



**Evaluation of the HillRAP Intervention in
Davie County Middle Schools**

2008-2010

Final Report

Submitted by

**Christina Christopoulos,
Katie Rosanbalm, &
David Rabiner**

**Center for Child and Family Policy
Duke University**

June 2011

Table of Contents

Executive Summary	4
Key findings.....	4
Summary and conclusions	6
Introduction.....	8
Purpose.....	8
Background.....	8
Methodology.....	9
School profiles	9
Student selection.....	10
Teacher training	11
Data sources	11
Achievement testing.....	11
Hill Center Database.....	12
Teacher form.....	12
Student form.....	13
Observation data	13
Results.....	13
Teacher Data	13
Demographics	13
Observations	13
Student Data.....	15
Demographics	15
WJ-III achievement test scores	16
Year 1 Outcomes.....	16
Year 2 Outcomes.....	18
Demographic predictors of performance on the WJ-III.....	19
Implementation predictors of performance on the WJ-III	21
End-of-Grade Assessment Results.....	21
Developmental Scale Scores.....	21
Achievement Levels.....	23
C-Scores.....	24
Growth Scores.....	26
Demographic and implementation predictors of performance on the EOG	28
Summary and Conclusions	28
References.....	32
Appendix A.....	33
Fidelity ratings from 4 th observation (after 1 year of HillRAP implementation)	33
Appendix B.....	36
Cohort 1, Year 1 demographic predictors of performance on the WJ-III.....	36
Pre-intervention WJ-III standard scores	36
Grade Level.....	37
Age.....	38
Sex.....	39
Free/reduced lunch eligibility	40

Full-scale IQ score	41
Implementation predictors of performance on the WJ-III	42
Appendix C	44
Cohort 1, Year 2 variability in WJ-III outcomes	44
Appendix D	45
Cohort 2 demographic predictors of performance on the WJ-III	45
Pre-intervention WJ-III standard scores	45
Full-scale IQ score	46
Appendix E	48
Cohort 2 demographic predictors of performance on the EOGs	48
Grade Level	48
Free/reduced lunch eligibility	49

Executive Summary

This report presents the final findings for a two-year evaluation of the Hill Center Reading Achievement Program (HillRAP) as implemented in a middle school setting from September 2008 to June 2010. The Center for Child and Family Policy at Duke University conducted this evaluation in collaboration with the Davie County Public Schools and the Mebane Charitable Foundation.

Key findings

The purpose of this evaluation was to examine the effectiveness of HillRAP in helping middle school students whose reading skills are compromised. More specifically, we addressed 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, teacher characteristics, and model fidelity (i.e., faithful implementation of HillRAP) on reading achievement?

The findings show that:

- The HillRAP program is successful in increasing the reading achievement of middle school students with reading difficulties. Students made improvements in reading achievement as evidenced both on a state-wide reading achievement test (i.e., EOGs) and on a nationally-normed achievement test (WJ-III). Two samples of HillRAP middle school students (Cohort 1 and Cohort 2) showed significant improvements in reading achievement on state-wide End-of-Grade testing (i.e., EOG standard scores, achievement levels, developmental scale scores, and growth scores) after one year of intervention despite the fact that they started with significant delays in reading. Cohort 1 students' EOG scores were already improving before the intervention started, but HillRAP participation seems to have accelerated their improvement. Cohort 2 showed declining EOG scores prior to HillRAP participation, but significant improvements following HillRAP. These students reversed their pattern of worsening performance and instead began to close the gap between themselves and their grade-mates. For the smaller number of students who received HillRAP for two years and took the EOGs in 07-08, 08-09, and

09-10 (n = 24), the same improvement trend was obvious, with even larger gains after Year 2.

On a nationally-normed achievement test (WJ-III), the biggest gains after one year of intervention were on the Reading Fluency subtest, followed by Letter-Word Identification. These gains suggest that after one year of HillRAP, students are better able to read individual words as well as sentences. Moreover, they are making these gains at a faster rate than their same-grade peers, thus closing the gap between themselves and the average student. For the smaller group of students who received the intervention for two years and were tested pre-intervention, post-year 1 and post-year 2 on the WJ-III (n=33), significant results were found for Letter-Word Identification, Reading Fluency and Word Attack. Passage Comprehension did not improve significantly, as was expected after the second intervention year, though a gradual upward trend was noted. Middle school students may need even longer interventions to show marked improvements in more complex reading skills.

- Due to the variability of gains in EOG and WJ-III standard scores, possible student-level predictors of improvement were examined (i.e., pre-intervention achievement scores, grade, age, sex, free/reduced lunch eligibility status, IQ score, grade retention history, LEP status and Exceptional Child status). A number of significant predictors of reading improvement emerged after one year of HillRAP. Looking across both cohorts of students, two findings were consistent:
 - 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.
 - Students with higher IQs improved their Word Attack scores at a rate faster than their peers, possibly suggesting that this is a group of youth who have the cognitive resources to achieve but struggle due to a fairly circumscribed learning disability. Once these students are taught ways to compensate for their disability, their phonics skills may increase rapidly. Word Attack assesses generalized phonics skills applied to made-up words rather than actual words. This requires more complex abstract thinking rather than sheer memorization, thus IQ may play a stronger role for scores on this subtest.

Other predictors of reading achievement showed less consistent patterns across cohorts, but deserve further study.

- In general, and consistent with findings of other HillRAP evaluations, younger students showed the strongest gains, suggesting that HillRAP will be most beneficial when started early in a student’s academic career. Specifically, for Cohort 1, students in lower grades made more improvement in Letter-Word Identification, and younger students made more gains in Word Attack than did older students. In contrast to these findings, however, Cohort 2 students in *higher* grades showed more gains on EOG scores. The effects of grade level on reading improvement will need further study with larger samples before firm conclusions can be drawn.
 - Females increased their performance on Reading Fluency at a faster rate than did males, but this was only true in Cohort 1. Other HillRAP evaluations have found similar, though nonsignificant, effects for Reading Fluency, though males have outperformed females on other reading subtests of the WJ-III.
 - Youth coming from higher socioeconomic backgrounds (i.e., not eligible for free/reduced lunch) made significantly higher gains on Word Attack and Passage Comprehension than did students from more underprivileged circumstances (Cohort 1) and showed more improvement on EOG scores following one year of intervention (Cohort 2). For Letter-Word Identification and Reading Fluency, however, the reverse pattern was found (though it was not significant). It is possible that youth from underprivileged homes require longer periods of remediation to develop the more complex skills of comprehension and generalized phonics.
- 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).
 - Teacher characteristics, implementation proficiency ratings, and student attendance did not predict changes in reading achievement. Teacher proficiency and student attendance were both consistently high across the sample, so there was minimal variation for observing effects on outcomes.

Summary and conclusions

Results from the evaluation of HillRAP in the Davie County middle schools suggest that this program improved the reading proficiency of struggling readers and began to close the

achievement gap between these students and their peers on a state-mandated achievement test (i.e., EOG) and a nationally-normed achievement test (i.e., WJ-III). These effects are stronger for students with more significant initial delays, higher IQs and who are from families with greater financial resources; the program did not have the same level of benefit for students with significant cognitive limitations or those who receive free/reduced lunch, particularly in the more complex areas of generalized phonics and comprehension. This pattern should be monitored in future evaluations to determine whether modifications of the program for these subgroups are necessary. The lack of improvement in Passage Comprehension suggests that older students who start with serious deficits in reading may need longer interventions for such improvements to appear.

While statistically significant, gains were highly variable among students. Given the lack of a comparison group, it is impossible to know what these students' scores would have looked like without intervention; it is possible that they would have fallen further behind, meaning that even those maintaining steady scores over time were benefitting from HillRAP participation. On the other hand, we cannot rule out the possibility that even without HillRAP, students would have maintained pace with their classmates or even shown some improvement. Thus, though results are promising, we cannot make strong conclusions about HillRAP as the agent of change in the observed improvements.

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.

Evaluation of the HillRAP Intervention in Davie County Middle 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 a middle school setting from September 2008 to June 2010. The Center for Child and Family Policy at Duke University conducted this evaluation in collaboration with the Davie County Public Schools and the Mebane Charitable Foundation.

Purpose

The purpose of this evaluation is to examine the effectiveness of HillRAP in helping middle school students whose reading skills are compromised. More 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, teacher characteristics, and model fidelity (i.e., faithful implementation of HillRAP) on reading achievement?

Background

The Hill Center was established in 1977 as the Hill Learning Development Center 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. Beginning in fall 2005, HillRAP implementation began in all six of the public elementary schools in Davie County. After initial success in the lower grades, the Mebane Charitable Foundation provided funds for the expansion of the HillRAP curriculum to the middle school level. This adapted program, which was first implemented in the fall of 2008 in the three Davie County middle schools, is the subject of the current evaluation.

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. Phonological awareness, phonics, fluency, vocabulary, and comprehension are the core of the daily instructional program that is ideally implemented in 45- to 60-minute sessions five days a week (note: for the Davie County Schools, HillRAP sessions were only 40 minutes long because they were held during students' extracurricular time). While students work in small groups of four, each student has an individualized curriculum to provide instruction in areas where there are demonstrated skill deficits in reading. 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. All 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.

Methodology

Because the Davie County Public Schools desired to make HillRAP available to all students who might benefit from it, the evaluation did not incorporate random assignment to an intervention or the use of a comparison group. Instead, the current evaluation followed a pre-post design in which students' achievement before intervention was compared to their achievement after participation in HillRAP. This design allows us to examine changes over time in reading performance; however, because we do not know with certainty how the students' scores would have changed over time without HillRAP participation, conclusions are more tentative than if a randomized-controlled trial had been conducted.

School profiles

All three middle schools (grades 6-8) in Davie County participated in the program. Demographics for these schools are presented in Table 1.

Table 1. School demographics

	Ellis	North Davie	South Davie
School Size 08-09 (number of students)	516	502	573
Percent Race/Ethnicity 08-09			
White	86	88	69
Black/African American	5	6	15
Hispanic/Latino	8	5	15
Other	1	1	0.3
Percent eligible for free/reduced lunch 08-09	25	27	48
Percent who passed the reading EOG in 07-08 (the year before program implementation)	70	65	57

Student selection

Davie County middle school principals selected students for participation in HillRAP based on the following criteria: previously failed EOGs (48.3%), risk for failing EOGs (66.4%), risk for grade retention (11.4%), LD identification (38.9%), EC identification with an IQ < 85 (11.0%), and/or an inability to read despite having completed other reading interventions (1.3%) (students may have been selected for multiple reasons). Once eligible students were identified, middle school personnel contacted families to inform them about the HillRAP program and evaluation and to obtain informed consent for participation. Over the two-year intervention, 155 consents were collected, 101 in Year 1 and 54 in Year 2. Because students were identified solely by the schools, we do not know how many eligible students declined participation.

Out of the 101 students consented in Year 1, 92 were tested both before and after the intervention. The remainder moved to new schools and were removed from the study. This group of students is described as Cohort 1, Year 1 below. Only 33 of the 92 students received a second year of intervention followed by a final assessment. The remainder either transitioned to high school or moved out of the school system. These students are described as Cohort 1, Year 2 below. To keep the classes full and allow the maximum number of students to receive HillRAP, new students were enrolled in HillRAP for 09-10 (Year 2). Of the 54 students newly consented in Year 2, 50 were tested both before and after the intervention. That group is described as Cohort 2.

