level). Fourth quarter scores were used as an indicator of end-of-the-year reading ability.
Twenty-three students from the current HillRAP sample completed the K-2 assessment in both 07-08 (1st grade, prior to the start of the HillRAP intervention) and 08-09 (2nd grade, after one year of intervention). Only one student completed the K-2 assessment in 09-10, so effects of two years of the HillRAP intervention on K-2 scores could not be assessed.

At the end of the 07-08 school year (1st grade), these students were reading at an average level of 7.8 (SD = 4.7). In comparison, the average reading level of 1st graders across Durham County was 20.1 (SD = 9.6). After a year of intervention, the HillRAP sample was reading at an average level of 16.4 (SD = 9.8), for an average improvement of 8.6 levels. Looking across Durham County, average readers increased 10 levels to 30.2 (SD = 9.5) after one year of universal reading curriculum instruction. Even when the Durham sample was limited to only those with low reading scores at the end of the 1st grade (mean = 7.5), reading improved by an average of 9.7 levels by the end of 2nd grade. These findings suggest that students did not show unusual gains on the K-2 assessment following HillRAP participation. Given that the K-2 assessment is not standardized and that our analyses examine a very small sample, these results should be viewed with caution.

**Hill Write Student Data**

**Demographics**

HillWrite is the Hill Center’s Writing Achievement program that was piloted in the Durham Public Schools in 08-09. Four teachers did one HillWrite group each. Of the 16 students who received the intervention, 11 were consented to be in the study. Table 4 describes the demographic characteristics of the HillWrite sample (for a description of the HillWrite vs. HillWrite comparison sample, see Appendix E).

<table>
<thead>
<tr>
<th>Table 4. Demographic Characteristics of HillWrite Study Sample</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>HillWrite pilot sample</strong></td>
</tr>
<tr>
<td><strong>(n=11)</strong></td>
</tr>
<tr>
<td><strong>Mean (SD)</strong></td>
</tr>
<tr>
<td><strong>Age</strong></td>
</tr>
<tr>
<td>10.2 (1.8)</td>
</tr>
<tr>
<td><strong>Attendance (Hours)</strong></td>
</tr>
<tr>
<td>81.5 (0.9)</td>
</tr>
<tr>
<td><strong>Full Scale IQ</strong></td>
</tr>
<tr>
<td>89.8 (10.5)</td>
</tr>
<tr>
<td><strong>Sex</strong></td>
</tr>
<tr>
<td>Male 72.7%</td>
</tr>
<tr>
<td>Female 27.3%</td>
</tr>
</tbody>
</table>
WJ-III achievement test scores

The 11 students who were a part of the HillWrite pilot participated in the HillWrite intervention for at least one year (6 students continued the intervention in Year 2) and completed the WJ-III reading and writing subtests at the beginning (September) and end (April/May) of the 08-09 school year and at the end of the 09-10 school year. At the beginning of 08-09, prior to the start of the HillWrite intervention, these students performed an average of 1.5 to 2 years below grade level on each of the writing subtests (and 1 to 2 years below grade level on the reading subtests).

Figure 9 shows the mean standard scores on all six WJ-III subtests (both reading and writing subtests) before and after HillWrite. By the end of Year 2, students who received HillWrite improved significantly on Writing Samples. This improvement indicates that following one to two years of intervention, HillWrite students were better able to write high-quality sentences with increasing passage length, level of vocabulary, grammatical complexity and level of
concept abstraction. This change on the Writing Samples subtest was the largest absolute improvement seen across measures and interventions in this study (7 of 11 students improved at least 10 standard score points), bringing students to near average writing achievement. This is a promising finding, suggesting that HillWrite may produce a substantial improvement in writing skills. Future evaluation with larger sample sizes is necessary to replicate this finding.

Significant gains in Reading Fluency were also found after Year 1, even though these students did not receive the HillRAP intervention and no gains in reading achievement were thus expected.

Figure 9. Changes in standard scores pre- to post-Year 2 HillWrite

*statistically significant difference (p < .05)  
^trend difference (p < .10)
Summary and Conclusions

The purpose of this evaluation was to examine the effectiveness of HillRAP in helping public school students whose reading skills are compromised. A preliminary evaluation of the HillWrite program, designed for students with writing difficulties, was also included. More specifically, we addressed the following questions:

- What are the effects of HillRAP instruction on reading achievement over time?

The present report examined the effects of the HillRAP program in a pre- to post-intervention design. When pre- to post-intervention group results were examined, findings indicated that the HillRAP program was moderately successful in increasing the reading achievement of students with reading difficulties on the WJ-III. Significant gains were evidenced after one year of HillRAP on the Word Attack, Reading Fluency and Letter-Word Identification subtests. These gains suggest that after one year of HillRAP, students are better able to use phonics and read individual letters, words and sentences. Moreover, they are making these gains at a faster rate than their same-age peers, thus closing the gap between themselves and the average student. After a second year of HillRAP, students maintained or increased their gains in Reading Fluency and Word Attack. Notably, a significant improvement on the Passage Comprehension subtest of the WJ-III was evident at the end of Year 2, suggesting that students need a firm foundation of basic reading skills (achieved during the first year) before their comprehension improves.

HillRAP students showed primarily small, nonsignificant improvements in reading achievement on a statewide reading achievement test (i.e., EOG for grades 3-8, K-2 assessment for grades K-2) from 07-08 to 08-09 and 09-10. Interpretation of the EOG results should take into account two factors: (a) the EOG test is aligned to the curriculum for each grade and measures what an average student in that grade knows, and (b) a large percentage of students in this sample took the Extend2 alternative version of the EOG. Students selected to participate in this study were performing below their grade level (some even 3-4 years below grade level), so the EOG test is not as sensitive in detecting changes in growth toward grade level achievement. In addition, the Extend2 test scores are not standardized, so they cannot be analyzed with the scores of the regular EOG tests. This resulted in small samples for the analyses of the EOG test scores. Because of these two factors, EOGs are not a strong measure of change for students with reading delays. The small but positive changes on the EOG achievement scores do suggest that HillRAP holds promise for contributing to gradual progress in EOG performance. This is encouraging when one considers that students who are considerably behind in reading may indeed be expected to fall farther behind without any intervention.
• How do program outcomes vary for different subgroups of children? Specifically, do the effects of the Hill Center programs vary by race/ethnicity, sex, age, grade level, English language proficiency, free/reduced lunch eligibility, exceptional children status, exceptional children classification, IQ, or history of grade retention?

There was considerable variability on both WJ-III scores and EOG test scores, suggesting that HillRAP may be more appropriate for certain groups of students.

A number of significant predictors emerged for the WJ-III subtests:

• Pre-intervention WJ-III scores predicted improvement in Letter-Word Identification, Reading Fluency, Word Attack, and Passage Comprehension in both Years 1 and 2. Students who were further behind their peers at baseline made gains at a faster rate than did students with achievement closer to the average.

