2015

READ 180 Evaluation: Balanced Literacy in a Low-Income, Underperforming Urban High School

Daniel Lombardi
Walden University

Walden University

COLLEGE OF EDUCATION

This is to certify that the doctoral study by

Daniel Lombardi

has been found to be complete and satisfactory in all respects,
and that any and all revisions required by
the review committee have been made.

Review Committee
Dr. Edward Behrman, Committee Chairperson, Education Faculty
Dr. Charles Elliott, Committee Member, Education Faculty
Dr. James Valadez, University Reviewer, Education Faculty

Chief Academic Officer

Eric Riedel, Ph.D.

Walden University
2015

Abstract

READ 180 Evaluation: Balanced Literacy in a Low-Income, Underperforming Urban High School

by

Daniel Lombardi

MA, Saint Peter's College 2003

BA, Montclair State University, 1999

Doctoral Study Submitted in Partial Fulfillment

of the Requirements for the Degree of

Doctor of Education

Walden University

April 2015

Abstract

This study investigated the effectiveness of the READ 180 balanced literacy program in addressing the problem of low reading achievement among urban minority high school students. Research has shown low reading achievement to negatively impact academic and economic success. Holdaway's theory of natural literacy, which suggests reading instruction should be purposeful and realistic, supports READ 180 as a remedial intervention. The specific purpose of the study was to evaluate READ 180 in relation to improving classroom reading achievement, standardized language arts test scores, and graduation rates among students in a low-income, high-minority urban high school. Subjects were 2 cohorts of students consisting of 619 enrolled in READ 180 during the school year of 2007-2008 and 358 students enrolled in READ 180 during the school year of 2010-2011 at an urban high school in New Jersey. Following a goals-based program evaluation design, a paired difference *t* test was used to evaluate classroom reading achievement; a chi-square test was used to evaluate graduation rates; and a multiple regression analysis controlling for initial status was used to evaluate performance on the standardized language arts test. Findings indicated moderate improvement in classroom reading achievement, no improvement in graduation rates, and strong improvement in standardized test scores only for English learners. Results suggested that a balanced literacy program such as READ 180 may provide effective reading remediation for English learners in low-income urban areas, thereby promoting social change through increased academic success and upward economic mobility.

READ 180 Evaluation: Balanced Literacy in a Low-Income, Underperforming Urban High School

by

Daniel Lombardi

MA, Saint Peter's College 2003

BA, Montclair State University, 1999

Doctoral Study Submitted in Partial Fulfillment

of the Requirements for the Degree of

Doctor of Education

Walden University

April 2015

Dedication

I dedicate this to my father, not only to his memory but to the lasting impact he had on so many. You live on as I hope to, in the good deeds done to those that are most fragile. I can only pray one day, to have such an impact on those in my life.

"I went to the woods because I wished to live deliberately…and not, when I came to die, discover that I had not lived." –Thoreau from *Walden*

Acknowledgments

I would like to thank my committee members for all their help and guidance throughout this process, especially Dr. Behrman. It was Dr. Behrman's attention to detail and vision that guided me during the most trying times of this process. For being an endless supply of support and understanding, I must thank the one person that made this most possible; I would not have had the opportunity or ability to perform this task without my wife, Sabrina. As hard as it was to sacrifice so much time with my son, for keeping me motivated with his limitless love and the joy he brings me, I must thank Daniel. It was your unyielding smile I could always count on during the hardest moments of this process. I can still remember, vividly, sitting down with my mother as she taught me to read before I was old enough for kindergarten. For instilling in me at the youngest of ages the ability to learn and desire to acquire knowledge for myself, I thank my mother.

Table of Contents

List of Tables

List of Figures

Section 1: The Problem

Introduction

Research has indicated that reading achievement can be increased in upper grades with a balanced literacy program that emphasizes both bottom-up processes such as decoding and fluency and top-down reading processes such as making meaning (Kim, Capotosto, Hartry, & Fitzgerald, 2011). In this evaluation study, I investigated the extent to which a balanced literacy program was effective as an intervention for high school students at a low-income New Jersey school with a high percentage of students from racial and ethnic minority groups. The local problem this study addressed was a high proportion of below-grade reading scores among students at the completion of 10th grade.

The specific classroom intervention used at the project site to supplement reading instruction was the READ 180 program, which supports reading skills development and reading comprehension with a multimodal approach in the form of whole-group instruction, small-group instruction, independent reading, and individualized work at computer stations (Schacter, 1999). Whole-group instruction was used to teach and model specific skills, strategies, or vocabulary; small-group instruction provided directed applications at students' ability levels; independent reading afforded undirected applications of reading strategies; and computer stations provided instructional software for word study, comprehension, vocabulary, and spelling (Schacter, 1999). However, over the 7 years of implementation at the project site, no formal evaluation of the program was ever conducted. Therefore, I intended this study to assist future curricular implementation decisions at the project site by evaluating the success of the balanced

literacy program in accomplishing three program goals: (a) improving reading achievement at the classroom level; (b) improving scores on state language arts tests; and (c) increasing graduation rates.

This study includes four sections. Section 1 contains the definition of the problem, evidence for the problem, introduction of the research question, and a review of the literature related to the problem. Section 2 addresses the methodology used to investigate the research question, including the research design, instrumentation, data collection, and data analysis. Section 2 also presents and discusses the research findings. In Section 3, I explain the project I created based on the findings. Section 4 contains my reflections on the project's strengths and limitations in addressing the problem. The appendix includes the white paper created from the outcomes of the study.

Definition of the Problem

The local problem this study addressed was low reading achievement in a low-income, underperforming urban high school in New Jersey. At the time of data collection, the State of New Jersey Department of Education (2013) classified the high school that served as the research site as a "focus school" in need of support because of ranking in the lowest 10% of the state based on low graduation rates, low student achievement, or disparity between high- and low-performing students. In addition, the research site's school district was rated among New Jersey's poorest districts under the provisions of the *Abbot v. Burke* decision, which required adequate funding be provided to poor districts on par with what wealthier districts receive (Education Law Center, 2013). Indeed, the district served a community with per capita income of only $20,000 and almost 20% of

residents below the poverty line (U.S. Census Bureau, 2013). Of 2,526 students at the high school as of 2013, 92% were Hispanic and 66% qualified for free school lunch (State of New Jersey Department of Education, 2013b), with about 20% English learners (State of New Jersey Department of Education, 2010a). As it is not uncommon among low-income school districts with high numbers of people from minority groups, reading achievement of high school students at this school was below expected standards. Low achievement was evidenced by performance on the 2010-2011 state language arts assessment indicating about one fourth of the students below proficient (State of New Jersey Department of Education, 2010a). In addition, the average verbal SAT score at the school for years 2008, 2009, and 2010 was 405, 406, and 410, compared to state averages of 494, 496, and 493 (State of New Jersey Department of Education, 2010a)

To address the problem at the project site, the school instituted the READ 180 Reading Intervention Program from 2005 to 2012. According to the school reading specialist (personal communication, September 20, 2013), READ 180 program goals were to (a) improve reading achievement, (b) improve language arts scores on the state High School Proficiency Assessment (HSPA), and (c) increase graduation rates. READ 180 is a balanced literacy model that emphasizes deriving meaning from text through student-centered, holistic reading experiences but also develops reading subskills through teacher-directed instruction. According to Scholastic (2013), its publisher, READ 180 emphasizes text comprehension, vocabulary, and spelling, while giving support when needed to phonemic awareness, phonics, and fluency. Instruction in each 90-minute lesson involves a balanced literacy approach with four principal delivery modes to

improve reading proficiency for struggling students: (a) whole-group direct instruction to develop background knowledge, model fluent reading, and teach vocabulary; (b) small-group instruction to provide teacher-directed applications of skills, strategies, and vocabulary, based on student needs; (c) independent reading at the student's ability level; and (d) adaptive instructional software to provide individualized practice in reading, spelling, vocabulary, and writing (Salinger, Moorthy, Toplitz, Jones, & Rosenthal, 2010). Students are given a reading inventory prior to instruction and are matched, based on a Lexile Framework for Reading, to independent reading texts and instructional software that progressively increase the difficulty of text to match student Lexile measurements (Scholastic, 2013). The publisher claimed that when properly implemented, increases of 2 to 5 years of Lexile growth are common, based on national data (Scholastic, 2013).

For school years 2007-08 and 2009-10, students in ninth grade were enrolled in READ 180. For school years: 2005-06, 2010-11, 2011-12, placement into READ 180 was reserved for 10th grade students. Selection for the intervention was based on students who were at risk of failing yet still showed the highest probability of passing the state test. The program ran three to four regular education English classes and one to two special education English classes. Priority placement went to students considered "on the bubble" or close to passing by scoring within 10 points of a passing grade on the Grade 8 Proficiency Assessment. Similarly identified students who failed ninth grade English were then moved into the READ 180 program as sophomores. In some years when funding was available for a practice version of the state HSPA administered to freshman, additional in need students were added based on their scores also being within 10 points

of passing. Where additional seats remained, students were enrolled into READ 180 by the reading specialist, based on test scores and grades in English. Thus there was the possibility for disparity in achievement between the higher scoring students and other enrollees with lower initial scores. (school reading specialist, personal communication, September 20, 2013).

General education and special education English teachers underwent training from the publisher to teach the READ 180 classes, and the reading specialist provided one-on-one instruction to students not responsive to whole-group instruction. Students received READ 180 instruction for the first half of the school year in a 90-minute block each day (as opposed to the standard 45-minute period for other English classes). In the second half of the year, the regular English course curriculum was followed in a 90-minute block, so that READ 180 students would fulfill requirements for graduation. In September 2008, the school added two literacy coaches to improve reading test scores school-wide. In addition to supporting instruction in non-READ 180 classes, coaches implemented a state test "boot camp" for all students 2 months before testing to target specific skills needed to pass the assessment. Coaches aligned lessons across all English classes regardless of being in the READ 180 program or not. All students used premade lesson packs that the coaches disseminated to the classroom teachers (school reading specialist, personal communication, September 20, 2013).