Teacher training

Twenty-one teachers participated in a 3-day training offered by Hill Center staff in August 2008. The training included a 1-day training in Phonics Breaking the Code, a 2-day HillRAP workshop and two half-day follow-up workshops. During the core workshop teachers focused first on the fundamentals of phonological awareness and phonics to ensure all teachers had a core base of knowledge in teaching reading to struggling students, and then incorporated intensive instruction and role play on the components of HillRAP using the precision teaching approach adopted by the Hill Center. In addition, training was offered on the Hill Center's Reading Assessment instrument, which helps create a profile of strengths and difficulties and allows 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). 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 of teaching of the English language across all teachers.

During 08-09, two additional 3-hour follow-up meetings were conducted in a group setting, in which teachers were trained in using the SMART novel units technology (the Mebane foundation has generously provided funds to create a high-tech infrastructure for all the Davie County schools, which allows them to take advantage of the SMART technology) 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) the first year of HillRAP. Students with two years of HillRAP were also tested a third time at the end of the second year of intervention. The testing was coordinated by the school district, the teachers, and the Hill Center and was carried out by professionals 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:

- 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.

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 curriculum-based and are specifically aligned to the North Carolina Standard Course of Study. 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.

Hill Center Database

To accompany the HillRAP intervention, The Hill Center developed a web-based database to facilitate data collection, student monitoring, and skill tracking. 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.

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 provided the following information on each participating student: reason for selection, previous achievement and IQ testing dates and scores, reading interventions received in the past two years, 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.

Observation data

Master Mentors conducted classroom observations at least four times for each teacher during the 08-09 school year, and one more observation was conducted in 09-10. Mentors rated HillRAP implementation quality and fidelity using a HillRAP observation form with a scale of 1 (not proficient) to 3 (proficient) or 0 (not observed). This 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 rated. At the completion of each observation, the mentor gave the teacher an overall proficiency rating from 1 (not proficient) to 3 (proficient) and provided feedback to the teacher.

Results

Teacher Data

Demographics

A total of 21 teachers across the three Davie County middle schools implemented the HillRAP intervention. Teachers included 16 Exceptional Children's teachers, 2 English Language Learners' teachers, 2 remediation teachers and 1 counselor. These teachers are all Caucasian and predominantly female (85.7%). Just under half of the teachers (47.6%) have a Master's degree, and the remainder have a Bachelor's degree. They have an average of 12.9 years teaching experience (range = 3.75 to 31 years), and 76% are EC teachers.

Observations

Based on observation data recorded by a Hill Center Master Mentor, teachers improved their implementation of HillRAP substantially between the first observation in October and the fourth observation in April. 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 2. Teacher Proficiency in HillRAP Implementation

HillRAP Component	1st Observation % with each rating	4th Observation % with each rating
Drill Activities	2 = 61.1% 3 = 38.9%	2 = 10% 3 = 90%
Phonological Awareness	2 = 16.7% 3 = 83.3%	3 = 100%
Word Attack	2 = 61.1% 3 = 38.9%	3 = 100%
Fluency		
Timed Reading from a Text	3 = 100%	3 = 100%
Timed Reading from a List	2 = 62.5% 3 = 37.5%	2 = 5.3% 3 = 94.7%
Vocabulary	2 = 66.7% 3 = 33.3%	3 = 100%
Comprehension	2 = 27.8% 3 = 72.2%	3 = 100%
Classroom Management	2 = 77.8% 3 = 22.2%	3 = 100%
Overall Proficiency	2 = 72.2% 3 = 27.8%	3 = 100%

In addition to the overall component ratings represented in Table 2, 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 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.

Student Data

Demographics

Table 3 describes the demographic characteristics of the sample. Ninety-two students from Cohort 1 were tested with the WJ-III both before and after the HillRAP intervention. Students had an average age of 13.1 (SD = 1.2, range = 11 to 15.8) and attended an average of 93 hours of HillRAP in the first year (SD = 19.0, range = 33 to 114 hours). The subset for whom IQ scores were available (n = 67) had an average IQ of 87.7 (SD = 14.0, range = 52 to 116), about one standard deviation below the national mean.

Because only approximately a third of Cohort 1 students received HillRAP during the second year, a new Cohort (Cohort 2) was recruited in 09-10. The fifty Cohort 2 students had an average age of 12.8 (SD = 1.0, range = 11.3 to 14.7) and attended an average of 102 hours of HillRAP (SD = 8.7, range = 69 to 121 hours). The 25 students for whom IQ scores were available had an average IQ of 92.4 (SD = 9.4, range = 74 to 108).

A comparison of the demographic characteristics of the two cohorts showed that there were significant differences between them on several key variables. First, grade at entry to HillRAP was significantly lower for Cohort 2 ($\chi^2(2) = 13.3, p < .01$); Cohort 2 had twice the percentage of 6th grade students and a corresponding lower percentage of 7th and 8th graders. Moreover, Cohort 2 included more students of lower socioeconomic status (free/reduced lunch eligibility, $\chi^2(1) = 4.8, p < .05$) and more Spanish speakers ($\chi^2(1) = 4.0, p < .05$). Finally, the two cohorts differed significantly on pre-test Letter-Word Identification ($t(140) = -2.4, p < .05$) and Reading Fluency ($t(140) = -2.5, p < .05$), with Cohort 2 showing higher baseline reading skills (Letter-Word = 88.1, Reading Fluency = 88.0) than Cohort 1 (Letter-Word = 82.8, Reading Fluency = 82.9). Despite these significant differences, the results of the two cohorts were consistent overall and are presented together below.

Table 3. Student Demographic Characteristics

	Cohort 1, Year 1 (n=92)	Cohort 1, Year 2 (n=33)	Cohort 2, Year 1 (n=50)
	Mean (SD)	Mean (SD)	Mean (SD)
Age	13.1 (1.2)	12.5 (1.0)	12.8 (1.0)
Attendance (hours)	93.3 (19.0)	102.8 (11.1)	102.1 (8.7)
Full Scale IQ	87.7 (14.0)	86.5 (14.0)	92.4 (9.4)

	Cohort 1, Year 1 (n=92)	Cohort 1, Year 2 (n=33)	Cohort 2, Year 1 (n=50)
	%	%	%
Sex			
Male	68.5	69.7	62.0
Female	31.5	30.3	38.0
Race			
Caucasian	70.7	57.6	73.9
African American	9.8	15.2	2.2
Latino	14.6	21.2	17.4
Multiracial	4.9	6.1	6.6
Grade			
6 th grade	29.4	0	60.0
7 th grade	32.6	51.6	14.0
8 th grade	38.0	48.4	26.0
Retention			
Not retained	53.2	44.4	75.0
Retained once	44.3	55.6	25.0
Retained twice	2.5	0	0
Exceptional Children Status			
Specific Learning Disability (SLD)	62.5	51.4	48.8
Educable mental disability (EMD/ID)	35.9	36.4	34.1
Other Health Impaired (OHI)	9.8	6.1	0.0
Serious Emotional Disability (SED)	4.4	3.0	6.8
	1.1	0.0	0.0
Free/reduced lunch eligibility	46.7	51.5	66.0
Limited English Proficiency	6.8	15.2	17.4

WJ-III achievement test scores

Year 1 Outcomes

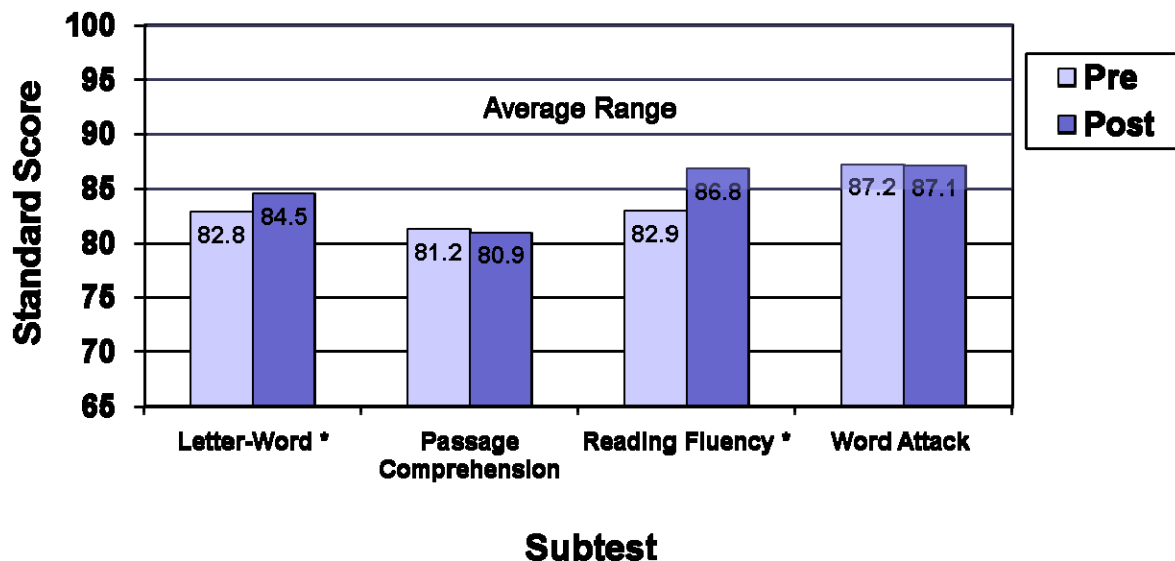
The WJ-III was used to measure changes in reading achievement from pre- to post-intervention. The grade-based WJ-III computerized scoring program yields raw and standard scores. A standard score is a transformation of the raw score based on national norms for students in a given grade. By using standard 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 (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 that grade level. On the other hand, students whose standard scores increase after they receive a year of HillRAP (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).

Ninety-two students from Cohort 1 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 2.5 to 3 years below grade level on each of the reading subtests.

Figure 1 shows the mean standard scores on all four WJ-III subtests before and after HillRAP. Students who received HillRAP improved significantly on two of the four WJ-III subtests, Letter-Word Identification ($t(91) = -2.5, p < .05$) and Reading Fluency ($t(91) = -5.0, p < .0001$). Word Attack and Passage Comprehension showed no change over time in standard scores, meaning that HillRAP students showed improvement in these areas equivalent to that of the average student in their grade, but did not begin closing the gap on these two subtests.

