Long-term Effectiveness of the Reading Recovery Early Reading Intervention Program in a Rural School District

by

Adam Warren Vaughan

A dissertation submitted to the Graduate Faculty of
Auburn University
in partial fulfillment of the
requirements for the Degree of
Doctor of Philosophy

Auburn, Alabama
December 12, 2011

Reading, Reading Recovery, Reading Programs, Interventions, End-of-Grade Tests

Approved by

Joseph A. Buckhalt, Chair, Professor of Special Education, Rehabilitation and Counseling
Craig Darch, Professor of Special Education, Rehabilitation and Counseling
Bruce A. Murray, Associate Professor of Curriculum and Teaching

Abstract

There are many programs that specialize in teaching students the necessary strategies for reading. But which ones will have the greatest impact and provide lasting skills to struggling students? The purpose of this study was to assess the effectiveness of the Reading Recovery early intervention program on the lowest performing first grade students in a rural North Carolina school district. This was accomplished by assessing their pre- and post performance in the program using the Reading Recovery assessment *An Observation Survey of Early Literacy Achievement*, specifically the *Text Reading Level* subtest (Clay, 2002) and tracking their subsequent progress via the NC EOG (End-of-Grade) test of reading. Students who participated in the program increased from a mean Text Reading level of 3 at the start of the program to a mean Text Reading level of 14 at the time of program completion. Long-term effectiveness of the program was less encouraging. A little less than half (44%) of the participants continued to maintain performance at the level of proficiency on the EOG in third grade and subsequent grades showed lower percentages. Once the participants had completed the Reading Recovery program in first grade they were performing on average with their peers. However, in third and subsequent grades the students were not performing equivalent to the average performance of the school district.

Acknowledgments

The author would like to extend his gratitude to Dr. Joseph Buckhalt for his instruction through the years and his assistance and continued support during this extended dissertation process. He has been a pillar for the School Psychology program at Auburn University. The author would also like to thank the Granville County school system and especially Dr. Gerri Martín, Assistant Superintendent for Curriculum and Instruction in Granville County schools, for her willingness to allow this important topic to be studied in her school system. I would especially like to thank Alan Lydiard and the technology staff in Granville County for their efforts in working to provide the appropriate data for this study. And finally, a debt of gratitude is owed to the author's wife and children who have been with him since the beginning and in every sense of the word have 'attended' graduate school together with him.

Table of Contents

List of Tables

List of Figures

List of Abbreviations

RR   Reading Recovery Program

TRL   Text Reading Level

EOG   End-of-Grade Test

CBM   Curriculum Based Measure

DRA   Developmental Reading Assessment

DIBELS   Dynamic Indicators of Basic Early Literacy Skills

RTI   Response To Intervention or Instruction

Introduction

Learning to read is a process that all children must go through in order to acquire the skills necessary not only to be successful in school, but also in life. Learning to read contributes to our children's overall well being. When a child fails to learn to read "hope for a fulfilling productive life diminishes" (Lyon, 2001, p. 14). Reading is not, however, as natural as learning to speak. Research over 25 years has not supported the idea that reading development is a natural process. In other words, natural exposure to literature does not naturally foster good readers the same way that natural exposure to speech and the spoken word (Lyon, 2001). In the end, it is apparent that we must explicitly teach our children to read.

In the course of teaching students to read many do not automatically develop the strategies necessary to become good readers. These reading strategies must be taught, often in addition to what is taught in the regular curriculum. Interventions must teach what Pinnell (1989) describes as the goal of helping children develop "in the head" processes that most students who are good readers develop and use naturally.

At any given time there are numerous reading instruction programs running concurrently in the schools. Each has its own approach to remediating those students who are not developing the strategies that will help them to become good readers. There are numerous programs that have been initiated with great fanfare over the years, often in response to a particular trend in education. They are sometimes maintained because the materials are paid for and the teachers have been hired or, as is more common, the programs fade into oblivion with much less fanfare than they were introduced (Slavin, 1989). Reading programs are especially plentiful and varied.

Teachers attend workshops and kits are purchased and in the end there are more than a few ways to address the problem of learning to read—many of which are effective. The difficulty arises in selecting a program which will best meet the needs of the broadest group of students struggling with reading. If their difficulty lies in letter-sound knowledge, fluency or comprehension or some combination of all three, which program, whether in or outside of the classroom, will most effectively address these problems, providing the most gain in the shortest amount of time?

As educational budgets become tighter and programs are eliminated, it is increasingly more important to determine which programs can be used to address the broadest array of areas of difficulty and do so in the most effective and efficient way. What is required is that each program be evaluated to determine what skills are being taught and how effectively students are learning them. What effect do these programs have on student reading ability when compared to students that remain in the regular classroom receiving instruction from their teacher? Do more intensive programs, in which the student is removed from the classroom, help those students who are the worst off?

The purpose of the study is to assess the effectiveness over time of one reading intervention program, Reading Recovery, on the lowest performing first grade students in a rural school district in North Carolina. Reading Recovery is an intensive one-to-one tutoring program for beginning first-graders, which focuses on reading and writing instruction that addresses the problems of struggling readers early on before these students become acutely poor readers. Lessons are taught by highly trained teachers over 15 to 20 weeks, with the goal of bringing low performing readers up to grade level (Pinnell, 1989). Specifically, this study will assess the students' overall reading achievement levels upon successfully completing the program and then in later grades to determine the sustained effects of this program as compared to the on-grade

level performance of their peers in the classroom. If Reading Recovery is effective in increasing low performing readers' reading levels and overall reading achievement, then one would also expect there to be an increase in classroom grades, benchmark, and standardized tests scores. Further, a majority of RR students would not need to be referred to special education for reading difficulties when compared to other low performing students in the classroom who were not enrolled in the RR program. This study will track student progress via the *Text Reading Level subtest* on the Reading Recovery assessment *An Observation Survey of Early Literacy Achievement,* (Clay, 2002) and gather extant data collected by the school district in the form of standardized tests such as the EOG (End-of-Grade) test of reading in third-grade and each subsequent grade afterwards through seventh grade.

The Reading Recovery program is a popular and widely used intervention program (Pinnell,1989, Pinnell, 1988, Slavin, 1987, Slavin & Madden ,1987) and yet it has its critics (Center, Wheldall, & Freeman, 1992, Grosson & Coulter, 2002). This research is an attempt to observe the effect of the Reading Recovery program on low performing students using the program's own measures and data collected by the school district in subsequent grades. Specifically, the purpose is to look at the program's effect on student reading achievement over time as compared to peers' on-grade level performance.

Reading Instruction

Reading is no more a natural process than riding a bike. We are not born to ride a machine, nor are we born to utilize a language which is both spoken and written with difficult rules and spelling or that associates specific sounds with arbitrary squiggly lines (phonetic as opposed to pictographic). Reading must be learned and therefore it must be taught. There are any

number of ways to teach reading (AFT, 1999) and for most students they often learn to read despite what is taught. However, when a student struggles, when they do not 'get' what is being taught or the way it is being taught, that is when instruction must be intensified and expertise increased.

Whether it is a teacher, a parent, or school psychologist—the goal is to find out where the student is struggling. It is not enough to say that "the student has a problem reading" because they have dyslexia or because they have a learning disability (Joseph, 2008). Rather, the goal is to find out what specifically the student is missing. Are they missing time spent on a particular task/skill? Are they missing a fundamentally key component that must be taught again or for the first time? Or do they specifically have difficulty with a particular concept, such as phonemic awareness, in which case a new and different way must be taught that will allow them to understand what many students were able to learn without explicit instruction.

We do not all learn to swim the same way. Some are thrown in and learn to get by, enough to stay afloat. Others are taught from day one the proper form and stroke and they learn later on to truly swim when their muscles develop. Still others never learn because of fears of the water. No child learns to read in the exact same way. But many learn to get by and keep afloat— often with great effort. The point is that there is no single solution to how to teach all of our children how to read. There is no one way to move them from learning-to-read to reading-to-learn (Joseph, 2008), but there are many programs that can be adapted and modified to meet the needs of many students.

Often our teachers follow a curriculum and infuse their own philosophies and ideals about literature and many of their students learn to read—at least as well as their peers. For those who struggle, good teachers often have a trick or two up their sleeve or an able-bodied assistant

who can spend some extra time helping them in weak areas. If the student continues to struggle, the resources of the school are taken into consideration. In many cases, there might be tutoring available, a small group with a reading teacher, or a referral may be made to the school based team that comes together to address student needs when a student is struggling. The question is how to determine what component the student is exactly struggling with and how best to address that need.

Historically there have been several approaches to explaining and classifying student difficulties and subsequent instructional approaches to address them. In the area of reading difficulties McEneaney (2006) highlights three approaches through the years, including the most recent, RTI (Response to Intervention or Response to Instruction) which is continuing to evolve. The first approach early on, grew out of studies involving students with specific brain injuries resulting in reading difficulties. It was assumed that with developmental problems there was an underlying dysfunction with the brain. Out of this mode of thought, grew the *process deficit models*, that relied on the idea of information-processing; reading was described as "the flow of information across cognitive processing systems" related to reading (e.g. visual perception, word recognition, and phonemic analysis) (McEneaney, 2006, p. 361).

The idea of process deficit models was that this allowed educators and researchers to distinguish between various types of disabilities and thereby design instruction based on the type of disability. The research however, did not support this. Attempts to divide students into disabled readers and everyday poor readers on the basis of processing deficits were unsuccessful (Lovell, Gray, & Oliver, 1964) according to McEneaney (2006). In fact, McEneaney reports that despite three decades of research the process deficit model is a case of "beautiful theory and ugly facts" (McEneaney, 2006, 362). There is not enough evidence to support the idea that process

deficits explain difficulties in reading and thereby provide instructional guidance for students with reading difficulties.

The *discrepancy model* came after the process deficit model, as a more statistically based approach. This approach views reading achievement as distributed normally (bell shaped) along a continuum. The majority of students fall in the middle and smaller numbers of readers are on the ends (tails) in the classic bell shape (Pearson 1999, Snow et al. 1998).

In an attempt to improve upon the identification of students who are Learning Disabled, the Department of Education issued a regulation stating that students who are learning disabled would be those that have a *severe discrepancy* between their achievement and performance (IQ) scores (U.S. Office of Ed., 1977). However, due to the many IQ and achievement tests available and the interpretations that states have given this regulation (difference between IQ and achievement, regression of IQ on achievement, amount of discrepancy—e.g. 1 SD vs. 2 SDs) has led to varying rates of identification between states (Dunn, 2007, Scruggs & Mastropieri, 2002).

Again the research fails to support this model. There is little evidence that there is a difference between even children who have severe reading difficulties. In other words, that they also represent readers in the lower range of the normal distribution and are indistinguishable from students identified as having a learning disability (Stanovich, 1991). Additional studies also have not shown significant differences between those readers diagnosed as low achieving and learning disabled (Fletcher et al., 1994, Vellutino, Scanlon, & Lyon, 2000). Concerns, or at least notable limitations, of the discrepancy approach were present even when it was first widely adopted in the Education of All Handicapped Children Act of 1975 and continue to be written about to the present (McEneaney, 2006, Stanovich, 2008).

The discrepancy model with its accompanying assessment and classification is flawed because it does not provide a clear, consistently reliable method of identifying LD students. In order to address this issue, the President's Commission on Special Education was convened (U.S. Department of Education, Office of Special Education and Rehabilitation Services [OSERS], 2002). The OSERS's focus was on the problem with IQ tests as assessment measures for special education eligibility and the practice of *wait-to-fail*. Studies found no difference in reading skills of students with reading disabilities with high IQs and low IQs (Tal & Siegal, 1996). The IQ tests did not help predict which students would gain from remediation (Kershner, 1990). This is further compounded by the *Matthew Effect*, which theorizes that reading difficulties may influence the development of language, knowledge, and vocabulary skills and therefore affecting performance on traditional IQ tests (Stanovich, 1988). The practice of waiting until third grade to see if a student is grasping/learning-retaining content material is making it more difficult for students in later grades (Dunn 2007, Lyon, Fletcher, Shaywitz, Torgesen, Wood et al., 2001).

RTI (Response To Intervention, also sometimes referred to as Response To Instruction) is offered as an alternative to the discrepancy model in order to identify students as having a learning disability. In the 2004 update to the law, the Individuals With Disabilities Education Improvement Act of 2004; IDEIA, introduced RTI as an alternative approach, but retained the process deficit model as its framework and discrepancy as an indicator of reading disabilities.

Within the RTI model, if a student is learning disabled then they are considered dually discrepant, in other words, they are not only low achievers, but are also making little-to-no progress within the three-tiered intervention program. The three tier programs typically consist of tier one, which is basically regular education and is targeted at all students and is based on the

school's core reading curriculum. At this first tier it is presumed that the instruction is research-based and student progress is monitored via benchmarks or at minimum in the beginning, middle and end of the school year on all students. A student who is dually discrepant (low achieving and making little progress) is considered at risk and moved to tier two. Tier 2 consists of data collected on interventions conducted individually or in small groups—essentially more intense instruction in the form of targeted programs to address specific needs. Progress monitoring is more frequent in Tier 2 than Tier 1. If progress is made then he/she returns to the regular education setting and is no longer considered at risk. If no progress is made—continues to fail to make adequate progress, then he/she most likely has a true disability (intrinsic, not lack of instruction) and thus needs to be moved to the third tier. The third tier represents special education and additional evaluations are conducted to determine disability identification and placement (Fuchs, Mock, Morgan, & Young 2003, Dunn 2007, Joseph 2008).

Even within the RTI model there are distinctions between various approaches. RTI can broadly be defined as "any set of activities designed to evaluate the affect of instruction, or intervention, on student achievement" (Christ, Burns, & Ysseldyke, 2005, p. 2). Even this fairly succinct definition is evolving and has been divvied up by those that take slightly different approaches. In their article on conceptual confusion, Christ et al. (2005) expound on several vernacular distinctions first noted by Fuchs et al. (2003), which represent varying philosophical approaches to RTI. According to the authors there are *early interventionists* advocating for a more standardized and validated treatment approach (Standard Protocol, RTI-SP) and there are those with a more behavioral approach who equate RTI with a more *problem-solving* approach (Problem Solving, RTI-PS). As the authors state, the confusion begins with the verbiage. The reality is that both fit within a problem-solving model, as do most RTI approaches. RTI, although

not entirely synonymous with problem-solving, both do represent "processes which may converge" (Christ, Burns, & Ysseldyke, 2005, p. 1).

The distinction between the two models lies in what occurs prior to the selection of the intervention and the subsequent steps. The RTI-SP approach relies more on standard protocols or procedures "empirically supported instructional approaches" that attempt to remediate and prevent problems with little analysis of the problem skill area (Christ, Burns, & Ysseldyke, 2005, p. 2). In this respect, RR is similar to this approach. Each RR lesson is standardized and includes a prescribed set of parts that are included in each session. Its goal also, is to remediate and prevent a student's reading difficulties from becoming a more severe problem.

The RTI-PS is more flexible and focuses on individualizing the intervention after analyzing the instructional environment and the target skill area. This process is systemized and its goal is to isolate the "skill deficits and shape targeted interventions" (Christ, Burns, & Ysseldyke, 2005, p. 2). RR also works in this manner by "roaming around the known" essentially getting to know where the student is in their reading development first, then modifying the lessons within the prescribed parameters to focus on areas of need focusing on a particular weakness such as phonemic awareness (Clay, 2005a, p. 33).

Given that research indicates 80% of student with a learning disability are learning disabled in the area of reading (Roush, 1995), a reading intervention program like Reading Recovery, fits within the RTI model and format. RTI and RR both entail one teacher and one student working together on activities for part of the school day over a specified period of time during which progress is monitored generating chartable data—the goal being to improve the student's academic performance enough to be able to return to the regular education classroom (Dunn, 2007). RR offers schools a research based intervention that may already be in use.

RR assesses student book levels (A, B, 1-30), when students do not progress sufficiently through the individually tailored daily lessons over the 20 weeks—it is determined that the student has impaired reading skills and further special education services are needed (Dunn, 2007, p. 34). RR is also able to point out areas of strength and continued weakness based on Observation Survey assessments.

Research conducted by Dunn (2007) was a retrospective study involving 155 students from third to fifth grade who had previously been enrolled in RR in first grade. Students were identified as having a reading disability or not and reading achievement data was analyzed compared to their RR scores from first grade (beginning text level, ending text level, and number of weeks' participation in RR) and free/reduced lunch status. The study looked at whether a connection existed between the scores obtained in the RR program and student's being identified as having a reading disability later on in third through fifth grade. Results indicated that ending text level was the most predictive—meaning a higher ending text level score, the less likely a student was to be later identified as having a learning disability. The inverse would also be true; students whose ending text level was low were more likely to be identified as having a reading disability. However, ending text level only explained 7 to 17% of the variance, therefore ending text level failed to explain the concept of having a reading disability in its entirety, but indicated the need for further evaluation (Dunn 2007).

The RR program is a good match for the goals of RTI's level 2: "prevent reading difficulty by delivering an intensive, and presumably effective, intervention that improves reading development" and "to assess the level of responsiveness to an instructional intensity from which most students' performance should improve" (Dunn 2007, p. 43). Although additional assessments would be needed, low RR ending Text Reading levels appear to be an

indicator of a reading disability. In other words, despite a scientifically research-based intervention the student continues to demonstrate a need in the area of reading and therefore should be evaluated to move on to tier three (special education). This also highlights the usefulness for schools to incorporate RR ending Text Reading levels in their procedures for identification of learning disabilities. This would allow students to be referred at the end of first grade instead of waiting for further difficulties in later grades.

Reading Recovery

Reading Recovery (RR) was developed by Dr. Marie M. Clay and is an early reading intervention for struggling readers. Targeted towards first-grade students who have the lowest reading achievement as they enter first grade, RR provides intensive one-to-one tutoring by specially trained RR teachers for 30 minutes each day for 12-20 weeks. RR focuses on providing students with reading and writing skills in order to bring them up to the average level of their peers. RR has also begun implementing a Spanish language version with initial reading instruction being in Spanish; Descubriendo la Lectura (DLL) (Reading Recovery, 2008).

Teachers who are involved in RR are specially trained and must participate in university level training for a year and then continue their training periodically with additional supports and in-service training. This is an important factor to the success of RR. Each school-based teacher is trained and supported, thereby providing each school with well-trained reading specialists for their schools.

Reading Recovery is currently being implemented each year in over 10,000 U.S. schools (National Data Evaluation Center—NDEC, 2008). Studies that have replicated the effects of RR are consistent with RR data collected and reported on each year and are published in reports

found on the RR data collection website (NDEC, 2008). The majority of students who successfully complete or are 'successfully exited' from the program are able to reach levels of achievement of their average peers and several studies have shown that these effects are sustained over time (Center, Wheldall, Freeman, Outhred, & McNaught, 1995; Iversen & Tunmer, 1993; Pinnell, 1989; Pinnell, Lyons, DeFord, Bryk, & Seltzer, 1994; Quay, Steele, Johnson, & Hortman, 2001; Schwartz, 2005).

There are three main components of RR that are at the core of its effectiveness: lessons, teacher training, and assessment. First, the tutoring sessions or lessons are structured in order to maintain consistency and focus, but allow for the RR teacher to adapt to the student's individual areas of need. The RR intervention program is intended to serve the lowest 20% of readers in first grade (RR, 2008). Selection for RR is based in part on recommendations from school staff—using prior reading achievement performance, diagnostic testing (*Clay Observation Survey of Early Literacy Achievement*) and teacher recommendations. RR teachers then compile a list of the lowest performing 20% and begin working with a few at a time (Clay, 2005a, Clay, 2005b).

The first ten one-on-one tutoring sessions act as screening and diagnostic tools. This is called *roaming around the known*. The following components are addressed in each lesson and form a flexible lesson framework for teachers to adapt on-the-fly. Each daily lesson begins with *familiar rereading*. A previously read book that the student is already familiar with is selected for the student to read. A *running record analysis* is conducted in which each student reads the previous day's new book (the familiar rereading) and the teacher codes reading behaviors using a running record. Reading is done by the student independently. The next component involves *writing a message*. The teacher assists the student in composing and then writing a message (1-2

sentences), which is then written word-for-word. This step provides opportunities for the teacher to assist the student in constructing words, analyzing sounds and representing them with letters. The message is read many times thereby increasing use and knowledge of high-frequency words. This is followed by *putting together a cut-up sentence*. The teacher then writes the message on a strip of paper and cuts it up, asking the child to reconstruct the message. This encourages rereading and searching for visual information.

Finally, a new book is read in *reading a new book*. A new book is selected (slightly more challenging) and the pictures are reviewed and the story introduced. Focus is on meaning, but the student may be required to locate some key words based on predicting the first letter. Then the teacher and child read the story, which then provides the basis for the familiar reading for the next day's lesson. (Clay, 2005a; Clay, 2005b; Pinnell et al., 1988; Pinnell et al., 1990).

Students progress through a maximum of 60 lessons. A student is said to have successfully completed RR (lessons are discontinued) when they are able to read on the average level for their peers in first grade (based on local averages)—occurring between 12 to 20 weeks in RR. Those students who make progress but are still not on average grade level after 20 weeks are referred for further evaluation (i.e. special education) (AFT, 1999). "Children who seem likely to fail, despite tutoring in RR—those not progressing at the desired pace after 10 lessons—may be referred to special education and removed from the program" (AFT, 1999).

The next critical component in Reading Recovery is its extensive teacher training. Each teacher involved in RR must participate in university level training for a year. This is key to the success of RR. Each school-based teacher is trained and supported by district- or site-level teacher leaders, who in turn have been trained by university trainers. Training occurs while teachers are continuing to work with children. One-way mirrors are used to observe and discuss

proper instruction with teachers. Teachers are taught to be sensitive to student's reading and writing behaviors in order to be able to make moment-by-moment analyses that inform teaching decisions (AFT, 1990; Clay, 2005a; Clay, 2005b).

Training continues after the initial year of training in the form of ongoing professional development called *continuing contact*. Observations of one another continue, as do discussions of intervention practices. This continuing contact provides opportunities to collaborate and continue to hone their skills, as well as receive support for especially difficult children and receive knowledge of new research (Smith-Burke, 1996).

This extensive training requirement of the RR program is sometimes seen as a critique of RR and an undue burden placed on schools who may already be poorly funded. However, this is actually one of its strengths. Two important points regarding this training were emphasized in a recent report from the National Research Council (1998):

> First, the program demonstrates that, in order to approach reading instruction with deep
>
> and principled understanding of the reading process and its implications for instruction,
>
> the teachers need opportunities for sustained professional development. Second, it is
>
> nothing short of foolhardy to make enormous investments in remedial instruction and
>
> then return children to classroom instruction that will not serve to maintain the gains they
>
> made in the remedial program. (p. 258)

The third component of Reading Recovery is its assessment tool, the *Observation Survey of Early Literacy Achievement* (Clay, 2006). The Observation Survey is used to assess the progress of students in the RR program and to determine appropriate discontinuation upon reaching the average level of reading for their peers. The Observation Survey measures decoding, letter knowledge and concepts about print. These tasks are designed to measure knowledge about

reading and writing in relation to literacy learning. Information from many of these subtests provides guidance in adapting instruction to the student's strengths and needs.

Assessments

The RR program utilizes multiple sources of data. A *running record* is kept for each student, which is maintained to monitor the student's oral reading and involves the student reading a passage of text and the teacher monitoring his/her accuracy. Progress is monitored on text reading, daily lesson records, students' writing, and change over time in reading and writing vocabulary. The RR teacher receives specific training on how to take a running record and record miscues.

The Observation Survey of Early Literacy Achievement measures six literacy tasks that describe each student's reading and writing progress: Letter Identification, Word Test (vocabulary), Concepts About Print, Writing Vocabulary, Hearing and Recording Sounds in Words (phonemic awareness, representing sounds in graphic form), and Text Reading.

Letter Identification assessment is accomplished by having the student correctly identify both upper- and lower-case letters. The student responds to 54 letter forms (26 uppercase and 28 lowercase). Extra lowercase forms include alternate forms of a and g. The student is allowed to respond with either a letter sound, letter name, or a word that begins with that letter. The maximum possible score is 54, Cronbach's alpha was .95. (Clay, 2002).

The **Word Test** assesses vocabulary knowledge and word recognition and is based on the *Ohio Word Test*. The student must correctly identify 20 sight (Dolch) words from graded lists compiled from basic reading texts and are scored on accuracy (i.e. number of words read correctly). The maximum possible score is 20, Cronbach's alpha was .92. (Clay, 2002).