Unfortunately, no analysis of program gains was ever conducted at the project site. The new district superintendent of schools chose to abandon READ 180 and replace it with a different reading intervention, the Reading Horizons phonics program, in

September 2012. Then, in September 2013, no formal reading program was used by the school at all, leaving teachers to create their own literacy interventions in the absence of any curricular mandate. Thus, there was a discernable gap in practice in which staff implemented reading intervention strategies without benefit of any evaluative data to guide curricular decisions, a circumstance compounded by the needs of students among the state's poorest and lowest achieving.

Therefore, the purpose of this study was to conduct a post hoc evaluation of the effectiveness of the ill-fated READ 180 Reading Program in relation to each of its stated goals: (a) to improve reading achievement at the classroom level; (b) to improve scores on state language arts tests; and (c) to increase graduation rates. The district established each goal as being directly related to a student's ability to exit school with reading skills necessary for opportunities similar to citizens in any other community. I hypothesized that a balanced literacy intervention involving direct instruction, independent reading, and skill-building reading software, with instruction differentiated by student ability level, could improve classroom reading achievement, language arts test scores, and graduation rates.

Rationale

Evidence of the Problem at the Local Level

Students at the project site consistently scored at the low end of state literacy assessments, most recently with only 71% of students scoring proficient and 4.1% scoring advanced proficient on the 2010-2011 state language arts/literacy test of the state HSPA (State of New Jersey Department of Education, 2010a). Further, three student

subpopulations tested lower than their statewide peers: 55% of students with disabilities scored below proficient, compared to the state average of 35%; female students scored 20.6% below proficient, compared to the state average of 7.3%; and male students scored 29.1% below proficient, compared to the state average of 12% (State of New Jersey Department of Education, 2010a). On the school's performance report, students at the project site met 0% of the state targets for college and career readiness and ranked at the 24th percentile compared to schools statewide (State of New Jersey Department of Education, 2012). Such targets are determined by the percent of scores aggregated for Scholastic Aptitude Test (SAT) and Advanced Placement (AP) testing, as well as participation in classes with rigor. For example, average Verbal SAT scores were 410, with the 25th percentile of students scoring 360 (State of New Jersey Department of Education, 2012), in contrast to the national mean of 500. On the School Report Card, students at the project site ranked at the 12th percentile statewide on the language arts portion of the HSPA (State of New Jersey Department of Education, 2012), used to gauge whether or not a child is gaining the basic core skills needed for graduation (State of New Jersey Department of Education, 2006). The school was recently downgraded from a statewide ranking of 318th to 320th out of 328 high schools in New Jersey (Schlager & Staab, 2012).

Evidence of the Problem from the Professional Literature

The problem of low reading achievement at the project site is consistent with a national problem of underachievement. U.S. schools have demonstrated an overall record of inadequacy, which can be represent by data from the National Assessment of

Educational Progress (NAEP): the average NAEP reading scale score for high school students in their senior year for the academic year of 2009 was 288, while a proficiency would require a score of 302 (National Center for Education Statistics, 2013). In urban schools where the population consists of mostly low-income earners, the problem of underachieving in reading is exacerbated. Students from schools that are economically disadvantaged score an average of 273 and urban students score on average of 283 (National Center for Education Statistics, 2013).

Across the United States, many high school students lack reading skills necessary for college level classes or employment opportunities for high wages (Carnegie, 2011). The number of students in the United States entering college lacking fundamental skills in both reading and writing continues to rise (Gruenbaum, 2012), with many students reading at only a level equivalent to a fifth grader as they begin their senior year of high school (Gruenbaum, 2012). In addition, high school graduation is not the indicator of preparedness for college it once was. For example, in Massachusetts only 46% of high school students were able to read at a 12^{th} grade level (Carmichael, 2012). Some states have improved graduation rates without improving reading skills, such as in Florida, where huge gains were made to increase high school graduation rates while the student average in reading skills simultaneously decreased to below the national average (Roth, 2010). Statistics have shown that of all high school graduates, close to half are not satisfactorily prepared for the reading they will encounter in college (ACT, 2013).

To overcome the lack of reading readiness, many colleges and universities offer remedial reading courses, with some colleges reporting numbers of incoming freshman in

remedial courses to be as high as 60% (Gruenbaum, 2012). For instance, at Westmoreland Community College in Pennsylvania in 2010, 65% of students took remedial English (Crawford, 2010). For the national graduating high school class of 2004, Kurlaender and Howell (2012) reported that 16.2% of incoming students at 4-year colleges and 26.9% of incoming students at 2-year colleges were enrolled in remedial reading. However, the actual number of students in need of remedial reading may be even higher because 33% of students who are recommended to enroll in remedial reading never enroll (Bailey, 2009). Unfortunately, remedial reading courses may not be enough to overcome the lack of requisite reading ability, as only 17% of students complete a bachelor's degree after taking a remedial reading course (Sladky, 2010).

The same reading skills needed for college success can also benefit first-time applicants to entry-level jobs, as reported by a study performed by ACT (Olson, 2006). The study analyzed student scores on the college admissions test and compared them to a database developed for the Department of Labor. Seven reading skill indicators were identified that were necessary to be successful for entry-level jobs. The study concluded that out of seven indicators for entry-level jobs, five were present on college admission tests. This result suggested that college success is crucial to postsecondary life (Olson, 2006).

Definitions

In this paper, the following terms are operationally defined.

Balanced literacy programs: Balanced literacy includes instruction intended to develop both top-down and bottom-up processes. Such programs are particularly

beneficial to students initially having trouble with reading and writing (Pressley, Roehrig, Bogner, Raphael, & Dolezal, 2002). Time in these programs is "balanced" between bottom-up activities such as phonics practice and top-down activities such as group discussion, inference, and writing reflections about the text (Duke & Pearson, 2008). Instructional methodology is also balanced between teacher-led and student-centered activities (Frey, Lee, Tollefson, Pass, & Massengill, 2005). Instructors attempt to immerse students in authentic reading activities that simulate context-bound situations with text that is relevant to those situations (Uzuner et al., 2011). Often bottom-up processes are the sole focus of intervention for students from minority groups with diverse language backgrounds. This singular focus may contribute to the achievement gap, whereas a balanced literacy approach may alleviate this shortfall of instructional practice (Au, 2009).

Cohort: Cohorts consist of a subgroup in the population sharing defining characteristics (Creswell, 2012). Cresswell (2012) stated cohorts longitudinally tracked over time can be used for studies regarding certain aspects. Cohort members consist of different individuals and no members exist at any time in another cohort.

Comprehension processes: Comprehension occurs with the simultaneous interaction of bottom-up and top-down cognitive processes in order to derive meaning from text (Rumelhart, 2004). Rumelhart (2004) described the act of reading as both a visual and cognitive action that uses parallel processes. He determined that comprehension of reading begins as the reader first visually takes in graphemic information. Next, parallel processes analyze and determine hypotheses to apprehend

meaning. The parallel processes then check and balance each other as the feature, orthographic, lexical, syntactic, and semantic knowledge within the reader combine and either discount, confirm, or amend hypotheses until meaning is achieved (Rumelhart, 2004).

Other researchers have supported this interactive theory. Stanovich, West, and Freeman (1981) concluded that the addition of a compensatory factor to this model would account for the unexplained results from experiments attempting to isolate the cognitive components of this theory. The compensatory assumption was that readers make up for weaknesses in one knowledge area by relying heavily on knowledge in other levels. Therefore, a reader could overcome a weak bottom-up skill-set with greater top-down functionality. Liu (2010) stated, in a broader sense, bottom-up processes begin as graphemes and turn into phonemes, which are then turned into a word as meaning is assimilated by the reader. According to Liu, top-down processes use prior reader knowledge of semantic and syntactic cues that may be useful to their own particular understanding of language. Finally, the reader uses knowledge to apply only the most needed cues to assimilate meaning.

Understanding can be developed by improving the reader's cognitive perception of how text consists of symbols that represent sounds and applying this to prior knowledge to gain meaning (Basaran, 2013). Automaticity of phonemic awareness is a requirement to having higher-order comprehension skills (Samuels & Flor, 1997). Readers who move sequentially from letter recognition to word recognition to units of meaning will struggle with comprehension, whereas readers whose word recognition is

automatized are more highly skilled at deriving meaning and moving toward higher order understanding (Samuels & Flor, 1997). It is only after a student is able to free up memory and cognitive resources due to automatizing the process of decoding words that he or she is able to put forth maximum effort towards comprehension (Samuels & Flor, 1997). Therefore, improving decoding skills may be especially important to allow beginning or struggling readers to move toward automaticity. Top-down contextual exercises such as prediction, main idea, and sequencing can help with higher-order comprehension skills only after the requisite decoding skills are in place (Dole, Duffy, Roehler, & Pearson, 1991). Ultimately, the convergence of top-down and bottom-up processes strengthens both (Rumelhart, 2004).

Reading achievement: Under recent iterations of federal and state education policy, reading achievement is most generally defined as the reader's ability to score competently on standardized assessments of reading ability (State of New Jersey Department of Education, 2014b). No Child Left Behind (NLCB) progress indicators are decided by scores on tests such as the HSPA. New Jersey performance indicators expand this to also include the SAT and PSAT measures of reading and math achievement. Acceptable achievement on the SAT and PSAT are determined by The College Board's benchmark, and HSPA achievement is determined by the New Jersey Department of Education (State of New Jersey Department of Education, 2014b). Reading achievement affects student placement into remedial programs, which in turn affects graduation rates. Up to 45% of remediated college students never graduate (The American Diploma Project, 2014). Teacher evaluation is also affected by reading achievement (State of New

Jersey Department of Education, 2014). Ratings of school effectiveness are also based on reading achievement, with ramifications for school funding and/or oversight (State of New Jersey, 2012). Furthermore, reading achievement may project towards future postsecondary learning and career readiness (Carnegie, 2011).

Reading comprehension: Reading comprehension is the ability to understand printed text at both denotative and connotative levels. Denotative comprehension is literal understanding of what is "on the lines" while connotative comprehension is inferential understanding of what is "between the lines" and applied understanding of what is "beyond the lines" (Duke & Pearson, 2008). Ability to comprehend at both denotative and connotative levels may be differentiated by text genre, style, or content area (Duke & Pearson, 2008).