Figure 1. Changes in standard scores pre- to post-Year 1: Cohort 1

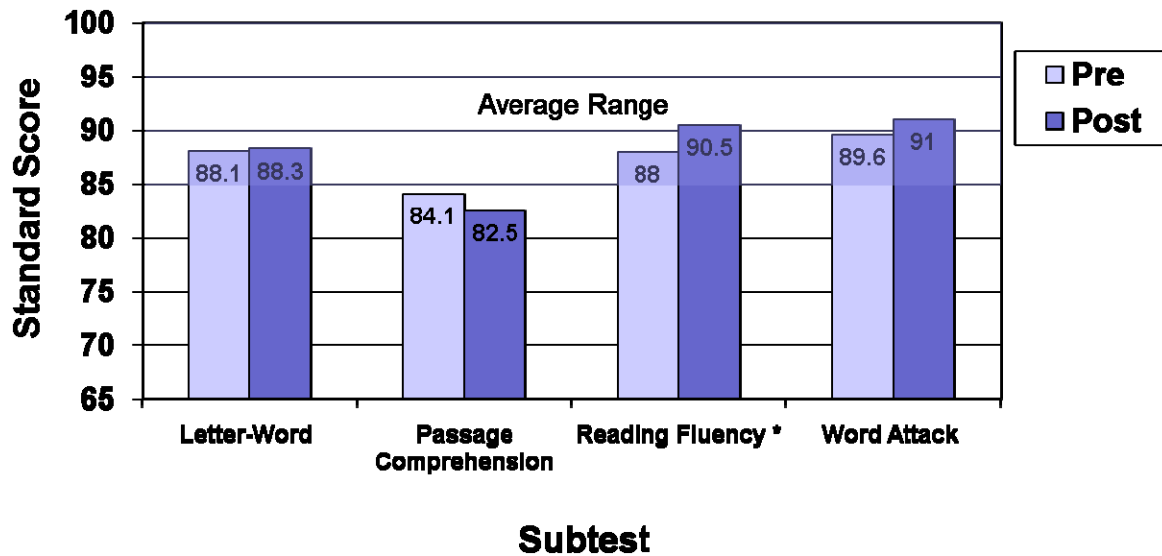


*statistically significant change from Pretest ($p < .05$)

All 50 students from Cohort 2 completed the WJ-III reading subtests at the beginning and end of 09-10. Prior to the start of the HillRAP intervention, these students performed an average

of 1.75 to 3.0 years below grade level on each of the reading subtests (making these students less delayed in reading than those in Cohort 1). Indeed, as seen in Figure 2, students in Cohort 2 were in the average range at pretest for three of the subtests. Significant improvement was seen only in Reading Fluency, which was also the area of the most growth for Cohort 1.

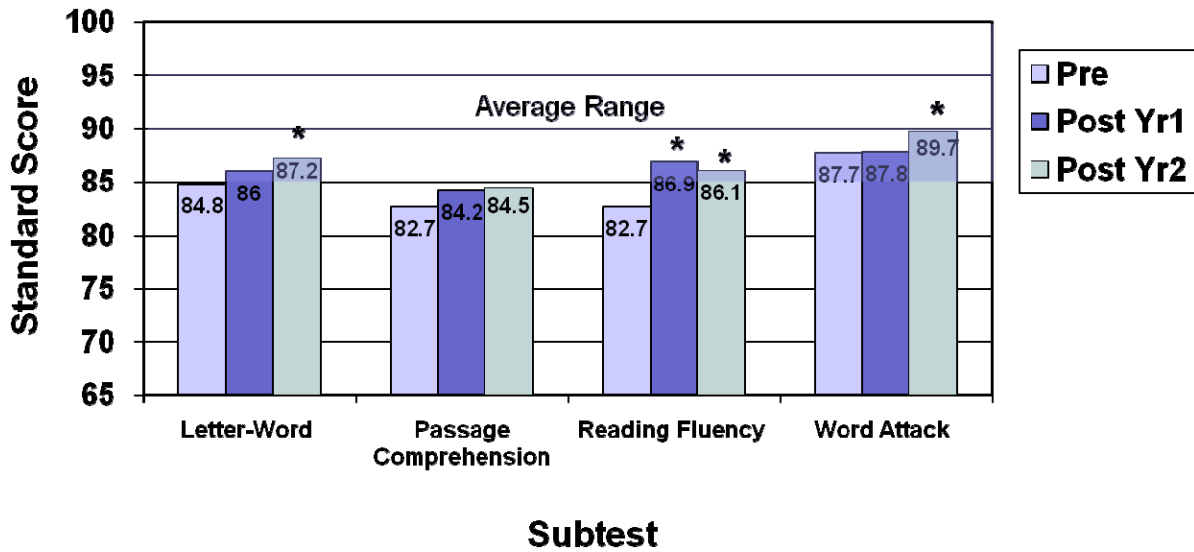
Figure 2. Changes in standard scores pre- to post-Year 1: Cohort 2



Year 2 Outcomes

Only thirty-three students from Cohort 1 completed a second year of HillRAP as well as a third WJ-III reading assessment at the end of 09-10 (after two years of intervention). These students made gains consistent with those of the students who received only one year of HillRAP, but in addition, they improved significantly on the Word Attack subtest (see Figure 3). Passage Comprehension did not improve significantly even after two years of intervention, suggesting that older students may require a longer intervention period for higher level skills to show substantial improvement.

Figure 3. Changes in standard scores pre- to post-Year 2: Cohort 1



*statistically significant change from Pretest ($p < .05$)

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. These analyses were completed only for Year 1 outcomes for the 92 students in Cohort 1 and the 50 students in Cohort 2. The sample of students receiving two years of HillRAP was too small for more in depth analyses. For a detailed discussion of analyses and the demographic predictors of performance on the WJ-III, including graphs of findings, see Appendices B, C, and D.

Two variables were found to be significant predictors of improvement on the WJ-III for students in both cohorts: pre-intervention achievement level and full-scale IQ score. Replication of findings between the two cohorts strengthens our confidence that these are true predictors of HillRAP outcomes.

For both cohorts, students with larger delays in reading at baseline showed the largest improvements on three of the four subtests of the Woodcock-Johnson. This suggests that HillRAP is appropriately targeted towards student with significant reading challenges.

Full-scale IQ emerged as a predictor of improvement in Word Attack for both cohorts. Students with above-average IQs made the most improvement. Children with higher IQs have

more cognitive aptitude for learning, and thus may be more responsive to intervention. They may also have stronger abstract thinking skills, which are necessary to apply phonics skills to the made-up words in Word Attack. In the current sample, children with the higher IQs likely represent those with traditional learning disabilities, as opposed to generally lower cognitive resources. Thus, HillRAP may be most effective in bringing students up to a level of functioning congruent with their aptitude.

Several predictors of WJ-III improvement were identified only in the Cohort 1 sample. This may be due to underlying differences in the two cohorts or to the smaller sample size in Cohort 2. On the other hand, it may indicate that these are spurious findings rather than true predictors of intervention outcomes. Though useful for hypothesis building, the validity of these predictors will need to be tested in future samples of students.

Grade level predicted outcomes on the Letter-Word Identification subtest of the WJ-III. Cohort 1 students in lower grades made larger gains on this subtest, emphasizing the need to start HillRAP as early as possible in a student's academic career in order to maximize achievement gains. When age was examined over and above the effect of grade, younger students showed more progress on Word Attack than did older students, with the students above 13 years of age making progress at a slower rate than their peers. 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. Similarly, older students are entering puberty and are in a different developmental phase, with more emphasis on peer interaction than on academics.

In the Cohort 1 group, females made significantly more gains on the Reading Fluency subtest than did males, which may reflect the difference in reading attitudes in girls and boys. Several large-scale studies have documented that girls as a group are more positive about reading than are boys, regardless of reading ability (e.g., McKenna, Kear & Ellsworth, 1995).

Socioeconomic status (SES), as measured by free-reduced lunch eligibility, was found to be related to gains made in Word Attack and Passage Comprehension for Cohort 1, with students of lower SES (with free-reduced lunch eligibility) progressing at a slower rate than same-grade peers. These findings may be driven by the predominance of single parenthood in low SES families (Caldas, 1999). These families may not spend as much time reading together and building comprehension skills. Also, these students may have a narrower range of experience and exposure to ideas, analytical discussions and vocabulary, which limits comprehension skills (Chall, 1983; Beals & Detemple, 1993). However, studies have found that small class sizes can ameliorate the negative effects of poverty on achievement (e.g., Howley & Bickel, 1999). This suggests that the low student-teacher ratio used in HillRAP should be an effective strategy in working with a low-SES student sample. It may be that more time is required to achieve

equivalent gains in complex reading skills for this higher-risk group; if this effect is identified in other samples, it should be monitored over time to see if it diminishes.

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.

No teacher characteristics were significant predictors of improvement on any subtest. Indeed, because of the high level of proficiency in HillRAP implementation, there was little variability in Overall Proficiency ratings. Likewise, attendance did not significantly predict improvement on any subtest, most likely because of the high attendance rates (average attendance across cohorts and years ranged from 93 to 103 hours).

End-of-Grade Assessment Results

The Department of Public Instruction reports EOG reading performance 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.

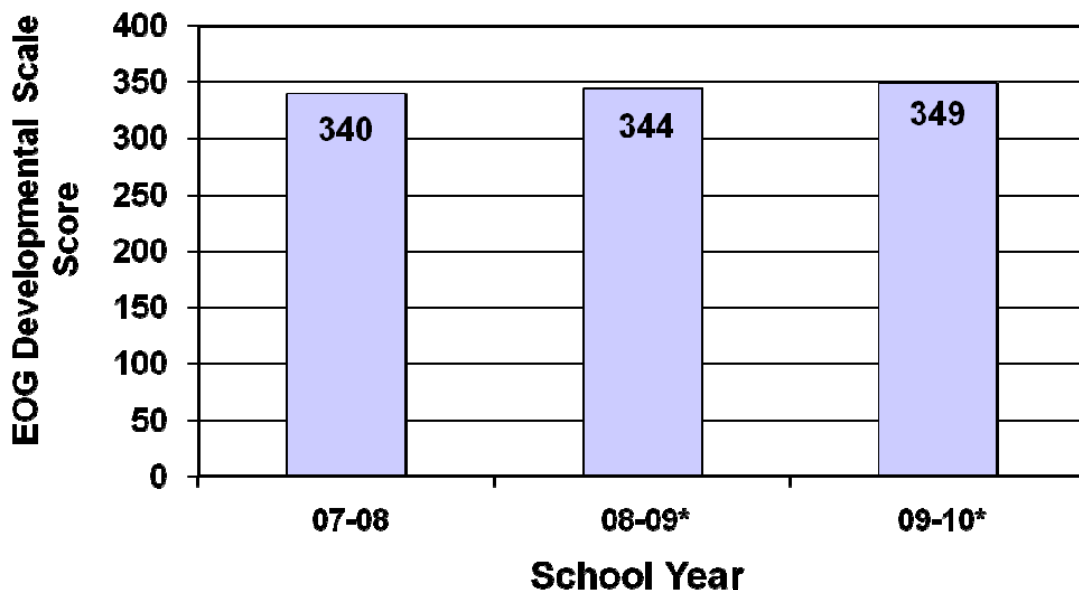
Developmental Scale Scores

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.

Sixty-nine of the 92 students from the Cohort 1 HillRAP sample completed the reading EOG assessment in both 07-08 (prior to the start of the HillRAP intervention) and 08-09 (after one year of intervention). The remainder either completed the alternative Extend2 assessment or were not in the Davie 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 343 to 347.

Twenty-four of these students completed a second year of HillRAP and took the reading EOG in 07-08 (prior to the start of the HillRAP implementation), 08-09 (after one year of intervention), and 09-10 (after two years of intervention). The remainder either completed the alternative Extend2 assessment or did not participate in HillRAP during 09-10 (many moved to high school). Those participating in Year 2 of HillRAP had the slightly lower EOG scores in 07-08 and 08-09 than did the larger sample from Year 1, suggesting this subset may have started even further behind. Change in EOG scores over time is equivalent between the two groups, however; both groups showed identical gains after one year of intervention. Again, as expected, developmental scale scores increased significantly during the second year of HillRAP. This improvement is depicted in Figure 4.