Students are assessed on **Concepts About Print** in which they demonstrate knowledge of 22 printed language concepts/conventions (e.g. front of book, back of book, text direction, word concepts, etc.). This is achieved in the form of booklets that the teacher reads and the child responds to questions or requests to manipulate a book. The maximum possible score is 24, Cronbach's alpha was .78, split-half reliability was .95. (Clay, 2002).

Writing vocabulary is assessed by having the student write as many words as he/she knows. Students are given 10 minutes to write as many words as they can on a blank sheet of paper. When needed, a standard set of prompts is used to encourage additional attempts to write. This activity is scored by counting the number of correctly spelled words. Test-Retest reliability was .62 and .97. (Clay, 2002).

Hearing and Recording Sounds in Words (HRSW) task is a dictation task. It assesses the student's phonemic awareness and representing sounds in graphic form. The student must correctly write (encode) words from a dictated passage. The teacher reads one of five passages, and then asks the student to write the words as it is read again by the teacher. If a child does not know a word, they are prompted to say the word slowly, thinking about what they heard and how to write it. The score is based on the number of correctly written phonemes (smallest unit of a word). The maximum possible score is 37, Cronbach's alpha was .96. (Clay, 2002).

Text Reading levels are obtained by having students read books leveled by difficulty and text characteristics (Peterson 1991). Books are drawn from a basal reading series. Students are assessed on error rate and self-correction, while a running record is kept. Leveled passages are read until their accuracy falls below 90%. This is used to compare the percentage of students scoring at the first grade level or higher with those who score lower than first grade. This performance is converted into a numerical reading level (level 1 to 30) and corresponding grade

levels (Clay, 2006).

The goal of RR is that the child achieve the average reading level of his/her peers, however, Schwartz (2005) outlines several sources for average reading level for first grade. For example, the Ohio Stanines for text level indicated an average of Level 2 for the fall of first grade and an average of Level 9 to 12 in the spring (Clay, 2002). However, a Level 20 was the average level indicated by the National Data Evaluation Center's random sample (Gómez-Bellengé & Thompson, 2005). Text level decisions are based on the running record, for which Clay (2002) reports are reliable over a two year interval across two scorings with a trained recorder (r = .98). This presents varying exit criteria depending on the average reading performance of a particular school.

Validity and reliability for all tasks of the Observation Survey have been documented (Clay, 2006; Denton, Ciancio, & Fletcher, 2006) and the Observation Survey highly correlates with the *Iowa Test of Basic Skills* (Rodgers, Gómez-Bellengé, Wang, & Schultz, 2005) and (Tang & Gómez-Bellengé, 2007). National norms have been developed to assist in interpreting scores (Gómez-Bellengé & Thompson, 2005).

In this study, in addition to the Observation Survey that is a part of the Reading Recovery program, data from the NC End-of-Grade (EOG) test in reading, was included in the analyses. The school district routinely administers the EOG annually for both Reading and Math. The EOG reading served as an objective measure of the students' reading achievement. The EOG measures reading achievement and comprehension via a series of reading passages followed by multiple choice questions on the content of the passages. The EOG is designed to measure reading and comprehension that align with the NC *Standard Course of Study*. In this school district, teachers use the End-of-Grade assessment (EOG) to evaluate and monitor student

reading progress each year (NCDPI).

Each student is given the EOG at the end of each school year, beginning in third grade, in order to determine their level of reading achievement, as well as measure progress made throughout the school year. The EOG produces two main scores and a percentile. The first is the Developmental Scale Score and the second is the Achievement Levels. The scale score is derived from the raw score which is composed of the number of questions the student answered correctly. Scale score ranges vary from grade to grade and depending on the year of the EOG, can vary from year to year (e.g. starting with 2007-08 school year, third grade Level I - <330, Level II – 331-337, Level III – 338-349, and Level IV - >350). The scale score represents growth in reading achievement from year to year which allows parents and the school system to measure each child's growth in reading.

The second score produced by the EOG is an Achievement Level. The Achievement Levels are a method of dividing the range of scale scores on the EOG into four levels (i.e. levels I, II, III, and IV). The levels are predetermined performance standards used to compare the performance of the students to grade-level expectations (NCDPI-DAS/NCTP, 2007).

Literature Review

Since its inception Reading Recovery has been the focus of many studies, some better than others. In an attempt to put rhyme and reason into the selection of studies for which to review for this research, two common criteria were utilized. The first set of criteria were those suggested by the United States Department of Education's Quality of Research Decision Tree published in a 2002 report. The report suggested states and local schools should view evidence for reading programs based on the following criteria: the theoretical base, evidence of effects, and evidence of replicability. These criterion were applied by the Reading Recovery Council of North America, North American Trainers Group Research Committee—RRCNA (2004). In an article authored by the research committee, a group of studies involving RR that followed the criteria set forth by the DOE was compiled and evaluated for effectiveness (RR Council, 2002). The DOE published a Quality of Research Decision Tree, highlighted in the research committee's article, in which they outline criteria that should be applied when reviewing research studies. The RRCNA compiled a list of five research studies that conformed to these criteria (Center, Wheldall, Freeman, Outhred & McNaught, 1995; Iversen & Tunmer, 1993; Pinnell, 1989; Pinnel, Lyons, DeFord, Bryk & Seltzer, 1993/94; Quay, Steele, Johnson, & Hortman, 2001).

The second criteria used in the selection of research articles, was the What Works Clearinghouse's review of Reading Recovery in which they reviewed 78 studies. The What Works Clearinghouse is a branch of the United States Department of Education (DOE) and the

Institute of Education Sciences (IES). The What Works Clearinghouse released a 3-year

independent review of the experimental research on Reading Recovery in March 2007. These

studies were selected based on 4 factors: quality of research design, statistical significance of

results, size of difference between participants in the intervention and comparison group, and

consistency of findings across studies (WWC, 2007). The result of the WWC applying their

criteria to approximately 78 studies was 5 studies that met WWC's stringent evidence-based

standards criteria. Four of the studies met the full WWC evidence standards (Baenen, Berhole,

Dulaney, & Banks 1997, Pinnell, DeFord, & Lyons 1988, Pinnell, Lyons, DeFord, Bryk &

Seltzer 1993/94, and Schwartz 2005) and one study (Iverson & Tunmer 1993) met WWC

evidence standards with reservations.

When the RRCNA and WWC lists were compared 2 studies (Iverson & Tunmer 1993,

and Pinnell, Lyons, DeFord, Bryk & Seltzer 1993/94) were selected by both groups for review.

Three studies were unique to WWC group, and four were unique to the RRCNA group.

In the following section, the two studies selected by both groups will be reviewed along

with comments from each of the reviewers (WWC and RRCNA). Then the remaining five

studies will be reviewed with commentary from their respective groups.

The Rhode Island school district study conducted by Iverson and Tunmer (1993)

examined the progress of three matched groups of first-graders at-risk for reading difficulties.

The study was reviewed by both RRCNA and WWC. The study used a quasi-random assignment

to three groups: RR, a modified RR group, and a standard intervention group (small group Title

I). There were a total of 96 students, 64 from 34 classrooms and 23 schools in the two RR groups

and 32 from 7 schools for the standard small group intervention. The students were administered

a battery of tests in the beginning and end of the school year, and once at the midpoint when RR

students were being discontinued. In addition, average students from the same classrooms were tested at the discontinue (midpoint) point.

The outcomes of students who received RR (n = 32) were compared with non-Reading Recovery students who received the "standard small group, out-of-class support services" (n = 32). Students were matched based on pretest scores. The third group (n = 32) received a modified version of RR which added explicit instruction in letter-phoneme patterns in lieu of RR's letter identification segment. This modification highlighted for students the fact that words with common sounds share patterns of spelling. This group was not included in the WWC review since it represented a modified version of RR, however, research which features modifications to RR are often enlightening as to which components are essential to RR's success. In this particular case this modification made little difference to RR's effectiveness.

Researchers for this study did include a former RR teacher; however this person was not presently involved in ongoing professional development with RR. The second investigator was a university researcher conducting independent, critical evaluation of RR. The affiliations of each researcher are important under these circumstances given the potential bias of a researcher who is closely involved with RR prior to beginning research of the program.

Iverson and Tunmer (1993) took advantage of the instruments for gathering data that RR utilizes, namely the six tasks from the Observation Survey. In addition, they used measures of phoneme segmentation, phoneme deletion, and phonological recoding.

The study found all three groups to be equally low on the pretest measures. Once they had successfully discontinued the program, both RR groups, the standard RR and the modified RR group, scored significantly higher on the outcome measures. On the Text Reading Level subtest from the Observation Survey (one of the assessments which had no ceiling) the

differences were even larger (over eight standard deviations on Text Reading Level and over two standard deviations on the Dolch Word Recognition Task). At this stage, students in both RR groups had profiles similar to average students in the classroom.

The review article by the RRCNA also noted that the two RR groups scored similarly—even on tasks that were included in the study to measure the effectiveness of the modified RR intervention—the phonemic awareness measures. The modified RR group did not actually do as well as the regular RR group on phoneme deletion measure. The article further stated that the true measure of benefit from the modified RR group came from the students' being able to successfully discontinue in fewer lessons (41.75 = RR modified, 57.31 = Standard RR). RR's standard program has since been modified in a similar manner as the modified RR group, not as a result of this particular study, but rather as a response to changes in the field of reading. At the end of the school year the standard and modified groups were similar, the modified group being only slightly higher on the test reading measure (19.56 = RR and 18.38 = modified RR) (Iverson & Tunmer 1993).

In the end, this study compared two versions of RR and small group Title I instruction. The students in the two RR one-to-one tutoring groups showed an advantage from having participated in the groups, including earlier discontinuation for the modified RR group. Both sets of procedures in the RR groups fostered phonemic awareness learning and applying that knowledge to text reading and writing.

This next study, Pinnell et al. (1993/94), was also selected for review by both the WWC and RRCNA. Dr. Pinnell, who is a leading researcher on RR along with her colleagues, designed a randomized controlled study involving 324 students in 33 schools. The study involved four groups: RR students (individual tutoring), Reading Recovery-like intervention (individual

tutoring with teacher trained in an alternative and shortened setting), RR-like small group, and basic skills small group, plus a randomized comparison group at each school. Students were assessed at the beginning of the year, midyear, end of year, and at the beginning of the following year (Pinnell et al., 1993/94).

Pinnell's study was funded by a grant from the John D. and Catherine T. MacArthur Foundation. In addition, the research was supervised by a national advisory board which was actively involved in every phase of the research (RRCNA, 2004). This provides good evidence that any potential conflict of interest or past experience/opinion of Pinnell's having researched RR would be checked by this advisory board.

Pinnell et al. (1993/94) measured intervention effectiveness with the Gates-MacGinitie Reading Test, the Woodcock Reading Mastery Test, and Text Reading Level and Hearing and Recording Sounds in Words (RR Observation Survey). They found statistically significant positive effects of RR for the RR group (individual tutoring with trained teachers) on the Gates-MacGinitie, the Dictation subtest of the Observation survey (Hearing and Recording Sounds in Words), Text Reading Level, and the Woodcock Reading Mastery Test – Revised. The following fall, significant mean effects were found on Text Reading with smaller effects on Hearing and Recording Sounds in Words. The RRCNA reviewers explained this as possible ceiling effects of the measures (RRCNA, 2004).

This study was well designed, was conducted over a year's time with random assignment to treatment and comparison groups with large numbers of students. The end result was that "Reading Recovery emerged as the most powerful of the interventions tested from the beginning of Year 1 through the beginning of Year 2 of the study" (RRCNA, 2004).

The studies reviewed by both the WWC and the RRCNA, (Iverson & Tunmer, 1993;

Pinnell et al., 1993/94) all used random assignment (although Iverson's and Tunmer's design was quasi-random). This is very difficult to achieve in the context of an educational setting such as the schools. All three showed positive effects of RR and used a variety of assessments to obtain these results.

The next three studies were ones that met the WWC's stringent evidence standards criteria, but were not included in the RRCNA's review. These three were among the final five selected from the original 78 reviewed by the WWC.

The first study (Pinnell, DeFord, & Lyons, 1988) was a randomized controlled study involving 187 first-graders distributed across 14 urban Columbus, Ohio schools. The students were randomly assigned either to a group that received regular classroom instruction and RR (n = 38) or a control group (n = 53) which received an alternate compensatory program. The study also included a third group which received RR, but was also taught in their regular classroom instruction by an RR teacher (n = 96). Although this was not included in the WWC ratings, it is notable as another attempt by researchers to constantly find ways to improve the delivery and content of RR.

Researchers utilized a writing assessment, the Reading Vocabulary subtest of the Comprehensive Test of Basic Skills (CTBS), and the Reading Comprehension subtest of the CTBS, in addition to five subtests of the Observation Survey (Letter Identification, Word Recognition, Concepts about Print, Writing Vocabulary, and Dictation).

The study was reviewed by the WWC based on how student outcomes were addressed in four domains: alphabetics, reading fluency, comprehension, and general reading achievement. The Pinnell et al. (1988) study looked at the effects of RR on the following constructs within the alphabetics domain: print awareness and phonics. The researchers found statistically significant

positive effects, but not on the Letter Identification subtest of the Observation Survey, also in the alphabetics domain.

In the comprehension domain, researchers found a positive statistically significant effect of RR on the Reading Comprehension subtest of the Comprehensive Test of Basic Skills (CTBS) and on the vocabulary construct on the Reading Vocabulary subtest of the CBTS. Fluency was not addressed, however, on the general reading achievement domain. Pinnell (1988) found positive and statistically significant effects of RR on Hearing and Recording Sounds in Words (Dictation) and Writing Vocabulary, two subtests of the Observation Survey.

The second study by Baenen, Bernhole, Dulaney, & Banks (1997) was another randomized controlled study. This one was conducted in Wake County, North Carolina and involved a total of 772 first-grade students. The study spanned from 1990 to 1994 and included four cohorts. Students who qualified were randomly assigned to either a RR group or comparison group. Students were evaluated at the end of first, second, and third grade. Only one cohort (1990-91) met criteria for WWC. This was due to the comparison group being made up of students who were no longer similar to RR with regard to achievement level. After attrition, the final sample included 147 first-grade students (72 = RR and 75 = non RR group) in the 1990-91 cohort. All 147 of the students were followed into second grade and 127 were included in the third-grade analysis.

All students in the study conducted by Baenen et al. (1997) were assessed for eligibility using three subtests from the Observation Survey—Text Reading Level (running record), Dictation, and Writing Vocabulary. At the end of first and second grade, grade retention was also measured and at the end of third grade the North Carolina End-of-Grade test in reading was used. The study also measured referrals to special education and Title I services, and a gauge of

25

teacher perception of student achievement. Results for the single cohort (1990-91) reviewed by the WWC fell under the general reading achievement domain, however no statistically significant effects were found on grade retention.

The Schwartz (2005) study was reviewed by the WWC and involved a randomized controlled study. The study included students from 14 states, 37 of whom were randomly assigned to RR the first-half of the year and were compared to the other group of 37 who were randomly assigned to receive RR the second-half of the year. The groups were then compared at mid-year before the second group had begun RR. Data was also collected on low average and high average students from the same classrooms as the at-risk students for comparison. The WWC excluded the low-average and high-average comparison groups as it only focused on those students eligible for RR in its review. However these groups provide a reference since the goal of RR is to assist participants in attaining average levels of performance (RR, 2008).

Students in the Schwartz (2005) study were assessed at the beginning of the year, at the transition from the first round of RR to the second, and at the end of the year. Measures used included six tasks from the Observation Survey and then at the transition period and end of the year students were assessed using the Yopp-Singer Phonemic Segmentation task (a sound deletion task), the Degrees of Reading Power test, and the Slosson Oral Reading test.

The Schwartz (2005) study found that the at-risk RR students performed significantly better at the end of their intervention period than students assigned to receive services later in the year. Large effect sizes were found especially during the transition period between first- and second-round intervention services, for Text Reading Level, the Ohio Word test, Concepts About Print, Writing Vocabulary, Hearing and Recording Sounds in Word and the Slosson Oral Reading Test-Revised. Also, upon comparing the RR students with the low-average and high-

average groups at the transition period, RR students had closed the achievement gap with the average group which is the goal of RR (Schwartz, 2005).

In reviewing the early reading interventions, as noted previously, the WWC developed a beginning reading protocol which looks at four domains for addressing student outcomes: alphabetics, reading fluency, comprehension, and general reading achievement. Each study reviewed by the WWC covers these domains and most constructs contained therein. WWC's review found that in the alphabetics domain, two studies met evidence standards and showed statistically significant positive effects (Pinnell, DeFord, & Lyons, 1988 and Schwartz 2005). An additional study met WWC evidence standards with reservations and reported significantly positive effects (Iverson & Tunmer, 1993). One study in the fluency domain was found to have demonstrated statistically significant positive effects (Schwartz 2005), as was one in the domain of comprehension and another study in the domain of comprehension resulted in an *indeterminate effect* ((Pinnell, DeFord, & Lyons, 1988; Schwartz, 2005; WWC, 2007).

Three studies in the general reading achievement domain had strong designs and statistically significant positive effects (Pinnell, DeFord, & Lyons, 1988; Pinnell et al., 1994; Schwartz, 2005). An additional study had *indeterminate effects* (Baenen et al., 1997) and another met WWC evidence standards with reservations (Iverson & Tunmer, 1993), while demonstrating statistically significant positive effects (WWC, 2007).

The RR group that reviewed the same RR studies as the WWC (Iverson & Tunmer, 1993; Pinnell et al., 1993/94) also reviewed three additional RR studies based on criteria from the Department of Education's quality of research decision tree (Center et al., 1995; Pinnell, 1989; Quay et al., 2001).

The first study reviewed by the RR Committee (Center et al., 1995) included a random

assignment of subjects into either a RR group (n = 31) or a no-intervention control group (n = 39) over 10 schools. Researchers also followed a comparison group of low-progress students (n = 39) from five matched schools which did not have RR. Students were assessed pre- and after 15 weeks (post-test), after another 15 weeks (short-term maintenance), and again 12 months after post-test (medium-term maintenance).

Researchers assessed the students using *Clay's book level test* (Observation survey), Burt Word Reading Test, Neale Analysis of Reading Ability, Passage Reading Test, Waddington Diagnostic Spelling Test, Phonemic Awareness Test, Cloze Test, and the Word Attack Skills Test. After 15 weeks (post-test), researchers reported that RR students "significantly outperformed" students in the control group on "tests measuring words read in context and in isolation, but not on some tests of metalinguistic skills" (Center et al., 1995, p. 252).

At the short-term maintenance stage (end of first grade) researchers reported that RR students continued to perform better than control group students on tests measuring word reading in context and phonemic awareness tasks, but did not perform significantly better on phonological recoding or syntactic awareness. According to the authors, however, these areas were not specifically addressed by the RR program. One year after RR intervention (medium-term maintenance) RR students were still performing higher than both the comparison group and the control group.

In this study we find additional support for the effectiveness of RR from researchers not connected with or affiliated with RR, whose research is a high-quality, independent evaluation of RR, showing significant, long-lasting effects (Center et. al., 1995).

In the next study Pinnell (1989) included six urban schools with a high number of low-income students. Essentially the study consisted of a group of RR students (n = 55) who were the

lowest in the RR designated classrooms and were taught by a RR-trained teacher and a comparison group (n = 55) who were the lowest students in the comparison classrooms and were taught by a non-RR trained teacher.

Assessment measures for the study included Text Reading Level, Observation Survey and the Stanford Achievement Test. Students were assessed in October, mid-year, end of year, and end of year following treatment. Analyses were conducted on four groups: RR students in program classroom, RR students in regular classroom, comparison students, random sample students (May assessments only). RR students from regular classrooms did better ($p < .05$) than comparison students on 7 out of the 9 assessments. Two assessments (Letter-Identification and the Word Test) resulted in ceiling effects. RR students in program classrooms did better ($p < .05$) than students in the comparison group on all assessments. RR students, both program and regular classroom, did equally well—in other words, whether they were instructed by a RR teacher in the classroom or not. Follow-up results a year later (end of year following treatment) showed RR students scoring significantly higher ($p < .05$) on all assessments than the comparison students.

These results were promising and were conducted by a research team implementing RR in the U. S. in its first year. Researchers from the Center for the Study of Reading, University of Illinois independently audited the results reporting to the Ohio Department of Education (Pinnell, 1989). Although this research was conducted some time ago, during RR's earliest introduction into the U.S. these results are similar to previously mentioned studies that were conducted more recently and with only relatively minor modifications to RR through the years.

The final study reviewed by the RRCNA (Quay et al., 2001) included a quasi-random assignment procedure in which a group of students in 34 schools were assigned to two groups: RR and a control group. Each school contained a classroom that was randomly designated the

classroom from which RR students would come and another classroom which was the control classroom. Both groups were given the Observation Survey and ITBS (Iowa Test of Basic Skills) in the fall and spring. The Gates-MacGinitie Reading Test and Classroom Teacher Assessment of Student Progress were administered in spring as well (Quay et al., 2001).

Analysis was conducted to show that the two groups had few differences on the ITBS (Iowa Test of Basic Skills) scales in the fall—this confirmed the groups' initial equivalence on reading achievement (Quay et al., 2001). RR students were in regular classroom with exception of 30 minutes for RR. The control group, not served in RR, was given access to other programs available at the school. Within the control group 66% participated in daily literacy groups conducted by the RR teachers which were offered at the school.

Researchers in this study relied on data from the RR Observation Survey, Iowa Test of Basic Skills (ITBS), Gates-MacGinitie Reading Test, Classroom Teacher Assessment of Student Progress, as well as retention rates. Researchers noted that the Classroom Teacher Assessment of Student Progress is "an instrument developed and used extensively in large-scale evaluations and demonstrating high test-retest reliability" (Quay et al., 2001, p.14). At the end of the school year, RR students in the study significantly outperformed the control group students on four out of the six subtests on the ITBS, all subtests of the Gates-MacGinitie, all tasks on the RR Observation Survey and all nine measures of the Classroom Teacher Assessment of Student Progress. A significantly higher percentage of RR students were also promoted at the end of first grade than control group students. Despite the relative shortcomings of the Quay study (lack of truly random assignment, disruptions during the study) results not only showed that the RR children performed significantly better on standard measures over the control groups, but the researchers also made efforts to assure equivalence of the groups prior to beginning the intervention and in

their analyses measured retention rates, which ultimately equal economic savings for schools.

In the end, the RRCNA committee, after applying the criteria set forth by the Department of Education (2002), concluded that all five studies showed positive effects for RR, were in-line with evaluation data gathered and reported each year by the Reading Recovery National Data Evaluation Center (NDEC) and were published in peer-reviewed journals. It should also be noted that two of the studies were conducted by researchers who have been critical of Reading Recovery and three by researchers associated with Reading Recovery.

WWC has reviewed approximately 20 interventions to date using the following categories to rate an intervention's effectiveness in a particular outcome domain: positive, potentially positive, mixed, no discernable effects, potentially negative, or negative. The WWC based these ratings on 4 factors: quality of research design, statistical significance of results, size of difference between participants in intervention and comparison group, and consistency of findings across studies. Only one intervention, Reading Recovery, had qualifying research evidence in all four domains (alphabetics, reading fluency, comprehension, and general reading achievement). Reading Recovery received the highest ratings of any of the 20 programs with two *positive effects* (+) ratings for alphabetics and general reading achievement and two *potentially positive effects* (+?) ratings for reading fluency and comprehension (WWC, 2007). This authoritative, independent assessment presented ample evidence that suggests Reading Recovery is an effective intervention based on the current scientific evidence.

This study's purpose was to compare data collected by RR and the school district over time on students who have been successfully discontinued from the RR program in order to measure the sustained effects of the RR program on students' with low reading achievement in subsequent grades. The previously reviewed research presents a strong case for the program's

31

effectiveness, however, additional assessments beyond what is administered by the RR program are needed to validate its effectiveness in raising first grade students' reading achievement levels in first grade and to answer the question of whether those gains are maintained over time. In addition, if RR is a good fit for RTI (Lose, Schmitt, Gómez-Bellengé, Jones, Honchell, & Askew, 2007), then this study will help in determining RR's effectiveness in raising poor readers' reading achievement levels in subsequent grades by way of an early reading intervention (i.e. Reading Recovery).