Skills-based reading programs: A skills-based reading program places primary emphasis on bottom-up reading processes that allow the reader to transpose printed text to oral language. Thus, skills-based reading programs devote considerable instructional time to decoding the relationship between letters or words and the sounds they represent (Sterzik & Fraser, 2012). Such decoding of linguistic units is commonly referred to as phonics (Sterzik & Fraser, 2012). In essence, skills-based reading programs attempt to develop one or more of the reading subskills presumed to correlate with reading comprehension. Decoding is used to search for letter-, then word-based meaning, which is expected to lead to sentence and ultimately paragraph meaning. Such programs are often highly effective at beginning reader intervention (Slavin, Cheung, Groff, & Lake, 2008). However, reading comprehension is not always best served by these programs in

upper grades due to the different nature of text complexity and meaning (Slavin, Lake, Chambers, Cheung, & Davis, 2009).

Top-down reading programs: In contrast to bottom-up programs, top-down reading programs place primary emphasis on connecting what the reader already knows from prior reading or personal experience to gain meaning from new text (Sterzik & Fraser, 2012). Such programs are based on an inductive approach (Sterzik & Fraser, 2012). Top-down programs rely heavily on reader questioning, hypothesizing, and inferencing to construct meaning (Sterzik & Fraser, 2012). Instruction does not focus on phonemic skills or word-level skills, instead allowing children to acquire their own strategies for gaining meaning (Ryder, Tunmer, & Greaney, 2008).

Underachievement: Underachievement in reading is evidenced by a reader's inability to score competently on a standardized assessment of reading ability. Typically, state departments of education have established norms or levels that signify underachievement on state-administered reading tests. For example, in New Jersey, students are classified as *advanced proficient*, *proficient*, and *less than proficient* (State of New Jersey Department of Education, 2006). Less than proficient is considered underachievement.

Significance

Social Significance

The social significance of this study may be traced to Dewey's (2013) concept of democracy and education, which holds that a society can be considered democratic only if its institutions treat all members equally. According to Dewey, reading and writing are

more valuable than personal interaction in assuring that society maintains its cultural traditions and builds upon them. Citizens capable of critical thinking, along with understanding quantitative information, are more likely to discern conflicting economic reports and compare them to what is happening in their own communities (Hicks, Jacobs, & Matthews, 2013). Unfortunately, to the extent that U.S. public schools are not successful in developing literacy skills equally across all student populations, the American educational system does not meet Dewey's definition of a democratic institution nor provide the basis for universal civic engagement

Furthermore, inequality of educational achievement has also been shown to parallel economic disparity. For example, workers with no high school diploma earn an average of $471 per week compared with $652 per week for those having a diploma (U.S. Department of Labor, 2013a). The disparity grows as education levels increase, with associate's, bachelor's, and master's degree holders earning $785, $1,066, and $1,300 per week respectively (U.S. Department of Labor, 2013a). Further, the highest earners are typically the highest educational achievers, as corporate CEOs on average possess superior schooling (Forbes & Piercy, 1991).

Several reports from the U.S. Department of Labor regarding job growth and decline for the decade 2010-2020 highlighted the link between educational attainment and economic success. Of the 30 jobs that will have the highest rate of decline, only three will require more than a high school diploma (U.S. Department of Labor, 2013b). Of the 30 fastest growing jobs, 26 will require more than a high school diploma (U.S. Department of Labor, 2013c).

Educational Significance

The educational significance of this study may be evidenced through an understanding of how the central authority of states and the federal government over local determination has affected reading and literacy instruction in the United States. The reserve clause of the U.S. Constitution embodied in the Tenth Amendment reserved control for schools with the states, which in general ceded control to local school districts (National Archives and Records Administration, 2013). However, the failure of schools to provide for satisfactory literacy and academic attainment led to educational control in the mid-20th Century reverting to the states, with the 1973 *San Antonio School District V. Rodriguez* court case upholding state authority over local control (Wong & Langevin, 2005). At about the same time, spurred by the threat to American exceptionalism by the Soviet launch of *Sputnik*, the federal government began providing funding to public schools under the National Defense Education Act (NDEA; Kessinger, 2011). Although states were free to forego the federal funding, states became dependent upon the new source of income (Wong & Langevin, 2005).

At the root of NDEA and subsequent federal educational legislation was the theory of educational essentialism, which promoted the idea of emphasizing "core" academic subjects along with testing to prove mastery (Kessinger, 2011). The Elementary and Secondary Education Act (ESEA) of 1965 furthered the essentialist agenda and exerted more control through the mechanism of categorical funding (Kessinger, 2011). For the first time specific categories, as opposed to general aid, could be tied to parts of an education reform model. The impetus for ESEA was to wage what then-President

Johnson referred to as a “War on Poverty” through education as a key to his vision of a Great Society. ESEA placed specific emphasis on improving literacy and mathematics achievement in poor schools through provisions of Title 1. Soon thereafter, a National Assessment of Education Progress was developed in 1969 to gauge educational attainment at a national level (Kessinger, 2011).

Subsequent iterations of federal influence over educational policy included the founding of a federal Department of Education in 1979; adding federal standards and mandated accountability under the reauthorization of ESEA in 1994; and providing for student testing to be linked to state takeovers of underperforming districts under the 2001 reauthorization of (NCLB) (Kessinger, 2011). A curricular effect of linking school accountability to student test scores was a more narrow definition of successful literacy based on improved test scores (Hunsberger, 2007). Instruction became more about following a preset curriculum designed to pass the state testing (Meyer, 2013). Pressures on local districts to meet federal mandates of test performance encouraged many districts to replace more comprehensive intervention programs such as READ 180 with unidimensional, skills-building programs intended specifically to improve standardized test scores (Teale, Paciga, & Hoffman, 2007). An adverse effect of narrowing the scope of literacy instruction to pass state benchmark exams has been the reduced emphasis on higher-level literacy development, as required for college and job success, in favor of test preparation (Teale et al., 2007).

ESEA has not been reauthorized due to a lack of ability by the U.S. Congress to agree upon any revisions to the current NCLB version of the legislation (U.S. Department

of Education, 2013). In the absence of a revised version of ESEA, Race to the Top (RTT) was funded based on an interpretation of the way the president could use a portion of the $787 billion from the American Recovery and Reinvestment Act (ARRA). The interpretation would allow $4.35 billion of ARRA to be directed towards the RTT initiative promising schools more accountability and incentives for teachers (Onosko, 2011). Accountability under RTT is largely the same as it was under NCLB, with high-stakes testing conducted in two core subjects, reading and mathematics. Unlike NCLB, which allowed states to design and set their own assessments and thresholds for passing, RTT will provide $361 million to two assessment companies to create them (Onosko, 2011). The two companies will also provide curriculum and class materials for instruction, marking the first time ESEA or similar federal regulation has so directly imposed pedagogy on local schools. Unlike earlier federal reforms that incentivized school districts by moving funding from general to categorical aid, in RTT teachers will receive payment directly as individuals in the form of merit pay (Onosko, 2011). In such an environment, balanced literacy programs such as READ 180, which do not cater to testing markets, may be less sought after.. Thus, an important educational outcome of the present study will be to report data-based evidence of whether the balanced literacy approach can bring about improved reading test scores without eliminating the emphasis on developing higher level literacy skills as is common among skills development programs.

Research Question

The social significance of this study is tied to a strong relationship between reading skills and having equal opportunities in a democratic society, while the educational significance of this study is based on combating the narrowed literacy curriculum imposed by federal efforts to link school accountability with test scores. While the research literature has provided important findings regarding balanced literacy reading programs in general, the effectiveness of the READ 180 at the project site was the focus of this study. A turnover in administration and lack of program oversight at the project site has left an important question unanswered. Although test scores indicated that student achievement was less than desired, school officials did not know if the READ 180 intervention had any beneficial effect on improving classroom reading achievement, language arts test scores, and graduation rates of program participants during the 7 years of its implementation. Therefore, this study was guided by the following research questions:

1. As a result of participation in the READ 180 program, did reading achievement change at the classroom level, as measured by the Scholastic Reading Inventory? Were there differences in performance by gender?
2. What is the relationship of students' language arts scores/performance on the state-administered HSPA (HSPA) for all students and students enrolled in the READ 180 program? Were there differences in performance by gender?

3. As a result of participation in the READ 180 program, did graduation rates change during the course of the program's implementation? Were there differences in graduation rates by gender?

Review of the Literature

Theoretical Framework

The conceptual underpinning for this study is drawn from the theory of literacy development developed by Holdaway (1979). The theory of literacy development is founded on the concept of reading as an act of natural developmental (Morrow & Tracey, 2012). Holdaway proposed that the process of reading is rooted during infancy as children observe their parents read stories to them. According to the theory of literacy development, infants develop literacy by emulating real-life applications rather than exposure to explicit teaching. Holdaway emphasized the importance of emulation at all times, even if it is only a mere approximation of the actual skill. The natural evolution of reading mimics the way children develop oral language skills (Morrow & Tracey, 2012). Ultimately, children emulate parents using a gross approximation of reading (Morrow & Tracey, 2012). Early reading-like behaviors are attempts by the child to replicate the joy he or she receives from parental reading. Holdaway emphasized that replication is not, as some would assume, purely rote learning lacking emotional response and deep meaning. Rather, it is the child's own encoding attempts that come from a deep understanding of what is being read (Holdaway, 1979).

An important precept of the theory of literacy development is what Holdaway (1979) called *native language transmission (NLT)*. Transmission of language is

considered universal: language can be learned successfully regardless of culture or race due to the natural environment that it is used in (Holdaway, 1979). Holdaway summarized seven major characteristics of NLT:

- The learner is immersed in an environment that uses the skill in ways that are purposeful.
- Examples of the skill are used and reinforced realistically.
- Proper responses by the learner are immediately rewarded.
- Improper attempts are not rewarded.
- Both time on task and speed of lessons are determined by the learner.
- The environment is safe, nurturing, and presents no threats.
- Learning occurs continuously in a structured order of sequence that varies for each learner. (Holdaway, 1979)

To provide a successful classroom for literacy instruction in line with natural development, Holdaway (1979) recommended a combination of teacher-led reading, independent reading, artistic work centers related to the reading, vocabulary games or puzzles, and student choice in rereading or selecting text. Such activities allow for four key categories of literacy instruction. *Observation* occurs when a child sees an adult reading. *Collaboration* occurs when an adult or peer interacts with the child and gives any help that is needed along with motivation and encouragement. *Practice* occurs when the child is without adult observation and experiments with reading and writing alone. Finally, *performance* occurs when the child takes what has been learned and shares it with adults who are supportive (Morrow & Tracey, 2012).