Figure 4. Change in EOG developmental scale scores Pre- to Post-Yr2



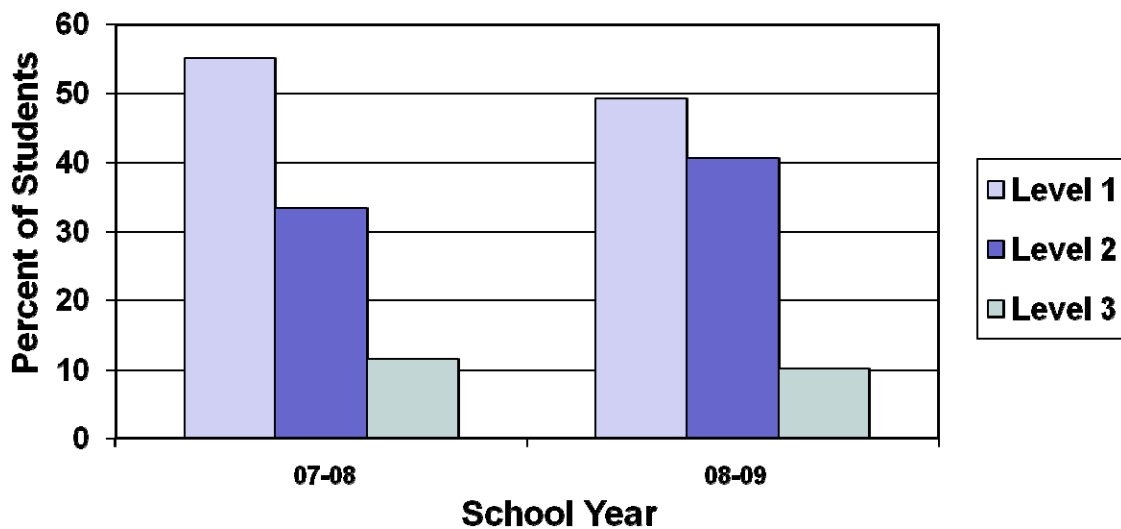
*statistically significant difference ($p < .01$)

Forty-one students from the Cohort 2 HillRAP sample completed the reading EOG assessment in 08-09 (prior to the start of the HillRAP intervention) and 09-10 (after one year of intervention). Consistent with findings from Cohort 1, students' developmental scale scores significantly increased after one year of HillRAP, moving from an average of 340 to 346.

Achievement Levels

Achievement levels provide a broader view of student performance. Students are grouped into 4 levels of performance based on their scale scores, with level 1 showing the lowest performance and level 4 showing the highest performance. Levels 3 and 4 are considered “passing” scores. Because achievement levels are based on developmental scale scores, which are not standardized, they are more variable between years due to changes in test difficulty, scoring standards, etc. The reading EOG was re-normed in 07-08, resulting in a shift in scale scores from the previous years; for data based on scale scores, only the years after re-norming can be compared. As seen in Figure 5, reading achievement levels improved significantly between the 07-08 and 08-09 school years ($\chi^2(4) = 28.1, p < .0001$). This is largely the result of students moving from level 1 to level 2.

Figure 5. Changes in EOG Reading Achievement Levels: Cohort 1



When achievement levels were examined for the twenty-four students who received HillRAP for two years, positive changes were found, though they were not significant. Following a second year of intervention, students shifted from a majority in achievement level 1 to a majority in achievement level 2.

Similar positive but nonsignificant changes in achievement levels were also found for the Cohort 2 students who took the EOG scores in 08-09 (prior to the intervention) and 09-10 (at the end of the first year of HillRAP). Interestingly, these students showed worsening achievement levels in the year prior to HillRAP participation. Following a year of HillRAP, however, 17% of the students moved from achievement levels 1 and 2 to achievement level 3.

C-Scores

In contrast to the scores presented above, c-scores are standardized scores that allow us 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:

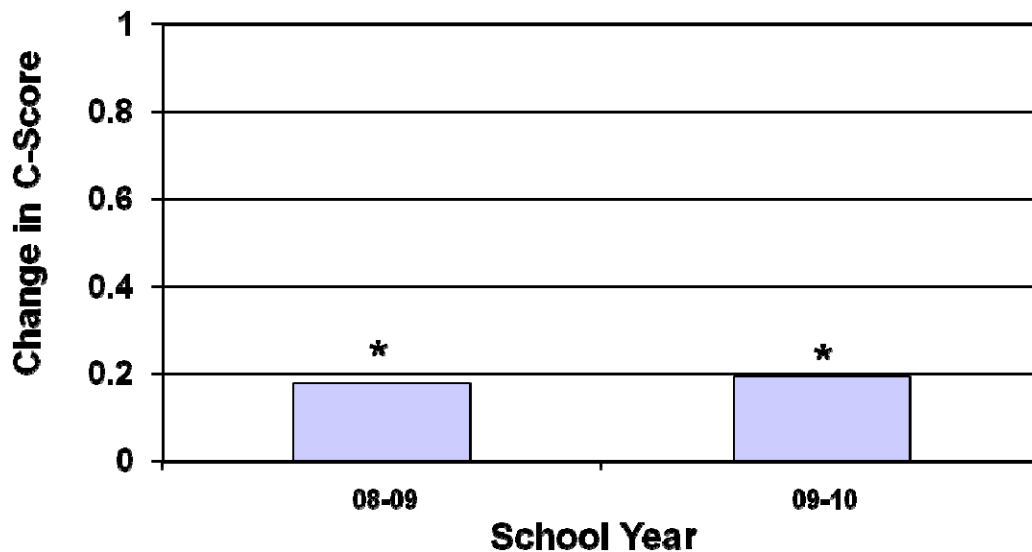
$$\text{c-score} = \frac{\text{Student score} - \text{State-level mean for the norming year}}{\text{State-level standard deviation for the norming year}}$$

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.

The 69 Cohort 1 students with reading EOG scores in both 07-08 and 08-09 had an average c-score of -1.2 (SD = 0.7) in 06-07, indicating that they were performing an average of 1.2 standard deviations below the state mean for students in their grades. In 07-08 (the year immediately prior to HillRAP implementation), this improved to an average c-score of -1.1. HillRAP students showed larger improvements in EOG reading c-scores between the 07-08 and 08-09 school years, after the implementation of HillRAP, increasing an average of 0.16 points (SD = 0.7). This post-HillRAP score increase was statistically significant ($t(68) = 2.03, p < .05$). These findings suggest that students in Cohort 1 were already making gains in reading achievement prior to the HillRAP intervention, but HillRAP participation may have accelerated their improvement. Not surprisingly given their initial level of deficits, however, they continued to perform below the level of the average student in the state.

Students with EOG c-scores in 09-10 showed additional improvements after the second year of HillRAP ($n = 24$). Figure 6 depicts changes to EOG c-scores from baseline (07-08) for students who completed two years of HillRAP and took EOGs for all years of the analysis.

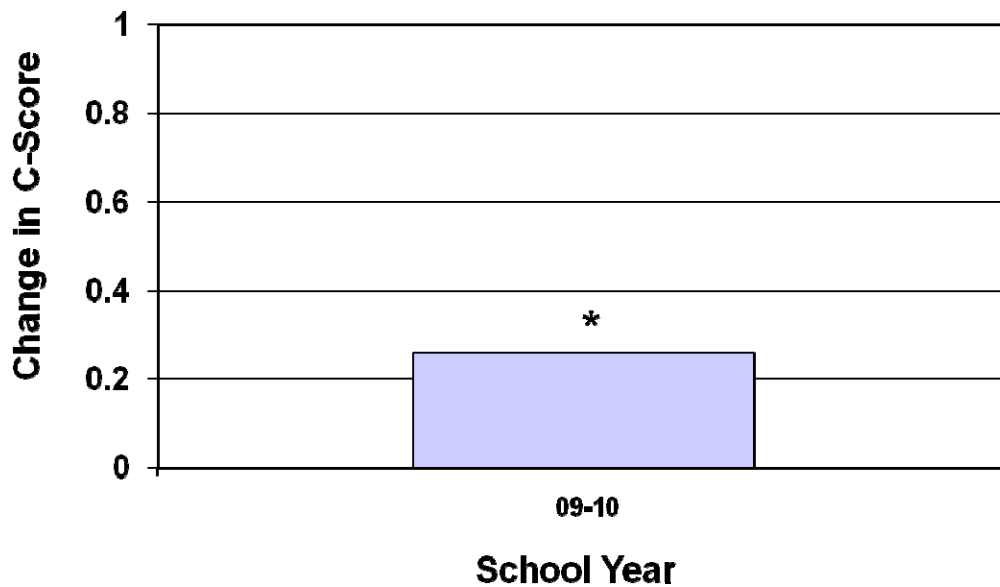
Figure 6. Changes in EOG c-scores from Baseline: Cohort 1 (n = 24)



*statistically significant difference ($p < .01$)

Cohort 2 HillRAP students showed a slightly different but overall consistent pattern on c-scores (see Figure 7). Prior to the start of HillRAP, the EOG reading scores of Cohort 2 were declining, moving from -1.1 in 07-08 to -1.2 in 08-09. After a year of intervention, their scores improved significantly ($t(40) = 2.3, p < .05$). These findings suggest that students in the Cohort 2 reversed their pattern of worsening performance and instead began to close the gap between themselves and their grade-mates.

Figure 7. Changes in EOG Reading c-scores from Baseline: Cohort 2 (n = 41)



*statistically significant difference (p < .05)

Growth Scores

Finally, the Department of Public Instruction 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:

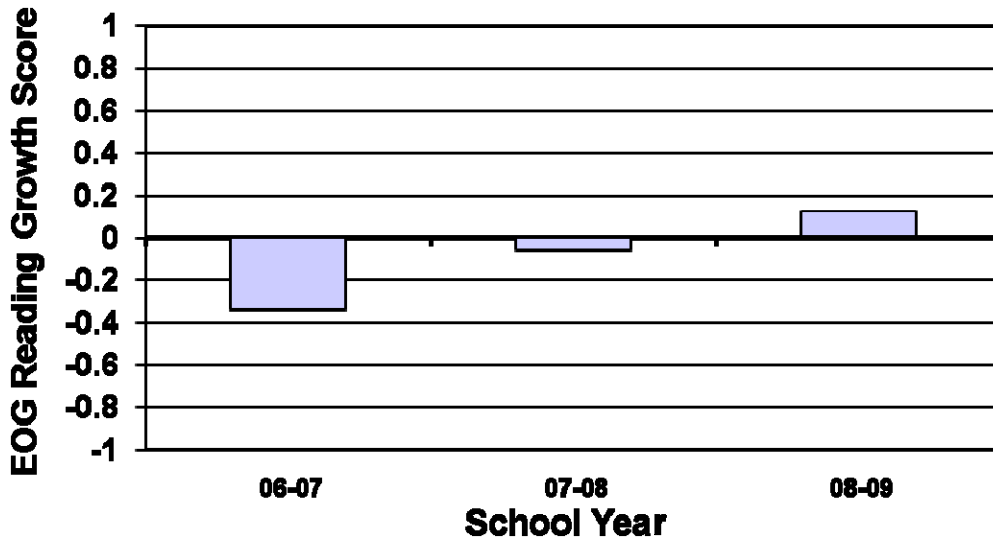
$$\text{Predicted c-score} = \text{average of past 2 c-scores} * 0.92$$

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.