This study adds to previous research in the following ways. First, there have been few RR related studies that have focused on longitudinal effects and district wide assessments (e.g. EOG)—in other words real-world, schoolwide measures of reading achievement/improvement and the question of sustained effects in later grades. Additional studies are needed to track students after participating in reading intervention programs and determine if the gains in the program are maintained over time. This would therefore result in poor readers becoming better readers and continuing to be improved readers in later grades maintaining those gains over time. Fuchs et al. (2003) laments the scarcity of research on intervention programs already in practice, including programs like Reading Recovery. Finally, this study further elucidates how RR and RTI might fit together. As Lose et al. (2007) noted, RR research provides anecdotal evidence of the need that RR fills as a scientifically research based intervention as part of the RTI process. This study shows the potential for Reading Recovery to be an effective intervention in the RTI process that could reduce future costs to school districts by reducing the need for additional services and referrals to special education in later grades.

Hypothesis and Research Question

Hypothesis:

Participation in the Reading Recovery intervention program will have a statistically

significant positive effect on the reading achievement of students performing in the

bottom 20% in reading in first grade, resulting in a statistically significant increase on the

Text Reading Level subtest of the Reading Recovery Observation Survey and

successfully exiting the Reading Recovery program.

Research Question:

Are the improved reading achievement effects of the Reading Recovery program

maintained in third grade and subsequent grades, as measured by performance on the

North Carolina End-of-Grade test in Reading when compared to district-wide average

performance on the EOG?

Methodology

The study is a retrospective longitudinal research study of the sustained effects of the Reading Recovery early intervention program. The reading achievement of RR participants who successfully discontinued the RR program their first-grade year was measured relative to their peers beginning in third grade. The purpose of the measurement was to assess the sustained effects of the RR intervention program on the students' reading performance. Former RR students who successfully completed the reading program, had their reading performance measured using their scores on the Reading portion of the state End-of-Grade test (EOG). The group consisted of five cohort years (school years 2002-03 to2006-07) each group's data ranged from third grade up to seventh grade spanning the 2004-05 to 2008-09 school years. Former RR student EOG scores were compared to average EOG scores for the school district for each respective year which reflected the average performance of their peers.

Participants

The target population for the study were students who successfully exited the Reading Recovery program in first grade from the 2002-03 to 2006-07 school years. The students were either currently, or were at one time, enrolled in the rural North Carolina school district. Elementary and middle schools in the study were comprised of a diverse population from varying socioeconomic backgrounds. District data identifying gender, race, and ethnicity collected along with the state administered EOG information, was also be included.

Study participants were drawn from the available pool of students who successfully

discontinued the RR program in first-grade in the school district and who were either still

enrolled or were enrolled during the study period in a school within the school district between

the 2004-05 and 2008-09 school years.

Measures

The dependent measure in this research study was reading achievement and was

measured using two assessments: the Text Reading Level subtest of the Reading Recovery

assessment the Observation Survey of Early Literacy Achievement and the North Carolina End-

of-Grade test of reading administered during the 2004-05 to 2008-09 school years (Beaver, 2006;

Clay, 2006; NCDPI-DAS/NCTP, 2007).

During the students' participation in the RR program, they were administered the

Observation Survey of Early Literacy Achievement, first developed by Clay (2002). The

Observation Survey is made up of a series of six subtests that describe each student's reading and

writing progress: Letter Identification, Word Test (vocabulary), Concepts About Print, Writing

Vocabulary, Hearing and Recording Sounds in Words (phonemic awareness, representing sounds

in graphic form), and Text Reading. This study focused primarily on the Text Reading level

subtest. On the Text Reading subtest levels are obtained by having the students read leveled

books based on difficulty and characteristics of the text (Peterson, 1991). Teachers assess

students on rate of errors and utilization of self-correction—all the while a *running record* is kept

to monitor their progress. The leveled passages are read by the student until their accuracy falls

below 90%. Their performance is then compared to the number of students who scored at the

first grade level or higher and those who perform lower than first grade. Their accuracy score is

then changed into a numerical reading level (level 0, 1 to 30) and corresponding grade levels (Clay, 2006).

The teachers administering the Reading Recovery program utilize the Observation Survey to assess the progress of students in the program and to determine appropriate discontinuation upon reaching the average level of reading for their peers. A running record is kept on each student's progress on text reading, daily lesson records, students' writing, and change over time in reading and writing vocabulary. Text Reading level decisions in the RR program are based on the running record. Clay (2002) reports a reliability over a two year interval across two scorings with a trained recorder of .98. The Text Reading levels present varying exit criteria depending on the average reading performance of a particular school. Validity and reliability for all tasks of the Observation Survey have been documented (Clay, 2006; Denton, Ciancio, & Fletcher, 2006) and the Observation Survey highly correlates with the *Iowa Test of Basic Skills* (Rodgers, Gómez-Bellengé, Wang, & Schultz, 2005) and (Tang & Gómez-Bellengé, 2007). National norms have been developed to assist in interpreting scores (Gómez-Bellengé & Thompson, 2005).

It was decided the Text Reading Level subtest would be the most indicative of actual reading ability at the time of successfully completing the RR program for the purposes of this study, based on previous studies that have successfully utilized the Text Reading level as an outcome measure. Several studies have utilized a number of subtests from Clay's Observation Survey. The Text Reading subtest is one of the more frequently used tools to measure the effectiveness of the RR program (Baenen, Berhole, Dulaney, & Banks 1997; Iversen & Tunmer, 1993; Pinnell, 1989; Pinnel, Lyons, DeFord, Bryk & Seltzer, 1993/94; Quay, Steele, Johnson, & Hortman, 2001; Schwartz 2005). This measure represents a real world measure of reading as it

involves the student actually reading a leveled passage of text in which they must combine the skills they have learned in the RR program. This skill, or set of skills are what the RR program is attempting to improve—helping the students become more proficient readers.

In North Carolina, beginning at the end of third-grade, students are required to take the state designed End-of-Grade test (EOG). The North Carolina End-of-Grade tests are designed to measure student performance on the goals, objectives, and grade-level competencies specified in the North Carolina Standard Course of Study (NCSCS). There are two portions: Reading and Math. The reading portion measures reading comprehension components of each grade's curriculum and English/Language Arts North Carolina Standard Course of Study (NCSCS). The test is made-up of eight reading passages, each followed by six to nine multiple-choice questions for each passage and was designed to measure reading, thinking and comprehension skills. There are four literary selections (two fiction, one nonfiction, one poem), three informational selections (two content and one consumer), and one embedded experimental selection, that may be fiction, nonfiction, poetry, consumer, or content (Baenen, Dulaney & Banks, 1997; NCSCS; NCDPI-DAS/NCTP, 2007).

The EOG test is administered to each student the last three weeks of the school year and contains 50 items (plus 8 experimental items). It is a measure of reading achievement, as well as a measure of progress made throughout the school year. The two sets of scores, the Developmental Scale Score and Achievement Levels, are derived from the raw scores which are composed of the number of questions the student answered correctly. Scale score ranges vary from grade to grade and also may vary from year to year depending on the year of the EOG administration (e.g. starting with 2007-08 school year, third grade Level I - <330, Level II – 331-337, Level III – 338-349, and Level IV - >350). Achievement Levels are essentially a method of

dividing the range of scale scores on the EOG into four levels (i.e. levels I, II, III, and IV) (NCDPI-DAS/NCTP, 2007).

Both measures, Text Reading Level and the EOG in reading, attempt to measure the reading achievement of the student. However, they differ in several important ways. The Observation Survey is individually administered to each RR student prior to, during and immediately upon completing the RR program. Therefore, the assessment is given very closely upon completing the instruction in the RR program. The EOG, on the other hand, is administered in a group format at the end of the school year following a year of classroom instruction. Much of the material on the EOG may have been taught several months earlier with only minimal review prior to the assessment. Also, at the time the End-of-Grade test in reading is administered the EOG test in math and in later grades science are also administered just days before or soon thereafter. Finally, the Observation Survey was designed to compliment and direct the RR program instruction and its purpose is to measure the progress of individual students during their participation in the RR program. Its function was to inform instruction and not necessarily measure group change, but rather give RR instructors the information necessary to adapt instruction to meet the needs of the student (Clay, 2006; Denton et al, 2006). The EOG is not directly tied to the instruction in the classroom. The EOG is an attempt to measure reading comprehension components of a particular grade's English/Language Arts NC Standard Course of Study (NCSCS).

Procedures

Data collection consisted of obtaining permission from the school district to gain access to the RR data for this study but without identifying information. Archival data from RR's National Data Evaluation Center in Columbus, Ohio was used by the school district to identify

Reading Recovery students who have successfully completed the RR program during the 2002-03 to 2006-07 school years. The school district compared this list to school records to identify students who were still attending schools in the district at the end of the 2008-09 school year. The treatment group consisted of students enrolled in first grade during the 2002-03 to 2006-07 school years and who continued to attend an elementary school in the district through the 2008-09 school years.

Data Analysis

Analysis of the research data involved a series of comparisons of the mean scores and performances, correlations, as well as *t* test for independent means and equivalence testing using one-sample *t* tests in order to compare the means and confidence intervals of test scores with district-wide yearly averages. These analyses will ultimately be used to compare the Reading Recovery students' performance on the Observation Survey subtest, Text Reading Levels, and the EOG score performance for subsequent grades, beginning in third grade. Data collected will be compiled into a database and statistically analyzed by the researcher using Microsoft Excel and the statistical program IBM® SPSS® Statistics (Statistical Package for the Social Sciences). Descriptive statistics (e.g. frequencies), correlations, *t* tests and equivalence testing were applied to the data to determine whether any observed differences in the means were statistically significant (i.e. not due to chance) between the RR students pre- and post program assessments and if student mean End-of-Grade scores in third grade and beyond are equivalent to the school district mean EOG scores.

Limitations

This study was limited by its scope of students. Only RR students who successfully completed the RR program were included in the dataset. This means that potentially, students

who began the program, but were not making adequate progress, may have been prematurely removed from the program. The assumption is of course that these students are removed from the RR program with the intent of providing a different, better fitting intervention to meet their needs. This has been a common criticism of RR and much of the research involving RR, due to the removal of students who are not successful in the program prior to analyzing the program's effectiveness based on the success of those who complete the program. However, it should be clear that for those lower performing students who do remain in the program, the program has an impact on their reading achievement. It was the goal of this research to determine the extent of that impact. The second limitation was found in the lack of a formal comparison or control group. Because the data for this study was extant data and was not collected for the sole purpose of this research study, a control group of either low performing peers or average performing otherwise equivalent peers was not included in this research. This data is routinely collected by the school district, but other more custom data was not available. In addition to the nature of the data collection, the multiple school years and various years of matriculation in the RR program in first grade and subsequent grades among the participants would have made for an overly complex and convoluted experimental design, in addition to the difficulty of matching control groups to this study population using only the data available from the school district. As it stands the data was limited in scope between the cohorts due to attrition and lack of sufficient data on every initial RR student.

Summary

This research provided a measure of the effectiveness of the Reading Recovery program with low performing first grade students in a rural elementary school. More specifically, the level

of the program's effectiveness and sustained effects over time through performance on third-grade and beyond End of Grade test reading achievement. In addition, the study provided a comparison between the internal measures that are utilized by the Reading Recovery program to measure progress and ultimately successful completion (discontinuation) of the program (e.g. Text Reading Level on the Observation Survey) and End of Grade test reading performance. This component of the research study helped address one of the critiques of the Reading Recovery program—that although Reading Recovery instruments show progress and an improvement in reading ability, often independent standardized measures do not show the same gains (Center, Wheldall, & Freeman, 1992). Another critique is that many of the gains that students make upon completing RR soon dissipate and are not sustained over time (Hiebert, 1994; Shanahan & Barr, 1995). This study tracked the reading achievement of students successfully exited from RR and monitored their reading achievement and subsequently measured their performance on the End of Grade test in third-grade and beyond.

Results

The intent of this research was to examine the effectiveness of the RR early intervention reading program—a program which is implemented only in first grade as a preventative measure for the lowest performing 20 percent of students in reading. The aim was to examine the students who had successfully completed the RR program in order to determine the short-term effectiveness of the program (e.g. the reading performance of the students upon completing the program as measured by the Observation Survey, specifically the Text Reading Level subtest) and then to track the reading achievement of the RR students in subsequent grades (third through seventh) using the standardized state assessment the NC End-of-Grade test in reading. The

students in this study consisted of 177 school age students from 19 schools in a rural North Carolina school district, who were enrolled in and successfully discontinued (i.e. successfully completed) the Reading Recovery program in first grade between the 2002-03 to 2006-07 school years. Subsequent North Carolina End-of-Grade (EOG) reading scores were also collected for the students during grades three through seven of the 2004-05 to 2008-09 school years. The students were studied in five cohorts pertaining to the school year in which they were enrolled in first grade between the 2002-03 and 2006-07 school years and then followed longitudinally into third grade through seventh grade (e.g. The 2002-03 cohort received Reading Recovery instruction during first grade which occurred that school year and End-of-Grade test scores were collected for subsequent grades, third grade for this cohort occurred in the 2004-05 school year, etc). Table 1 demonstrates the stratification of the cohorts visually, so that the reader might better understand the data collection schedule and how the year in which a student is enrolled in first grade and participates in the RR program affects the year in which that student is first administered the End-of-Grade test in third and subsequent grades.

Table 1

Five cohort years spanning the 2002-03 to 2008-09 school years and years in which they participated in RR and the subsequent years/grades for which EOG test data are available.

| | | | | | EOG Scores Re-Normed | |
2002-2003	2003-2004	2004-2005	2005-2006	2006-2007	2007-2008	2008-2009
2002-03 COHORT 1 1st Grade Reading Recovery	*No Data (2nd Grade)*	**2002-03** Cohort **3rd Grade** EOG Data	**2002-03** Cohort **4th Grade** EOG Data	**2002-03** Cohort **5th Grade** EOG Data	**2002-03** Cohort **6th Grade** EOG Data	**2002-03** Cohort **7th Grade** EOG Data
	2003-04 COHORT 2 1st Grade Reading Recovery	*No Data (2nd Grade)*	**2003-04** Cohort **3rd Grade** EOG Data	**2003-04** Cohort **4th Grade** EOG Data	**2003-04** Cohort **5th Grade** EOG Data	**2003-04** Cohort **6th Grade** EOG Data
		2004-05 COHORT 3 1st Grade Reading Recovery	*No Data (2nd Grade)*	**2004-05** Cohort **3rd Grade** EOG Data	**2004-05** Cohort **4th Grade** EOG Data	**2004-05** Cohort **5th Grade** EOG Data
			2005-06 COHORT 4 1st Grade Reading Recovery	*No Data (2nd Grade)*	**2005-06** Cohort **3rd Grade** EOG Data	**2005-06** Cohort **4th Grade** EOG Data
				2006-07 COHORT 5 1st Grade Reading Recovery	*No Data (2nd Grade)*	**2006-07** Cohort **3rd Grade** EOG Data
					EOG Scores Re-Normed	

The analysis sample ranged in size from 177 to 22 depending on the analysis conducted or cohort year due to missing or unavailable data and the fact that some student cohorts had only recently completed third, fourth, fifth, sixth, and seventh grade at the time of this study and subsequent grade information did not exist. The entire sample was made up of 39.5% female, 60.5% male, 46.9% African American, 31.1% Caucasian, 20.3% Hispanic, and 1.7% multi-racial. The school district's racial composition is made up of 57.4% Caucasian, 35.2% African American, 5.0% Hispanic, and 0.9% multi-racial (NCDPI, n.d.).

It should be noted, that the NC State Department of Public Instruction allows districts to administer the EOG to students more than once in the same grade based on their initial performance, in an effort to improve a student's test performance. For example, if a student scores a level II, then the school might decide to remediate those students and have them retake the EOG a few days later. This often results in a higher EOG score than their original performance, although a few student scores decreased slightly. In this sample of students, this was not always the case in every school (some students who scored a level II were not retested), but many of the students who initially scored a level II on the EOG were retested resulting in two, sometimes three separate EOG scores for a particular year/grade. In this study it was important to capture the students' best performance once they had exited the RR program, therefore it was decided to use their highest score on the EOG for each grade whether that was their last score for a given year after re-testing or in rare cases an earlier score for that grade that they performed better on.

Before moving on to the results, it should also be noted that this particular set of data encompasses an anomaly in the End-of-Grade test scores and Achievement Levels. The NC Department of Public Instruction rescaled the EOG scoring prior to the 2007-08 school year (NCDPI-DAS/NCTP, 2007). Score ranges were rescaled for the four Achievement Levels (e.g. I, II, III, & IV). The result was that the calculation of the number of correct items render a different score changing the score range for a III in third grade, for example, from a 240-249 to 338-349 (see Table 2). Table 2 presents the range of scores and their corresponding Achievement Levels and grades prior to the score change (prior to 2007-08) and then after the score changes (starting in the 2007-08 school year). This is important to note since, depending on the grade a student is in and the year in which they were in that grade, their developmental scale score on the EOG

44

may be a Level II or a Level III. This not only makes comparison of the EOG scores from year-to-year difficult, but also makes analyses across year cohorts more complex than simply comparing scores.

Table 2

EOG Scale Score ranges shifted for Achievement Levels prior to 2007-08 and then after the shift starting with the 2006-07 school year. Data retrieved from the NC Dept. of Public Instruction website (NCDPI-DAS/NCTP, 2007).

Subject/Grade		Level I	Level II	Level III	Level IV
Reading					
Prior to	3	≤ 229	230-239	240-249	≥ 250
2007-08	4	≤ 235	236-243	244-254	≥ 255
School Year	5	≤ 238	239-246	247-258	≥ 259
	6	≤ 241	242-251	252-263	≥ 264
	7	≤ 242	243-251	252-263	≥ 264
	8	≤ 243	244-253	254-265	≥ 266
Reading	3	≤ 330	331-337	338-349	≥ 350
Starting with	4	≤ 334	335-342	343-353	≥ 354
2007-08	5	≤ 340	341-348	349-360	≥ 361
School Year	6	≤ 344	345-350	351-261	≥ 362
	7	≤ 347	348-355	356-362	≥ 363
	8	≤ 349	350-357	358-369	≥ 370

Program Effectiveness

The first research question is whether low performing students improved in their overall reading achievement after participating in the Reading Recovery program. To answer this

question we first looked at the assessment tool that is used in the Reading Recovery program. The Observation Survey of Early Literature Achievement is used in the RR program to measure student progress while in the program. This study specifically focused on the Text Reading level subtest which is one of several indicators used by researchers to measure the efficacy of the RR program. Scores on the Text Reading level subtest range from 0, 1 to 30. Mean Text Reading levels for the group of RR students in this study are presented in Figure 1.

Mean Scores on Text Reading Level subtest of the Observation Survey
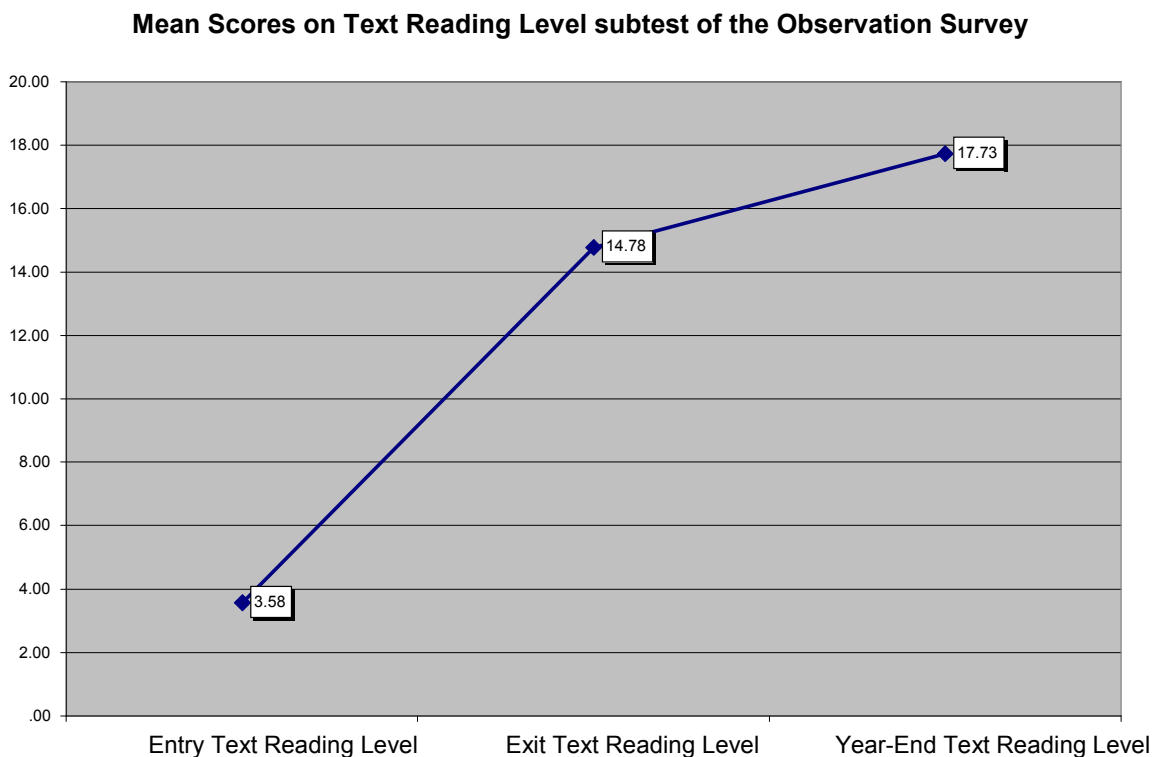
Figure 1. Mean Scores on the Text Reading Level Subtest of the Observation Survey at Entry into the RR program, upon Exit and then at Year-End.

The mean Text Reading level for the students as a whole in this study at the time of entry into the RR program was a level 3 ($M = 3.58$, $SD = 2.78$). Individual student scores at the time of entry varied from 0 to 14 with most (81.1%) receiving a level of 5 or lower. Students' Text

46

Reading levels were monitored throughout their participation in the program. Upon successfully

exiting the RR program the group mean Text Reading level had jumped to a level 14 (M = 14.78,

SD = 3.88). This is an increase of 11 levels in the 20 weeks or less that the participants were in

the RR program. This consisted of 16.0%, 28.2%, and 24.9% of the students ending on levels 10,

16, and 18 respectively. The students were evaluated again at the end of the school after exiting

the program. The mean Text Reading level for the group was a level 17 (*M* = 17.73, *SD* = 3.46),

an increase of on average three additional levels from the time of exiting the program until the

end of the year and an average increase of 14 levels since they first entered the RR program. At

year's end 75.9% of the students had scores at level 16 or level 18. This finding demonstrates

that in the short-term the RR students appeared to continue improving their reading achievement

by applying the skills they learned in the RR program.

When we look at the individual year cohorts similar results emerged. Table 3 summarizes

the results for the individual cohorts spanning the 2002-03 school year through the 2006-07

school year.

Table 3

Entry, Exit, & Year-End Mean Text Reading Levels Divided into Year Cohorts.

2002-03 Cohort	Mean	Std. Deviation	Minimum	Maximum	N
Entry Text Reading Level	3.00	2.612	0	7	35
Exit Text Reading Level	13.77	3.647	8	24	35
Year-End Text Reading Level	17.37	2.157	14	24	35
2003-04 Cohort	**Mean**	**Std. Deviation**	**Minimum**	**Maximum**	**N**
Entry Text Reading Level	4.03	3.640	0	12	29
Exit Text Reading Level	14.76	4.050	8	24	29
Year-End Text Reading Level	17.52	2.487	12	24	29
2004-05 Cohort	**Mean**	**Std. Deviation**	**Minimum**	**Maximum**	**N**
Entry Text Reading Level	3.19	3.167	0	14	32
Exit Text Reading Level	14.75	4.355	8	26	32
Year-End Text Reading Level	17.84	4.978	9	30	32
2005-06 Cohort	**Mean**	**Std. Deviation**	**Minimum**	**Maximum**	**N**
Entry Text Reading Level	3.78	1.874	0	7	49
Exit Text Reading Level	15.53	3.641	9	24	49
Year-End Text Reading Level	18.29	3.596	14	30	48
2006-07 Cohort	**Mean**	**Std. Deviation**	**Minimum**	**Maximum**	**N**
Entry Text Reading Level	3.90	2.820	0	12	30
Exit Text Reading Level	14.77	3.775	9	22	30
Year-End Text Reading Level	17.33	3.417	8	30	30

Each cohort increased on average 10 to 14 Text Reading levels from their initial reading

levels while in the Reading Recovery program. This is in addition to improvement on other areas

of the Observation Survey not examined here. This finding demonstrates consistent performance

across cohort years, indicating that once implemented in this school district, the program's

effectiveness on this subtest remained the same from year-to-year at least over the years of this

study.