• Grade was a significant predictor of Word Attack and Passage Comprehension in both Years 1 and 2, suggesting that HillRAP should start as early as possible in order to maximize achievement gains. Grade also predicted Letter-Word Identification in Year 1 and Reading Fluency in Year 2.

• Overall, Latino students made the largest gains across subtests. Their reading delays may have been caused at least in part by language difficulties. Improvement in English skills may have therefore accelerated their reading growth. More specifically, Latino students’ improvement approached significance over both African American and Caucasian students in Passage Comprehension and over African American students in Letter-Word identification for both intervention years. Caucasian students increased their performance on Reading Fluency at a faster rate than African Americans after one year of intervention. Differences in program effectiveness by race should be further assessed in future HillRAP evaluations.

• Even though it did not reach significance, the results on gender showed some interesting trends. Both males and females made gains across all subtests during both years except for Passage Comprehension. On average, females lost ground in Passage Comprehension in Year 1, but regained the lost ground in Year 2. Overall, males made bigger gains across all WJ-III subtests than did females.

• How well does the Hill Center teacher training prepare teachers to deliver HillRAP in a public school setting? What is the level of intervention fidelity for these teachers?

Overall, the Hill Center teacher training appears to prepare teachers to deliver HillRAP in a public school setting with a high level of fidelity. Some revision of the fidelity instrument is recommended to allow more accurate assessment of implementation fidelity, however. When skills are not observed, the instrument does not currently provide a clear indication of whether the skills were not relevant to the specific unit taught (i.e., not
applicable) or whether they were relevant but not taught (i.e., a break in fidelity). There were several specific skills related to Vocabulary and Reading Comprehension that were infrequently observed, but these items were not mandatory components and not appropriate for all units or stages of learning. If revision to the fidelity measure reveals that these areas are being omitted more often than appropriate, training or mentoring guidance in efficient allocation of time and skill prioritization in the areas of vocabulary and reading comprehension may be beneficial.

- What are the effects of attendance and model fidelity (i.e., faithful implementation of HillRAP) on reading achievement?

   Overall, higher attendance resulted in higher scores on both state-mandated and nationally-normed achievement tests, thus highlighting the importance of attendance for increased gains in reading achievement. Student attendance significantly predicted EOG reading c-scores and performance on Letter-Word Identification and Reading Comprehension in Year 1 and Reading Fluency in Year 2. More marginal results were found in both years for the other subtests. The threshold for attendance level necessary to see clear gains varied by subtest—more basic skills showed growth with a modest level of attendance, but more advanced skills (i.e., passage comprehension) required a high level of attendance to produce improvements. This is consistent with the finding that, across students, Passage Comprehension did not significantly improve until Year 2. Comprehension is a more complex skill to develop, and clearly requires a longer period of time to address deficits and strengthen basic reading skills before it can improve.

   The Hill Center recommends 108 hours of HillRAP for optimal results, but believes that improvements can be seen with fewer hours. HillRAP sessions are not usually 60 minutes long, but rather 45. Consequently, 108 hours is equivalent to 144 sessions. The current sample attended 77.5 hours of HillRAP on average, which is the equivalent of 103 class periods (Figure 10). This means that students, on average, completed 72% of the optimal HillRAP dosage. Despite that, improvements were seen on a nationally-normed achievement test after both years and on state-mandated tests after Year 1. It may be that if all students received the expected amount of HillRAP sessions, significant results on state-mandated tests would also be evident after the second year. However, some consideration needs to be given to the fact that in order to receive 144 sessions of HillRAP, students will need to attend the program on 144 of 180 school days (80% of all school days in a year). Whether that is a realistic expectation merits further discussion.
Figure 10. Actual versus Recommended Yearly Attendance

![Bar chart showing actual versus recommended yearly attendance](chart.png)

Model fidelity as evidenced by higher average Overall Proficiency scores for teachers marginally predicted student Passage Comprehension in Year 1. This finding suggests that more experience and skill are required to teach the higher level skill of reading comprehension. Moreover, teachers with lower proficiency scores may spend the bulk of their time on other HillRAP components and have less time to devote to comprehension instruction. This effect disappeared in Year 2, perhaps because teachers had additional practice and mentoring and were more effective in implementing HillRAP. Alternately, several students switched teachers in Year 2, muddying the waters with respect to teacher effects on performance.

- What preliminary information can we deduce about the effectiveness of the HillWrite program?

The preliminary examination of the small number of students who participated in the HillWrite program showed some promising results on the students’ reading and writing achievement on the WJ-III. Students in HillWrite significantly improved their Reading Fluency scores in Year 1 and showed large, significant improvement on the Writing Samples subtest of the WJ-III in Year 2. This was the largest absolute improvement found in the entire study suggesting that further study of the HillWrite program with larger sample sizes is warranted.

Results from the evaluation of HillRAP in the Durham Public schools show that this program significantly improved the reading proficiency of struggling readers as evidenced on nationally-normed achievement tests (i.e., WJ-III), but not on the state-mandated achievement tests (i.e.,...
EOGs and K-2 Assessment), with the latter limited by small sample sizes. All effects observed were stronger for younger students and students with more significant initial achievement delays. As expected, students’ comprehension improved significantly only after basic reading skills were well established. Hispanic students showed higher rates of improvement overall. These subgroup patterns should be monitored in future evaluations to determine whether any modifications need to be explored. A comparison of the current study with the one conducted by RTI suggest that, despite the demographic differences of the two samples, the students in both showed significant improvements on the Woodcock-Johnson Test of Achievement. In addition, both studies point to the importance of high attendance. Both studies also found significant improvements in developmental scale scores, but given their limited validity as a measure of intervention effectiveness, no conclusions can be drawn from them. It is unfortunate that the RTI study did not have c-scores and growth scores available. Furthermore, the RTI study concluded that the impact of HillRAP is comparable across student groups. The present study found some subgroup differences and as a result we would recommend that this question is studied further with a larger sample.

Considering the improvement of the HillRAP students on the WJ-III, but the general lack of results on the EOGs/K-2 assessment, we cannot draw strong conclusions about the effectiveness of the HillRAP program in Durham County. However, the significant gains observed in WJ-III scores over the two-year period and the improvement in reading comprehension in the second year are encouraging findings, especially when consideration is given to the fact that selected students were experiencing severe, and some maybe extreme, challenges (e.g., IQ scores below 50, autism diagnosis). Moreover, effects on all tests including the EOGs were much stronger for students who attended at least 78 hours of HillRAP, suggesting that much larger gains might be seen if all students had attended HillRAP at this minimal level.

**Recommendations**

Given the promising but inconclusive current results, we recommend the following:

1. Conduct a follow-up study using random assignment and a larger sample so we can draw definitive conclusions about the effectiveness of the program.