Growth scores for 06-07 through 08-09 in Cohort 1 are presented in Figure 8. Prior to HillRAP, students performed slightly more poorly than predicted, with negative growth scores indicating worsening EOG performance. Following HillRAP, students performed slightly better than predicted, with a positive growth score (though this change is not significant). Growth score results are consistent with c-score results for Cohort 1; students began showing improvement prior to intervention (from 06-07 to 07-08), but HillRAP boosted scores into the positive growth

range (08-09). Students with a second year of HillRAP showed even more improvement, with growth scores of 0.18.

Figure 8. Changes in EOG Growth Scores: Cohort 1



Findings for Cohort 2 were consistent with those for Cohort 1. In the year prior to HillRAP intervention, their average growth score was -0.22, indicating worsening performance. Following HillRAP, the average growth score improved to 0.08.

Based on the precision of the EOG assessment, growth scores between -0.5 and 0.5 are considered to be within the expected range (i.e., only growth scores higher than .5 are considered to represent “true” growth in performance). Thus, while there is a measured improvement following HillRAP, the change is quite small on average and does not move students into the “clearly improved” range.

Taken together, EOG scores indicate that students did show improvement on standardized statewide testing following participation in the HillRAP program. Improvements were small, however, and students remain behind their peers as a group. Importantly, there was considerable variability from student to student on changes in EOG scores. Across cohorts, over half (58.2%) improved c-scores by at least 0.1 after the first year of HillRAP, whereas 10.0% remained at a consistent c-score level and 31.8% got lower c-scores in 08-09. Likewise, 29.6% had growth scores greater than 0.5, while 51.9% had growth scores in the expected range (-0.5 to 0.5) and 18.5% had growth scores less than -0.5.

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 the

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 their 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 is reduced.

Demographic and implementation predictors of performance on the EOG

To examine why some students improved while others did not, we explored possible predictors of improved EOG scores. The following demographic variables were examined: student sex, race, and English language proficiency; free/reduced lunch eligibility; exceptional children status and classification; full-scale IQ and discrepancy between IQ and baseline achievement; and history of grade retention. None of these variables significantly predicted change on the EOG reading assessment for students in Cohort 1. In contrast, Cohort 2 students showed greater improvement in scores as grade level increased (contrary to findings for the WJ-III). Moreover Cohort 2 students not eligible for free/reduced lunch showed greater improvement in EOG scores (see Appendix E for a more detailed discussion).

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. None of these variables significantly predicted change on the EOG reading scores for either Cohort.

Summary and Conclusions

The purpose of this evaluation was to examine the effectiveness of HillRAP in helping middle school students whose reading skills are compromised. More specifically, we addressed the following questions:

1. *What are the effects of HillRAP instruction on reading achievement over time?*

Two samples of HillRAP middle school students (Cohort 1 and Cohort 2) showed significant improvements in reading achievement on a state-wide reading achievement test (i.e., EOG standardized c-scores, achievement levels, developmental scale scores, and growth scores) after one year of intervention despite the fact that they started with significant delays in reading. Cohort 1 students' EOG scores were already improving before the intervention started, but HillRAP participation seems to have enhanced their improvement. Cohort 2 showed declining EOG scores prior to HillRAP participation, but

significant improvements following HillRAP. For the smaller number of students who received HillRAP for two years and took the EOGs in 07-08, 08-09, and 09-10 (n = 24), the same improvement trend was obvious, with even larger gains after Year 2.

On a nationally-normed achievement test (WJ-III), the biggest gains after one year of intervention were on the Reading Fluency subtest, followed by Letter-Word Identification. These gains suggest that after one year of HillRAP, students are better able to read individual words as well as sentences. Moreover, they are making these gains at a faster rate than their same-grade peers, thus closing the gap between themselves and the average student. For the smaller group of students who received the intervention for two years and were tested pre-intervention, post-year 1 and post-year 2 on the WJ-III (n=33), significant results were found for Letter-Word Identification, Reading Fluency and Word Attack. Passage Comprehension did not improve significantly, as was expected after the second intervention year, though a gradual upward trend was noted. Middle school students may need even longer interventions to show marked improvements in more complex reading skills.

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?*

There was considerable variability on both the EOG test scores and the WJ-III scores, suggesting that HillRAP may be more appropriate for certain groups of students. A number of significant predictors of reading improvement emerged after one year of HillRAP. Looking across both cohorts of students, two findings were consistent:

- Pre-intervention WJ-III scores predicted improvement across subtests. Students with more substantial initial delays made gains at a faster rate than did students with achievement closer to the average.
- For the subset with IQ scores available, IQ was found to be a significant predictor of improvement on Word Attack. Students with higher IQs improved their Word Attack scores at a rate faster than their peers, possibly suggesting that this is a group of youth who have the cognitive resources to achieve but struggle due to a fairly circumscribed learning disability. Once these students are taught ways to compensate for their disability, their phonics skills may increase rapidly. Word Attack assesses generalized phonics skills applied to made-up words rather than sight words. This requires more complex abstract thinking rather than sheer memorization, thus IQ may play a stronger role for scores on this subtest.

Other predictors of reading achievement showed less consistent patterns across cohorts, but deserve further study.

- In general, and consistent with findings of other HillRAP evaluations, younger students showed the strongest gains, suggesting that HillRAP will be most beneficial when started early in a student's academic career. Specifically, for Cohort 1, students in lower grades made more improvement in Letter-Word Identification, and younger students made more gains in Word Attack than did older students. In contrast to these findings, however, Cohort 2 students in *higher* grades showed more gains on EOG scores. The effects of grade level on reading improvement will need further study with larger samples before firm conclusions can be drawn.
- Females increased their performance on Reading Fluency at a faster rate than did males, but this was only true in Cohort 1. Other HillRAP evaluations have found similar, though nonsignificant, effects for Reading Fluency, though males have outperformed females on other reading subtests of the WJ-III.
- Youth coming from higher socioeconomic backgrounds (i.e., not eligible for free/reduced lunch) made significantly higher gains on Word Attack and Passage Comprehension than did students from more underprivileged circumstances (Cohort 1) and showed more improvement on EOG scores following one year of intervention (Cohort 2). For Letter-Word Identification and Reading Fluency, however, the reverse pattern was found (though it was not significant). It is possible that youth from underprivileged homes require longer periods of remediation to develop the more complex skills of comprehension and generalized phonics.

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?*

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.

4. *What are the effects of attendance, teacher characteristics, and model fidelity on reading achievement?*

Teacher characteristics, implementation proficiency ratings, and student attendance did not predict changes in reading achievement. Teacher proficiency and student attendance were both consistently high across the sample, so there was minimal variation for observing effects on outcomes.

Results from the evaluation of HillRAP in the Davie County middle schools suggest that this program improved the reading proficiency of struggling readers and began to close the achievement gap between these students and their peers on a state-mandated achievement test (i.e., EOG) and a nationally-normed achievement test (i.e., WJ-III). These effects are stronger for students with more significant initial delays, higher IQs and who are from families with greater financial resources; the program did not have the same level of benefit for students with significant cognitive limitations or those who receive free/reduced lunch, particularly in the more complex areas of generalized phonics and comprehension. This pattern should be monitored in future evaluations to determine whether modifications for these subgroups should be explored. The lack of improvement in Passage Comprehension suggests that older students who start with serious deficits in reading might need longer interventions for such improvements to appear.

While statistically significant, the gains in reading achievement that were found were modest and highly variable. Given the lack of a comparison group, it is impossible to know what these students' scores would have looked like without intervention; it is possible that they would have fallen further behind, but also possible that they would have maintained pace with their classmates or even shown some improvement. Thus we cannot make strong conclusions about HillRAP as the agent of change in the observed improvements.

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 more definitively conclude that their improvement was attributable to HillRAP. 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.

References

- Beals, D., & Detemple, J. (1993). Home contributions to early language literacy development. In D. Leu & C. Kinzer (Eds.), *Forty-second yearbook of the National Reading Conference* (pp. 207-216). Chicago: National Reading Conference.
- Caldas, S. J. (1999). Multilevel examination of student, school, and district-level effects on academic achievement. *Journal of Educational Research, 93*, 91-101.
- Capraro, M. M., Capraro, R. M., & Wiggins, B. B. (2000). *An investigation of the effect of sex, socioeconomic status, race and grades on standardized test scores*. Paper presented at the meeting of the Southwest Educational Research Association, Dallas, TX.
- Chall, J. S. (1983). *Stages of reading development*. New York: McGraw-Hill.
- Entwisle, D. R., & Alexander, K. L. (1996). Family type and children's growth in reading and math over the primary grades. *Journal of Marriage and the Family, 58*, 341-355.
- Good, R. H., & Jefferson, G. (1998). Contemporary perspectives on Curriculum-Based Measurement validity. In M. R. Shinn (Ed.), *Advanced applications of Curriculum-Based Measurement* (pp. 61-88). New York: Guilford.
- Howley, C., & Bickel, R. (1999). *The Matthew project: National report*. Randolph, VT: Rural Challenge Policy Program.
- Mather, N., & Woodcock, R. W. (2001). Examiner's Manual. *Woodcock-Johnson III Tests of Achievement*. Rolling Meadows, IL: Riverside Publishing.
- McKenna, M. C., Kear, D. J., & Ellsworth, R. A. (1995). Children's attitudes toward reading: A national survey. *Reading Research Quarterly, 30*, 934-956.
- National Reading Panel (2000). *Teaching Children to Read: An Evidence-Based Assessment of the Scientific Research Literature on Reading and its Implications for Reading Instruction*. Rockville, MD: National Institute of Child Health and Human Development. Retrieved August 27, 2009, from <http://www.nichd.nih.gov/publications/nrp/smallbook.cfm>.
- NCDPI Division of Accountability Services/North Carolina Testing Program (2007). *Assessment Brief: Understanding North Carolina End-of-Grade Testing*. Retrieved September 2, 2009, from <http://www.ncpublicschools.org/docs/accountability/testing/eog/AssessmentBriefEOG04.pdf>
- Tindal, G., Marston, D., & Deno, S. L. (1983). *The reliability of direct and repeated measurement* (Research Rep. 109). Minneapolis, MN: University of Minnesota Institute for Research on Learning Disabilities.

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 possible 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.
- 95.2% were fully proficient in asking questions suitable in level, type, and structure.
- 100% charted responses accurately.
- 95.2% provided an appropriate pace for each student.
- 100% made strong eye contact with the students.
- 90.5% were observed to correct incorrect responses. It is unclear whether the remaining 9.5% failed to correct responses, or simply did not have any incorrect responses.
- However, 38.1% did not have students practice drill questions correctly and ask them again. Only 47.6% were fully proficient at this component.
- 100% used positive constructive reinforcement.
- 100% used drill effectively for practice instead of instruction.

Phonological Awareness

- Only three teachers were observed to cover phonological awareness. It is likely, given the students' ages, that this skill was already mastered by most students.
- Of those who covered phonological awareness, 100% were fully proficient in all components.