The multimodal model of READ 180 follows a pattern that is consistent with Holdaway's (1979) four key categories of literacy instruction by incorporating whole-group, small-group, and independent topology within each lesson. *Observation* occurs during the onset of every lesson, with a whole-group instructional setting in which students view a short video accompanied by narration and music of the story they will soon read. The shared reading principle of opening warm-up is also fulfilled at this stage (Holdaway, 1982). *Collaboration* occurs at the end of this whole-group activity, when there is opportunity for shared discussion on what children think they will read in the story. Teachers take this opportunity to "think-aloud" with the students what to expect based on the discussion and video. Students then have a chance to apply the newly observed behavior with encouragement and help (Morrow & Tracey, 2012). *Practice* occurs as the lesson moves next to independent stations (either book reading or computerized instruction), where students work without supervision and can experiment with newly learned skills. Students have choice of what to read at either the reading stations or the computer station. The computer stations also have language games, which are also part of shared-reading strategy (Holdaway, 1982). The output for these stations is highly self-corrective and self-monitoring as students must monitor their own work, another component of shared reading. *Performance* occurs when students move to small-group instruction with peers and their teacher. They now have a chance to demonstrate what they have learned with encouragement and help of the teacher. Each paragraph of the story has places to stop and discuss newly learned skills. This discussion is also another opportunity for collaboration and fulfills the shared-reading strategy of *new story*.

The shared reading strategy of experiential meaningfulness occurs as there are many opportunities to connect events from the story to prior ones (Holdaway, 1982). Each story is included due to its high level of reliability to adolescents with embedded vocabulary. This small-group interaction is also an opportunity to perform creative and persuasive writing based on the readings. Another shared reading strategy occurs here, as there are places where reading is paused so students may predict and analyze plot events. During this step, students can self-regulate and self-correct, with intrinsic motivation based on peer acceptance.

READ 180 is also consistent with the theory of literacy developments emphasis on providing a nurturing literacy environment with low levels of fear (Morrow & Tracey, 2012). READ 180 encourages a low-stress classroom by grouping students according to ability. It also uses computerized instruction that matches students to tasks by ability level as determined by the Lexile Framework. In addition, stress is reduced because students may retry all lessons until a passing grade is achieved. Overall, the READ 180 balanced literacy classroom provides opportunities for teacher-led reading, independent reading, work centers, games, and student choice, all in low-stress setting, consistent with the principles of natural literacy development.

Review of Related Research

In the following review of related research, I first discuss the effectiveness of the READ 180 reading program and other balanced literacy programs. Then, to contrast effectiveness of balanced literacy reading programs with other reading programs, the review addresses the effectiveness of skills-based and top-down reading programs.

Search methodology. To obtain publisher-sponsored or authored READ 180 evaluation reports, I went to the Scholastic Inc. website and selected the "Research & Results" link. Next, I downloaded the *2011 Compendium of Research*. I also searched the Education Research Complete database and toggled full-text and peer-reviewed options after typing *READ 180*. To gather other balanced literacy program research I went to the Education Research Complete database. Once there I typed *mixed methods reading* and found no results. I tried *balanced literacy* and found 49 articles. I selected the full text only, publication years 2007 to present, and peer review filters. I also typed *literacy programs* with the same filters and found more results. As articles were found, I would use the reference list at the end of the article to obtain more potential sources. To obtain skills-based programs I used the same filters and typed *Phonics Program*. I also typed *Wilson Reading Program* into Google Scholar to locate more skills-based articles. I once again used the reference lists at the end of each article whenever possible to identify more articles.

READ 180. Studies of the effectiveness of READ 180 have appeared in (a) publisher-sponsored reports, (b) the What Works Clearinghouse, (c) doctoral dissertations, and (d) in limited number, professional journals or other publications. Overall, this body of literature concludes while many schools achieved positive gains due to the program, other schools have met disappointing results. Successful usage appears to be related to clear entrance criteria based on outcomes that can be aligned to the program. In general, students who are closer to grade-level reading standards seem to benefit more than students farther below grade level. Teacher training by the publisher also appears to

contribute to program success. Thus, schools that have had less success often display one or more similar attributes: little rationale over student participation and selection, local goals not aligned with program outcomes, limited teacher training, poor fidelity of implementation, and lack of program coordination.

Publisher-sponsored reports. A number of district-based evaluation reports supporting the effectiveness of READ 180 have been either sponsored or authored by Scholastic Inc., the program publisher. Unfortunately, these evaluation reports have not undergone the rigorous peer review required for publication in professional journals. Notwithstanding the potential bias of publisher-sponsored research, according to the most recent compendium of research reports (Scholastic, 2011), many schools achieved positive program gains. For example, the Colton (CA) Joint Unified School District reported significant gains of 26 scale-score points on the California Standards Test English Language Arts for the 2008-2009 school years with 212 freshman students selected for the program based on poor SRI and CST ELA scores. In 2012, the San Antonio (TX) Independent School District reported a significant improvement among 663 special education students from grades 4-11 on the TAKS exam after only one year of the program. The James A. Garfield Senior High School in Los Angeles (CA) reported in 2012 that struggling readers were now on track for college as CST ELA scores rose for students in the program in the categories of proficient and advanced proficient by 26% and 3% respectively. 232 ninth grade and 183 tenth grade students participated. Rochester (NY) East High School reported in a case study of several autistic students, from 2011 through 2012, that the program was able to improve reading based on lexile

growth for half of the learners. Hernando County (FL) School District, reported that in 2010, freshman (436) with learning disabilities exceeded yearly growth expectations on the FCAT reading exam due to the program. Huntington Beach (CA) Union High School District reported in 2009 that students placed into the program for being below grade level in reading and vocabulary actually outperformed students not in the program and deemed *at* grade level. These students outperformed the general education students by a statistically significant margin. In 2009 Springfield (MD) and Chicopee (MD) public schools reported that their READ 180 students outperformed their control group by a statistically significant margin on the Stanford Diagnostic Reading Test. 489 Students were selected based on below grade level reading scores. 159 students were used as a control group. 175 students were placed into the READ 180 program while 155 were placed in an alternative intervention. Students in READ 180 outperformed the control group by 5.28 scale score points, while the alternative intervention group only scored less than 1 scale score point over the control group. Findings of this publisher-sponsored compendium indicate that READ 180 works for high school students who are either in general education or special education programs.

What Works Clearinghouse. A second series of READ 180 evaluation reports, subject to a higher level of peer review, were submitted to the What Works Clearinghouse (WWC) evidence of standards review. Of 101 studies submitted to WWC, seven studies were sufficient to pass review, albeit "with reservation" ("READ 180," 2009). Haslim, White, and Klinge reported that program participants in Austin (TX) showed significant gains on the TAKS reading exam from 2004-2005. Participants were

614 seventh through eighth grade students selected due to poor TAKS scores prior to intervention with 94% being Hispanic. In 2002 Interactive Inc., conducted a study across seven school districts in the United States. Each school provided students based on a ranking of poor reading ability. Schools initially agreed to a random assignment to control and intervention groups, but ultimately violated this agreement based on the assertion that it would be harmful to not allow certain students to have the intervention. The random assignment was maintained only by Columbus (OH), Dallas (TX), and Houston (TX) for eleven schools. It was determined that a positive and substantively important, but not statistically significant, effect was found towards reading comprehension. Seminole County Public Schools (FL), Lang, Torgesen, Vogel, Chanter, Lefsky (2008) conducted a randomized control trial of 1,265 9th-10th grade students. Students were selected based on their 8th grade scores, of moderate to high-risk, for the Florida Comprehensive Assessment Test. Students were randomly assigned to either a control group or one of three interventions. Findings indicated a positive and statistically significant difference was found for READ 180 participants. WWC found this study's findings with an additional *reservation* since there was no way to determine if differential attrition occurred as well as the use of multiple imputations of missing data. Desert Sands Unified School District (CA) in 2008 conducted a quasi-experimental study under Scholastic Inc. Participants were in grades 6, 7, and 9 and were 64% Hispanic. Five-hundred-seventy students were selected for the study based on having below basic or basic levels of reading comprehension as defined by the California Standards Test of English Language Arts. Fifty-eight percent of the students could be defined as being

Language Learners. Half of the students were placed in the intervention and the other half was given the standard curriculum. The standard curriculum was either a Holt Literature and Language Arts curriculum or a Prentice Hall Literature Curriculum. A statistically significant difference was found for intervention participants as defined by WWC for the general literacy achievement domain. The Phoenix Union High School District (AZ) in 2006 conducted a quasi-experimental study by White, Haslam, and Hewes. Three cohorts of participants were selected over the course of the 2003-2004 (cohort 1), 2004-2005 (cohort 2), and 2005-2006 (cohort 3) school years. Students were selected based on reading one or more grade levels below grade level. A comparison group was matched with similar reading level and demographics for each cohort. The Terra Nova test was used to measure achievement levels. A positive and statistically significant result was found for READ 180 intervention participants of cohort 2 only. New York Community School District (NY) in a quasi-experimental study by White, Williams, and Haslam (2005) on students in grades 4-8 compared test outcomes of students in READ 180 to peers not in the program. Sixteen schools participated with the majority of students (86%) being African American as well as on free and reduced lunch (90%). Participants were compared based on the New York State Department of Education End-of-Year Test in English Language Arts. Scholastic reported the findings to be statistically significant while WWC found no statistical significance either with results individually or averaged. Virginia (2006) Woods conducted a quasi-experimental study on three cohorts of students at an urban middle school. Fifty-eight participants were selected for the intervention based on pretesting and guidance counselor recommendations and were

matched to peers in a control group. Findings for the study were indeterminate. Overall, the effectiveness report of WWC was performed on the seven studies, and in some instances recalculated equations to determine accuracy of findings, to determine the affect on reading domains of comprehension and general literacy achievement. The report findings would appear to suggest that READ 180 is beneficial to comprehension and literacy since three studies showed positive or statistically significant effects on comprehension, while two showed statistically significant and positive effects on general literacy.