Follow-up studies using random assignment and larger samples will be necessary to draw any definitive conclusions about the effectiveness of the program. Random assignment is of particular importance in future evaluations of HillRAP in order to ensure clear results. Through random assignment of students to HillRAP or a comparison group, evaluators could create two groups of students who are similar on every aspect other than intervention type. As a result, if HillRAP students’ reading achievement improved significantly over that of the comparison group, evaluators could definitively conclude
that their improvement was attributable to HillRAP and nothing else. Random assignment is the only way to ensure group equivalence and therefore unbiased findings. Given the promising early results of HillRAP, rigorous evaluation is warranted to further examine program effectiveness.

2. Do not reach conclusions about the effectiveness of the HillRAP program based on EOG scores.

The lack of statistically significant improvements in EOG scores should not be considered a measure of the HillRAP effectiveness. The End-of-Grade test is a criterion-referenced test based on the North Carolina Standard Course of Study for the curriculum at each grade level. As a result the items on the test are focused on skills and content specific to each grade. The children who participated in HillRAP were selected because of their delays in reading, and many were performing two or more years below their actual grade level. Consequently, it does not make sense to expect them to perform acceptably on material written at their grade level even after a year of intensive intervention. The lack of statistically significant improvements on the EOG after one or even two years of intervention is not surprising, though EOG scores are moving marginally closer to those of their grade peers. An additional concern is that only a small number of students took the EOG test on consecutive years. Because the content of EOG test is grade-specific and overly challenging for students with significant delays, many instead take the Extend2, an alternative nonstandardized assessment. Such dramatic loss of data could have certainly contributed to the current results, but also speaks to the inappropriateness of the EOG to adequately measure achievement for these students.
References


Appendix A

Fidelity ratings from 4th observation (after 1 year of HillRAP implementation)

Based on observation data recorded by a Hill Center Master Mentor, teachers improved their implementation of HillRAP slightly between the first observation in October and the fourth observation in April, but were highly proficient at both observations. The Master Mentors also rated a number of specific teaching skills within each component area. Ratings indicated a very high level of overall proficiency, though a few specific skills (in Reading Comprehension and Vocabulary) stand out as areas for improvement that may need ongoing practice and monitoring. These skills were rated as not observed, but it is impossible to ascertain whether they were not observed because they were not applicable, or whether they were truly a break in fidelity (and as such, requires remediation). The individual skills and their fidelity ratings are listed below.

Drill Activities
- 100% were fully proficient in starting with students in the ready position.
- 85.7% were fully proficient in asking questions suitable in level, type, and structure.
- 100% charted responses accurately.
- 100% provided an appropriate pace for each student.
- 100% made strong eye contact with the students.
- 100% were observed to correct incorrect responses.
- However, 28.6% did not have students practice drill questions correctly and ask them again. The remaining 71.4% were fully proficient at this component.
- 92.9% used positive constructive reinforcement.
- 100% used drill effectively for practice instead of instruction.
- 61.5% used visual cues for support (visual cues may not have been needed for the remainder).
- 71.4% allowed additional time for those who process slowly (this may have been unnecessary for the remainder).

Phonological Awareness
- 100% used questions/activities appropriate to the age and skill level of the group.
- 50% used manipulatives proficiently, 7.1% used them somewhat proficiently, and 42.9% did not use them (likely because they were not needed).
- 100% made eye contact with the students during this component.
- 100% used positive constructive reinforcement during this component.
- 85.7% were fully proficient at using an appropriate pace of instruction.
Word Attack
- 92.9% were fully proficient at choosing suitable word lists.
- 78.6% were fully proficient at providing effective pacing of instruction.
- 100% gave positive constructive reinforcement.
- 100% had each student read 3-4 words before moving on to the next student.
- 92.9% kept students engaged in some stage of this component at all times (decoding, marking syllables, practicing).
- 71.4% were fully proficient at marking the students’ responses.

Fluency: Timed Reading from a List
- 100% were fully proficient in handling this fluency component efficiently and systematically.
- 92.9% were fully proficient in accurately charting test results.
- 100% selected the correct timed test lists.
- 85.7% corrected errors and practiced correct responses (it may be that there were no errors for other observations). Of these, 75% were fully proficient at practicing correct responses.
- 100% were fully proficient at setting and monitoring goals.
- 100% gave positive constructive reinforcement.

Fluency: Timed Reading from a Text
- 88.9% were fully proficient in handling this fluency component efficiently and systematically.
- 88.9% were fully proficient in accurately charting test results.
- 77.8% were fully proficient at selecting the correct text.
- 77.8% corrected errors and practiced correct responses (it may be that there were no errors for other observations).
- 100% gave positive constructive reinforcement.

Vocabulary
- 84.6% reinforced vocabulary words through drill.
- 100% taught vocabulary words within the context of the reading text.
- 100% gave positive constructive reinforcement.

The following vocabulary skills were not observed consistently across teachers, and will be an area for increased focus (though all teachers who did incorporate these components were fully proficient):
- Teaching prefixes, suffixes, and/or roots
- Identifying tiered words
Reading Comprehension
- 85.7% were fully proficient at choosing appropriate text.
- 92.9% spent equitable time with the students.
- 85.7% asked comprehension questions at the appropriate skill level.
- 100% offered questions in an oral and written form.
- 85.7% gave a variety of comprehension questions (literal, inferential, prediction, sequencing, etc.).
- 85.7% gave positive constructive reinforcement.

The following reading comprehension skills were observed infrequently, though all teachers who did incorporate these components were fully proficient. Importantly, teachers are given a choice about including these skills—they are not mandatory.
- Having students make predictions before and during reading
- Having students make text to self or text to world connections
- Having students able to summarize what they read
- Having students think aloud while reading
- Having students able to visualize the passages read
- Having students identify text structures
- Having students who could generate questions about the text

Classroom Management
- 92.9% had materials well-organized and ready for class.
- 100% demonstrated proper posture and articulate questions and directions.
- 100% made smooth transitions between class components.
- 85.7% started class promptly.
- 85.7% had students on task throughout the class period.
- 85.7% needed/used a point/reward system. Of these, 91.7% were fully proficient at using the point/reward system effectively.
- 71.4% implemented multi-sensory techniques. Of these, 90% were fully proficient at using these techniques effectively.
Appendix B

**Year 1 demographic predictors of performance on the WJ-III**

Individual change over time in Woodcock-Johnson scores showed considerable variability. Table B1 presents the proportion of students for each subtest who improved their standard scores by at least 2 points (generally the level of improvement needed to reach significance in the current sample), maintained their scores, or declined by at least 2 points relative to their peers. Word Attack was the area where the most students showed improvement, but Reading Fluency was where the largest average improvement occurred.