Word Attack

- 100% chose suitable word lists.
- 85.7% introduced students to the pattern, rule, and/or syllable type of each word list from which they read.
- 90.5% were fully proficient at providing effective pacing of instruction.

- 100% gave positive constructive reinforcement.
- 95.2% had each student read 3-4 words before moving on to the next student.
- 90.5% kept students engaged in some stage of this component at all times (decoding, marking syllables, practicing).
- 100% marked the students' responses.

Fluency

- 90% were fully proficient in handling the fluency component efficiently and systematically.
- 95% were fully proficient in accurately charting test results.
- 100% selected the correct timed test lists.
- 95% corrected errors and practiced correct responses.
- 85% were fully proficient at setting and monitoring goals.
- 100% gave positive constructive reinforcement.
- Timed reading from a text was observed for only three teachers, but is only intended to be implemented once per week. The three observed for timed reading from a text were fully proficient.

Vocabulary

- 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):

- Reinforcing vocabulary words through drill
- Teaching vocabulary words within the context of the reading text
- Teaching prefixes, suffixes, and/or roots
- Identifying tiered words

Reading Comprehension

- 100% chose appropriate text.
- 100% spent equitable time with the students.
- 100% asked comprehension questions at the appropriate skill level.
- 100% offered questions in an oral and written form.
- 100% gave a variety of comprehension questions (literal, inferential, prediction, sequencing, etc.).
- 100% 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

- 100% had materials organized and ready for class.
- 100% demonstrated proper posture and articulate questions and directions.
- 100% made smooth transitions between class components.
- 100% started class promptly.
- 85.7% had students on task throughout the class period.
- Only 9.5% needed/used a point/reward system, but in these cases it was used effectively.
- Only 19.1% implemented multi-sensory techniques.

Appendix B

Cohort 1, Year 1 demographic predictors of performance on the WJ-III

Individual change over time in Woodcock-Johnson scores showed considerable variability. For Cohort 1, Year 1, 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 relative to their peers. Reading Fluency was clearly the area where the most students showed improvement, as well as where the largest average improvement occurred.

Table B1. Individual-level changes in standard scores pre- to post-Year 1 (08-09)

WJ-III Subtest	Improved by 2 or more points	Maintained standard score	Declined by 2 or more points
Letter-Word	46.7%	21.8%	31.5%
Passage Comprehension	39.1%	20.7%	40.2%
Reading Fluency	65.2%	20.7%	14.1%
Word Attack	37.0%	23.9%	39.1%

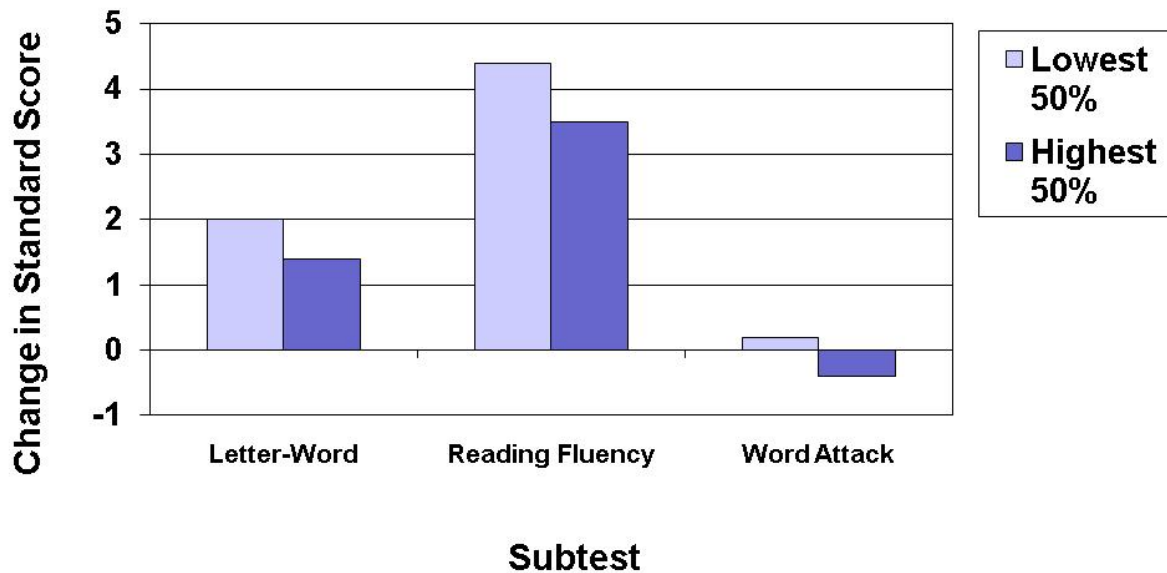
When results show such a high level of variability, it is helpful to explore possible reasons for the differences in change over time. To accomplish this, we examined whether the impact of the HillRAP program varied by key student characteristics. The student variables examined were: pre-intervention achievement scores, grade, age, sex, free/reduced lunch eligibility status, IQ score, grade retention history, LEP status and Exceptional Child status. General linear models predicting standard score change over time were conducted, controlling for school, days 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 over time on three of four subtests (Letter-Word: $F(1, 84) = 6.8, p < .01$, Reading Fluency: $F(1, 84) = 5.7, p < .05$, and Word Attack: $F(1, 82) = 18.3, p < .0001$). As expected, students with lower initial standard scores on these subtests demonstrated more gains in these areas than did 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 Letter-Word, Reading Fluency, and Word Attack were 86, 83, and 88, respectively. Figure B1 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 Letter-Word and Reading Fluency, both groups showed improvement, though students starting below the median improved more. For Word Attack, students starting below the median improved slightly, while those above the median lost ground relative to their peers. These results suggest that the HillRAP intervention is appropriately targeted to students with the most substantial delays in reading achievement, but may have benefits for students with less significant delays as well.

Figure B1. Pre-intervention achievement scores as a predictor of improvement

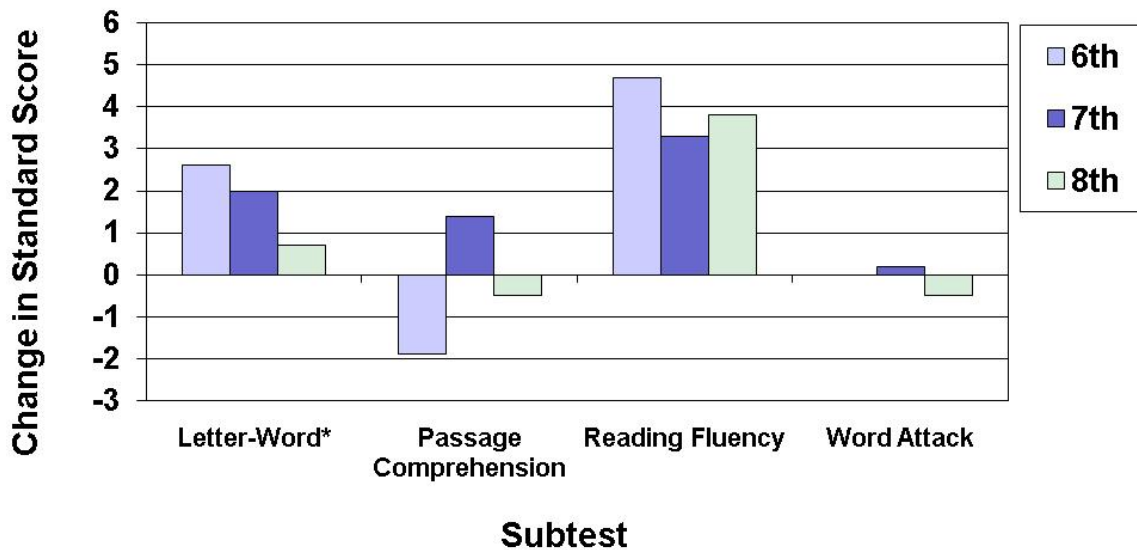


Grade Level

Student grade level significantly predicted improvement on the Letter-Word Identification subtest ($F(1, 84) = 4.7, p < .05$). Students of all three grades improved at a rate faster than expected; however, students in lower grades made larger gains. 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. Similarly, older students are entering puberty and are in a different developmental phase, with more emphasis on peer interaction than on academics. This finding emphasizes the need to start HillRAP as early as possible in a student's academic career in order to maximize achievement gains.

Figure B2 depicts the findings by grade level for all four subtests, but only Letter-Word was significantly affected. Results for all predictor variables are presented in a similar way to provide complete information, however only subtests noted as significant demonstrate statistically significant effects that are likely to be reliable.

Figure B2. Grade as a predictor of Letter-Word Identification

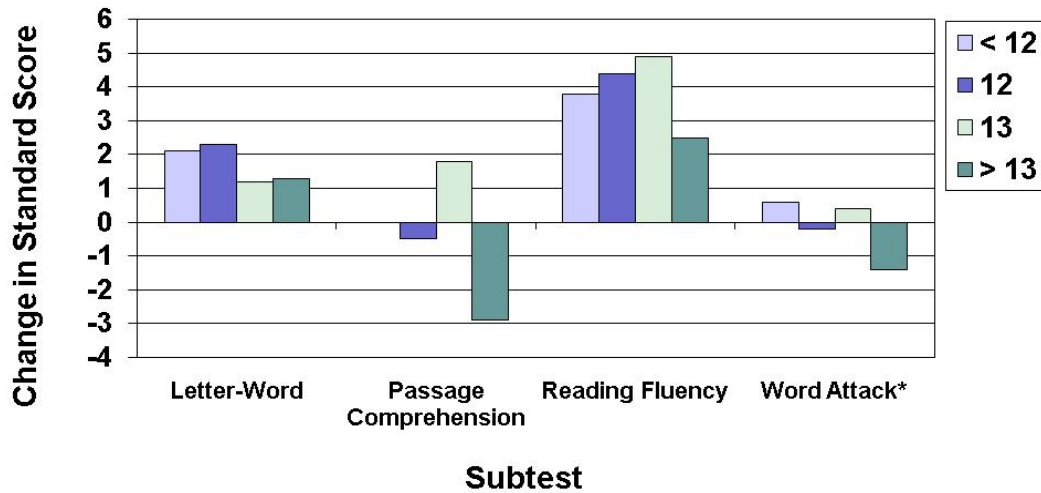


*statistically significant ($p < .05$)

Age

Controlling for grade level, the continuous variable for age was included in the analysis to determine whether it was significantly predictive of reading improvement. Indeed, age did predict improvement in Word Attack above and beyond the effects of grade ($F(1, 82) = 5.7, p < .05$). For a clear depiction of this relationship we used the following age categories in Figure B3: less than 12 years of age, 12 years old, 13 years old, and older than 13 years of age. As with grade, the younger students made more progress in Word Attack across the year than did older students, with the students above age 13 making progress at a rate slower than their peers. Again, this may be related to increasing frustration with reading difficulties, more ingrained reading attitudes and beliefs, or shifting priorities accompanying the onset of puberty.