Doctoral dissertations. Several doctoral dissertations have reported on the efficacy of the READ 180 program related to either minority, learning disabled, or English language learner adolescent populations. Key findings support the program's ability to help students increase comprehension in both basic and academic comprehension in a wide range of sub-groups if they first have a requisite skill set. Findings also show students need more than one year of intervention if they are to make strides beyond the basic comprehension skill set.

Nave (2007) studied 160 fifth and sixth grade students in Sevier County, Tennessee, schools for the 2004-05 school years. At-risk students were selected based on being in the bottom quartile of the Tennessee Comprehensive Assessment Program (TCAP). Scores were then compared with an ex post facto research design for students with similar results that were not in the READ 180 program. Students enrolled in the intervention had significantly higher results. It was determined that neither gender nor socio-economic status impacted achievement results.

Vogel (2013) in a study of twenty-one 9th grade students at a Title I school (CA) to determine comprehension levels and reading attitudinal effects of the program. Students taking part in the program were found to have more enjoyment during independent reading. Writing ability was increased at the basic level but not advanced. No students were able to reach the 70% benchmark score deemed successful by the program. The study concluded READ 180 is successful in part, when modified to fit individual students' needs.

Wu (2009) conducted a case study (FL) of four adolescent English Language Learners (ELL) perceptions regarding the READ 180 program. Student needs were analyzed with constructivist grounded theory. Researchers found that with varying levels of English language learner 's (White, Black, Asian, Hispanic) it was difficult for the program and instruction to match individual needs. Students had a difficult time inferencing or extending content beyond literal interpretation due to a lack of conversational skills from lower level English language learner 's. This made it difficult for students to benefit fully from small-group activities, while a disconnect from the cultural content prevented them from engaging texts more fully. Yet, the repeated practice of skills and scaffolding of comprehension strategies during group reading of texts was seen as beneficial for students' basic comprehension.

Bishop-Kallmeyer (2008) performed a study in Ohio to determine READ 180 effectiveness at increasing reading achievement among special education and English language learners. Both groups showed a positive impact. A comparison group was used to determine pre and post test results in a quasi-experimental design. Participants totaled

47 seventh graders and 34 eighth graders between both groups. Special education students showed no gains from the program, while English language learners showed statistically significant gains in inference.

Robby (2008) performed a program evaluation in California of 260 READ 180 participants compared to a teacher generated intervention of 280 participants with quasi-experimental design. Both groups were freshman reading below grade level and deemed at-risk. Inconsistent cohort affects prevented clear analysis of one year's data, but the successive year's data indicated that a statistically significant gains for English language learners were present.

Parker (2011) studied READ 180 in Texas and compared its effectiveness to another reading intervention program Voyager Journeys III. Freshman students made up the cohorts at an urban high school. READ 180 students were found to have a statistically higher TAKS score then Voyager Journeys III participants, while the latter had statistically greater SRI results than the former.

McWhorter (2009) performed a quasi-experimental study of below average READ 180 students to determine if the program was not only effective at increasing reading comprehension but also math achievement. Four-hundred-fifty-four ninth grade students from a Title 1 high school took part in this study. The comparison group was constituted of average reading ability students. Findings indicated that no statistical significance was found for READ 180 participants.

Nelson (2008) attempted to determine if students' characteristics could be used as predicative factors for success in use of READ 180. The researcher examined grade level,

pretest scores, and English language learner status. Thirty-seven students participated and were in grades six through eight. All participants achieved positive gains. Students scoring in the lowest 1/3rd of the pretest had the greatest gains on the posttest. Factors of being a language learner or of a certain grade level made no impact.

Sigears (2008) in a quasi-experimental study explored students with learning disabilities gains in both READ 180 and traditional reading instruction courses. Two middle schools participated in the study for a total of 44 students. Researchers evaluated test results and determined minority status special education students participating in READ 180 were found to have a statistically significant gain over non-minority special education students participating in READ 180.

Professional journals or other publications. Peer-reviewed studies of READ 180 have also recently appeared in professional journals, though in limited quantities. Melekoglu (2011) investigated both reading achievement and reading motivation among 38 middle and high school students. The study found that students with learning disabilities (LD) had comprehension gains on the posttest compared to the pretest. Students without LD also scored significant gains in comprehension. Motivation among LD students was not higher after the program, but non-LD students did increase motivation. The study concluded that READ 180 contributed to gains in reading. It also determined that due to the small group studied and the low self-concept of LD students going into the study, it would take longer to increase their motivation as compared to non-LD struggling readers. Findings of this study suggest that despite the longer time LD

students require to show equivalent gains from intervention, READ 180 is successful at increasing comprehension.

Kim, Samson, Fitzgerald, and Hartry (2010) used an experimental model to isolate the effects READ 180 had on 294 children in grades 3-5. The READ 180 version had a modification to fit a 60-minute period (as opposed to the usual 90) and the whole group instruction segment was not used. Members of the READ 180 treatment group had increases in comprehension and word reading proficiency. They also outperformed the control group with statistical significance in regards to oral fluency. The control group received a district-made after-school enrichment program and also showed gains in comprehension and word reading proficiency. The study concluded that both programs had similar designs and instructional materials for literacy improvement, which explained why the control group had similar gains. Findings of this study seem to indicate that students benefit from balanced literacy instruction in regards to comprehension and word reading proficiency, but READ 180 students benefit greater at oral proficiency.

Kim, Capotosto, Hartry, and Fitzgerald (2011) used an experimental model design to identify the impact READ 180 had on children in grades four through six. The findings were that reading comprehension as well as vocabulary increased with statistical significance over all grade levels through the course of the study. A further examination showed that students who participated more actively scored better on vocabulary and comprehension than students who did not. The study concluded that students of moderate risk in the 40-45th percentile (Stanford Achievement Test) can benefit from the balanced literacy approach of READ 180. This study would appear to show that READ 180

students that are moderately at risk benefit from how much time they spend on program tasks.

A meta-analysis by Frattura and Capper (2006) showed READ 180 to also fit an Integrated Comprehensive Services (ICS) model. ICS models of reform focus on equity, having equitable structures of education services, implementing change, and providing high quality teaching and learning. These models attempt to support students *before* they fail, and have a structure of non-segregated services seamlessly woven into the structure of the school for all students' needs. The study found that when students are placed in reading intervention programs due to labeling and are pulled out of regular class settings they do not benefit the same as in an ICS model. Students with special needs perform at a higher level when in an equitable structure. The conclusion of this study was that when students chose to go to the READ 180 program on their own, and it was not a segregated model, it would be part of an effective ICS reading program. Findings of this study suggest that READ 180 works with special needs students in ICS models helping them to perform at a higher level.

Slavin, Cheung, Groff, and Lake (2008) evaluated reading programs based on a best-evidence synthesis of reading research studies. The study evaluated several READ 180 studies and used statistical analysis to determine efficacy and effect size for each. READ 180 as a balanced literacy reading program had *moderate* effect on students, while no other program attained the higher designation of *strong* effect. The conclusion was there may be confounding variables because students in this program receive many more minutes of reading (90) per day than control group counterparts. The study also

concluded READ 180 is an effective balanced literacy program. This study suggests that not only the program itself but the additional time of student exposure to intervention can impact results.

Papalewis (2004) in a two-year quasi-experimental study of 8th grade students at an urban school looked at 537 students. Participants were selected based on SAT-9 scores, subject grades, and teacher recommendation. To qualify students needed at least a "D" or worse in English class and a failing grade on the district writing exam. The majority of participants (78%) were Hispanic. Intervention students made significantly higher gains than comparison group students. READ 180 participants showed gains of over three normal curve equivalents in reading and two normal curve equivalents in Language Arts on the SAT-9. Further analysis showed that the 42% of Hispanic students that were classified as Limited English Proficient (LEP) scored gains identical to non LEP students. It can then be determined that READ 180 is as beneficial to LEP students as to non-LEP students. This study suggests LEP minority students can benefit from READ 180, even if they are below average readers.

Mims, Lowther, Strahl, and Nunnery (2006) performed a one-year program evaluation of the Little Rock School District (AR). Participants were drawn from 5 middle schools (6-8 grade) and 5 high schools (9th grade) for a total of 1,000 students. Literacy Scale scores and Proficiency Levels from the Arkansas Benchmark examination and the Iowa Test of Basic Skills (ITBS) were used. No significant effects were found for grade 6th students. A statistically significant negative effect was found for 7th grade students on the ITBS with no effect on the Benchmark. A statistically significant negative

effect was found for the ITBS writing assessment for 8th grade students. A statistically significant negative effect was found for 9th grade students on the ITBS for vocabulary, reading comprehension, and total reading. Overall conclusions were that the READ 180 program had negative effects on the ITBS results as well as the Benchmark. Evaluators discovered that the students were not randomly assigned and that selection bias may be at work. Another conclusion was that the state exams used to assess achievement may not be aligned with program outcomes. Also, room observations by researchers determined that teachers were not using the program materials more than 60% of the time. Computer portions of the lessons were not adequately completed by students due to technical issues and time constraints of class schedules. Furthermore, it was found that only 18 of the teachers were considered certified by READ 180 based on training, while 269 were not. Program fidelity appears to be a factor in the final student achievement results according to this evaluation. Report findings appear to indicate that a lack of program fidelity, either staff based or logistical, is of major importance to program results. Alignment of program outcomes with expectations of the measurement instrument also seems to impact findings.

Hasselbring and Goin (2004) performed a study in the Orange County Public Schools (FL). Sixty-three students from grades 6-8 who were designated as *most disabled* were chosen for the intervention while 62 students from grades 6-8 designated as *below grade-level readers* were chosen as the control group. The intervention used was the Peabody Literacy Lab, which was the prototype for the computerized portion of the READ 180 program. The Stanford Diagnostic Reading Test was used to measure growth

first in the fall then again in the spring. Significant gains were found for the sub-tests of Auditory Vocabulary, Literal Comprehension, Inferential Comprehension and Total Reading Comprehension. Teachers who participated in the program were interviewed and conveyed an opinion that strongly backed the program based on their observations as instructors. Findings by this study seem to show that READ 180 was successful for several categories of comprehension for participants below grade level.