These first set of results address one of the research questions for this study. The

hypothesis states that participation in the RR program will have a statistically significant positive

effect on the reading achievement of the lower performing students who participated in the RR

program, which will result in a statistically significant increase in their Text Reading level. The

analyses of these data resulted in a statistically significant effect for the RR program, $t(174) =$

54.21, $p < .05$, with the students significantly improving their Text Reading Level after successfully completing the early interventions program.

Long-term Effectiveness

The next research question is whether the improved reading achievement effects of the RR program are maintained in subsequent grades, as measured by students' performance on a district wide standardized assessment, in this case the North Carolina End-of-Grade test in Reading, when compared to district-wide average performance on the EOG. The initial analysis begins with the end of third grade which is the first time the NC EOG test is administered. Had these former RR students maintained the gains that they had made in the RR program?

In order to answer this research question the analysis was focused on the former RR students' mean Developmental Scale Scores on the End-of-Grade test in reading. This was compared to the mean EOG score performance of the school district at the time of the EOG's administration. The mean EOG score, confidence interval, and standard deviation were calculated for each of the five cohorts. The analyses are divided into individual cohorts since each cohort had a unique year in which they were enrolled in third grade (and subsequent grades) and therefore were compared to the specific third grade (and successive grades) mean EOG score for the school district for that particular year.

Table 4 focuses on the third grade mean EOG scores for each cohort which is the first EOG administration since the RR program. This table also includes the 95% confidence intervals and standard deviations. Equivalence testing was performed for the third grade EOG with confidence intervals of ± 5 points as the *zone of clinical indifference* (a predefined range of equivalence) in this case around the district mean EOG score for that grade. Essentially this is plus or minus the mean school district EOG score as the interval around the *standard* against

which we compare the confidence intervals (CI) of each of the RR cohorts. In this situation, when comparing a sample with a "standard comparator" it is essential to show that the sample is sufficiently similar to the standard to be "clinically indistinguishable" (Cleophas, Zwinderman, & Cleophas, 2006, p.63). If 95% CIs of the sample fall completely within the zone of indifference it can be concluded that equivalence is demonstrated and therefore the RR students have maintained their reading achievement levels from first grade when they were achieving average to their peers upon exiting the RR program. Confidence intervals completely outside the zone of indifference are considered not equivalent to the standard. If confidence intervals cross into the zone of indifference (i.e. part in, part out), equivalence cannot be determined.

When the results are reviewed in Table 4, it is clear that the 95% confidence intervals from each of the cohorts in third grade do not fall completely within the zone of indifference for each of the district mean EOG scores, the predefined standard. For example, for the 2002-03 Year Cohort the zone of clinical indifference was established as 242.3 and 252.3. The 95% CI [238.76, 242.79] for the 2002-03 cohort does not fall completely within the zone of indifference and therefore equivalence cannot be established (although there is some overlap).

Table 4

Third (3rd) Grade Mean and Confidence Intervals for Year Cohorts compared to School District

Mean EOG scores.

Cohort	Grade	Year	Mean EOG	SD	Confidence Interval		Zone of Indifference		N
2002-03	3rd	2004-05	240.78	5.96	238.76	242.79	242.30	252.30	36
2003-04	3rd	2005-06	242.17	4.69	240.42	243.92	242.70	252.70	30
2004-05	3rd	2006-07	238.97	7.40	236.30	241.64	242.20	252.20	32
2005-06	3rd	2007-08	331.16	9.53	328.43	333.90	332.80	342.80	49
2006-07	3rd	2008-09	326.76	7.95	323.74	329.78	332.00	342.00	29

Note: EOG = End-of-Grade; SD = Standard Deviation.

Moving forward into the fourth grades of each of the cohorts a similar pattern emerges of close approximation to the district mean, but only small amounts of overlap, if at all, with the ± 5 points zone of clinical indifference.

Table 5

Fourth (4th) Grade Mean and Confidence Intervals for Year Cohorts compared to School District

Mean EOG scores.

Cohort	Grade	Year	Mean EOG	SD	Confidence Interval		Zone of Indifference		N
2002-03	4th	2005-06	243.43	6.65	241.14	245.71	246.10	256.10	35
2003-04	4th	2006-07	243.89	7.05	241.16	246.63	247.10	257.10	28
2004-05	4th	2007-08	337.09	8.23	334.13	340.06	338.80	348.80	32
2005-06	4th	2008-09	337.72	8.10	335.09	340.34	338.70	348.70	39
2006-07	-	-	-	-	-	-	-	-	-

Note: EOG = End-of-Grade; SD = Standard Deviation.
[a]No 4th grade data were obtained for the 2006-07 Cohort as they had not matriculated in this grade at the time of this study.

Table 6

Fifth (5th) Grade Mean and Confidence Intervals for Year Cohorts compared to School District

Mean EOG scores.

Cohort	Grade	Year	Mean EOG	SD	Confidence Interval		Zone of Indifference		N
2002-03	5th	2006-07	**247.32**	7.56	**244.69**	**249.96**	**251.30**	**261.20**	34
2003-04	5th	2007-08	**341.25**	6.88	**338.58**	**343.92**	**343.50**	**353.50**	28
2004-05	5th	2008-09	**343.68**	7.64	**340.29**	**347.07**	**344.00**	**354.00**	22
2005-06	-	-	-	-	-	-	-	-	-
2006-07	-	-	-	-	-	-	-	-	-

Note: EOG = End-of-Grade; SD = Standard Deviation.
^aNo 5th grade data were obtained for the 2005-06 and 2006-07 Cohorts as they had not matriculated in this grade at the time of this study.

Table 7

Sixth (6th) Grade Mean and Confidence Intervals for Year Cohorts compared to School District

Mean EOG scores.

Cohort	Grade	Year	Mean EOG	SD	Confidence Interval		Zone of Indifference		N
2002-03	6th	2007-08	**342.00**	7.79	**339.32**	**344.68**	**346.50**	**356.50**	35
2003-04	6th	2008-09	**345.08**	6.71	**342.37**	**347.79**	**346.10**	**356.10**	26
2004-05	-	-	-	-			-	-	-
2005-06	-	-	-	-			-	-	-
2006-07	-	-	-	-			-	-	-

Note: EOG = End-of-Grade; SD = Standard Deviation.
^aNo 6th grade data were obtained for the 2004-05, 2005-06, and 2006-07 Cohorts as they had not matriculated in this grade at the time of this study.

Table 8

Seventh (7th) Grade Mean and Confidence Intervals for Year Cohorts compared to School

District Mean EOG scores.

Cohort	Grade	Year	Mean EOG	SD	Confidence Interval		Zone of Indifference		N
2002-03	7th	2008-09	349.56	6.45	347.00	352.11	349.30	359.30	27
2003-04	-	-	-	-			-	-	-
2004-05	-	-	-	-			-	-	-
2005-06	-	-	-	-			-	-	-
2006-07	-	-	-	-			-	-	-

Note: EOG = End-of-Grade; SD = Standard Deviation.
^aNo 7th grade data were obtained for the 2003-04, 2004-05, 2005-06 and 2006-07 Cohorts as they had not matriculated in this grade at the time of this study.

Achievement Levels Comparison

In light of the results of these last analyses it should be emphasized that, although equivalence was not established in any of the cohorts, the purpose of the analysis was to determine if the RR students had maintained their gains made during the RR program in first grade. The metric by which the school district and the state of North Carolina measure proficiency in students is by looking at the number of students whose scores result in Achievement levels of III or IV. On the EOG, there are four possible Achievement Levels; I, II III, and IV, each representing a range of scores depending on the year in which the EOG was administered and the grade the student is in (see previously Table 2). Students who obtain a level III or IV are considered to be performing at proficiency per the NC Dept of Public Instruction (NCDPI).

Table 9 shows the percentage of RR students who achieved a level III or IV beginning in third-grade through seventh-grade. In third grade 44% of RR students scored a III or IV on the

EOG. In fourth grade 40% of RR students scored a III or IV. In subsequent grades the percentage dropped to 35% in fifth grade, then to 20% in sixth and 26% in seventh grade. It should also be noted that the number of participants who had completed these subsequent grades is also smaller.

Table 9

Third-grade Through Seventh Grade EOG Results for RR Students (All Cohorts Combined): Percentage of Students Scoring At or Above Grade Level (Levels III and IV).

Year	N	Level I	Level II	Level III	Level IV	Levels III & IV
3rd Grade	176	29.0	27.3	39.2	4.5	**43.7%**
4th Grade	134	25.4	34.3	37.3	3.0	**40.3%**
5th Grade	84	32.1	33.3	33.3	1.2	**34.5%**
6th Grade[a]	61	52.5	27.9	19.7	-	**19.7%**
7th Grade[a]	27	37.0	37.0	25.9	-	**25.9%**

Note: EOG = End-of-Grade; RR = Reading Recovery
[a]No Level IV's were obtained by these groups in sixth and seventh grade.

The following tables show the percentage of students scoring at or above grade level in grades three through seventh grade, broken down into year cohorts spanning the 2002-03 school year to the 2006-07 school year. Table 10 shows the 2002-03 cohort year. In third grade, 67% of the RR students achieved a level III or IV. In fourth grade, the percentage dropped to 52% and then dropped significantly to only 14%, then up to 26% in seventh grade.

Table 10

2002-03 Cohort Year Third-grade Through Seventh Grade EOG Results: Percentage of Students

Scoring At or Above Grade Level (Levels III and IV).

Year	N	Level I	Level II	Level III	Level IV	Levels III & IV
3rd Grade	36	5.6	27.8	58.3	8.3	66.6
4th Grade	35	14.3	34.3	48.6	2.9	51.5
5th Grade	34	11.8	32.4	52.9	2.9	55.8
6th Grade[a]	35	51.4	34.3	14.3	-	14.3
7th Grade[a]	27	37.0	37.0	25.9	-	25.9

Note: EOG = End-of-Grade; RR = Reading Recovery
[a]No Level IV's were obtained by these groups in sixth and seventh grade.

Table 11 shows the 2003-04 cohort year. In third grade, 70% of the RR students achieved a level III or IV. In fourth grade, the percentage decreased to 54% and then decreased significantly to only 14%, then rose to 27% in seventh grade.

Table 11

2003-04 Cohort Year Third-grade Through Seventh Grade EOG Results: Percentage of Students

Scoring At or Above Grade Level (Levels III and IV).

Year	N	Level I	Level II	Level III	Level IV	Levels III & IV
3rd Grade	30	-	30.0	63.3	6.7	70.0
4th Grade	28	17.9	28.6	50.0	3.6	53.6
5th Grade	28	53.6	32.1	14.3	-	14.3
6th Grade[a]	26	53.8	19.2	26.9	-	26.9
7th Grade[a]	-	-	-	-	-	-

Note: EOG = End-of-Grade; RR = Reading Recovery
[a]No Level IV's were obtained by these groups in sixth and seventh grade.

Table 12 shows the 2004-05 cohort year. In third grade, 59% of the RR students achieved

a level III or IV, less than the previous two cohorts. In fourth grade, the percentage dropped to

31% and then dropped to 27% in fifth grade.

Table 12

2004-05 Cohort Year Third-grade Through Seventh Grade EOG Results: Percentage of Students

Scoring At or Above Grade Level (Levels III and IV).

Year	N	Level I	Level II	Level III	Level IV	Levels III & IV
3rd Grade	32	12.5	28.1	56.3	3.1	**59.4**
4th Grade[a]	32	34.4	34.4	31.3	-	**31.3**
5th Grade[a]	22	36.4	36.4	27.3	-	**27.3**
6th Grade[b]	-	-	-	-	-	-
7th Grade[b]	-	-	-	-	-	-

Note: EOG = End-of-Grade; RR = Reading Recovery
[a]No Level IV's were obtained by these groups in fourth and fifth grade.
[b]No data was available for this cohort for sixth and seventh grade as they had not matriculated in those grades at the time of this study.

Table 13 shows the 2005-06 cohort year. In third grade, 23% of the RR students achieved

a level III or IV, lower than any previous cohort. In fourth grade, the percentage was slightly

higher, 28%.

Table 13

2005-06 Cohort Year Third-grade Through Seventh Grade EOG Results: Percentage of Students

Scoring At or Above Grade Level (Levels III and IV).

Year	N	Level I	Level II	Level III	Level IV	Levels III & IV
3rd Grade	49	55.1	22.4	18.4	4.1	**22.5**
4th Grade	39	33.3	38.5	23.1	5.1	**28.2**
5th Grade[a]	-	-	-	-	-	-
6th Grade[a]	-	-	-	-	-	-
7th Grade[a]	-	-	-	-	-	-

Note: EOG = End-of-Grade; RR = Reading Recovery
[a]No data was available for this cohort for fifth through seventh grade as they had not matriculated in those grades at the time of this study.

Table 14 shows the final cohort, the 2006-07 cohort year. In third grade only 7% of the RR students in this cohort scored a level III, there were no level IVs in this group.

Table 14

2006-07 Cohort Year Third-grade Through Seventh Grade EOG Results: Percentage of Students

Scoring At or Above Grade Level (Levels III and IV).

Year	N	Level I	Level II	Level III	Level IV	Levels III & IV
3rd Grade[a]	29	62.1	31.0	6.9	-	**6.9**
4th Grade[b]	-	-	-	-	-	-
5th Grade[b]	-	-	-	-	-	-
6th Grade[b]	-	-	-	-	-	-
7th Grade[b]	-	-	-	-	-	-

Note: EOG = End-of-Grade; RR = Reading Recovery
[a]No Level IV's were obtained by these groups in fourth and fifth grade.
[b]No data was available for this cohort for fourth through seventh grade as they had not matriculated in those grades at the time of this study.

If the percentage of just third grade EOG achievement levels (I, II, III, IV) by cohort is shown in Table 15, a difference can be seen in the year cohort performance on the EOG Reading. The earlier cohorts (2002-03, 2003-04, and 2004-05) have a higher percentage of students who scored either a III or IV. The later cohorts (2005-06, 2006-07) had fewer level III's and IV's and a higher percentage of level I's. The 2006-07 cohort had a slightly smaller sample size (n=29), but the 2005-06 cohort had 49 students, more than any other cohort. This difference in percentage of IIIs and IVs in the 2005-06 and 2006-07 cohorts was directly related to the district (and statewide) shift or re-norming of the range of scores that correspond with the various achievement levels mentioned previously. The 2005-06 cohort was enrolled in third grade during the first year after the re-norming and the 2006-07 cohort was in third grade the following year.

Table 15

Third-Grade EOG Results: Percentage of Students Scoring At or Above Grade Level (Levels III and IV).

Year Cohort	N	Grade	Level I	Level II	Level III	Level IV	Levels III & IV
2002-03	36	Third	5.6	27.8	58.3	8.3	**66.6**
2003-04[a]	30	Third	-	30.0	63.3	6.7	**70.0**
2004-05	32	Third	12.5	28.1	56.3	3.1	**59.4**
2005-06	49	Third	55.1	22.4	18.4	4.1	**22.5**
2006-07	29	Third	62.1	31.0	6.9	-	**6.9**

Note: EOG = End-of-Grade; RR = Reading Recovery
[a]No Level I's were obtained by this group in third grade.

Third-Grade EOG Results: Percentage of Students Scoring At or Above Grade Level (Levels III and IV)
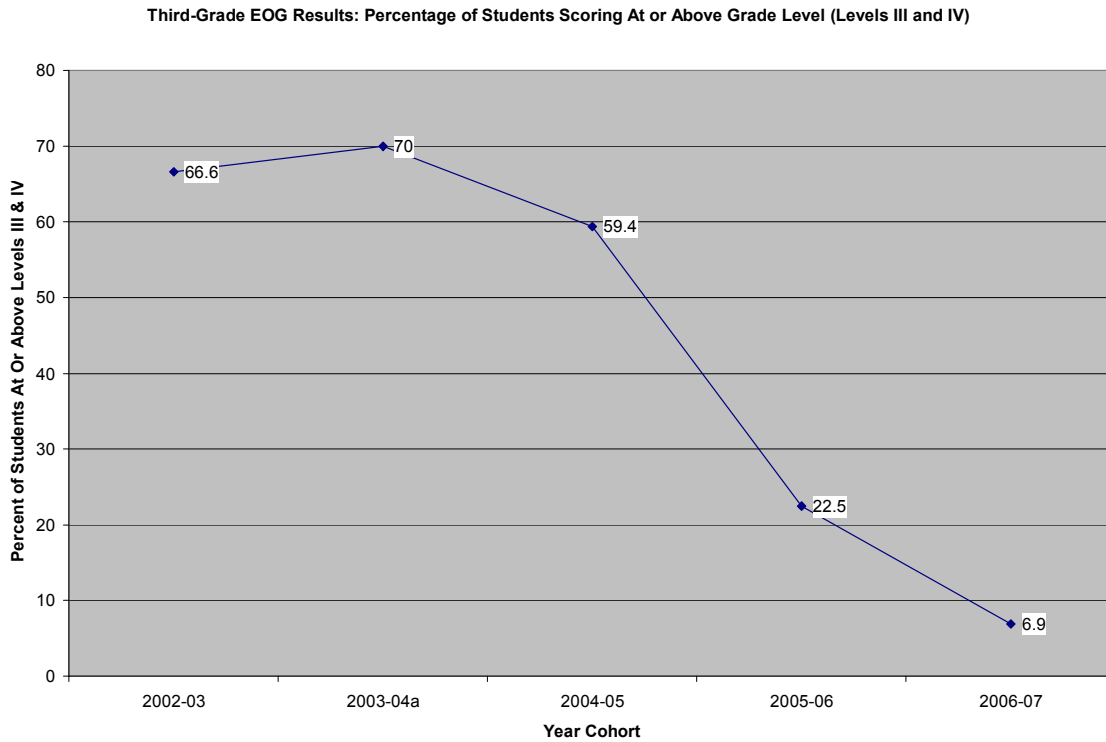

Figure 2. Third Grade EOG results: Percentage of RR students at or above grade level (Levels III and IV).

Upon reviewing the data made available by the NC Department of Public Instruction for those years there is a clear difference in overall percentage of students across the district scoring in the III or IV range shown in Figure 3.
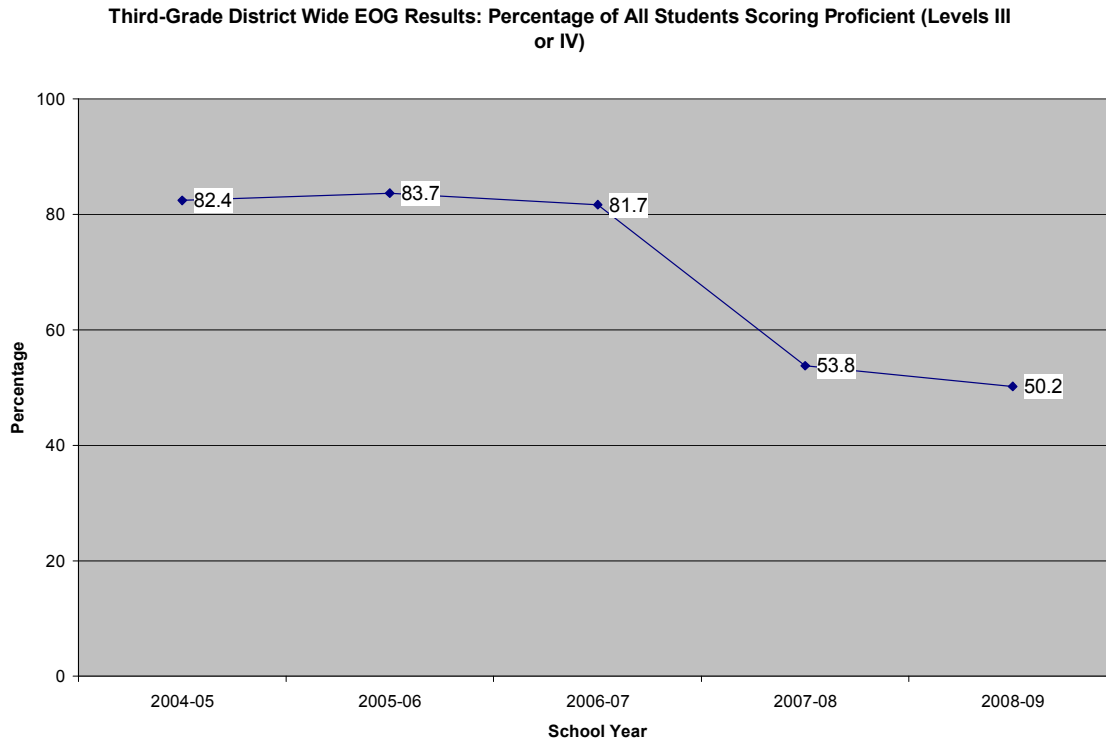
Third-Grade District Wide EOG Results: Percentage of All Students Scoring Proficient (Levels III or IV)

Figure 3. Third Grade EOG results district wide: Percentage of all students scoring proficient (Levels III and IV).

Correlations

In addition to the hypothesis and research question posed by this study—it would also be advantageous to know if the Text Reading Level subtest could be used as an indicator of future reading achievement (e.g. on the EOG). The Text Reading levels of the Observation Survey range from 0, 1 to 30. Upon successfully exiting the RR program the participants in this study were on levels ranging from 8 to 30. Given this range, was there a connection between Exit Text Reading levels and EOG scores in reading in later grades? In other words, did a higher Text Reading level increase chances of higher EOG scores in reading? Also, is the inverse true; for students whose ending Text Reading levels were low, were they also more likely to have low

EOG scores in reading in later grades?