<table>
<thead>
<tr>
<th>WJ-III Subtest</th>
<th>Improved by 2 or more points</th>
<th>Maintained standard score</th>
<th>Declined by 2 or more points</th>
</tr>
</thead>
<tbody>
<tr>
<td>Letter-Word</td>
<td>53.3%</td>
<td>12.5%</td>
<td>34.2%</td>
</tr>
<tr>
<td>Passage Comprehension</td>
<td>38.8%</td>
<td>11.2%</td>
<td>50.0%</td>
</tr>
<tr>
<td>Reading Fluency</td>
<td>53.3%</td>
<td>18.4%</td>
<td>28.3%</td>
</tr>
<tr>
<td>Word Attack</td>
<td>57.9%</td>
<td>14.5%</td>
<td>27.6%</td>
</tr>
</tbody>
</table>

When results show such a high level of variability, it is helpful to explore possible reasons for the differences in change over time. To accomplish this, we examined whether the impact of the HillRAP program varied by key student characteristics. The student variables examined were: pre-intervention achievement scores, grade/age, sex, free/reduced lunch eligibility status, IQ score, IQ/achievement discrepancy, grade retention history, limited English proficiency status/English as a second language (LEP and ESL), and specific learning disability status. General linear models predicting standard score change over time were conducted, controlling for hours of HillRAP attendance, pre-intervention achievement scores, grade level, and sex (where these were not the predictors of focus). Interactions were examined for all significant variables. The student characteristics that were statistically significant predictors of improvement in academic achievement are described below.

**Pre-intervention WJ-III standard scores**

Pre-intervention achievement scores on the WJ-III significantly predicted improvement during Year 1 on all four subtests (Letter-Word: F(1, 147) = 34.3, p < .0001, Passage Comprehension: F(1, 147) = 56.6, p < .0001, Reading Fluency: F(1, 147) = 19.7, p < .0001, and
Word Attack: F(1, 147) = 66.3, p < .0001). As expected, students with lower initial standard scores demonstrated more gains in each area than did students with relatively higher pre-intervention achievement scores.

To show this effect graphically, a median split was used for pre-intervention scores on each subtest (i.e., students were split into the lower 50% and the higher 50% based on the distribution of scores in the current sample). Median scores for these subtests were: 77 for Letter-Word, 74 for Passage Comprehension, 73 for Reading Fluency, and 81 for Word Attack. Figure 2 shows the average change in WJ-III standard score for students scoring below the median and for students scoring above the median for this sample. For Reading Fluency, both groups showed improvement, though students starting below the median improved more. For Letter-Word and Word Attack, students starting below the median improved significantly while those starting above the median generally maintained their scores. Finally, for Passage Comprehension, students with lower pre-test scores showed improvement at post-test, while those starting above the median lost ground relative to their peers.

Figure B1. Pre-intervention Achievement Scores as a Predictor of Improvement

![Graph showing change in standard score for different subtests.]

*statistically significant difference

Grade Level

Student grade level significantly predicted improvement on the Letter-Word (F(1, 147) = 18.1, p < .0001), Passage Comprehension (F(1, 147) = 14.3, p < .001), and Word Attack (F(1, 147) = 18.3, p < .0001) subtests. As shown in Figure B2, students in lower grades made larger gains.
Interestingly, the Reading Fluency subtest (which measures the ability to quickly read and comprehend short, simple sentences), showed the opposite pattern. Though not significant, students in the later grades improved slightly more in reading fluency than did younger students. Differences in HillRAP implementation for middle school students should be explored to understand this discrepancy. It may be that students in later grades focus more heavily on reading statements as opposed to phonics skills, given that these students need to be prepared for a much higher level of reading in content classes. These findings suggest that HillRAP instructors may need to increasingly emphasize phonics skills with older students to ensure they have a strong foundation for reading.

*Figure B2. Grade as a Predictor of Improvement*

Race

Race significantly predicted improvement over time on Reading Fluency (F(1, 138) = 3.2, p < .05), with significant differences only between African American and Caucasian subgroups. Race was related to both Letter-Word (F(1, 138) = 2.8, p = .06) and Passage Comprehension (F(1, 138) = 3.0, p = .05) at the trend level. Letter-Word differences were significant only between African-American and Latino students, whereas for Passage Comprehension, Latino students improved more than both African-American and Caucasian students. Overall, Latino students showed more improvements in reading than either of the other two subgroups.
**Figure B3. Race as a Predictor of Improvement**

![Bar graph showing change in standard scores across different domains for African American, Latino, and Caucasian students.](image)

*statistically significant difference (p < .05)  ^trend difference (p < .10)

**Sex**

Student sex predicted improvement on Letter Word (F(1,147) = 2.8, p = .09) and Passage Comprehension subtests (F(1,147) = 3.6, p = .06) at the trend level. Figure B4 shows that both males and females made gains in Letter-Word Identification (though males improved significantly more), whereas only males improved in Comprehension (females lost ground relative to peers).
To assess the effects of teacher-specific traits (e.g., years of experience), fidelity to the HillRAP curriculum, and attendance, a multi-level model was constructed with students nested within teachers. Separate analyses were run for each subtest of the WJ-III, including baseline standard scores and the significant predictors identified in the general linear modeling described above. Teacher-level variables included: years of teaching experience, level of education (Bachelor’s or Master’s), and average Overall Proficiency scores across the four observations. The student-level variable added for this multi-level model was hours of attendance.

**Teacher characteristics**

Teacher education level and years of teaching experience were not significant predictors of improvement on any subtest. Average Overall Proficiency scores, however, did tend to predict improvement on the Passage Comprehension subtest at the trend level (F(1, 12) = 4.0, p =.07), despite low variability in Overall Proficiency ratings (i.e., all teachers had high proficiency scores). Teachers who received higher proficiency scores had students whose Passage Comprehension showed more improvement than students of teachers who were less proficient.
Student attendance

Student attendance significantly predicted improvement over time on two of the four subtests. For Letter-Word (F(1, 147) = 15.8, p < .001), each hour of attendance in HillRAP was related to an average corresponding increase of 0.14 in Letter-Word identification from pre-test to post-test. In other words, for each 7 additional hours of HillRAP that a student attended, his/her change in Letter-Word identification was 1 standard score better than a student with 7 fewer hours. This does not mean that students improved by 1 standard score from baseline with each 7 hours they attended, however; those with the lowest attendance tended to lower their scores, losing ground relative to normative peers. Additionally, this ratio of attendance to improvement applies only to the observed range of attendance in this sample and cannot be extrapolated beyond this range.