In addition to articles in professional journals, two other evaluation reports have been published by respected research organizations. In a report prepared for the Milken Family Foundation, Schacter (1999) examined reading programs for grades pre-k through 4th grade. Schacter reviewed Hasselbring and Goin's earlier work with the Peabody Literacy Program. He determined the gains found by Hasselbring did not account for the lack of a control group in his earliest work before publishing his paper with Goin. Schacter then identified another 10,000 students originally participating in an Orange County Public Schools (FL) study by Hasselbring and Goin (2004). Students gained 33 percentile points on reading achievement on the Degrees of Reading Power test. Schacter determined the students participating began at a very low level, which amplified results (1999). This report appears to support the idea that much of the research on balanced literacy programs such as READ 180 is in need of refinement, while also confirming a positive effect for participants using program resources.

The American Institutes for Research performed a study for The Council of The Great City Schools and Scholastic, Inc. analyzing the implementation and results of READ 180 use in five poor urban districts (Salinger, Moorthy, Toplitz, Jones, &

Rosenthal, 2010). Each participating school was kept anonymous, but geographic and demographic data was included. All schools had the majority of students as minorities, all were mid-sized to large, two were in the southwest, one was in the southeast, one was in the south, and one was in the northeast. Implementation was seen as a key to the program's success. The researchers also determined a clear entrance and exit criteria for students were present in schools with positive results. In schools that did not have favorable results this was often seen as one of the factors. Not following the recommended time for each instructional rotation was also seen as a factor that negatively impacted results for schools. Districts with a dedicated staff person to monitor, collect, and analyze program data known as intermediaries used program data more effectively. Researchers also discovered that a lack of systematic information loops for free and easy flow of information between the administrative and classroom levels was a hindrance to successful implementation. The highest-achieving district also had the most systematic information loops and intermediaries. The district with the most time in rotations and using the computer stations also had the highest gains. This study seems to indicate having proper implementation, with dedicating a staff member, and using proper information flow coupled with following the recommended time in group rotations is of major importance to balanced literacy program success.

Other balanced literacy programs. In addition to research studies that focus specifically on the effects of READ 180, several studies have reported on other balanced literacy programs. Wonder-Mcdowell (2010) evaluated 133 second grade students in an urban setting transitioning from a separate skills-based program into an integrated

program that was aligned to collaborative classroom practices and curriculum was examined. The findings of the study were that statistically significant differences were found for students undergoing this new approach. It was concluded that after years of having separate skills being taught by programs in isolation, the restructuring of the intervention to collaborative integrated classroom instruction was best for student remediation. This study would appear to suggest that balanced literacy approaches benefit from collaborative integrated instruction.

Vaughn et al. (2012) examined 28 eighth graders who demonstrated resistance to intervention (RTI). RTI students in this study were part of a three-year longitudinal individualized intervention due to their specific needs. They concluded students can succeed when they are first identified by prequalifying criteria for individualized instruction for comprehension. The program compared standardized and individualized treatments that assisted with building phonics, word reading, fluency, vocabulary, and comprehension based on the needs of each student as evidenced through testing. The intervention classes were given daily in groups of two to four. Students were able to receive any combination of these treatments based on what they needed for the individualized group. The standardized treatment group was given all treatments in successive phases. Treatments were given with collaborative grouping (fluency), individualized computer instruction (decoding), and teacher-led small-group reading (comprehension). Study findings were that skills such as phonological awareness, rapid naming ability, and verbal ability were key identifiers in students that succeeded or failed in either intervention. The study also found that students with severe reading problems

were able to maintain progress with peers only if support was continued. Students that did not have severe reading problems were able to demonstrate statistically significant growth with either treatment for comprehension compared to the comparison group. There was no evidence to support a more successful treatment between groups for reading sub-skills. Findings of this report indicate students can succeed in a collaborative or standardized balanced literacy program when they are first identified by prequalifying criteria, and students with severe reading problems require continued intervention to show gains. Also, students without such problems were able to show significant gains in a much shorter time.

Savage et al. (2013) studied the Abracadabra reading program as part of a balanced intervention randomized control trial of 1,068 k-2 students. The program was used a web-based literacy system for 2 hours a week over 10-12 weeks. Teachers used the program in their regular classrooms as part of lessons. During the Abracadabra lessons teachers used whole-group instruction, small-group collaborative assignments as well as the web based system for reading skills and comprehension. The study found that reading sub-skills increased significantly for students but not in comprehension or oral fluency. Researchers concluded that the wide differences of implementation due to the teacher-created extension activities could have impacted the results. Findings of this study appear to show that reading sub-skills can be improved with balanced literacy instruction.

Hausheer, Hansen, and Doumas (2011) performed a study of the Read Right program and evaluated its efficacy by examining 24 fourth through sixth grade students.

Students were in groups no larger than five. The program consisted of students using computer audio readings of expertly read passages followed by a coached reading with instructor feedback, and an independent reading session with a critical thinking component. The study found that the reading comprehension and fluency based program, which patterned strategies it determined as constructivist, significantly increased both skills. The small sample size limits the generalizability of the findings and the researchers recognize some of the gains may have been due to maturation. A future study is recommended with a control group. It would appear that this study shows balanced literacy instruction can increase skills in reading comprehension and fluency.

Lawrence, McNeal, and Yildiz (2009) studied 12 urban High School students and analyzed the effects of using an alternative reading summer program involving graphic novels to connect reading, writing, and technology while improving their own media and critical literacy. Students were given whole-group lessons of key skills needed and then worked in small groups to complete tasks based on the whole group lessons and individually. Several of the lessons and tasks were computer based. The study found that students were able to complete all tasks and demonstrate media and critical literacy. The researchers concluded that students can increase media and critical literacy if teachers scaffold the work in groups, allow students to work collaboratively, and connect the work to prior knowledge. The study also concluded that the technology component was not effective in increasing students' ability to critique texts. Findings of this study would appear to suggest that balanced literacy instruction is effective in increasing media and

critical literacy, and technology does not necessarily have to be part of the intervention structure.

Parker, Holland, and Jones (2013) studied effectiveness of the Voyager Journeys III reading program with 114 ninth grade students. The reading comprehension program has whole-group, teacher-led instruction, small-group instruction, and computer-based individualized instruction. The study found that students had a statistically significant growth in lexile score from the pre and post test, but did not have a statistically significant better score on the Texas Assessment of Knowledge and Skills reading component compared to the READ 180 reading program. The study concluded that since the TAKS is a criterion-referenced test and the lexile measurement was norm-referenced, further study is needed with a control group. Findings of this study would indicate that program outcomes of balanced literacy interventions should be looked at before implementation to see if it is aligned with district desired outcomes. Alignment is necessary since lexile growth does not always positively correlate to criterion-referenced testing for all balanced literacy programs.

Denson (2008) in a study for Cambium Learning, Inc., publisher of the Voyager Journeys program, evaluated ninth grade students in the Dallas Independent School District (TX) high school using this balanced literacy program. The majority of the students were Hispanic (78%) and LEP (63%) with 68% receiving free or reduced lunch. Students in the program had a statistically significant growth compared to a control group. Placement of students was determined by standardized test scores of the TAKS in reading below the 40th percentile as well as bilingual program designation. Findings of

this study are in contrast to the Parker, Holland, and Jones (2013) study, which determined that Voyager did not increase criterion-referenced test results. It could be inferred from these two reports that a key difference and factor between increase in results shown in the Denson study and lack of increase in Parker, Holland, and Jones (2013) would be that students were in the 40th percentile and had clear entrance criteria. Kim, Capotosto, Hartry, and Fitzgerald (2011) support that the 40th percentile is a factor in successful implementation and the AIR (2010) report supports the need for clear entrance criteria.

Skills-based reading programs. In contrast to research on READ 180 and other balanced literacy reading programs, several recent studies have reported on the effectiveness of more targeted, skills-based reading programs that emphasize one or more reading sub-skills. Such programs may be effective at building decoding skills and denotative (literal) comprehension, yet lengthy transition is needed for program exit, and connotative (inferential, applied) comprehension is rarely increased. Kostewicz, Lemons, Mrachko, and Paterra (2012) studied eight children with intellectual disabilities (ID) who were placed in a phonics skill-based reading program. Participants showed moderate gains in decoding and showed more than double the rate of learning for decodable and sight words than a previous study. Significant gains were shown in all program specific measures. Phonological awareness was shown not to be improved. The study also concluded that near transfer skills showed no sign of increasing as part of the intervention. This study would seem to suggest that ID students can achieve gains in phonetic skills while not increasing phonological awareness or skills transfer.

Duff, Hayiou-Thomas, and Hulme (2011) examined 59 children aged 5-6 from eight primary schools in England that were not classified as learning disabled, but were recommended by teachers as being able to benefit from reading intervention. The program addressed literacy skills of letter knowledge, phonological awareness, spelling, grammar and vocabulary. Daily phonics instruction was augmented with weekly or bi-weekly independent, shared, or guided reading. The findings of the treatment groups displayed significant greater progress in early word reading, phonetic spelling, and phoneme awareness. Yet, six months after the conclusion of the program students were retested and found to have no difference from the control group with their progress of letter knowledge, early word recognition and spelling. Students even developed a slower progress regarding phoneme awareness, phonetic spelling and word reading than before. The conclusion of the study would be that students should be phased out of intervention programs to better maintain retention. Findings of this study show that phonics skills can be increased in the short term, but to maintain long-term retention students must slowly transition out of the program. The shared and guided reading component of this program also may indicate top-down factors were involved, yet when suddenly stopping one of those components all skills regressed. Only in conjunction was the intervention program successful.

Shaw and Davidson (2009) studied the Phono-Graphix phonics reading program and the use of this skills-based program on 4 primary 2 school age children in Scotland from economically disadvantaged settings. The children had all been previously receiving intervention for reading before the study began and two of them scored low enough to be

considered non-readers. The study determined that the children all made increases in sub-literacy skills taught by the program but only half were able to transfer those skills into reading comprehension and reading accuracy score improvements. The study concluded that the children were helped at a greater rate than they would have because of this program. Study findings suggest that the intervention factored into a skills increase but comprehension and accuracy did not always follow this increase. One could then infer that reading comprehension and accuracy may involve more than sub-skills.