This analysis first focuses on the correlation between the Text Reading levels at the end of the RR program with all of the participants combined. Figure 4 compares the Exit Text Reading Level of all of the RR participants with their combined third grade EOG scores. As you can see there is a distinct split between the data points. This is due to the re-norming of the EOG scores mentioned earlier. This comparison includes all of the third grades from the various year cohorts which include the years prior to the re-norming (2004-05 through 2006-07) and the years following the change (2007-08 through 2008-09). Since the change involved essentially shifting the score ranges for a I, II, III, and IV upwards approximately 100 points, the correlation shown in Figure 4 divides into two clusters. One cluster representing the third grades for cohorts prior to 2007-08 school year (2002-03, 2003-04, 2004-05) when the score scale change was made and the other cluster representing the cohorts whose third grades occur in the 2007-08 and 2008-09 school years (cohorts 2005-06 and 2006-07) after the shift. When all of the cohorts are combined in this manner a weak correlation of $r(174) = .132, p > .05$ is found between the Exit Text Reading level and the third grade EOG scores represented in Figure 4.
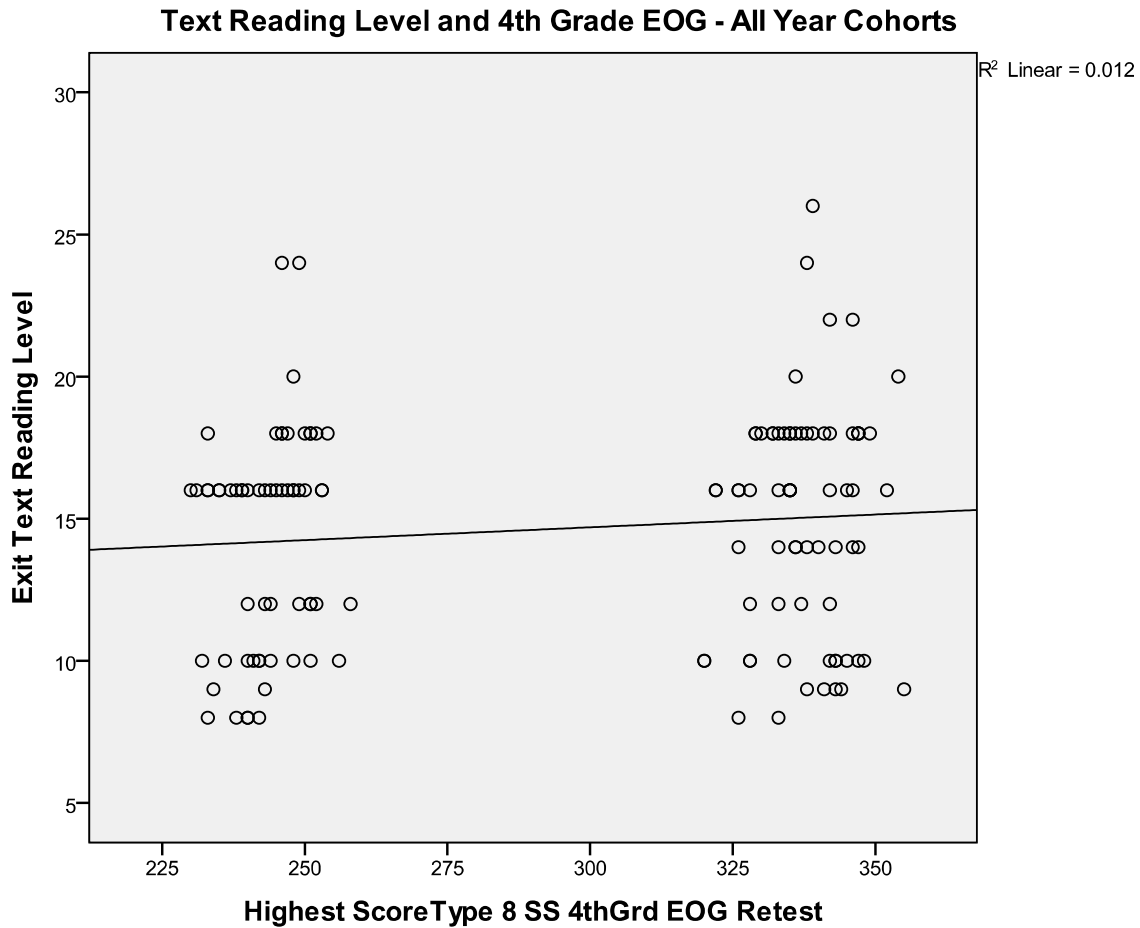
Text Reading Level and 3rd Grade EOG Scores - All Year Cohorts
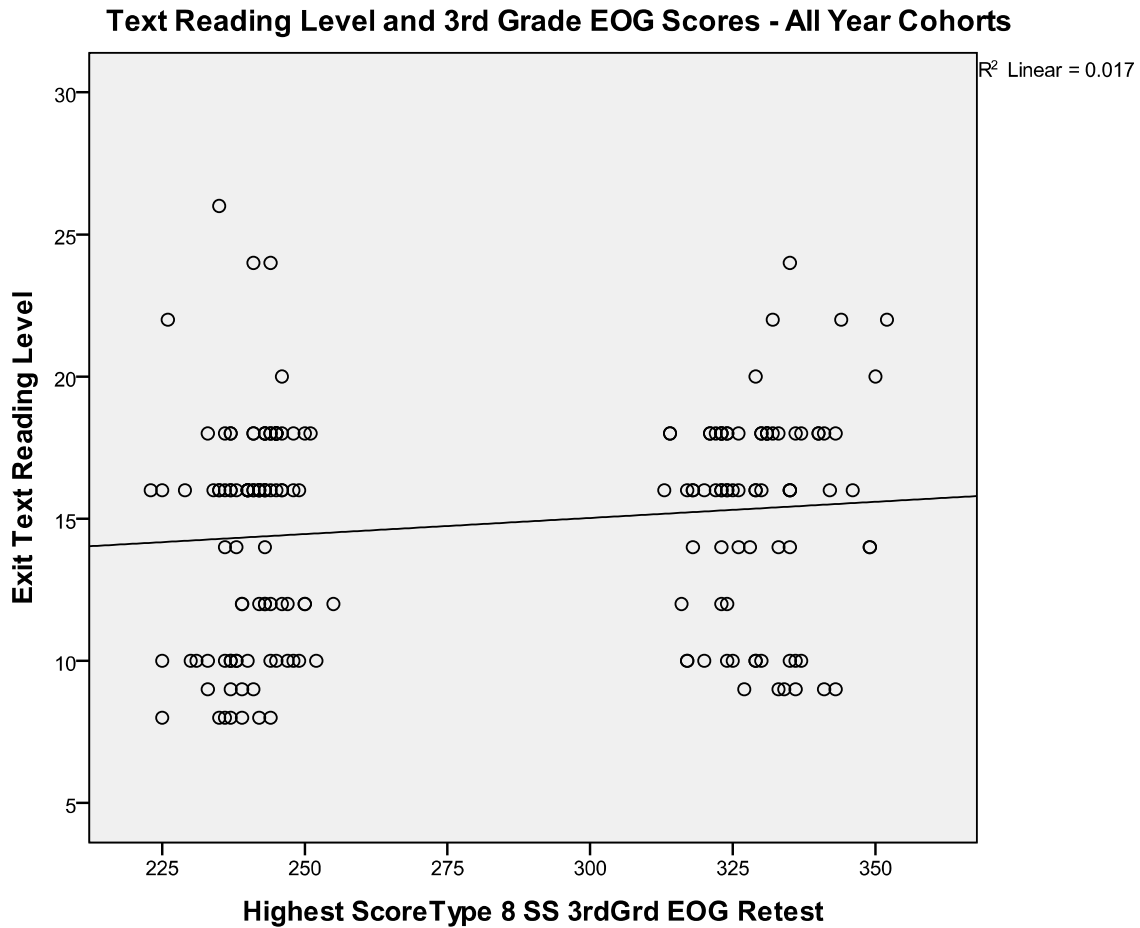
Figure 4. Scatterplot of Text Reading Level and 3rd Grade EOG Scores for All Year Cohorts.

When a correlation was conducted with all of the cohorts in their fourth grade year the same split occurred in the data represented in Figure 5. Figure 5 compares the Exit Text Reading Level of all of the RR to their combined fourth grade EOG scores. When all of the cohorts were combined in Figure 5 a weak correlation of $r(130) = .108, p > .05$, was found between the Exit Text Reading level and the fourth grade EOG scores.

Figure 5. Scatterplot of Text Reading Level and 4th Grade EOG Scores – All Year Cohorts.

In Figure 6, a correlation with all of the cohorts in their fifth grade year is represented and the effects of the scale score shift were still seen, as well. Figure 6 compares the Exit Text Reading Level of all of the RR students to their combined fifth grade EOG scores. A weak correlation of $r(80) = .137, p > .05$ can be seen between the Exit Text Reading level and the fifth grade EOG scores.

Text Reading Level and 5th Grade EOG - All Year Cohorts
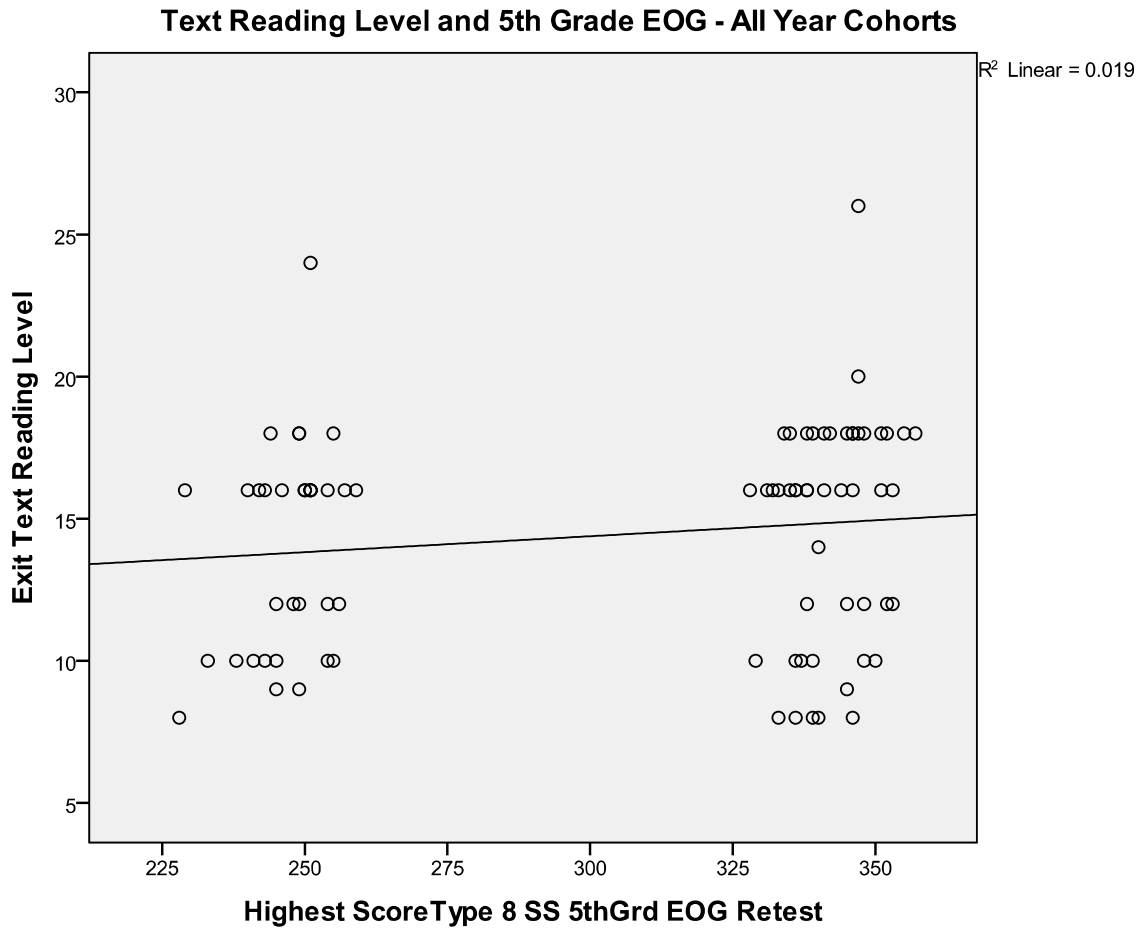
Figure 6. Scatterplot of Text Reading Level and 5th Grade EOG Scores – All Year Cohorts.

In Figure 7, Text Reading level upon exit of the RR program and sixth grade EOG scores for all of the cohorts combined were compared and the effects of the scale score shifts are no longer seen. This is due to the fact that the available sixth grade data only occurred after the scale score shift in 2007-08 and 2008-09 and only included the two longest running cohorts (2002-03 and 2003-04). Figure 7 compares the Exit Text Reading Level of all of the RR students to their combined sixth grade EOG scores and only a weak correlation of $r(57) = .124$, $p > .05$ between the Exit Text Reading level and the sixth grade EOG scores is observed.

Text Reading Level and 6th Grade EOG - All Year Cohorts
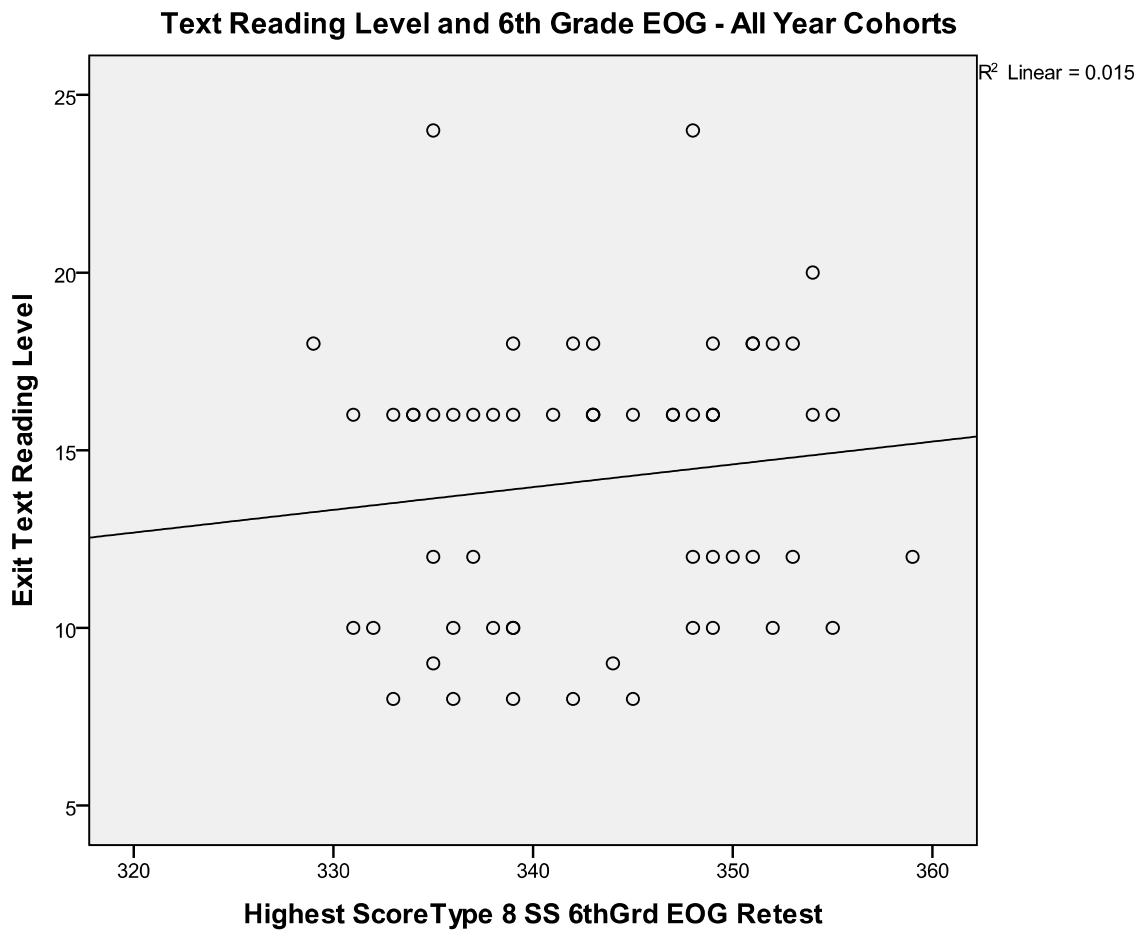
Figure 7. Scatterplot of Text Reading Level and 6th Grade EOG Scores – All Year Cohorts.

The final comparison was the seventh grade year and only the 2002-03 cohort had seventh grade data available, since they were the only ones who had completed seventh grade at the time this data was collected. Figure 8 shows the Exit Text Reading Level of the RR students compared with their seventh grade EOG scores. The result was a weak, but somewhat higher, correlation of $r(24) = .314$, $p > .05$. This was most likely due to the fact that this analysis included only one cohort, 2002-03, which had matriculated in that grade. As will be seen in later

analyses, this particular cohort had a higher correlation between Exit Text Reading level and EOG scores than any other cohort.

Text Reading Level and 7th Grade EOG - All Year Cohorts
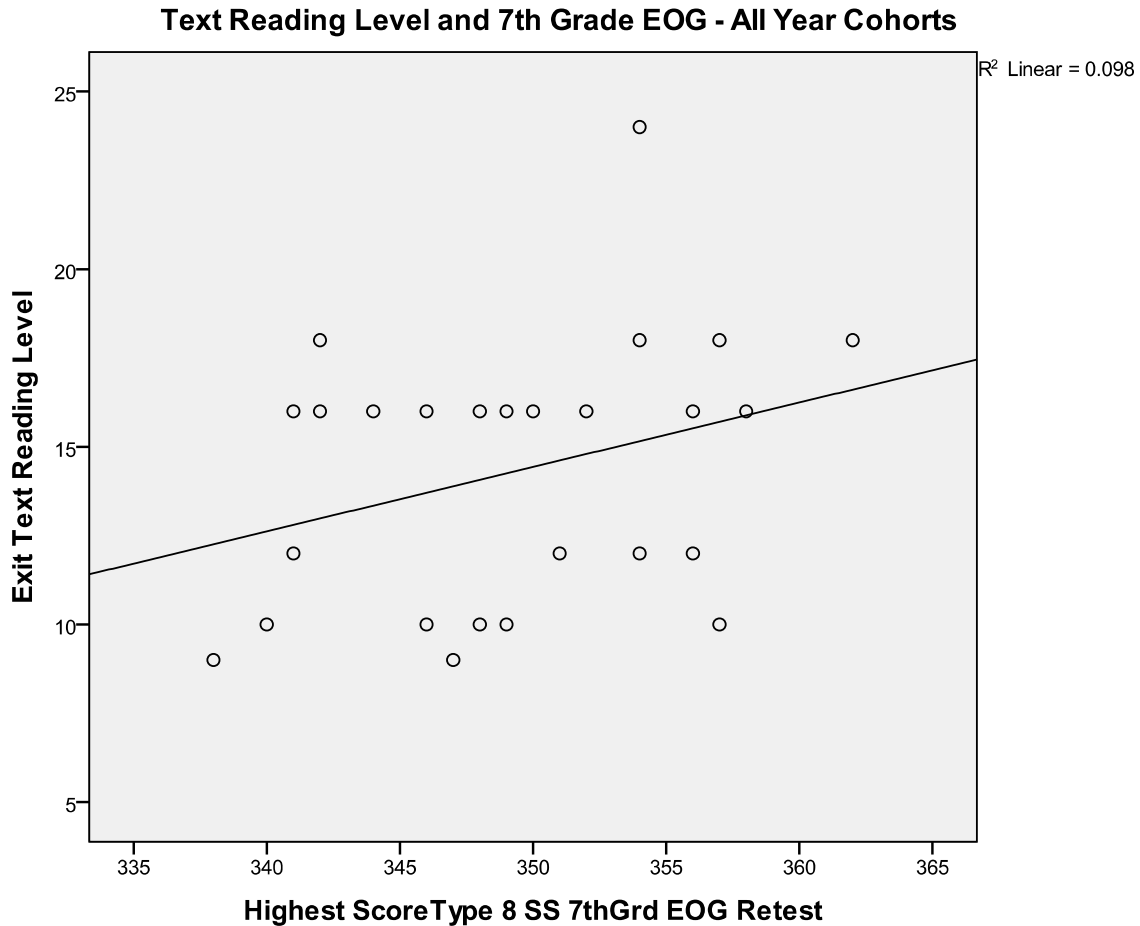
Figure 8. Scatterplot of Text Reading Level and 7th Grade EOG Scores – All Year Cohorts.

Overall, when the study data was considered in its entirety, there was little correlation between the Text Reading level upon exiting RR and the third through seventh EOG scores. The third grade EOG scores were the first standardized assessment available after the students completed the RR program. Because they were administered less than two years after the student's participation in the program, it would stand to reason that they would also be the most

likely to show residual predictive effects of the RR program. Unfortunately there was very little correlation among these variables. If the third grade EOG scores are divided into two groups, one composed of the former scale score range, the Pre-Re-Norm group, and another group made up of the newer re-normed scale score range, Post- Re-Norm group, the two groups separately continued to show a weak correlation between the Text Reading level and third grade EOG [$r(94) = .092, p > .05$ and $r(76) = .105, p > .05$ respectively] (See Appendix A).

In an attempt to further determine the predictive power of the Text Reading level scores, the correlations by year cohort were examined and some subtle distinctions could be seen. The first cohort was the 2002-03 Year Cohort. This was the earliest implementation of the RR program with regards to the other year cohorts in this study and was the cohort for which there was the most data covering the most school years all the way through seventh grade. Figure 9 shows a weak correlation $r(33) = .246, p > .05$ between the 2002-03 cohort's Exit Text Reading level and the third grade EOG scores. Scatterplots for the remaining grades for the 2002-03 year cohort and the subsequent cohorts (2003-04, 2004-05, 2005-06, and 2006-07) can be found in Appendix B.
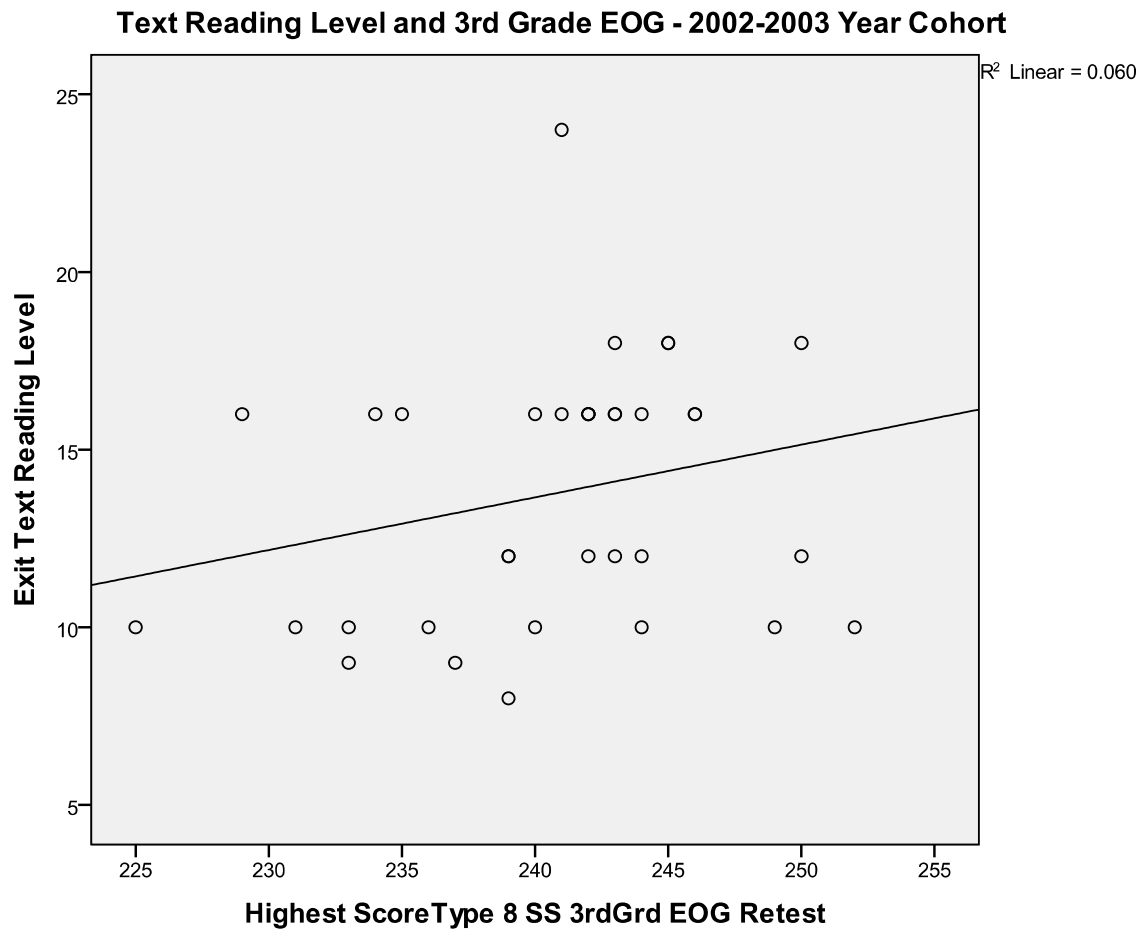
Text Reading Level and 3rd Grade EOG - 2002-2003 Year Cohort

Figure 9. Scatterplot of Text Reading Level and 3rd Grade EOG Scores – 2002-03 Year Cohorts.

This study was composed of 177 students who participated in the RR program in first grade and whose subsequent performance on the NC EOG was tracked beginning in third grade. The aim of the study was essentially twofold: Determine whether the RR program improved reading achievement in this study sample and establish whether or not the RR students maintained those gains over time. The results indicated that the RR students' reading achievement improved after having been in the program (mean Text Reading Level of 3 at the start of the program to a mean level of 14, an increase of eleven points) and continued to

improve even after the program's completion (mean increase of 3 additional levels by year's end). Based on equivalence testing the study participants' EOG scores and the mean EOG scores for the district were not equivalent, in third grade and beyond. In other words, by not being equivalent the results indicated that the study participants were no longer performing on the average level with the school district population despite having been on the average level of their peers upon completion of the RR program. Using the standard set by the school district (i.e. scores of III or IV indicating proficiency) 44% of the RR students were performing at proficiency in third grade. Subsequent grade levels have increasingly lower percentages of students performing at proficiency levels. Individual cohorts faired better with some achieving as high as 70% proficiency in third grade and some as low as 7% proficiency in third grade. Finally, a correlation was investigated between the TRL and future EOG performance, but this relationship proved to be weak.

Conclusion

This concluding section will briefly review and summarize this research, its methodology and discuss the findings and their implications. The results will be discussed along with their significance and ultimately indicate the direction that this researcher and further research should take.

The goal of this research was to determine how effective a particular early reading intervention was at raising poor readers' reading achievement, in this case the Reading Recovery early intervention program, and to determine if those acquired skills carried the child into third grade and beyond while still performing on average with their peers. The researcher was given a set of data comprised of scores for RR students who successfully completed the program. Based on their performance on the RR program's own measure, the Observation Survey, the students were considered to be performing on average with their peers upon completion of the program. Using the Text Reading Level subtest of the Observation Survey, the students pre- and post reading levels were compared to determine the amount of growth in the program. The data also included scores for each RR student on the district-wide administered assessment, the NC End-of-Grade (EOG) test in reading, beginning in third grade. The aim was to compare RR student performance on the EOG beginning in third grade with the school district average and determine if the students maintained their on average performance which was the result of having participated in the RR program in first grade.