For Passage Comprehension (F(1, 147) = 13.6, p < .0001), every extra hour of attendance corresponded with an average increase of 0.21 in Passage Comprehension performance. Attendance at 4.8 additional hours of HillRAP corresponded with a shift of 1 standard score in Passage Comprehension. A trend relationship was found for Reading Fluency (F(1, 147) = 2.9, p = .09), with each extra hour of attendance corresponding to a 0.08 increase in Reading Fluency scores. Attendance at 12.5 additional hours of HillRAP corresponded with a shift of 1 standard score in reading fluency. Figure B6 divides attendance into quartiles to demonstrate this effect.
Figure B6. Hours of Attendance as a Predictor of Improvement

*statistically significant difference (p < .05) ^trend difference (p < .10)

Year 1 demographic predictors of performance on the EOG

There was considerable variability from student to student on changes in EOG scores. Almost half (47.9%) improved c-scores by at least 0.1 between 07-08 and 08-09, 25% remained at a consistent c-score level, and 27.1% got lower c-scores in 08-09 (by at least 0.1). For growth scores, which include expected regression towards the mean, only 9.3% improved their scores by at least 0.5, while 66.7% remained at a consistent growth score and 24% showed a negative growth score trajectory.

To examine why some students improved while others did not, we explored possible predictors of improved EOG scores. The following variables were examined: student sex, race, and English language proficiency; free/reduced lunch eligibility; exceptional children status and classification; and history of grade retention. No significant demographic predictors of EOG scores emerged.

Year 1 implementation predictors of performance on the EOG

To assess the effects of teacher-specific traits (e.g., years of experience), fidelity to the HillRAP curriculum, and attendance, a multi-level model was constructed with students nested within teachers. Separate analyses were run for each subtest of the WJ-III, including baseline
standard scores and the significant predictors identified in the general linear modeling described above. Teacher-level variables included: years of teaching experience, level of education (Bachelor’s or Master’s), and average Overall Proficiency scores across the four observations. The student-level variable added for this multi-level model was hours of attendance.

Teacher education level and years of teaching experience were not significant predictors of improvement on any subtest. Overall Proficiency scores were not predictive of achievement on the EOG.

HillRAP attendance, on the other hand, predicted improvement on the EOG reading c-scores between 07-08 and 08-09 ($F(1, 45) = 4.5$, $p < .05$). Every increase in one hour of attendance in HillRAP was related to a corresponding increase of .01 in the EOG reading c-score. In other words, attendance at 100 additional hours of HillRAP corresponded to an average improvement of 1 c-score (equal to 1 standard deviation) on the reading EOGs. Figure B7 divides attendance into quartiles to demonstrate this effect. This finding shows very clearly the importance of high attendance to produce effects that can be observed based on EOG scores. The more hours students attended, the higher their improvement in EOG scores.

**Figure B7. Attendance as a Predictor of Change in EOG Reading C-scores**

![Figure B7](image-url)
Appendix C

**Year 2 demographic predictors of performance on the WJ-III**

Individual change in WJ-III scores continued to show variability in Year 2. Table C1 presents the proportion of students for each subtest who improved their standard scores by at least 2 points, maintained their scores, or declined by at least 2 points relative to their peers. After 2 years, the majority of students improved in Word Attack and Passage Comprehension.

<table>
<thead>
<tr>
<th>WJ-III Subtest</th>
<th>Improved by 2 or more points</th>
<th>Maintained standard score</th>
<th>Declined by 2 or more points</th>
</tr>
</thead>
<tbody>
<tr>
<td>Letter-Word</td>
<td>46.8%</td>
<td>15.4%</td>
<td>37.8%</td>
</tr>
<tr>
<td>Passage Comprehension</td>
<td>54.9%</td>
<td>10.9%</td>
<td>34.2%</td>
</tr>
<tr>
<td>Reading Fluency</td>
<td>39.6%</td>
<td>27.1%</td>
<td>33.3%</td>
</tr>
<tr>
<td>Word Attack</td>
<td>54.9%</td>
<td>18.1%</td>
<td>27.0%</td>
</tr>
</tbody>
</table>

To further explore the reasons for this variability, we examined whether the impact of the HillRAP program varied by key student characteristics. The student variables examined were: pre-intervention achievement scores, grade, age, sex, free/reduced lunch eligibility status, IQ score, IQ/achievement discrepancy, grade retention history, limited English proficiency status/English as a second language (LEP and ESL), and specific learning disability status. General linear models predicting standard score change over time were conducted, controlling for hours of HillRAP attendance, pre-intervention achievement scores, grade level, and sex (where these were not the predictors of focus). Interactions were examined for all significant variables. The student characteristics that were statistically significant predictors of improvement in academic achievement are described below.

**Pre-intervention WJ-III standard scores**

Pre-intervention achievement scores on the WJ-III significantly predicted improvement in Year 2 on all four subtests (Letter-Word: $F(1, 106) = 15.7, p < .001$, Passage Comprehension: $F(1, 106) = 142.7, p < .0001$, Reading Fluency: $F(1, 106) = 61.1, p < .0001$, and Word Attack: $F(1, 106) = 62.2, p < .0001$). Students with lower initial standard scores demonstrated more gains in each area than students with relatively higher pre-intervention achievement scores. To depict
this effect, a median split was used for pre-intervention scores on each subtest (i.e., students were split into the lower 50% and the higher 50% based on the distribution of scores in the current sample). Median scores for these subtests were: 80 for Letter-Word, 73 for Passage Comprehension, 78 for Reading Fluency, and 83 for Word Attack. Figure C1 shows the average change in WJ-III standard score for students scoring below the median and for students scoring above the median for this sample. Students with lower pre-test scores showed improvement at the end of Year 2, while those starting above the median lost ground relative to their peers.

Figure C1. Pre-intervention Achievement Scores as a Predictor of Improvement at the end of Year 2

![Bar chart showing change in standard score for different subtests with lowest 50% and highest 50% categories.](chart.png)

*statistically significant difference (p < .05)

**Grade Level**

Student grade level significantly predicted improvement on the Passage Comprehension (F(1, 106) = 15.7, p < .001), Reading Fluency (F(1, 106) = 7.6, p < .01), and Word Attack (F(1, 106) = 7.9, p < .01) subtests. As shown in Figure C2, students in lower grades made larger gains. In contrast to Year 1 findings, however, middle school students in Year 2 did not show gains in Reading Fluency. Indeed, after two years of intervention, these students showed little to no improvement in any measured area of reading.
Students in elementary school and middle school showed different patterns of change over time. Elementary students showed improvement in all four WJ-III subtests, whereas middle school students showed improvement only in Reading Fluency in Year 1, which returned to baseline at Year 2. Again, these findings emphasize the importance of targeting HillRAP to younger students to remediate problems prior to middle school. Systematic differences in middle school, both educationally and interpersonally, may make intervention more difficult in these later grades.
Race

Race significantly predicted improvement over time on Passage Comprehension (F(1, 98) = 3.1, p < .05), with Latino students improving significantly more than either African American or Caucasian students. Race had a trend-level effect on Letter-Word (F(1, 98) = 2.7, p = .07), with significant differences only between African-American and Latino students. Overall, Latino students again made the largest gains across subtests in Year 2, only surpassed by Caucasians in Reading Fluency.