Figure B3. Age as a predictor of Word Attack

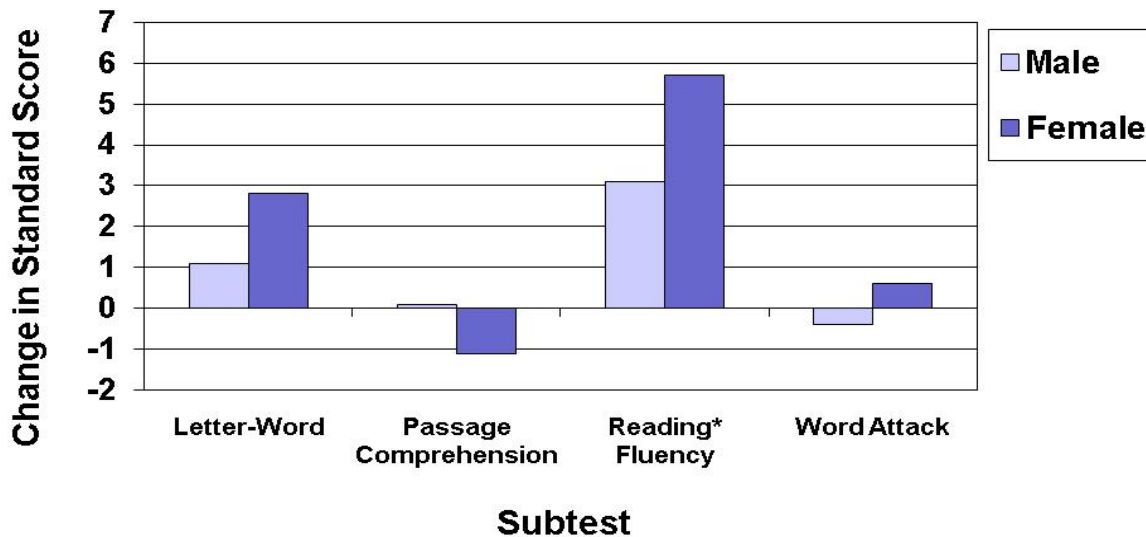


*statistically significant (p < .05)

Sex

Student sex (male vs. female) significantly predicted improvement in Reading Fluency ($F(1,84) = 4.32, p < .05$). As shown in Figure B4, all students improved on Reading Fluency at a faster rate than their peers. Females, however, made more gains than did males. This may be partly explained by the differences in reading attitudes in girls and boys. Several large-scale studies have documented that girls as a group are more positive about reading than are boys (e.g., McKenna, Kear, & Ellsworth, 1995). This is true across grade levels and is unrelated to actual reading ability. It is likely that this difference in attitudes results in different levels of effort and accomplishment in reading interventions such as HillRAP.

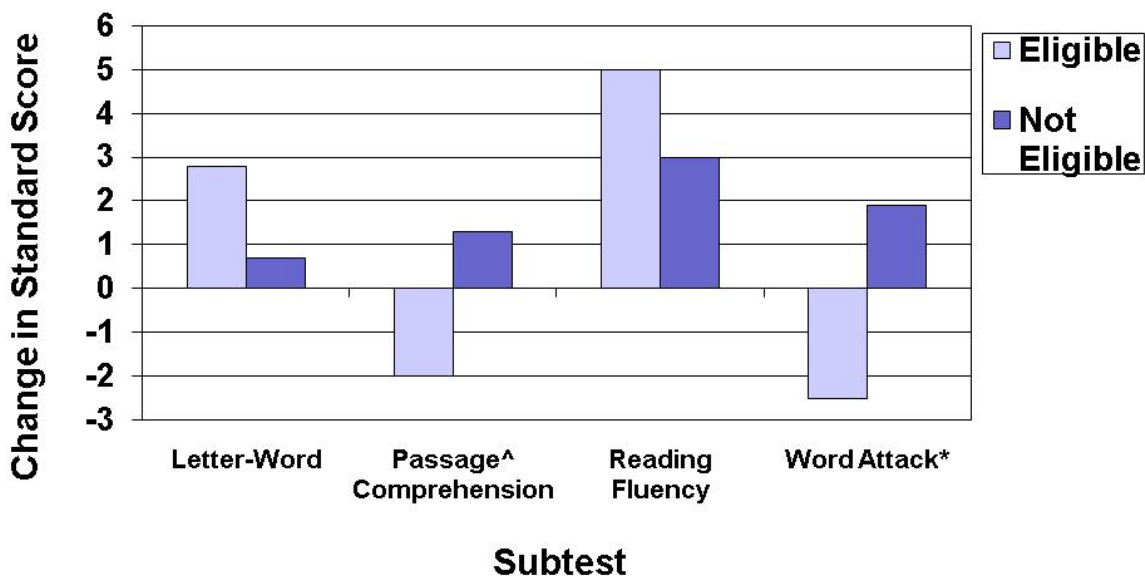
Figure B4. Sex as a predictor of Reading Fluency



Free/reduced lunch eligibility

Free/reduced lunch eligibility was used as an indicator of socio-economic status (SES). As shown in Figure B5, students of lower SES (with free/reduced lunch eligibility) progressed at a slower rate than their same-grade peers in Passage Comprehension ($F(1,83) = 3.3, p = .07$) and Word Attack ($F(1,82) = 12.1, p < .001$), whereas students of higher SES made gains at a faster rate than their peers. These are the two subtests that showed no overall improvements from pre- to post-HillRAP. Clearly, however, students with higher family income were able to show improvement in these areas. This finding is consistent with a large body of literature showing a strong relationship between SES and academic achievement (e.g., Capraro, Capraro, & Wiggins, 2000). Interestingly, though findings were not significant, students eligible for free/reduced lunch showed *stronger* gains in Letter-Word identification and Reading Fluency. These subtests are more concrete and may improve with memorization of sight words and practice reading short phrases, whereas Passage Comprehension and Word Attack require more abstract thinking and generalization of phonics skills to made-up words.

Figure B5. Free/reduced lunch eligibility as a predictor of Passage Comprehension and Word Attack



*statistically significant difference ($p < .05$)

[^]trend difference ($p < .10$)

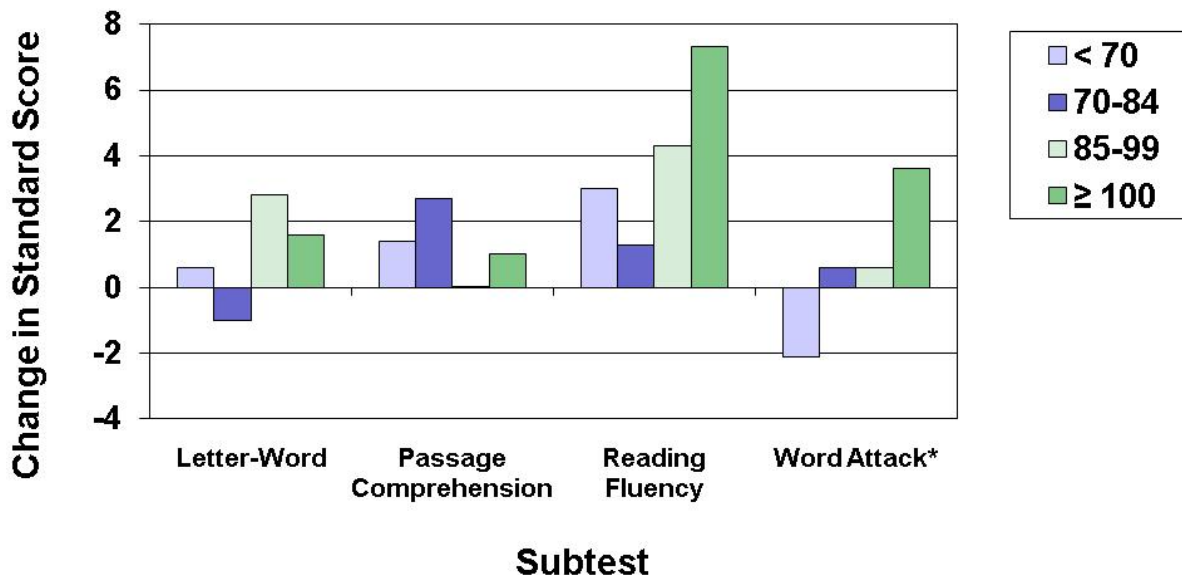
In part, findings of poorer performance in Passage Comprehension and Word Attack may be driven by the predominance of single parenthood in low SES families (Caldas, 1999). With only one parent, there is less time and energy available for supporting and emphasizing learning. These families may not spend as much time reading together and building comprehension skills. Also, these students may have a narrower range of experience and exposure to ideas, analytical discussions and vocabulary, which limits comprehension skills (Chall, 1983; Beals & Detemple, 1993). However, studies have found that small class sizes can ameliorate the negative effects of poverty on achievement (e.g., Howley & Bickel, 1999). This suggests that the low student-teacher ratio used in HillRAP should be an effective strategy in working with a low-SES student sample. It may be that more time is required to achieve equivalent gains in complex reading skills for this higher-risk group; this effect should be monitored over time to see if it diminishes.

Full-scale IQ score

A student's IQ score as assessed by the school system was also found to be a significant predictor of Word Attack performance ($F(1,58) = 10.74, p < .01$; note that IQ was only available for about two-thirds of the sample). To depict this relationship, student IQ scores were divided into four categories: IQ score less than 70 (> 2 SDs below the mean), between 70 and 84 (1-2 SDs below the mean), between 85 and 99 (0-1 SDs below the mean), and IQ score above 100 (at or above the mean). Children with an IQ of 70 or above made progress at a faster rate than their peers (see Figure B6), while those with IQs below 70 lost ground relative to their peers. Those with above-average IQs made the most improvement. Children with higher IQs have more

cognitive aptitude for learning, and thus may be more responsive to intervention. They may also have stronger abstract thinking skills, which are necessary to apply phonics skills to the made-up words in Word Attack. In the current sample, children with the higher IQs likely represent those with traditional learning disabilities, as opposed to generally lower cognitive resources. Thus, HillRAP may be most effective in bringing students up to a level of functioning congruent with their aptitude.

Figure B6. IQ as a predictor of Word Attack



*statistically significant difference (p < .05)

Achievement is only expected to reach a level commensurate with one’s IQ. It would make sense, then, that students with larger discrepancies between IQ and achievement would show greater gains in response to the HillRAP intervention. To examine this, we included child IQ-achievement discrepancy as a predictor of gains on the WJ-III. Surprisingly, though, this discrepancy was not a significant predictor. Indeed, 30-40% of the sample scored higher than their IQ level on each WJ-III subtest following HillRAP implementation.

Implementation predictors of performance on the WJ-III

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), teacher type (EC or not), and average Overall Proficiency scores across the four observations. The student-level variable added for this multi-level model was hours of attendance.

No teacher-level variables were significant predictors of improvement on any subtest. Indeed, because of the high level of proficiency in HillRAP implementation, there was little variability in Overall Proficiency ratings. These findings suggest that HillRAP instruction was equally effective for all teachers regardless of education level or teaching history. This speaks to the quality of HillRAP training as well as the strong portability of the program itself due to its concrete structure and implementation guidelines.

Likewise, attendance did not significantly predict improvement on any subtest. In Davie County, attendance was quite high across participants, with Cohort 1 students attending an average of 93 hours in Year 1 and 103 hours in Year 2, and Cohort 2 students attending an average of 102 hours. Only 12 students across the cohorts attended fewer than 75 hours in a year. This small variability in attendance makes it difficult to identify attendance effects on performance, even if they exist.