Program Effectiveness

The results indicate that prior to entering the RR program, the students involved in this study were clearly some of the lowest performers in reading achievement. According to the tenets of RR, these students were chosen because of their poor performance at the end of kindergarten and in the beginning of their first grade year. These lowest 20% scored on average at a level 3 on the Text Reading level subtest which essentially ranges from 1 to 30. A level 0 is in fact possible, but basically designates non-readers. Although it is true that some children are prematurely exited from the program for failure to make sufficient progress, those who do successfully complete the program are truly some of the lowest reading performers in their grade. In this study they increased on average 10-14 levels and another 3 levels by the end of the school year even after they no longer were receiving the RR instruction.

For these first graders this program was significantly effective. Participation in the RR program had a statistically significant positive effect on the students' reading achievement as demonstrated by the statistically significant increase on the Text Reading Level subtest and the students' successful exit from the program. The RR program does what it sets out to do, which is bring the lowest readers up to the average level of their peers. Administrators would do well to at least consider this program for their schools. It targets a very specific group of readers—the kindergartener/first graders who are performing well below their peers who are on the cusp of learning to read. The program intervenes before the emergent readers have a chance to fail instead of simply addressing their needs once they've begun to fail and feel discouraged about the challenges of learning to read. In the short-term, this intervention clearly has an impact on their reading achievement.

There are many considerations for school systems when choosing a reading program,, such as the cost of training, which populations are targeted, and the number of students who can be served at once. Reading Recovery has a large cost associated with training teachers with graduate level instruction, it is specifically designed for struggling students in first grade only, and it is a one-on-one program which limits the number of students that can participate in the program during a year's time. All of these factors must be considered by the school district prior to implementing the program. This final characteristic, the fact that RR is a one-one-one program, especially limits the volume of students who can be served and forces schools to select only the lowest students. This can have the unfortunate effect of excluding less severe students who might benefit from simply participating in the RR program. On the other hand, it also allows for a tool in the school's arsenal that focuses on an especially difficult population that will be affected most severely by failing to learn to read at an early age. Nevertheless, if the goal of an administration is to address the struggling emergent readers of their schools early on while still in first grade, then the RR program appears to function in that role—at least in the short term.

Long-term Effectiveness

As we look beyond the short-term effectiveness of the RR program, the long-term impact of the program will tell administrators and other educators whether the RR program is worth the time and investment. Are the improved reading achievement effects of the Reading Recovery program maintained in third grade and subsequent grades, as measured by performance on the North Carolina End-of-Grade test in Reading when compared to district-wide average performance on the EOG? The results are mixed. The first analysis consisted of testing for equivalence between the mean EOG score of each cohort, its corresponding 95% CI and the

school district mean EOG score and its corresponding interval of ± 5 points. The results did not allow for equivalence to be demonstrated in any of the cohorts. Although they were certainly close and clearly performing well above their original standing in first grade as the lowest 20% of readers, they were not similar enough to the school district population mean in later grades. In real terms, this does not mean that they had not maintained the gains made during the RR program, but only that the study analysis did not detect the relationship between the study sample and the school population.

The next analysis looked at a slightly different metric, one that is used by the school district and the state of NC to gauge student proficiency. Instead of comparing the RR students' scaled scores with the District mean EOG scores, the percentage of RR students who were scoring at *proficiency* was examined and the question asked, what percentage of RR students scored an Achievement Level of III or IV? The data showed that 44% of the former RR students in third grade scored either a III or IV. This is less than half of the former RR group. In fourth grade there were 40% and in subsequent grades the percentage dropped to 35% in fifth grade, then to 20% in sixth and 26% in seventh grade. Looking at the group as a whole, the return on the investment from the administrators' standpoint is an immediate increase following the initial program participation and the beginnings of a positive learning trajectory by the end of the year. Then in two years we had less than half of the group performing at proficiency. The group of proficient performers shrinks with each subsequent grade, making for an early intervention investment with little long-term gain or impact on the student performance. Admittedly there is much that can happen in those two years. The question for further research is why and when did half of the former RR students begin to struggle and get lost academically?

Some of the individual cohorts faired much better. The 2002-03 cohort retained 67% of the students in the proficiency category, scoring either a III or IV. The 2003-04 cohort faired even better with 70% of the students maintaining proficiency levels in third grade. The 2004-05 cohort did not fair as well, with only 59% of the students remaining on the proficient levels. Other research questions that need to be analyzed further are: Why is there such a contrast in percentages maintaining proficiency from cohort to cohort? Were there problems with fidelity in the administration of the program, procedural changes, or less support for the program in subsequent years? The remaining two cohorts, 2005-06 and 2006-07, had only 23% and 7% of its students scoring in the proficient range respectively. The 2006-07 cohort had no students receiving level IVs in third grade.

One apparent explanation for the sharp contrast between these last two cohorts and the others is due to the shift or re-norming of the EOG scoring which occurred just prior to the 2007-08 school year (highlighted earlier), which is the same year that the 2005-06 cohort was in third grade and the 2007-08 cohort was in third grade the following year. Clearly there is a problem with using the EOG as a metric in these longitudinal studies, in this case due to the shifting in scores and difficulty in making comparisons. Future research will require measures that will provide the researchers with more control, but that will also cover a broader range of reading skills. The EOG in reading essentially measures comprehension via a reading passage and multiple choice questions and is not administered until the end of third grade. On the EOG it is difficult to pinpoint an issue like fluency or letter-word identification because you simply cannot know why a student missed a particular question. Additionally, there is a period of two years after participation in the RR program in which there is little information gathered on the reading achievement of the students. There are many curriculum based measures (CBM) that are utilized

by school districts that can measure a variety of tasks and can be implemented prior to, during, and immediately after, as well as the following year after participation in an intervention program. These measures make it possible to see the immediate affects of an intervention, like RR, and whether it is having the desired effect.

Further examination of the relationship between the independent measure, the Text Reading Level of the RR Observation Survey, and the dependent variable, student performance on the EOG, was undertaken to see if the Text Reading Level (TRL) subtest might function as an early indicator of the students' later performance on the EOG.) However, this was not the case. The highest correlation $r(33) = .246, p > .05$ was between the Exit TRL of the 2002-03 cohort and their third grade EOG scores. This is considered a weak correlation. There does not seem to be a strong enough relationship between the TRL at the conclusion of the RR program and the first EOG in third grade almost two years later. There are simply too many variables unaccounted for between first grade and third grade and beyond. There is not sufficient variability of scores on the Exit Text Reading Level subtest to offer indicators of future performance. Eighty percent of the scores were either a level 16 or level 18 by year's end. Furthermore, the journey for the students from the end of first grade to the end of third grade is fraught with obstacles which interfere with their ongoing reading development. In the end, this may be the best explanation for the varied results of this study, at least with regards to the longitudinal effects of the RR program. There are simply too many factors that interfere with learning both intrinsically within the learner and externally in the school environment. As stated before, closer monitoring with quicker more efficient Curriculum Based Measures (CBM) that are administered earlier will give researchers and schools the data to know how effective their programs are.

The students who participate in the program show substantial gains and even retain those skills and continue to grow through the end of the first grade school year. After many months, a year, two years the students have not maintained those gains according to the results. Future research should focus on tracking the students more closely with curriculum based measures that are unlike the Observation Survey in that they are designed to measure student progress against criterion measures. Studies in which the researcher is allowed to chart the students' growth more frequently, especially during those critical second and third grade years, will produce a clearer picture of what occurs to these students once the RR program and first grade are completed. Given that the End-of-Grade test only measures reading comprehension after a year of classroom instruction, measures like the Developmental Reading Assessment (DRA) or the Dynamic Indicators of Basic Early Literacy Skills (DIBELS) offer not only more frequent assessments but also tap into other critical areas such as fluency. Perhaps the regression of the once improved low performers is due to unaddressed fluency issues that, although improved after the RR instruction, were not maintained and later became a problem for some of these students. Their other skills having continued to mature, fluency or inability to read and simultaneously comprehend the passages of the EOG may have made it difficult if not impossible to recall the information once presented with the comprehension questions.

Continued research on this subject will also need to replicate the analysis with other components of the Observation Survey and look at those subtests' ability to predict future performance and highlight future needs. Although the students upon exiting the RR program are considered to be on average with their peers, they are not yet proficient readers and therefore have weaknesses, some hidden, that may manifest themselves in the future as greater demand is placed on the student's reading. The current analyses indicated that Text Reading level was not a

good indicator of future performance, but perhaps one of the other subtests of the Observation Survey, or a combination of two or more, can assist classroom teachers in continuing to monitor and address weak areas beyond the RR program. A study which focused on the transmittal of information from the RR program to the first grade teacher and the second and the third grade teachers about the student's performance during the RR program and the results of the Observation Survey might yield markers or traits that would indicate areas in which the student needed additional instruction. These indicators could then be addressed in the classroom but using other interventions or programs (Tier 2 of the RTI process). In an education landscape where the schools are already accused of over-testing, what is required is more frequent assessments, not simply "tests", but smarter, quicker, less-intrusive measures that give a truer picture of where a student has been and what their areas of need are. These measures will assist any effective reading intervention program in tracking student progress, which will ultimately determine the long-term effects that that program has on student achievement.

References

American Federation of Teachers (AFT). (1999). *Building on the best, learning from what works: Seven promising reading and English language arts programs.* Washington, DC. Retrieved August 5, 2008 from http://www.aft.org/pubs-reports/downloads/teachers/remedial.pdf

Baenen, N, Bernhole, A., Dulaney, C., & Banks, K. (1997). Reading Recovery: Long-Term progress after three cohorts. *Journal of Education For Students Placed At Risk, 2*(2), 161-181.

Beaver, J. M. (2006). *Teacher guide: Developmental Reading Assessment, Grades K– 3, Second Edition*. Parsippany, NJ: Pearson Education, Inc.

Briggs, C. & Young, B. (2003). Does Reading Recovery work in Kansas? A retrospective longitudinal study of sustained effects. *Journal of Reading Recovery, 3*(1), 59-64.

Center, Y., Wheldall, K., & Freeman, L. (1995). Evaluating the effectiveness of Reading Recovery: A critique. Educational Psychology, 12(3-4), 263-274.

Christ, T. J., Burns, M. K., & Ysseldyke, J. E. (2005). Conceptual confusion within response-to-intervention vernacular: Clarifying meaningful differences. *Communiqué, 34*(3), 1–8.

Clay, M. M. (1987). Learning to be learning disabled. *New Zealand Journal of Educational Studies, 22*, 155-173.

Clay, M. M. (2001). *Change over time in children's literacy achievement.* Portsmouth, NH: Heinemann.

Clay, M. M. (2002). *Reading Recovery: A guidebook for teachers in training*. Portsmouth, NH: Heinemann.

Clay, M. M. (2005a). *Literacy lessons designed for individuals part one: Why? When? and how?*

    Portsmouth, NH: Heinemann.

Clay, M. M. (2005b). *Literacy lessons designed for individuals part two: Teaching*

    Portsmouth, NH: Heinemann.

Clay, M. M. (2006). *An observation survey of early literacy achievement*. Portsmouth,   NH:

    Heinemann.

Cleophas, T. J., Zwinderman, A. H., & Cleophas, T. F. (2006). Equivalence Testing., *Statistics*

    Applied to Clinical Trials. (pp. 59-65). Springer Netherlands.

Denton, C. A., Ciancio, D. H., & Fletcher, J. M. (2006). Validity, reliability, and utility of the

    Observation Survey of Early Literacy Achievement. *Reading Research Quarterly, 41*(1),

    8-34.

Dunn, M. W. (2007). Diagnosing a reading disability: Reading Recovery as a component of a

    response-to-intervention assessment method. *Learning Disabilities: A Contemporary*

    Journal 5(2), 31-47.

Fletcher, J. M., Shaywitz, S. E., Shankweiler, D. P., Katz, L., Liberman, I. Y., Steubing, K. K.,

    Francis, D. J., et al. (1994). Cognitive profiles of reading disability: Comparisons of

    discrepancy and low achievement definitions. *Journal of Educational Psychology, 86*, 6-

    23.

Fuchs, D., Mock, D. Morgan, P., & Young, C. (2003). Responsiveness-to-instruction:

    Definitions, evidence, and implications for learning disabilities construct. *Learning*

    Disabilities Research & Practice, 18(3), 157-171.

Gómez-Bellengé, F. X. & Thompson, J. R. (2005). Twenty years of data evaluation: A brief

    history of the national data collection. *The Journal of Reading Recovery, 4*(2), pp 66-69.

Gómez-Bellengé, F. X. & Thompson, J. R. (2005). *U.S. norms for tasks of an observation survey of early literacy achievement*. (Rep. No. NDEC 2005–02). Columbus: The Ohio State University, National Data Evaluation Center. http://www.ndec.us.

Hiebert, E. (1994). Reading Recovery in the United States: What difference does it make to an age cohort? *Educational Researcher, 23*(9), 15–29.

Iverson, S. & Tunmer, W. E. (1993). Phonological processing skills and the reading recovery program. *Journal of Educational Psychology, 85*(1), pp. 112-126.

Joseph, L. M. (2008). Best practices on interventions for students with reading problems. In A. Thomas & J. Grimes (Eds.), *Best practices in school psychology, Vol. 4*, (pp. 1163-1180). Bethesda, MD: National Association of School Psychologists.

Kershner, J. R. (1990). Self-concept and IQ as predictors of remedial success in children with learning disabilities. *Journal of Learning Disabilities, 23*, 368-374.

Lose, M. K., Schmitt, M. E., Gómez-Bellengé, F. X., Jones, N. K., Honchell, B. A., & Askew, B. J. (2007). Reading Recovery and IDEA Legislation: Early Intervening Services (EIS) and Response to Intervention (RTI). *The Journal of Reading Recovery*, 6(2), 44–49. Retrieved September 12, 2008, from http://www.readingrecovery.org/pdf/reading_recovery/SPED_Brief-07.pdf.

Lovell, K., Gray, E. A., & Oliver, D. E. (1964). A further study of some cognitive and other disabilities in backward readers of average non-verbal reasoning scores. *British Journal of Educational Psychology, 34*, 275-279.

Lyon, G. R. Fletcher, J. M., Shaywitz, S. E., Shaywitz, B. A., Torgesen, J. K., Wood, F. B., Schulte, A., & Olsen, R. (2001). Rethinking learning disabilities. In C. E. Finn, C. R.

Hokanson, & A. J. Rotherham (Eds.), *Rethinking special education for a new century* (pp. 259-287). Washington, DC: Thomas B. Fordham Foundation.

McEneaney, J. E. (2006). Agent-based literacy theory. Reading Research Quarterly, 41(3), 352–371.

National Data Evaluation Center (NDEC). (2008). Reading Recovery overview. Ohio State University. Retrieved on August 13, 2008 from http://www.ndec.us/AboutRR.asp.

North Carolina Department of Public Instruction – Division of Accountability Services/North Carolina Testing Program (NCDPI-DAS/NCTP) (2007). North Carolina End-of-Grade Tests, Technical Report. Raleigh, NC: Author. Retrieved August 12, 2008 from http://www.dpi.state.nc.us/accountability.

North Carolina Department of Public Instruction (NCDPI). (n.d.). North Carolina State Testing Results. Retrieved August 2, 2011, from http://report.ncsu.edu/ncpublicschools/AutoForward.do?forward=eog.pagedef.

National Research Council (1998). *Preventing Reading Difficulties in Young Children.* Washington, D.C.: National Academy Press.

Pearson, P. D. (1999). A historically based review of Preventing Reading *Difficulties in Young Children*. Reading Research Quarterly, 34, 231-246.

Peterson, B. (1991). Selecting books for beginning readers. In D.E. DeFord, C.A. Lyons, and G.S. Pinnell, eds. *Bridges to literacy: Learning from reading recovery* (pp. 119–147). Portsmouth, NH: Heinemann.

Pinnell, G S. (1989). Reading Recovery: Helping at-risk children learn to read. *The Elementary School Journal, 90*, 161-183.

Pinnell, G. S., Lyons, C. A., DeFord, D. E., Bryk, A. S., & Seltzer, M. (1993/94). Comparing

    instructional models for the literacy education of high risk first graders. *Reading*

    Research Quarterly, 29, pp. 8-39.

Pinnell, G.S., Deford, D.E., & Lyons, C.A. (1988). Reading Recovery: Early intervention for at-

    risk first graders. Arlington, VA: Educational Research Service.

President's Commission on Excellence in Special Education (2002). *A new era:*

    Revitalizing special education for children and their families. Washington, D.C.: Author.

Quay, L. C., Steele, D. C., Johnson, C. I., & Hortman, W. (2001). Children's achievement and

    personal and social development in a first-year Reading Recovery program with teachers

    in-training. *Literacy Teaching and Learning: An International Journal of Early Reading*

    and Writing, 5, 7–25.

Reading Recovery Council of North America (RR) (2008). Reading Recovery: Basic facts.

    Retrieved August 12, 2008

    http://www.readingrecovery.org/reading_recovery/facts/index.asp.

Reading Recovery Council of North America, North American Trainers Group Research

    Committee (2004). Five Reading Recovery studies: Meeting the criteria for scientifically

    based research. Retrieved on February 25, 2008, from

    http://www.readingrecovery.org/sections/research/index.asp.

Rodgers, E., Gómez-Bellengé, F., Wang, C., & Schultz, M. (2005). Examination of the validity

    of the Observation Survey with a Comparison to ITBS. Paper presented at the Annual

    Meeting of the American Educational Research Association Montreal, Quebec.

Roush, W. (1995). Arguing over why Johnny can't read. Science, 267, 1896-1898.

Schwartz, R. M. (2005). Literacy learning of at-risk first-grade students in the Reading Recovery

      early intervention. *Journal of Educational Psychology, 97*(2), 257-267.

Scruggs, T., & Mastropieri, M. (2002). On babies and bathwater: Addressing the problems of

      identification of learning disabilities. *Learning Disability Quarterly*, *25*(3), 155.

Shanahan, T., & Barr, R. (1995). Reading Recovery: An independent evaluation of the effects of

      an early instructional intervention for at-risk learners. *Reading Research Quarterly, 30*,

      958–996.

Slavin, R. E. (1987). Making chapter 1 make a difference. *Phi Delta Kappa, 69*, 110-119.

Slavin, R. E. (1989). Pet and the pendulum: Faddism in education and how to stop it, *Delta*

      Kappan, 70, 752-758.

Smith-Burke, M. T. (1996). Professional development for teacher leaders: Promoting program

      ownership and increased success. Network News, 1-4, 13, 15. Retrieved on May 23,

      2008, from http://www.readingrecovery.org/development/archives/smith-burke.asp

Snow, C. E., Burns, M. S., & Griffin, P. (1998). *Preventing reading difficulties in young*

      children. Washington, DC: National Academy Press.

Stanovich, K. E. (1988). Explaining the differences between the dyslexic and garden variety poor

      reader: The phonological-core variance-difference model. *Journal of Learning*

      Disabilities, 21, 590-604.

Stanovich, K. E. (1991). Discrepancy definitions of reading disability: Has intelligence led us

      astray? *Reading Research Quarterly, 26*, 1-29.

Stanovich, K. E. (2005). The future of a mistake: Will discrepancy measurement continue to

      make the learning disabilities field a pseudoscience? *Learning Disability Quarterly, 28,*

      103-106.

Tal, N. F. & Siegal, L. S. (1996). Pseudoword reading errors of poor, dyslexic and normally achieving Readers on multisyllable pseudowords. *Applied Psycholinguistics, 17*, 215-232.

Tang, M. & Gómez-Bellengé, F. (2007). Dimensionality and concurrent validity of the Observation Survey of Early Literacy Achievement. Paper presented at the 2007 American Educational Research Association Conference, Chicago, IL.

United States Department of Education. (n.d.). Institute of Education Sciences, National Center for Education Statistics (NCES). Retrieved August 2, 2011, from http://nces.ed.gov/surveys/sdds/2010/sprofile.aspx?id1=97000US3701800.

United States Department of Education (2002). Scientifically based research and the Comprehensive School Reform (CSR) program (pp. 17–18). Washington, DC: Government Printing Office.

United States Office of Education. (1977, December 29). Assistance to states for education of handicapped children: Procedures for evaluating specific learning disabilities. *Federal Register, 42*(250), 65082-65085. Washington, DC: U.S. Government Printing Office.

Vellutino, F. R., Scanlon, D., & Lyon, G. R. (2000). Differentiating between difficult-to-remediate and readily remediated poor readers: More evidence against IQ-achievement discrepancy of reading disability. *Journal of Learning Disabilities, 33*, 223-238.

What Works Clearinghouse (2007). *WWC intervention report: Reading Recovery*. Washington, DC: U.S. Department of Education, Institute of Education Sciences.

Woodcock, R., McGrew, K., & Mather, N. (2001). Woodcock-Johnson Tests of Achievement-Third Edition. Itasca, IL: Riverside Publishing.

Data from Figure 4 presented in two separate scatterplots correlating the relationship between

Text Reading level upon exiting RR and third grade EOG scores divided into years prior to state