Figure C5. Race as a Predictor of Improvement at the end of Year 2

*statistically significant difference (p < .05)  ^trend difference (p < .10)
Sex

Student sex predicted improvement on the Word Attack (F(1,106) = 3.0, p = .09) subtest at the trend level. Figure C6 shows that both males and females made gains in all four subtests. Males showed larger gains on all subtests, but only the Word Attack difference approached significance.

Figure C6. Sex as a Predictor of Improvement at the end of Year 2

IQ Score

Full-scale IQ score significantly predicted improvement over time on Passage Comprehension (F(1, 66) = 5.5, p <.05). Those with lower IQs showed the largest improvement in Passage Comprehension after two years of intervention.
To assess the effects of teacher-specific traits (e.g., years of experience), fidelity to the HillRAP curriculum, and attendance, a multi-level model was again constructed with students nested within teachers. Separate analyses were run for each subtest of the WJ-III, including baseline standard scores and the significant predictors identified in the general linear modeling described above. Teacher-level variables included: years of teaching experience, level of education (Bachelor’s or Master’s), and average Overall Proficiency scores across the four observations. The student-level variable added for this multi-level model was hours of attendance.

**Teacher characteristics**

Teacher education level and years of teaching experience were not significant predictors of improvement on any subtest following the second year of HillRAP implementation. Although overall proficiency predicted Year 1 Passage Comprehension scores, proficiency did not predict improvement after Year 2. It may be that teacher differences in implementation evened out by the second year of HillRAP instruction. Similar results were found in the RTI study. Teachers’ overall proficiency scores did not have a significant impact on student achievement after two years of intervention, possibly because by the end of the second year all teachers were highly proficient.
**Student attendance**

Student attendance across the two years of intervention significantly predicted improvement over time on Word Attack ($F(1, 106) = 4.5$, $p < .05$) and had a trend effect on Letter-Word ($F(1, 106) = 3.3$, $p = .07$) and Reading Fluency ($F(1, 106) = 2.9$, $p = .09$). For all three subtests, every increase in one hour of attendance in HillRAP was related to a corresponding increase of 0.05 in standard scores from pre-test to end of Year 2 post-test. In other words, for each 20 additional hours of HillRAP that a student attended, his/her change in score for Letter-Word, Reading Fluency, and Word Attack was 1 standard score better than a student with 20 fewer hours. It is not clear from this study that the rate of improvement in achievement would be constant with higher attendance.

Figure C8 divides attendance across the two years using a median split to demonstrate this effect. Again, this finding emphasizes the importance of strong student attendance through at least one full year of daily intervention.

**Figure C8. Hours of Attendance as a Predictor of Improvement at the end of Year 2**

![Image](image.png)

*statistically significant difference ($p < .05$)  ^trend difference ($p < .10$)

**Year 2 demographic predictors of performance on the EOG**

Considerable variability from student to student was also seen on changes in EOG scores. Almost half (54.5%) improved c-scores by at least 0.1 between 07-08 and 09-10, 6.1% remained at a consistent c-score level, and 36.4% lowered their c-scores by at least 0.1. For growth scores, which include expected regression towards the mean, 23.5% showed growth of at least 0.5 (the
minimum growth that is considered reliable), while 47.1% remained at a consistent growth score and 29.4% showed a negative growth score (smaller than -0.5).

To examine why some students improved while others did not, we explored possible predictors of improved EOG scores. The following variables were examined: student sex, race, and English language proficiency; free/reduced lunch eligibility; exceptional children status and classification; and history of grade retention. With this small subset of students, no significant predictors were identified.

**Year 2 implementation predictors of performance on the EOG**

Implementation predictors of Year 2 EOGs were examined, but because the samples were too small, they are not reported here.
Appendix D

Year 2 results of performance on the WJ-III for students receiving only one year of intervention

Of the students who participated in Year 1 of HillRAP, twenty-nine remained in the Durham Public School system but did not receive a second year of intervention. Generally this was because students moved to a school without HillRAP. These students were re-assessed using the WJ-III at the end of 09-10 to examine their retention in reading gains over time.

Figure D1. Changes in Standard Scores Pre- to Post-Year 2:
Students with no Year 2 HillRAP

Students with no Year 2 of HillRAP started off with lower scores on the WJ-III than did their counterparts who participated in two years of intervention. This may reflect the fact that a larger proportion of these students were older, with the majority in middle school by Year 2. Many of these students moved from a HillRAP elementary school in Year 1 to a non-HillRAP middle school in Year 2. As is the case for all middle school students, these students did not show strong gains following Year 1 of intervention. Furthermore, their scores declined after a second year without HillRAP.
Appendix E

Hill Write Comparison Study Student Data

Demographics

We were able to identify and consent five students that were close matches with HillWrite students on demographic variables and reading achievement. Table E1 describes the demographic characteristics of the matched HillWrite and HillWrite Comparison samples.

Table E1. Demographic Characteristics of the HillWrite comparison samples

<table>
<thead>
<tr>
<th></th>
<th>Matched HillWrite sample (n=5)</th>
<th>HillWrite Comparison sample (n=5)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean (SD)</td>
<td>Mean (SD)</td>
</tr>
<tr>
<td>Age</td>
<td>10.9 (2.5)</td>
<td>11.1 (2.3)</td>
</tr>
<tr>
<td>Attendance (hours)</td>
<td>81.8 (1.1)</td>
<td>--</td>
</tr>
<tr>
<td>Full Scale IQ</td>
<td>90.6 (12.2)</td>
<td>83.0 (22.6)</td>
</tr>
<tr>
<td>%</td>
<td>%</td>
<td>%</td>
</tr>
<tr>
<td>Sex</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>80%</td>
<td>80%</td>
</tr>
<tr>
<td>Female</td>
<td>20%</td>
<td>20%</td>
</tr>
<tr>
<td>Race</td>
<td></td>
<td></td>
</tr>
<tr>
<td>African American</td>
<td>40%</td>
<td>40%</td>
</tr>
<tr>
<td>Latino</td>
<td>40%</td>
<td>40%</td>
</tr>
<tr>
<td>Caucasian</td>
<td>0%</td>
<td>0%</td>
</tr>
<tr>
<td>Multiracial</td>
<td>20%</td>
<td>20%</td>
</tr>
<tr>
<td>Grade</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3rd grade</td>
<td>40%</td>
<td>40%</td>
</tr>
<tr>
<td>4th grade</td>
<td>20%</td>
<td>20%</td>
</tr>
<tr>
<td>7th grade</td>
<td>40%</td>
<td>40%</td>
</tr>
</tbody>
</table>
Matched HillWrite sample (n=5) | HillWrite Comparison sample (n=5)
---|---
Exceptional Children Status | 100% | 60%
Specific Learning Disability (SLD) | 100% | 40%
Written Learning Disability (SLD) | 80% | 20%
Intellectual Disability (EMD/ID) | 0% | 20%
Other Health Impairment (OHI) | 0% | 0%
Serious Emotional Disability (SED) | 0% | 0%
Visual Impairment (VI) | 0% | 0%
Hearing Impairment (HI) | 0% | 0%
Speech/Language Impairment (S/LI) | 0% | 20%

Free/reduced lunch eligibility | 80% | 80%
Primary Language not English | 20% | 25%

Relative to comparison students, gains in both writing subtests are evident (see Figures E1 and E2).