Appendix C

Cohort 1, Year 2 variability in WJ-III outcomes

Individual change over time once again showed considerable variability following year 2 of HillRAP. Table C1 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 relative to their peers by the end of Year 2. Reading Fluency showed the most improvement over all W-J subtests. Given the small sample size, predictors of WJ-III improvement were not examined for Year 2.

Table C1. Individual-level changes in standard scores pre- to post-Year 2 (09-10)

WJ-III Subtest	Improved by 2 or more points	Maintained standard score	Declined by 2 or more points
Letter-Word	60.6%	9.1%	30.3%
Passage Comprehension	57.6%	6.0%	36.4%
Reading Fluency	59.4%	18.8%	21.9%
Word Attack	51.5%	30.3%	18.2%

Appendix D

Cohort 2 demographic predictors of performance on the WJ-III

Table D1 presents the proportion of Cohort 2 students for each subtest who improved their standard scores by at least 2 points, maintained their scores, or declined relative to their peers by the end of one year of HillRAP intervention. Reading Fluency showed the most improvement over all W-J subtests.

Table D1. Individual-level changes in standard scores pre- to post-HillRAP for Cohort 2 (09-10)

WJ-III Subtest	Improved by 2 or more points	Maintained standard score	Declined by 2 or more points
Letter-Word	40%	24%	36%
Passage Comprehension	34%	20%	46%
Reading Fluency	60%	12%	28%
Word Attack	50%	22%	28%

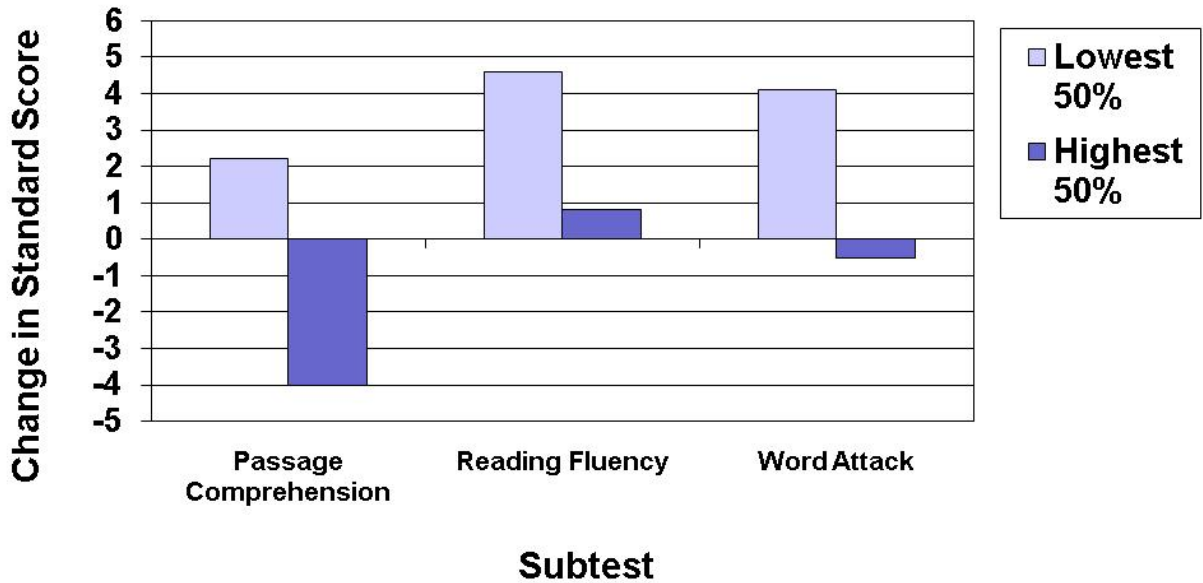
Student characteristics were again explored as possible predictors of improvement on the WJ-III. Ideally, results from this second cohort can be used to replicate Cohort 1 findings to see if identified predictors hold with different groups of students despite some baseline differences. The student variables examined were: pre-intervention achievement scores, grade, age, sex, free/reduced lunch eligibility status, IQ score, grade retention history, LEP status and Exceptional Child status. General linear models predicting standard score change over time were conducted, controlling for school, days 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 predicted improvement over time on Passage Comprehension ($F(1, 45) = 20.1, p < .0001$), Reading Fluency ($F(1, 55) = 3.8, p = .06$), and Word Attack ($F(1,45) = 7.5, p < .01$). Students with lower initial standard scores demonstrated more gains on all three subtests than did students with relatively higher pre-intervention achievement scores. To depict this effect, a median split was used for pre-intervention scores on each subtest. Median scores for these subtests were 86, 86, and 89,

respectively. Figure D1 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.

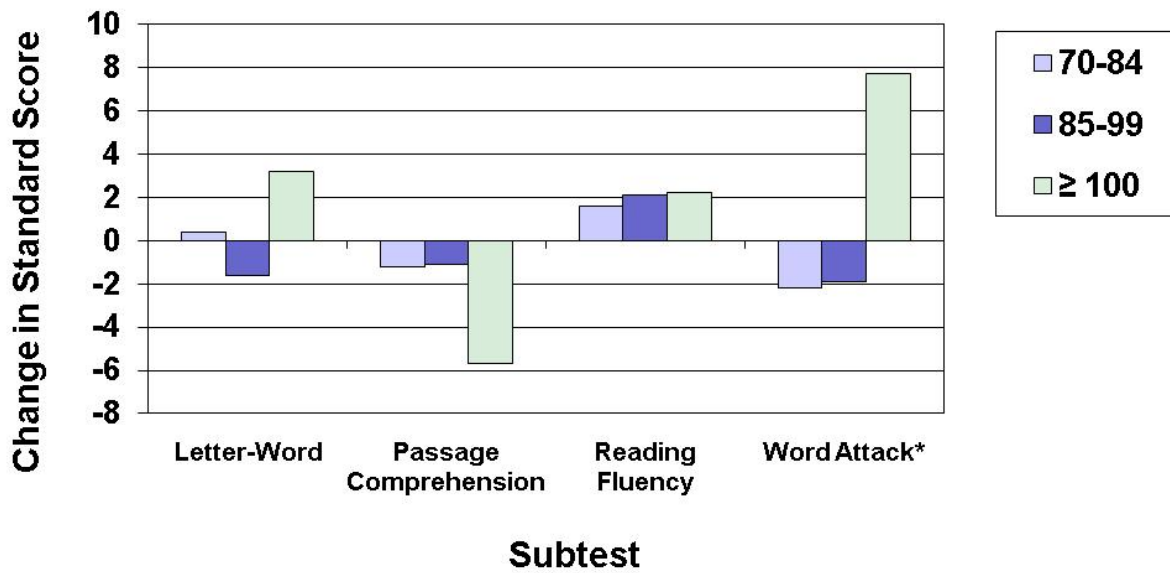
Figure D1. Pre-intervention achievement scores as a predictor of improvement



Full-scale IQ score

A student’s IQ score as assessed by the school system was also found to be a significant predictor of Word Attack performance ($F(1,19) = 4.0, p < .001$; note that IQ was available for only half of Cohort 2). To depict this relationship, student IQ scores were divided into three categories (no students in Cohort 2 had IQs lower than 70): IQ score between 70 and 84 (1-2 SDs below the mean), between 85 and 99 (0-1 SDs below the mean), and IQ score equal to or greater than 100 (at or above the mean). Children with an IQ of 100 or above made progress at a faster rate than their peers (see Figure D2), while those with IQs below 100 lost ground relative to their peers. This finding is similar to that of Cohort 1, that more intelligent children were more capable of making larger gains on Word Attack, in particular.

Figure D2. IQ as a predictor of Word Attack



*statistically significant difference ($p < .05$)

Appendix E

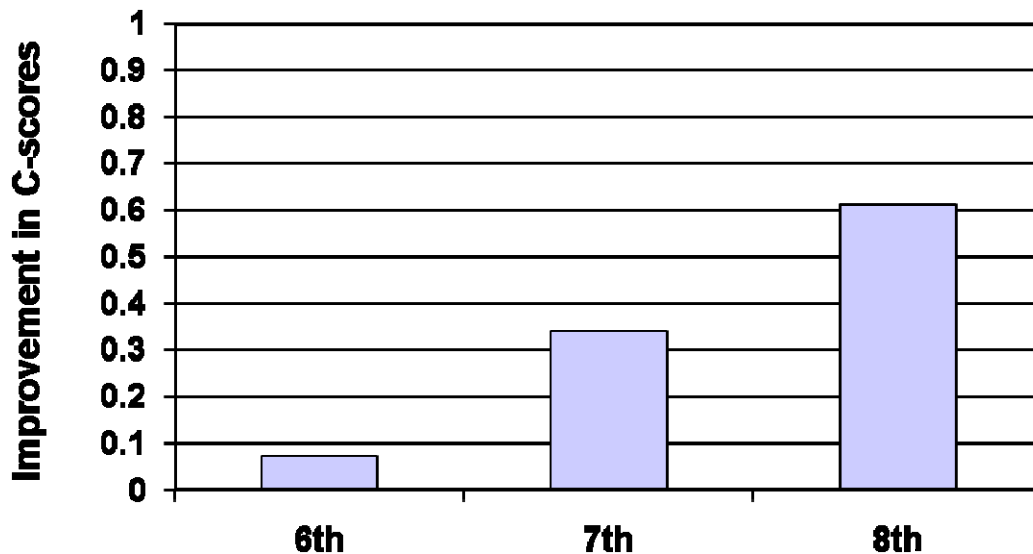
Cohort 2 demographic predictors of performance on the EOGs

Student characteristics were explored as possible predictors of improvement on the EOG c-scores following one year of HillRAP. The student variables examined were: grade, age, sex, free/reduced lunch eligibility status, IQ score, grade retention history, LEP status and Exceptional Child status. General linear models predicting standard score change over time were conducted, controlling for hours of HillRAP attendance, pre-intervention EOG c-scores, grade level, and sex (where these were not the predictors of focus). Interactions were examined for all significant variables. No characteristics were significantly predictive of EOG improvement for Cohort 1 students, but predictors did emerge for Cohort 2 as described below.

Grade Level

Student grade level significantly predicted improvement on EOG c-scores pre- to post-HillRAP for Cohort 2 students ($F(1, 35) = 4.8, p < .05$). Students of all three grades improved EOG performance; however, students in higher grades made larger gains. This is contrary to findings for the WJ-III and was not seen in Cohort 1 students. It may be a spurious finding, but is encouraging in that older students were able to demonstrate significant benefit. Figure E1 depicts the findings by grade level for EOG c-scores.

Figure E1. Grade as a predictor of improvement in EOG c-scores



Free/reduced lunch eligibility

Free/reduced lunch eligibility was used as an indicator of socio-economic status (SES). As shown in Figure E2, Cohort 2 students of lower SES (with free/reduced lunch eligibility) maintained consistent EOG c-scores following a year of HillRAP intervention, whereas students with higher SES improved their c-scores. This difference approached significance ($F(1,83) = 3.3$, $p = .07$). This finding is consistent with WJ-III findings for Passage Comprehension and Word Attack.

Figure E2. Free/reduced lunch eligibility as a predictor of improvement in EOG c-scores