Boltzmann, Rüsseler, Zheng, and Münte (2013) used an experimental quantitative design to identify the impact of the *Alpha Plus* reading program on phonological awareness, fluency, graphemes, and phonemes. The participants were 25 adult illiterates aged from 25-58 years with average to below average non-verbal Culture Fair Test I.Q. scores. The study found that activity in areas of the brain responsible for grapheme-phoneme conversion, words, pseudo-words, and orthography were all increased. Also, basic reading and writing ability (gauged by spelling) had increased. This study seems to indicate that illiterates have underdeveloped regions of the brain that utilize oral language, speech production, and phonological processing. It would go further to suggest proper sub-skills interventions can retrain these areas and increase literacy.

Top-down reading programs. Peer-reviewed research has also been conducted on a third category of reading programs that may be best described as top-down reading models that emphasize connecting text to prior knowledge. Whitehurst, Arnold, Epstein, Angell, Smith, and Fischel (1994) studied low-income children at a daycare center (NY). Children were either read to with dialogic reading techniques by teachers and parents,

teachers only, or in a control group. Dialogic reading is a shared reading experience where children take on active roles in the story-telling process as adults use probing open-ended questioning techniques (Whitehurst et al., 1994). Children in both treatment conditions demonstrated significant increases in verbal fluency and vocabulary. Findings of this study suggest that in beginning readers fluency and vocabulary are affected by active shared reading techniques.

Hargrave and Sénéchal (2000) studied 36 preschool children below age level for expressive vocabulary in Ottawa (CA.). Children were placed in a shared reading condition and a shared with dialogic reading condition. The researchers validated the purpose of this study with the correlation of vocabulary skills in first graders being correlated very strongly to reading ability in later grades. Children in both groups showed positive gains in vocabulary by the end of the intervention program. Children in the dialogic reading group had significant gains in expressive vocabulary. This study would appear to suggest that there is a link between vocabulary development and reading development. It suggests further that active shared reading experiences increase vocabulary sets in participants, which in turn may increase their future reading ability.

Whitehurst, Zevenbergen, Crone, Schultz, Velting, and Fischel (1999) performed a study of 280 poor urban children from pre-k through grade 2 in a Head Start center (NY). Children were randomly assigned and a pretest was given. A posttest was given at the end of pre-k, first grade, and second grade as the cohorts were tracked. Children were of poor emergent literacy compared to their peer norms upon entry to the program. Study findings were that children started at the 12th percentile and moved to the 40th by the end

of second grade. Researchers concluded participants were catching up a fast rate to their peers and were no longer falling behind. Researchers also concluded that a lack of generalizability between results and reading scores in first or second grade was due to a misalignment of testing and program outcomes. The testing used by schools was a phonics based assessment of sub-skills, while the methods used by researchers were semantic and narrative knowledge based. This study would seem to suggest that dialogic reading can help close achievement gaps in literacy for poor urban students, despite not being obvious due to a narrow use of reading assessments by public schools.

Top-down reading models are also used in reading for academic interpretive purposes. In a study by Applebee, Langer, Nystrand, and Gamoran (2003) of middle and high school across 5 states the use of dialogic reading for academic literacy, or readings with a need to interpret beyond the textual meaning, was examined. Within each state one urban and one suburban school was selected, within each district one middle school and one high school was selected, for a total of 20 schools. One middle school dropped out before they could complete the study. Four classes were selected in each school over a broad range of levels. Due to low enrollment in some classes only 64 remained throughout the study for a total of 1,111 students. Study findings were that literacy for academic reading showed higher literacy performances in classrooms engaging in dialogic reading and discussion methods. It was also found that based on GPA and overall academic performance *better* students benefited the same as *poorer* students from the model. Findings of this study would appear to show that top-down reading methods successfully transfer oral language expressiveness of inter-text concepts.

Discussion. The studies reviewed in this section indicate that balanced literacy programs, sub-skills reading programs, and top-down reading programs may all be effective in achieving program goals, but program goals are differentiated by type of reading program. Specifically, balanced literacy programs promote development of denotative and connotative comprehension, and are effective when students have attained, or are supported in improving, foundational reading sub-skills. Sub-skills reading programs promote decoding and fluency, but attainment of these sub-skills does not always translate into denotative comprehension gains. Top-down reading programs promote development of connotative comprehension and appear effective in bolstering individual interpretation of text but may not support foundational reading skills or denotative comprehension.

Students with very limited ability at decoding words appear to need skills-based instruction before they can analyze a text structure (Wonder-Mcdowell, 2010). This conclusion could also be supported by the fact that in the study by Kim et al. (2011) only students who scored at the 40-45th percentile on the SAT 9 or SAT 10 comprehension test showed significant gains in reading comprehension. Students scoring in that range are just above foundational reading and just below grade level. As Vaugn et al. (2012) found, having the foundation of phonological awareness, rapid naming of words and verbal ability were key in identifying students succeeding in a balanced literacy comprehension intervention. It appears that blending and decoding skills are necessary before a student can become a highly fluent reader and comprehender (Savage et al. 2013). Therefore, the

balanced literacy, comprehension-based approach may succeed if instruction is geared toward individual student abilities.

Yet, even students with adequate decoding ability may not progress to academic literacy, or interpretive comprehension. Whitehurst et al. (1994) showed that expressive ability and vocabulary are important factors in interacting with and understanding text. Phonics can increase word recognition but not necessarily vocabulary or expression of meaning. They then showed how a top-down literacy intervention could close the academic comprehension gap in students, even if they were lower SES. Applebee, Langer, Nystrand, and Gamoran (2003) identified students that were literate and could decode words yet were unable to grasp comprehension beyond the literal meaning. It was then demonstrated how a top-down literacy intervention can help with this aspect of literacy regardless of student GPA level.

Any intervention method would seem to need lengthy application duration with a slow transfer strategy out of the program. Having continued support would appear to prevent the regression shown by the research and avoid students being placed in reading programs that do not fit their needs. It would also seem to help them from leaving programs before they can apply skills with long-term effectiveness on their own. This continued support of students is what Vaugn et al. (2012) also concluded was needed because of many students losing gains if they leave a program. The research seems to show that students have trouble transferring literacy sub-skills from a sub-skills based program into regular classroom text reading in a short time frame. Students also seem to have a hard time maintaining gains made from the program after leaving. The Duff et al.

(2011) study showed that not continuing balanced literacy reading support had students regress after making gains in a skills-based intervention. Shifting from intervention settings to regular coursework as shown in the study by Kostewicz et al. (2012) also showed little ability to develop near transfer of skills based achievements. Frattura and Capper (2006) support the idea of skills not transferring as being a key component of failure and address it with the use of reading intervention programs in the ICS model. Shaw and Davidson (2009) study seemed to demonstrate that having a skills-based program in a pull-out environment can provide gains to student skill levels but those skills don't necessarily transfer to reading comprehension at the same rate or even at all for some children. Boltzmann et al. (2013) demonstrated gains in sub-skills, and like the Shaw and Davidson (2009) study, only the sub-skills taught by the highly focused program showed gains. The effect of sub-skills not transferring to comprehension appears to show that sub-skills programs are good at what they do, but that does not translate into higher reading skills if programs are not allowed to run their course. For learners at a sufficiently low level of skill, a literacy sub-skill program seems to be able to benefit them for those skills, but a top-down program appears to be best for learners that have already grasped the basics of blending and decoding. There seems to be a progression that should be followed with successive interventions as clearly demonstrated by research, and rarely if ever is any program a silver bullet.

Implications

Determining a reading improvement program of value to poor urban school districts would be a great aid to districts at a local level. Financially poor districts often

do not have the proper resources to pay for independent program evaluators. Poor districts also often end up paying more in the longer term with resources being put into ineffective programs until an effective one can be found. The findings of this evaluation study will help those districts by giving them more information they can use to make those decisions. At the conclusion of the study, if the program is proven to be effective, an executive summary report to the school board may be created. A version of this report that is readily available for community stakeholders and easy to digest for non-educators may be produced into a PowerPoint format via uploading to the school website for ease of dissemination. Also, a policy paper may be written to the school board on the importance of collaboration between the school and parents and the need to pursue new and modern reforms. If the findings of the evaluation determine the program is not effective, a pamphlet may be created for parents on the importance of creating a literacy rich environment for children. This pamphlet may include some key statistics on early exposure to literacy and simple tips on how to engage with pre-readers and emergent readers within the home based on Holdaway's Natural Literacy theory.

Summary

Section 1 has introduced the local problem of low reading achievement in a poor urban school. Reading underachievement affects individuals by limiting opportunities for academic and economic advancement. Reading underachievement further limits opportunities for full participation in a democratic society. Communities are impacted by a cycle of dependency, limiting self-determination of local authority over educational resources. Section 2 will describe the methodology for evaluating a reading program

intended to improve classroom reading achievement, improve language arts test scores, and increase graduation rates at an underperforming urban high school. Section 2 will also present results of the evaluation.

Section 2: The Methodology

Introduction

This evaluation study reported and analyzed quantitative evidence of the efficacy of the READ 180 intervention program in achieving three program goals: (a) improve classroom reading achievement, (b) improve language arts tests scores, and (c) increase graduation rates. Section 2 explains the design of the study as well as the rationale for using a goals-oriented program evaluation design. Detailed in Section 2 are research design, sample selection, instrumentation, data collection, data analysis, protection of human subjects, and results.

Research Design

Program evaluations help to gauge accountability and efficacy for staff to implement interventions successfully (U.S. Department of Education, 1993). This study followed a goals-based summative program evaluation model (Spaulding, 2008) in order to assess the alignment between program goals and outcomes. This model of program evaluation was appropriate because I intended the study to determine the impact READ 180 had on students' reading comprehension. With all data gathered being archival, evaluation methods that would fit under the participant-oriented model would not be possible. Such methods would require ongoing student observation and interaction that was not possible with a defunct program (Spaulding, 2008). Other objectives-oriented approaches such as Provus's discrepancy model also rely heavily on an ongoing program. This model also requires large systems support and staffing to monitor programs, and

both resources were unavailable for this evaluation study (Fitzpatrick, Sanders, & Worthen, 2012).