re-norming of the EOG (Pre- Re-Norm) and years after the shifting in score ranges (Post- Re-
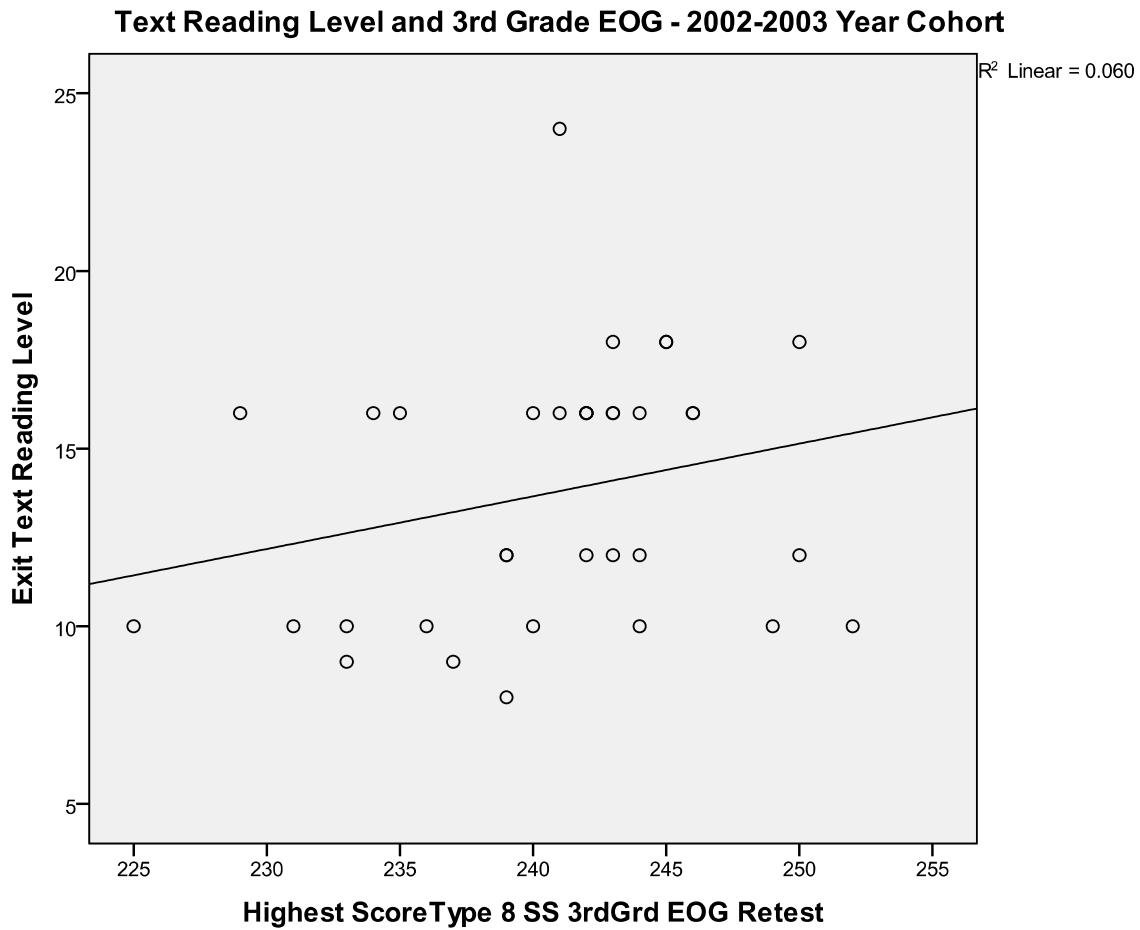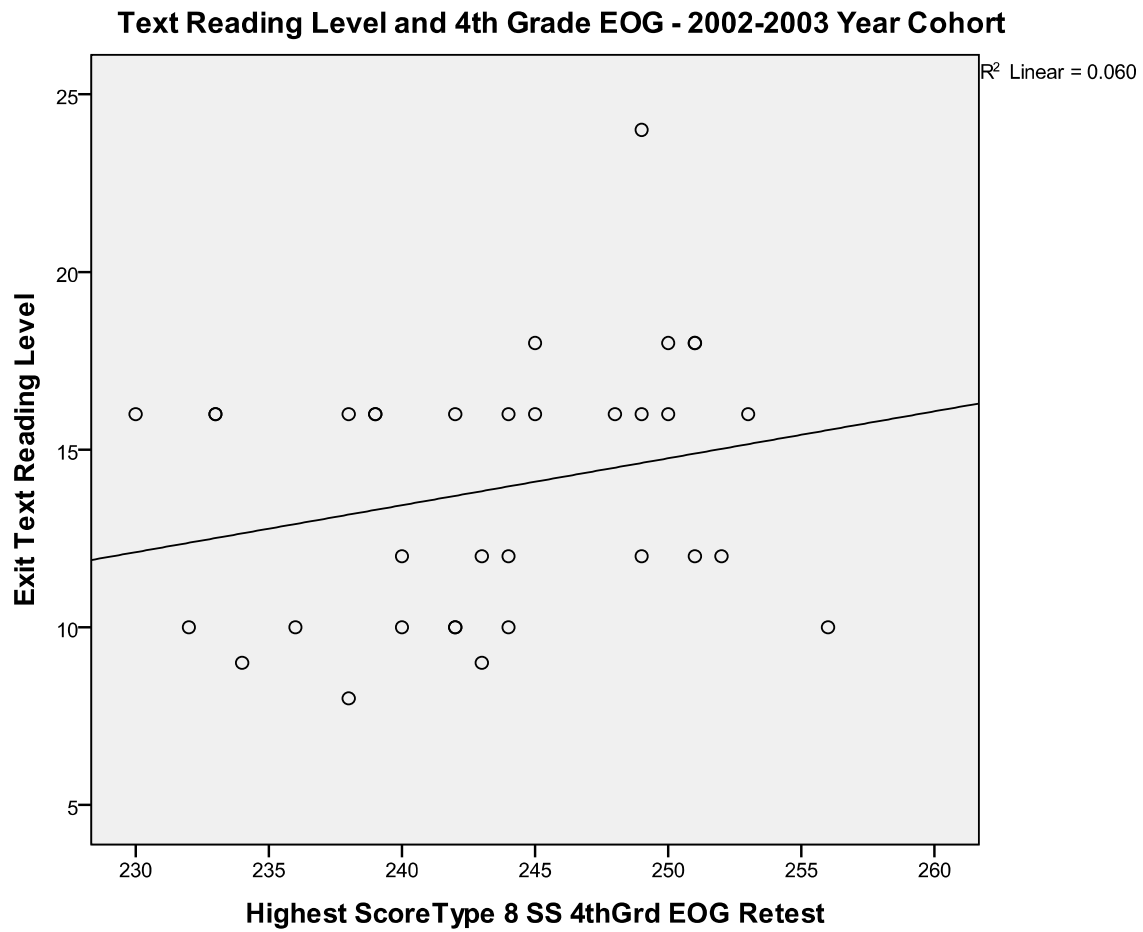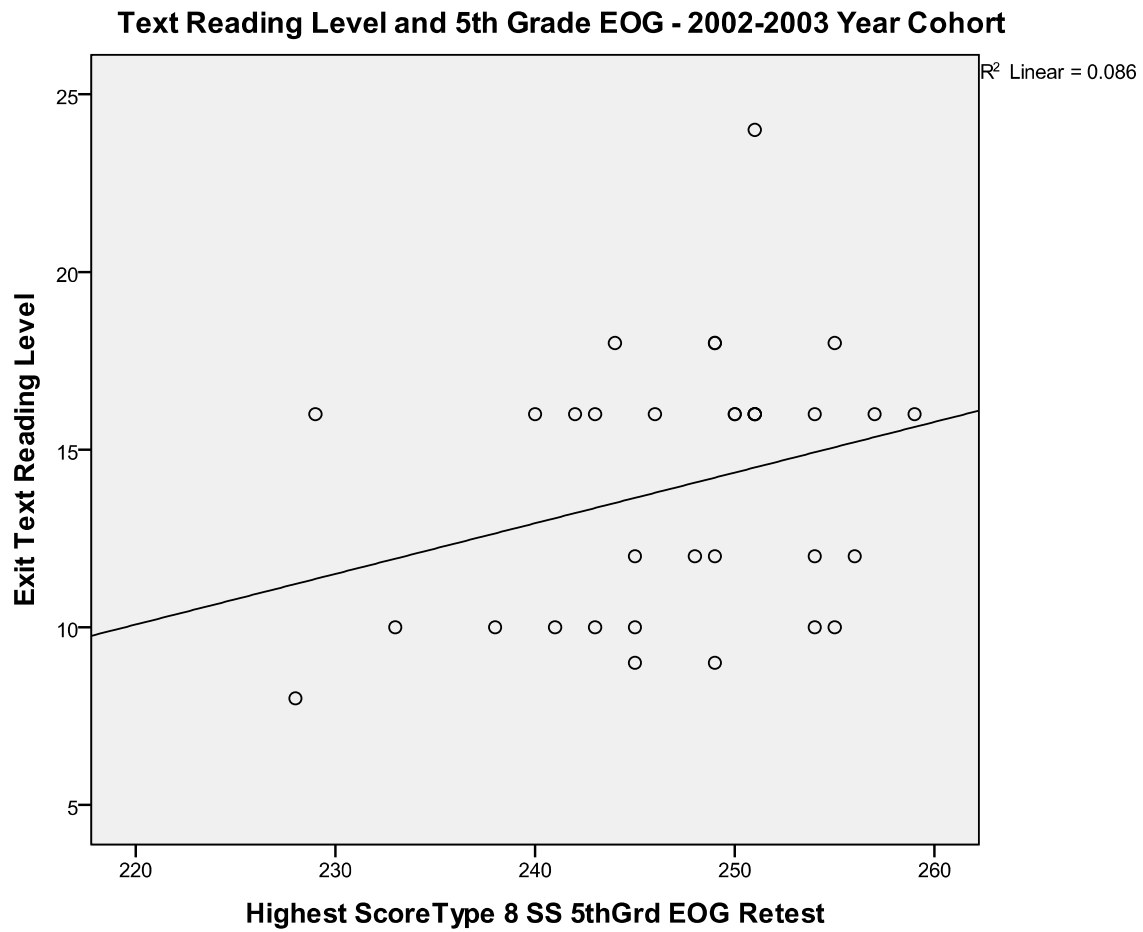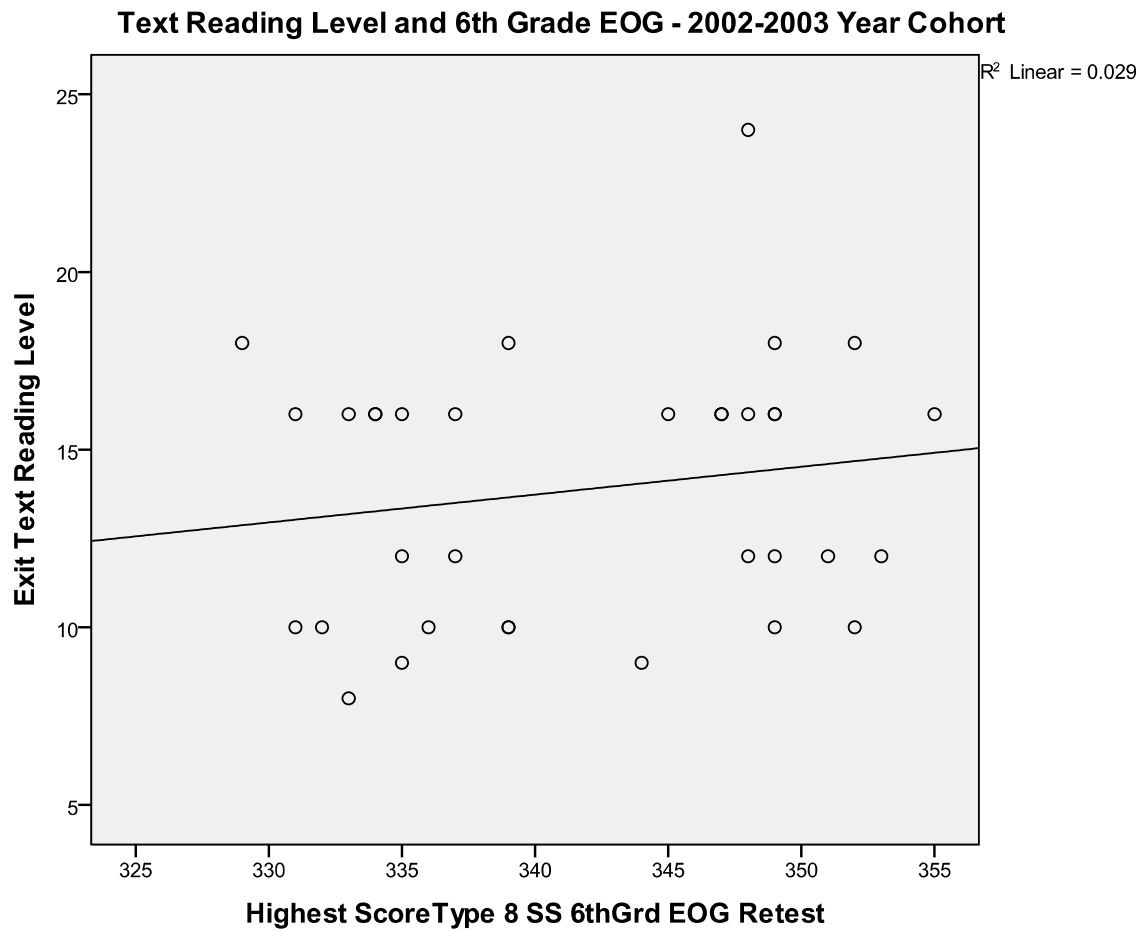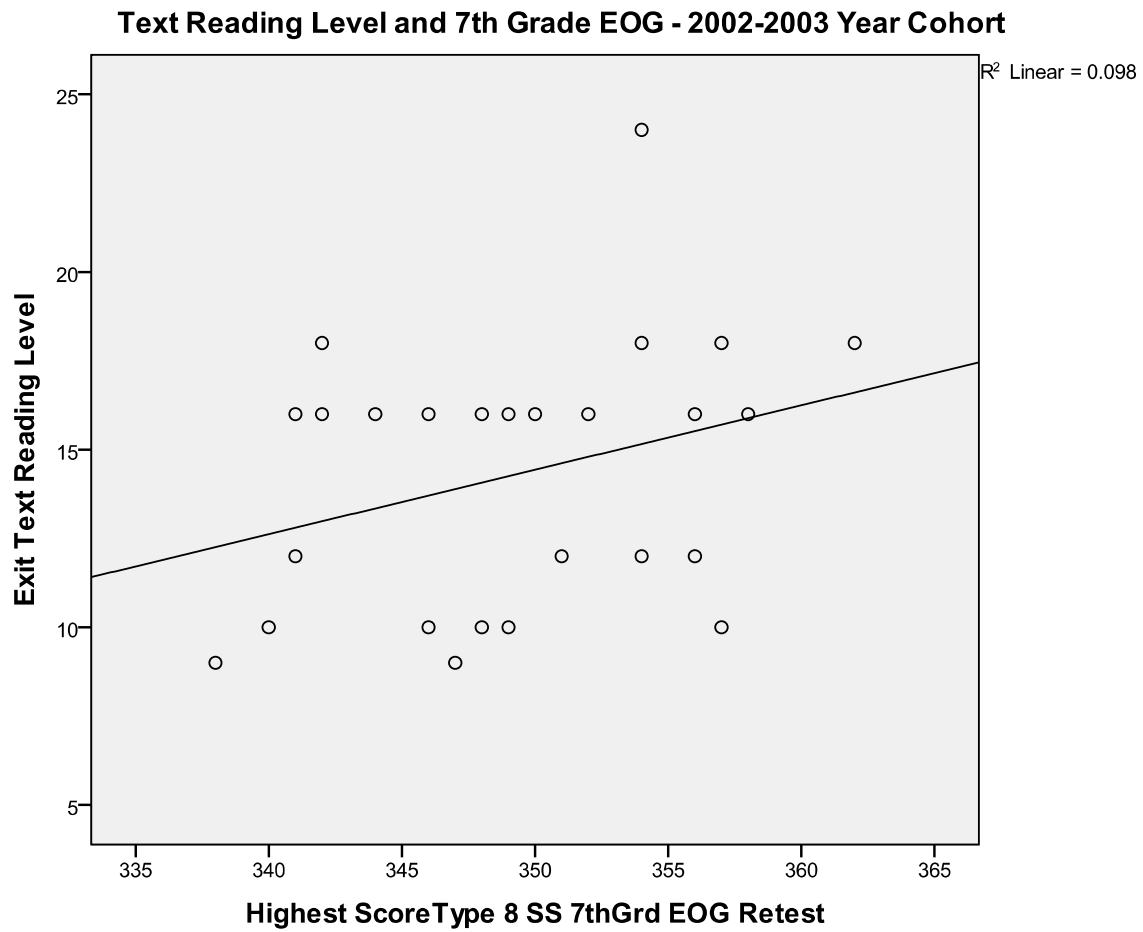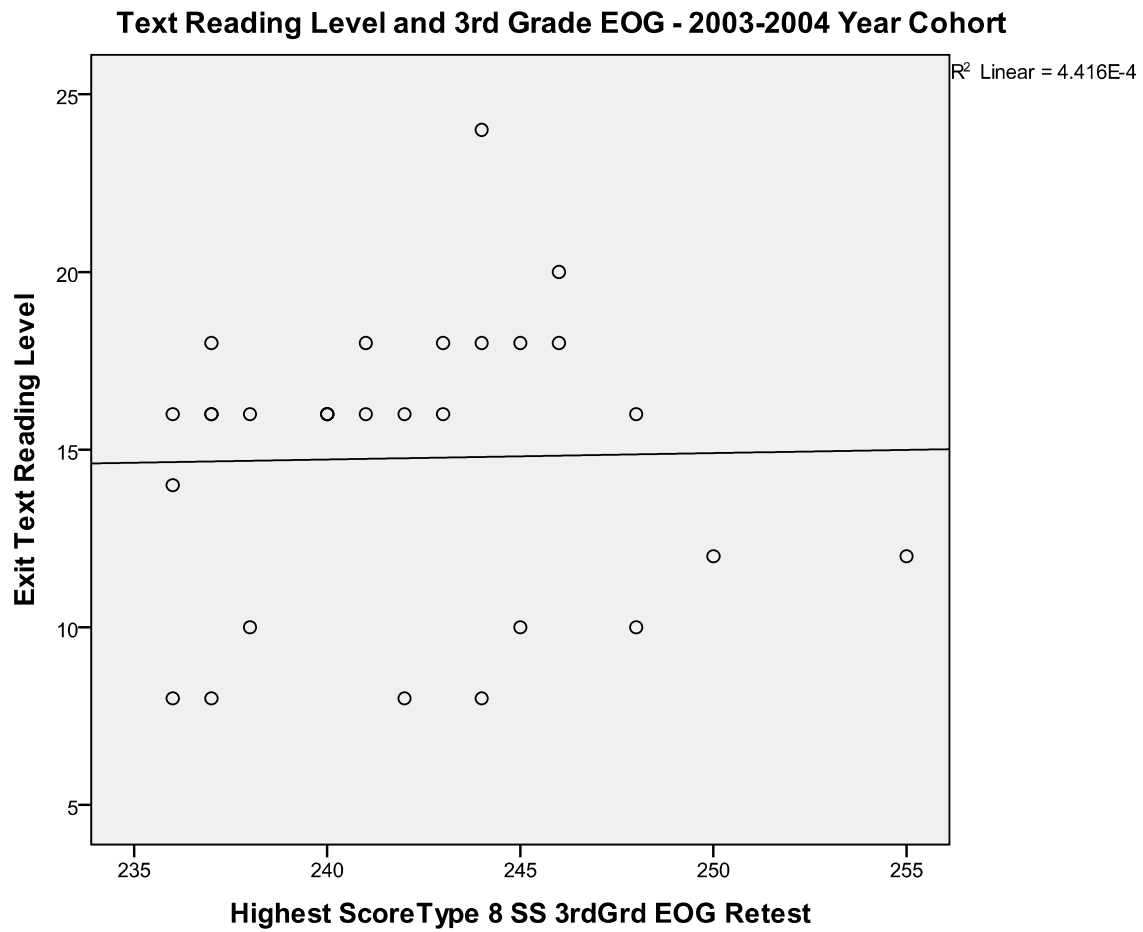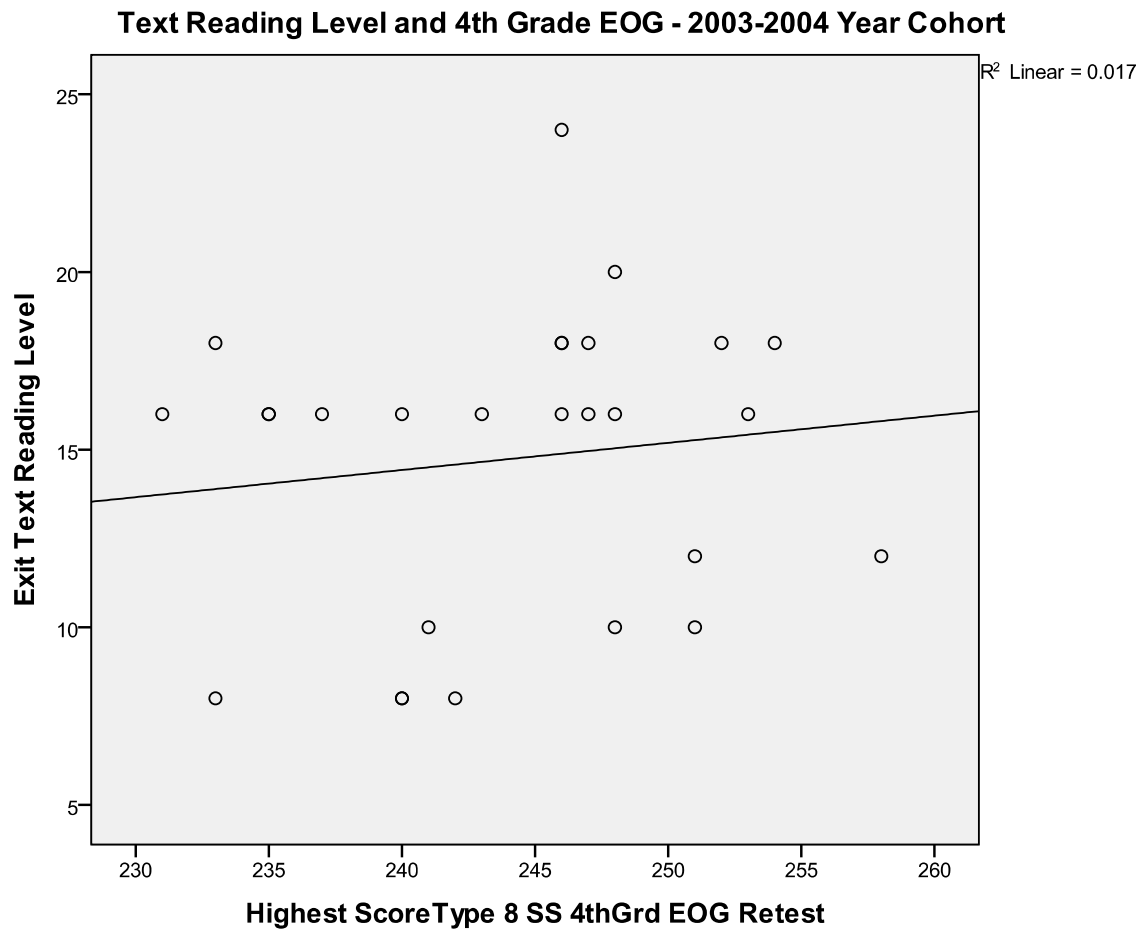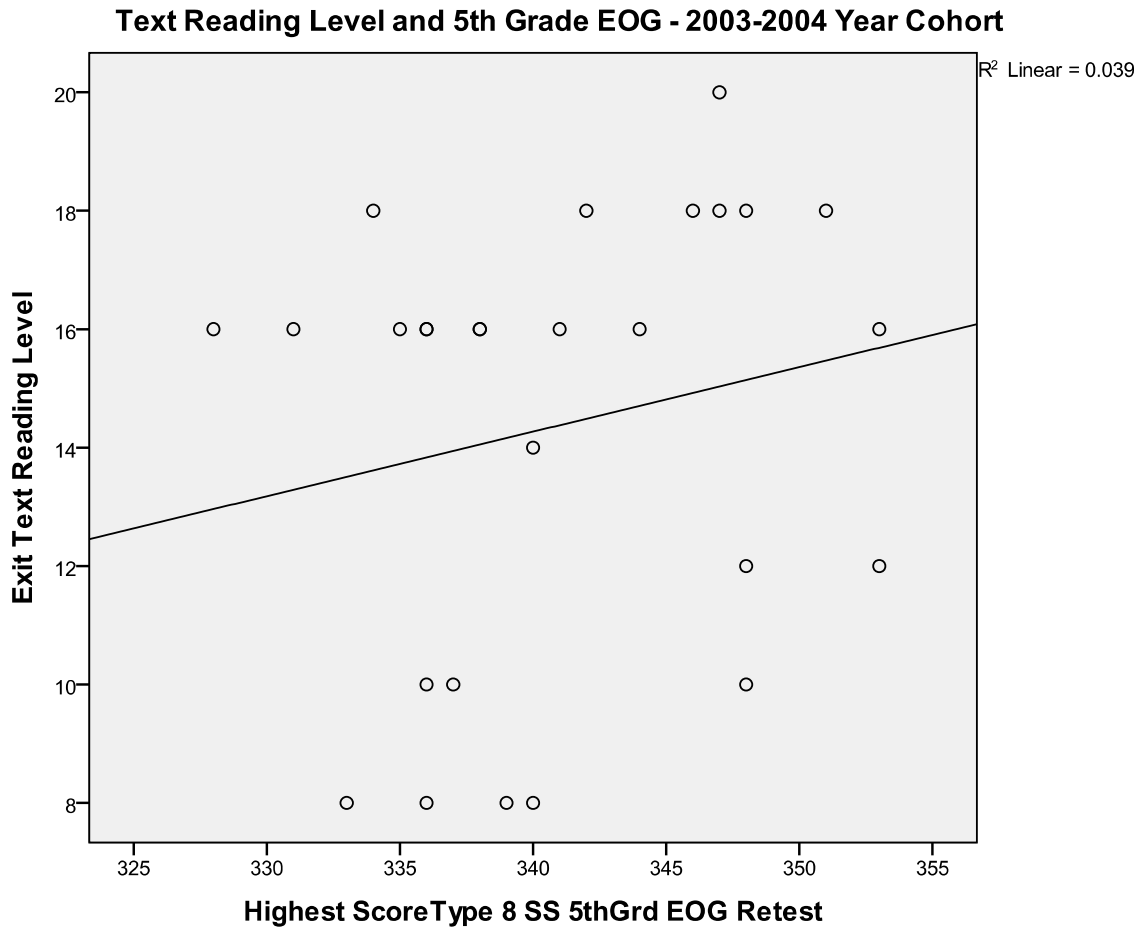
Norm).

Figure A1. Scatterplot of Text Reading Level and 3rd Grade EOG Scores – Pre- Re-Norm

Cohorts.

Figure A2. Scatterplot of Text Reading Level and 3rd Grade EOG Scores – Post- Re-Norm Cohorts.

Scatterplots for each of the year cohorts and their respective grades showing the relationship

between Text Reading level upon exiting RR and third through seventh grade EOG scores (when

available)

Figure B1. Scatterplot of Text Reading Level and 3rd Grade EOG Scores – 2002-03 Year

Cohorts.