Figure E1. Changes in Writing Fluency: Intervention versus Comparison Students

![Figure E1. Changes in Writing Fluency: Intervention versus Comparison Students](image-url)
Figure E2. Changes in Writing Samples: Intervention versus Comparison Students
Appendix F

HillRAP Comparison Study Student Data

Demographics

The HillRAP evaluation was designed to include a matched comparison sample to examine how HillRAP students changed relative to similar students who did not receive the intervention. A comparison group of students was selected from all non-HillRAP schools in Durham County. The evaluation team matched 122 intervention students with students from comparison schools based on sex, race/ethnicity, grade, Exceptional Children’s status, 2007-08 EOG score or K-2 assessment score, and free/reduced lunch status. From the comparison students identified, 75 consents were collected, 74 students were tested in April/May 2009, and 66 were re-tested in April/May 2010.

Unfortunately, the delay in the consenting process for the comparison group resulted in the absence of WJ-III achievement testing before the intervention started. As a result changes in the comparison students achievement scores over the first year could not be tested. However, W-J III testing was completed for both the intervention and control groups at the end of the first and second years. In Year 2, changes in WJ-III and in End-of-Grade scores were examined for those students with a second year of HillRAP relative to the matched comparison students.

Of the students assessed, 65 matched pairs of intervention and comparison students had sufficient data to be used for analysis at the end of Year 1 (i.e., the intervention student in the pair completed at least 20 sessions of HillRAP in Year 1 and both students in the pair had WJ-III assessments at the end of 08-09). In Year 2, only 43 matched pairs were included in analyses. The remainder were excluded for one of the following reasons: (1) the intervention student did not receive HillRAP in Year 2, (2) the comparison student did receive HillRAP in Year 2, or (3) one student in the pair did not have a WJ-III assessment in 09-10. This subsample may not be representative of the full comparison group sample.

Demographics for the groups are presented in Table F1. Matched HillRAP and comparison samples were significantly different on Exceptional Children (EC) status ($\chi^2(1) = 11.9, p < .001$) and Other Health Impaired (OHI) status ($\chi^2(1) = 5.9, p < .05$). Despite our efforts to match the HillRAP and comparison students as closely as possible, there were more comparison students with an EC diagnosis. This difference is attributable to the higher number of HillRAP students with an OHI diagnosis. The majority of students under that status group received an Attention Deficit Hyperactivity Disorder diagnosis.
Table F1. Demographic Characteristics of HillRAP and Comparison Group Samples

<table>
<thead>
<tr>
<th></th>
<th>Yr 1 Matched HillRAP sample (n=65)</th>
<th>Yr 1 Comparison sample (n=65)</th>
<th>Yr 2 Matched HillRAP sample (n=43)</th>
<th>Yr 2 Comparison sample (n=43)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean (SD)</td>
<td>Mean (SD)</td>
<td>Mean (SD)</td>
<td>Mean (SD)</td>
</tr>
<tr>
<td>Age</td>
<td>10.3 (1.8)</td>
<td>10.3 (1.8)</td>
<td>10.8 (1.7)</td>
<td>11.0 (1.7)</td>
</tr>
<tr>
<td>Attendance (hours)</td>
<td>76.8 (13.8)</td>
<td>--</td>
<td>81.2 (18.2)</td>
<td>--</td>
</tr>
<tr>
<td>Full Scale IQ</td>
<td>81.6 (12.8)</td>
<td>82.6 (17.0)</td>
<td>83.0 (12.4)</td>
<td>82.2 (16.6)</td>
</tr>
<tr>
<td>Q4 Reading Score, K-2 Assessment '07-'08</td>
<td>11.4 (7.6)</td>
<td>12.1 (8.0)</td>
<td>11.5 (8.2)</td>
<td>12.1 (8.2)</td>
</tr>
<tr>
<td>Reading C-Score, EOG Assessment '07-'08</td>
<td>-1.8 (0.4)</td>
<td>-1.8 (0.4)</td>
<td>-1.9 (.5)</td>
<td>-1.9 (0.5)</td>
</tr>
<tr>
<td>%</td>
<td>%</td>
<td>%</td>
<td>%</td>
<td>%</td>
</tr>
<tr>
<td>Sex</td>
<td>Male 70.8%</td>
<td>69.2%</td>
<td>65.1%</td>
<td>65.1%</td>
</tr>
<tr>
<td></td>
<td>Female 29.2%</td>
<td>30.8%</td>
<td>34.9%</td>
<td>34.9%</td>
</tr>
<tr>
<td>Race</td>
<td>African American 63.1%</td>
<td>64.6%</td>
<td>62.8%</td>
<td>65.1%</td>
</tr>
<tr>
<td></td>
<td>Latino 29.2%</td>
<td>29.2%</td>
<td>30.2%</td>
<td>30.2%</td>
</tr>
<tr>
<td></td>
<td>Caucasian 4.6%</td>
<td>4.6%</td>
<td>2.3%</td>
<td>2.3%</td>
</tr>
<tr>
<td></td>
<td>Multiracial 3.1%</td>
<td>1.5%</td>
<td>4.7%</td>
<td>2.3%</td>
</tr>
<tr>
<td>Grade</td>
<td>1&lt;sup&gt;st&lt;/sup&gt; grade 1.5%</td>
<td>6.2%</td>
<td>0.0%</td>
<td>0.0%</td>
</tr>
<tr>
<td></td>
<td>2&lt;sup&gt;nd&lt;/sup&gt; grade 23.1%</td>
<td>15.4%</td>
<td>2.3%</td>
<td>14.0%</td>
</tr>
<tr>
<td></td>
<td>3&lt;sup&gt;rd&lt;/sup&gt; grade 27.7%</td>
<td>32.3%</td>
<td>25.6%</td>
<td>11.6%</td>
</tr>
<tr>
<td></td>
<td>4&lt;sup&gt;th&lt;/sup&gt; grade 18.5%</td>
<td>18.5%</td>
<td>37.2%</td>
<td>39.5%</td>
</tr>
<tr>
<td></td>
<td>5&lt;sup&gt;th&lt;/sup&gt; grade 13.9%</td>
<td>12.3%</td>
<td>18.6%</td>
<td>18.6%</td>
</tr>
<tr>
<td></td>
<td>6&lt;sup&gt;th&lt;/sup&gt; grade 7.7%</td>
<td>7.7%</td>
<td>4.7%</td>
<td>4.7%</td>
</tr>
<tr>
<td></td>
<td>7&lt;sup&gt;th&lt;/sup&gt; grade 7.7%</td>
<td>7.7%</td>
<td>2.3%</td>
<td>7.0%</td>
</tr>
<tr>
<td></td>
<td>8&lt;sup&gt;th&lt;/sup&gt; grade 0.0%</td>
<td>0.0%</td>
<td>9.3%</td>
<td>4.7%</td>
</tr>
</tbody>
</table>
Retention