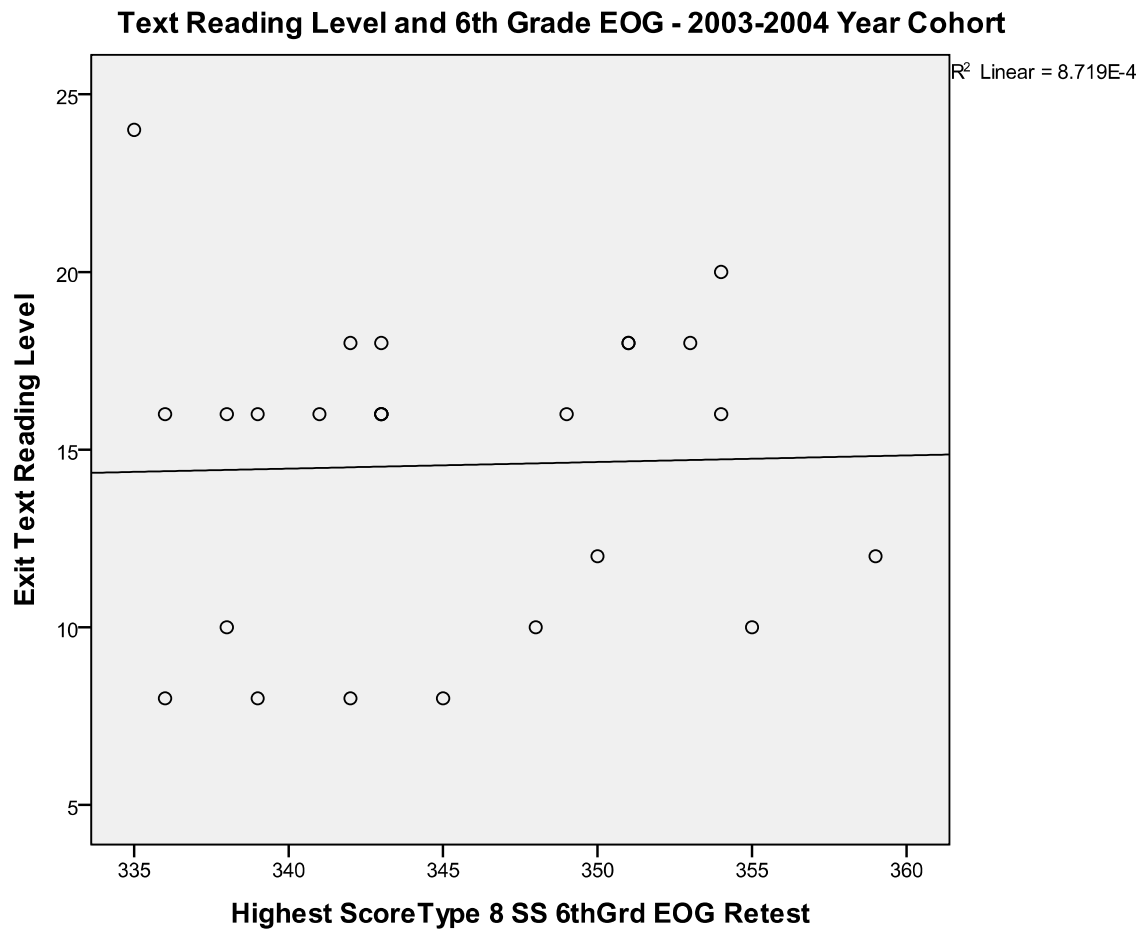**Text Reading Level and 4th Grade EOG - 2002-2003 Year Cohort**

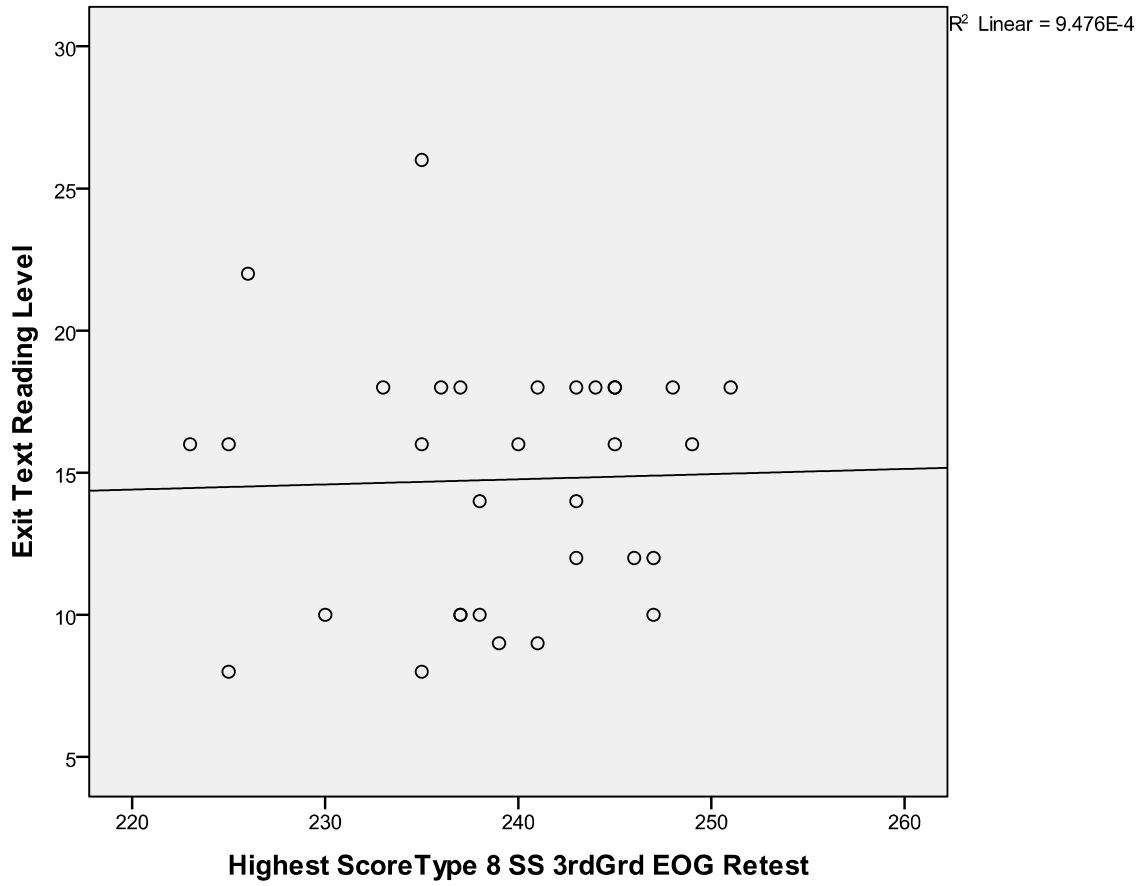*Figure B2.* Scatterplot of Text Reading Level and 4th Grade EOG Scores – 2002-03 Year Cohorts.

Text Reading Level and 5th Grade EOG - 2002-2003 Year Cohort

Figure B3. Scatterplot of Text Reading Level and 5th Grade EOG Scores – 2002-03 Year

Cohorts.

Text Reading Level and 6th Grade EOG - 2002-2003 Year Cohort

R^2 Linear = 0.029

Figure B4. Scatterplot of Text Reading Level and 6th Grade EOG Scores – 2002-03 Year

Cohorts.

Figure B5. Scatterplot of Text Reading Level and 7th Grade EOG Scores – 2002-03 Year Cohorts.

Text Reading Level and 3rd Grade EOG - 2003-2004 Year Cohort

Figure B6. Scatterplot of Text Reading Level and 3rd Grade EOG Scores – 2003-04 Year

Cohorts.

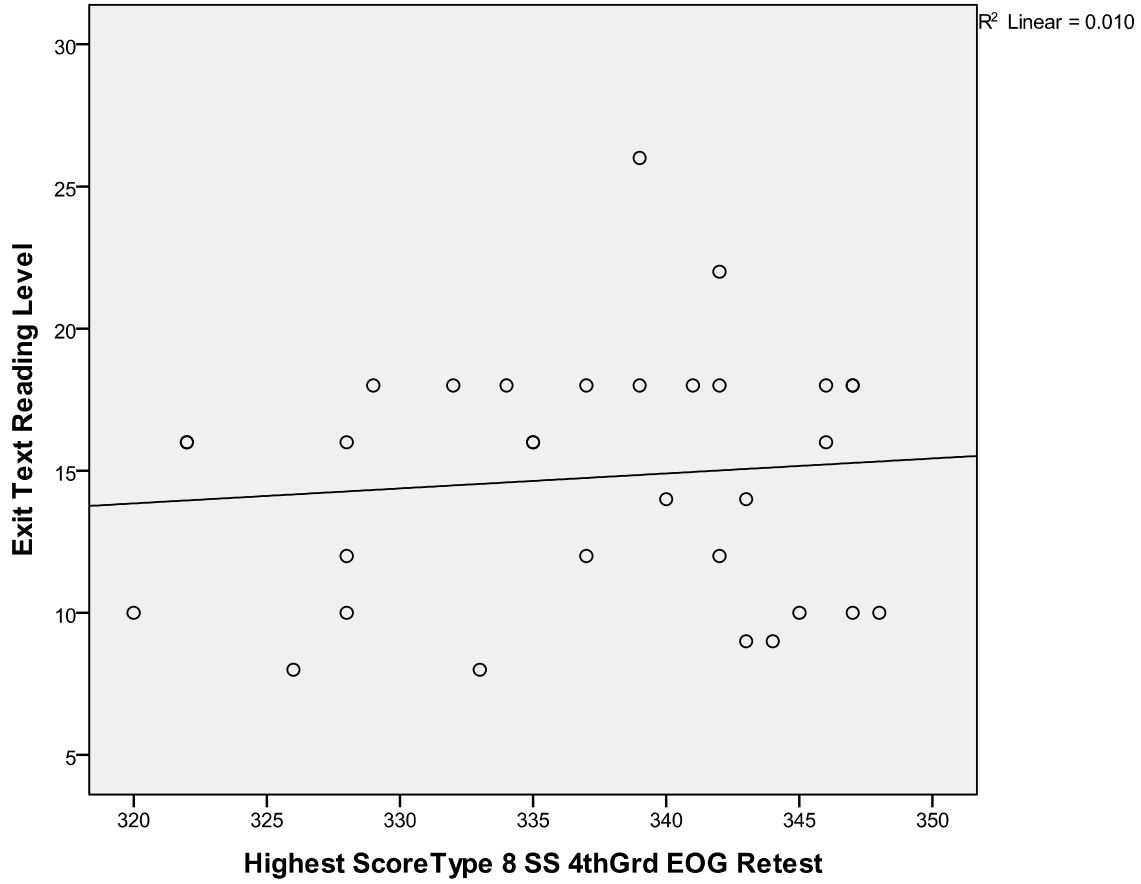**Text Reading Level and 4th Grade EOG - 2003-2004 Year Cohort**

R^2 Linear = 0.017

Figure B7. Scatterplot of Text Reading Level and 4th Grade EOG Scores – 2003-04 Year

Cohorts.

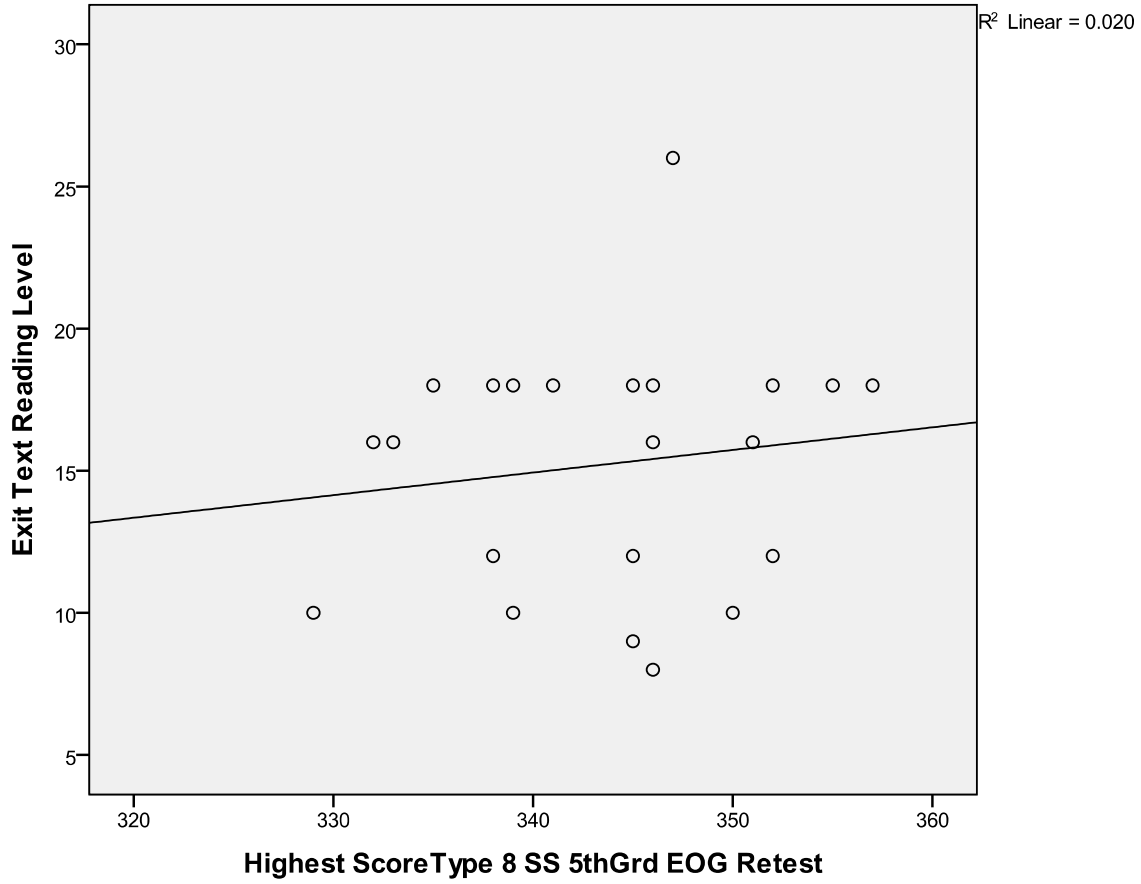**Text Reading Level and 5th Grade EOG - 2003-2004 Year Cohort**

R^2 Linear = 0.039

Figure B8. Scatterplot of Text Reading Level and 5th Grade EOG Scores – 2003-04 Year

Cohorts.

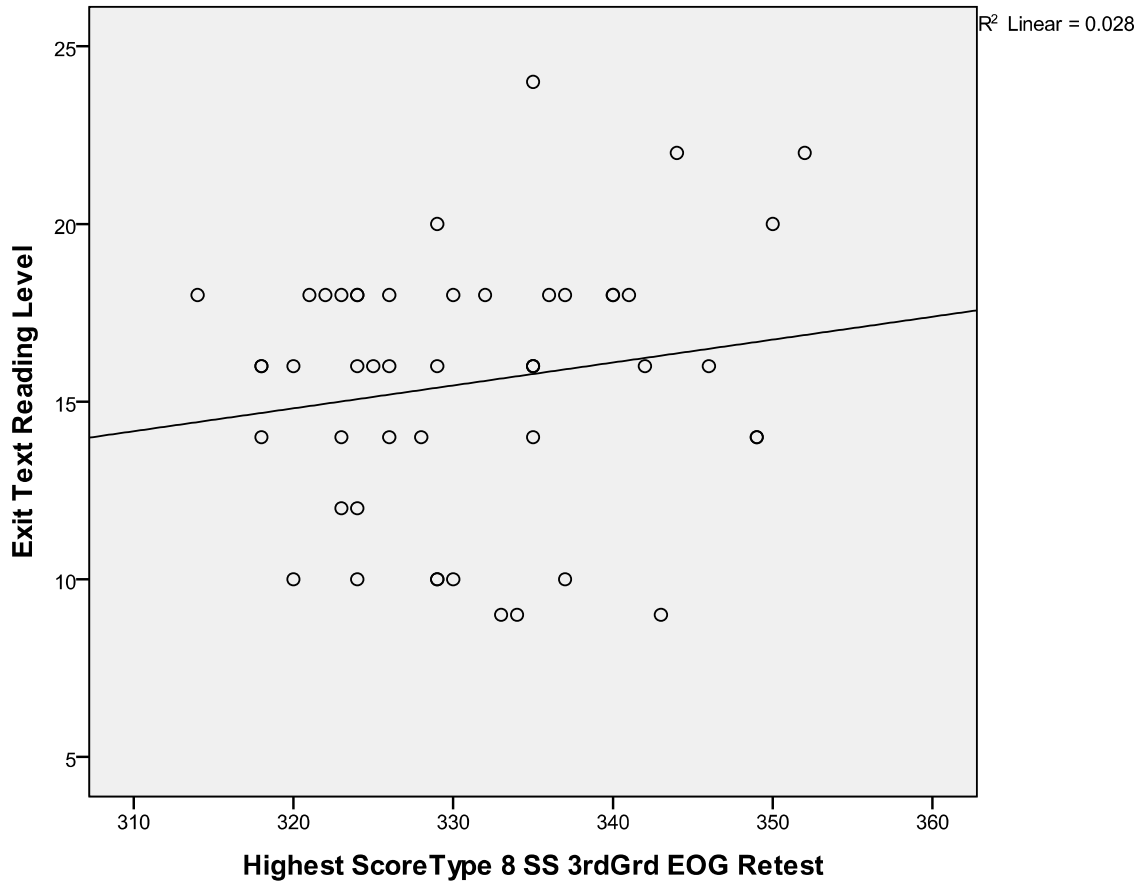**Text Reading Level and 6th Grade EOG - 2003-2004 Year Cohort**

R^2 Linear = 8.719E-4

Figure B9. Scatterplot of Text Reading Level and 6th Grade EOG Scores – 2003-04 Year Cohorts.

Figure B10. Scatterplot of Text Reading Level and 3rd Grade EOG Scores – 2004-05 Year

Cohorts.

Figure B11. Scatterplot of Text Reading Level and 4th Grade EOG Scores – 2004-05 Year

Cohorts.

Figure B12. Scatterplot of Text Reading Level and 5th Grade EOG Scores – 2004-05 Year

Cohorts.

Figure B13. Scatterplot of Text Reading Level and 3rd Grade EOG Scores – 2005-06 Year

Cohorts.

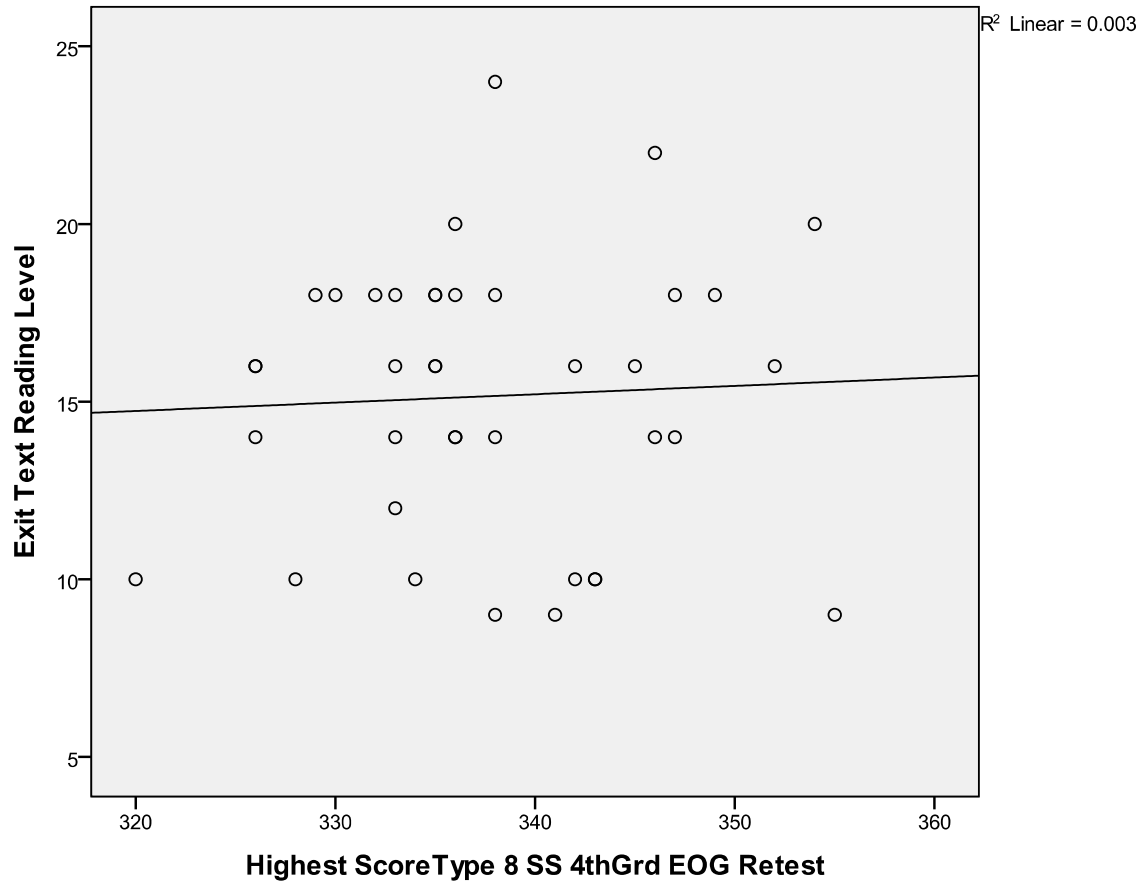**Text Reading Level and 4th Grade EOG Scores - 2005-2006 Year Cohort**

R^2 Linear = 0.003

Figure B14. Scatterplot of Text Reading Level and 4th Grade EOG Scores – 2005-06 Year

Cohorts.

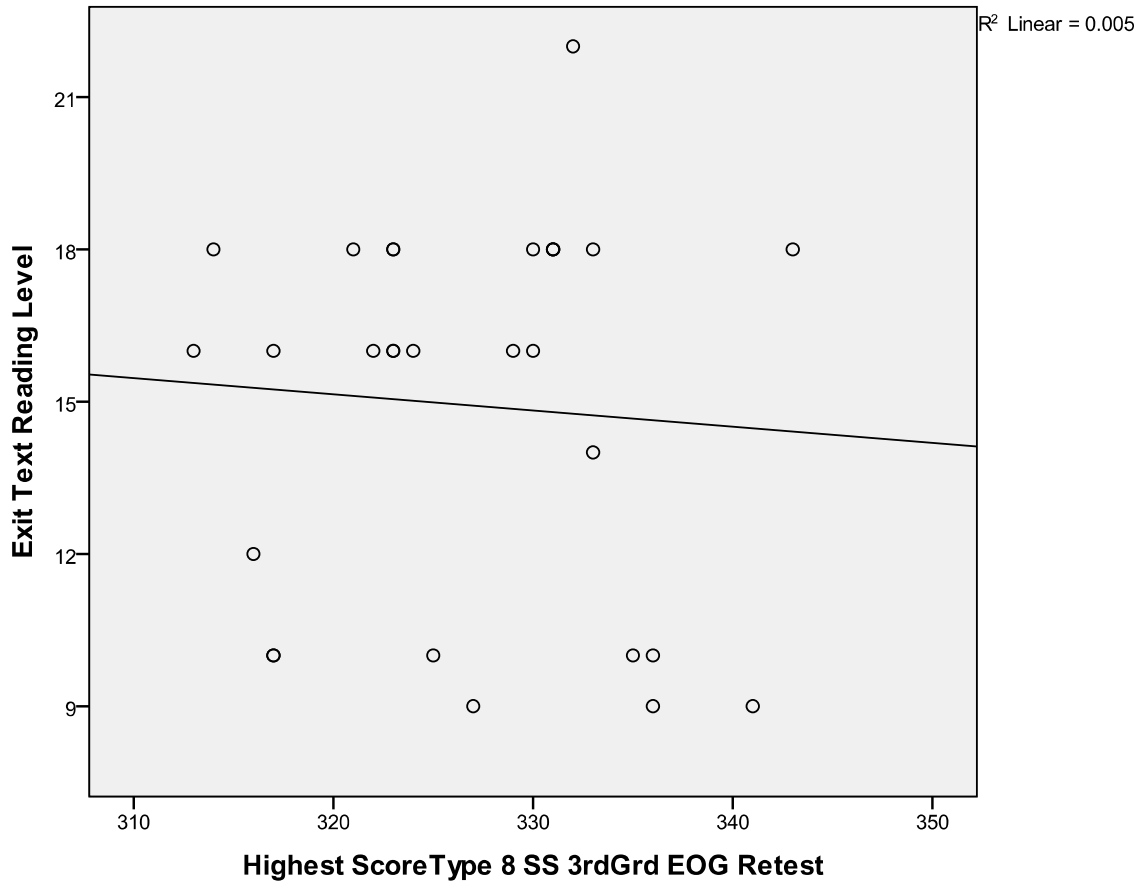**Text Reading Level and 3rd Grade EOG Scores - 2006-2007 Year Cohort**

Figure B15. Scatterplot of Text Reading Level and 3rd Grade EOG Scores – 2005-06 Year Cohorts.