Not retained 77.3% 53.3% 78.6% 66.7%
Retained once 20.5% 33.3% 17.9% 22.2%
Retained more than once 2.3% 13.3% 3.6% 11.1%

Exceptional Children Status

<table>
<thead>
<tr>
<th>Status</th>
<th>Study 1</th>
<th>Sample 1</th>
<th>Year 1</th>
<th>Outcomes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Specific Learning Disability (SLD)</td>
<td>47.7%</td>
<td>44.6%</td>
<td>46.5%</td>
<td>44.2%</td>
</tr>
<tr>
<td>Reading Learning Disability</td>
<td>43.1%</td>
<td>40.0%</td>
<td>39.5%</td>
<td>41.9%</td>
</tr>
<tr>
<td>Intellectual Disability (EMD/ID)</td>
<td>18.5%</td>
<td>10.8%</td>
<td>18.6%</td>
<td>11.6%</td>
</tr>
<tr>
<td>Other Health Impairment (OHI)</td>
<td>23.1%</td>
<td>7.7%</td>
<td>25.6%</td>
<td>9.3%</td>
</tr>
<tr>
<td>Serious Emotional Disability (SED)</td>
<td>0.0%</td>
<td>0.0%</td>
<td>0.0%</td>
<td>0.0%</td>
</tr>
<tr>
<td>Visual Impairment (VI)</td>
<td>1.5%</td>
<td>0.0%</td>
<td>0.0%</td>
<td>0.0%</td>
</tr>
<tr>
<td>Hearing Impairment (HI)</td>
<td>4.6%</td>
<td>0.0%</td>
<td>4.7%</td>
<td>0.0%</td>
</tr>
<tr>
<td>Speech/Language Impairment (S/LI)</td>
<td>6.2%</td>
<td>12.3%</td>
<td>4.7%</td>
<td>9.3%</td>
</tr>
</tbody>
</table>

Free/reduced lunch eligibility 80.0% 83.1% 83.7% 90.7%

Primary Language not English 27.7% 27.7% 27.9% 25.6%

Comparison Study Sample Year 1 Outcomes

Of the 65 matched pairs, only 16 pairs had EOG assessments in 07-08 and 08-09 for both the intervention and the comparison students. For these pairs, differences in 08-09 EOG reading c-scores were examined, controlling for 07-08 reading c-scores and for EC status (because more intervention students were classified as EC). There were no significant differences in the change in reading c-scores at the end of Year 1 for the HillRAP versus comparison students after controlling for the above variables (see Figure F1). Comparison of 08-09 reading growth scores also showed no significant difference between the two groups. Comparison students on average improved more than the HillRAP intervention students, but this difference was attributable to disparity in the rates of EC status. For this subset, only 12.5% of the comparison students were diagnosed with reading learning disabilities, whereas 43.8% of the intervention students had this diagnosis.
Ten matched pairs had reading K-2 assessment data available for both 07-08 and 08-09. Again, differences were examined controlling for pre-intervention K-2 reading levels and for EC status. Comparison students had a larger raw improvement rate, but after controlling for differences in EC status, no significant differences in post-intervention K-2 reading levels were found (see Figure F2).
Taken together these results suggest that the HillRAP program did not improve children’s reading achievement scores on state-mandated measures of reading achievement as compared to a group of students who received the ‘business as usual’ reading curriculum. There are, however, valid reasons why this may be so. First, with the much larger sample of EC students (in particular those with OHI/ADHD and reading disabilities) in the intervention group, these students would not be expected to improve at the same rate. Second, only 25% of student pairs were tested using the EOG in both years and as such were included in the comparison (for the remainder, one member of the pair completed the Extend2 measure, left the school system for one of the testing years, or was too young for EOG assessment in at least one of the years). Likewise, because of the switch from K-2 assessments to EOG assessments when children transition to the third grade, the K-2 analysis examined a very small subset of student pairs (15%). Such small sample sizes make these findings less stable and more difficult to interpret.

**Comparison Study Sample Year 2 Outcomes**

For Year 2 analyses, changes in WJ-III scores could be compared between the intervention and comparison groups. Additionally, the Oral Reading Fluency subtest of the DIBELS was administered to more fully assess reading fluency. Forty-three pairs had sufficient data for these comparisons. Figures F3-F7 show average scores on each subtest at the end of 08-09 (Year 1) and 09-10 (Year 2). For each subtest, the comparison students started out significantly above the intervention students. However, change over time was equivalent for the two groups.
Figure F3. Changes over time on the DIBELS

![Graph showing changes over time on the DIBELS for Intervention and Comparison groups between Year 1 and Year 2.]

Figure F4. Changes over time on Letter-Word Identification

![Graph showing changes over time on Letter-Word Identification for Intervention and Comparison groups between Year 1 and Year 2.]

Evaluation of HillRAP in the Durham County Schools
Figure F5. Changes over time on Passage Comprehension

Figure F6. Changes over time on Reading Fluency
Figure F7. Changes over time on Word Attack

Of the 43 matched pairs, none had K-2 assessment data across all three years (most students were 3rd-grade or older by the end of the study). Only 6 pairs had EOG assessments in 07-08, 08-09, and 09-10 for both the intervention and the comparison student. Clearly, this is an extremely small subset of students, thus results may not be representative of the full sample. Indeed, with such a small sample, statistical analyses were not conducted—only descriptive results are provided, in Figures F8 and F9 below. In this small subset, intervention and comparison students showed similar changes in reading EOG scores from 07-08 to 09-10.
Examination of the outcomes for HillRAP versus comparison group students on the EOG and K-2 assessments showed no differences between them. Given that the comparison group did not get tested at the beginning of the school year on the WJ-III, it is impossible to ensure the equivalence of the two groups; in fact, despite the evaluation team’s best efforts to match the comparison students as closely as possible to HillRAP students, the comparison group had a significantly smaller number of children with an EC classification and, more specifically, an OHI classification. The lack of a significant difference between the two groups on any of the achievement tests could be a result of the non-equivalence of the two groups, but also of the very small sample sizes.