Program evaluation is used often as a decision-making tool. While formative program evaluation focuses mostly on program improvement, goals-based summative evaluation can determine the worth or merit of a program as it pertains to important criteria (Fitzpatrick et al., 2012). Goals-based program evaluation can be useful due to its ability to focus on program purposes that the evaluator identified before the evaluations starts. The evaluator then proceeds to analyze how those purposes have or have not been achieved (Fitzpatrick et al., 2012). Goals-based program evaluation can be supported by the use of program benchmarks, quantitative goals the program has determined should be attained by participants (Lodico, Spaulding, & Voegtle, 2010). In the present study, goals-based evaluation was facilitated by access to archival data of student reading achievement.

This goals-based evaluation of the READ 180 program utilized summative data from both classroom-level reading tests and a state-administered language arts test. On the other hand, an experimental or quasi-experimental design was not possible because students could not be randomly assigned to study and control groups nor could matched groups be determined with the use of existing archival data. Thus, any inference of causality was limited due to the lack of comparison groups, introducing potential threats to internal validity such as maturation and history (Creswell, 2012).

Population and Sample

The target population to which this study generalized was underachieving readers in a predominantly Hispanic, low-income urban high school. The intended sample for the study was all participants in the READ 180 program at the research study school, in New Jersey, from 2005-2012. Unfortunately, for several of these years a matched set of test data for the start and completion of the program was not available, therefore making only two cohorts (2007-2008 and 2010-2011) usable for this study. Further, data for the 2007-2008 cohort were incomplete, so that complete analyses were conducted only for the 2010-2011 cohort. For 2010-2011, the data were first analyzed for the cohort as a whole and then analyzed by English language proficiency (English proficient versus non-English proficient).

Selection of Participants

The study participants comprised all students enrolled in the READ 180 program during the 2007-2008 and 2010-2011 school years at the research study school. Typically, there were between 90 and 130 students enrolled in the program each year. Only students with complete pretest and posttest data were included in the study. Therefore, students who lacked complete data sets due to transfers, absenteeism, or other reasons were excluded from the analysis.

Participant Protections

Approval from the Walden University's Institutional Review Board (IRB) was necessary to avoid undue harm to participants resulting either intentionally or unintentionally from the research.

Great care was taken as the participants in the study were (or were at the time of testing) minors. Despite the fact that this was an archival data set, privacy had to be respected. One issue I encountered in this study was the need for anonymity of participants. Students were de-identified and a numerical code was substituted. De-identification occurred at the point of data extraction, which means consent was required from parents or students. Files were in a locked drawer while in my possession and no one else was given access. At the conclusion of the study, the data will be held on to until such time as it needs to be destroyed as per IRB guidelines. All pertinent school administrators were notified and permission to conduct the research was granted from the local board of education and district superintendent as well.

Data Collection Instruments

Scholastic Reading Inventory

To measure reading achievement at the classroom level, I used the Scholastic Reading Inventory (SRI), a computer-adaptive assessment designed to measure comprehension in terms of a Lexile Framework (Metametrics, 2004). The Lexile Framework assigns to text a level of text difficulty based on semantic difficulty and syntactic difficulty (Metametrics, 2004). Semantic difficulty is determined by a word's frequency in standard text rather than length or number of syllables. Syntactic complexity is determined by the length of sentences. Using these two factors, texts are scanned with software to determine their lexile range (Metametrics, 2004). A student is assigned a lexile based on having successfully read text at that level with 75% comprehension (Metametrics, 2004). Thus, student reading ability is determined by the highest lexile

achieved and progress can be evaluated by movement from lower to higher lexiles. Lexile scores range from 100 to 1300, which roughly approximate first grade through 12th grade reading ability.

Construct validity of the SRI can be linked to its widely adopted use among 15 standardized tests, studies were then conducted using lexiles, such as the Stanford Achievement Test, Metropolitan Achievement Test, and the Terra Nova (Scholastic, 2007). Validity is also backed by a correlation between 11 basal series and their lexile equivalence (see Table 1). The average correlation was .995 (Scholastic, 2007).

Table 1

SRI Test Correlation Validity

Basal Series	Number of Units	r_{OT}	R_{OT}	R'_{OT}
Ginn Rainbow Series (1985)	53	.93	.98	1.00
HBJ Eagle Series (1983)	70	.93	.98	1.00
Scott Foresman Focus Series (1986)	92	.84	.99	1.00
Riverside Reading Series (1986)	67	.87	.97	1.00
Houghton-Mifflin Reading Series (1983)	33	.88	.96	.99
Economy Reading Series (1986)	67	.86	.96	.99
Scott Foresman American Tradition (1987)	88	.85	.97	.99
HBJ Odyssey Series (1986)	38	.79	.97	.99
Holt Basic Reading Series (1986)	54	.87	.96	.98
Houghton-Mifflin Reading Series (1986)	46	.81	.95	.98
Open Court Headway Program (1985)	52	.54	.94	.97
Total/Means	660	.839	.965	.995

A study by Stenner, Smith, Horabin, and Smith (1987) examined 1,780 test items related to reading comprehension that were found on nine national normed tests to identify the accuracy of lexile ratings. It was determined that non-continuous prose and poetry were the only items the lexile framework was inaccurate in rating. Since the SRI uses continuous prose and has no poetry, the construct validity of this system is validated. The data bank of prose was built from 1986 through 2003 and contains works from textbooks, periodicals, and literature (Scholastic, 2007). According to Scholastic, Inc. the

passages were included based on the following criteria: it must developed one main idea, the passage can be understood independently despite what came before or after it, and no prior knowledge not contained in the passage is required to understand it.

Questions for the SRI are similar to the fill-in-the-blank format and are referred to as *embedded completion format* by the publisher (see Figure 4). The questions are written to gauge logical connections of ideas and inferences of each passage, which is 30-150 words in length (Scholastic, 2007). The questions themselves are written by former classroom teachers familiar with a variety of student reading levels. To be included on the exam, all items went through a two-stage review process with specialists in the areas of editing, curriculum, and testing (Scholastic, 2007).

At Lunch, Gilberto was hoping he'd see Maddie. He wanted to tell her his good news and also see how she was doing. But the lunch period passed, and he saw no sign of her. Gilberto decided that after school he would job over to her house and check on her.

He was _____for her.

- Cooking
- Studying
- Reading
- Looking

Luis examined the bugle and noticed that the brass was dirty and green from age. He decided to give it to a good polishing, and when he finished it was gleaming bright. He practiced standing erect and putting the bugle to his lips. He practiced blowing it to bring out its mellow tones.

He _____the bugle.

- Cleaned
- Sold
- Dropped
- Covered

Figure 1. SRI sample questions.

The standard error of measurement of the SRI varies depending on how long it has been since a student last took the test. Through the use of algorithms the computer adapts questions based on prior answers in an effort to predict future question difficulty. Therefore, the more times a student takes the test the less error in questions assigned (Scholastic, 2007). Through this manner measurement of error for a student can be reduced by as much as 418 percent (Scholastic, 2007). To further support SRI consistency, Scholastic, Inc. administered the SRI to all students from 3rd through 10th grade in the 2004-2005 school year in a large urban school district. Students were tested

over a four month period. The reader consistency correlation for all students was .894 (see Table 2).

Table 2

Reader Consistency Correlation

Grade	N	Reader Consistency Correlation
3	1,241	.829
4	7,236	.832
5	8,253	.854
6	6,339	.848
7	3,783	.860
8	3,581	.877
9	2,694	.853
10	632	.901
Total	33,759	.894

The SRI is unlike traditional reading level measurements, which do not always use a common absolute scale. The use of a common absolute scale enables the Lexile Framework to match text and student reading ability in the same manner (Scholastic, 2007). This coupled with the fact that the Lexile Framework is an equal-interval scale gives teachers a growth level that is easier to use for statistical analysis (Metametrics, 2004). Students participated in the 90-minute balanced literacy READ 180 program and were evaluated with pre- and posttests based on the Lexile Framework test of reading level.

Scores from the SRI provide an approximation of reading comprehension level. The grade level equivalence is reported in scale form based on the publisher's guidelines. Grade-level equivalencies are given to overlap, with decidedly large intervals due to the difficulty in assigning a grade level to a specific reading comprehension range. This means that a student scoring a 500 may fit in two different grade levels simultaneously, as shown in the chart above. To approximate grade-level equivalence, SRI uses the middle 50% of all text-based materials as the median for the grade level (Scholastic, 2007). Thus, 50% of the students are reading above or below the set grade level range. The program also correlates this 50%, or interquartile range, with ongoing results of published studies on text complexity. This means the student scoring a 500 is technically grade 2 and 3 but is actually in the middle 50% of Grade 2 while being at the bottom 50% of grade 3 (Scholastic, 2007).

HSPA

To measure language arts achievement on a state-administered reading test, the study used the HSPA (HSPA). The HSPA is a standardized test given to all eleventh graders in New Jersey to identify whether or not students have a minimal set of skills required to graduate (State of New Jersey Department of Education, 2006). The test comprises of a language arts and mathematics section. The test results are used by the New Jersey Department of Education to satisfy the requirements of the NCLB Act. No student may graduate from a New Jersey high school without demonstrating proficiency on this exam (State of New Jersey Department of Education, 2006).

The language arts literacy section consists of two passages measuring Working with Text and Analyzing Text skills. To measure skills two different reading passages are provided, narrative and persuasive, with each being followed by 10 multiple-choice questions and two open-ended questions (short answer). To measure student ability to construct meaning in a written response for a sustained duration there are two extended response, or essay questions, based on two different tasks and types of writing: expository and persuasive. Reading passages and Open-Ended questions are assigned to either the Working with Text or Analyzing Text skill categories. Neither writing task (expository or persuasive) is assigned to a category. Points can be earned in this section with scores increasing in half-point increments along the following scale: 1 point for each multiple choice, 4 points for each open-ended, 6 points for Expository writing, and 12 points for the Persuasive writing (C. Blue, personal communication, May 8, 2014).

Scoring of the results is done by the publisher's machine for multiple-choice questions, while open-ended and essay questions are scored by hand using a holistic scoring rubric. The holistic scoring rubric was developed by the New Jersey Department of Education and is known as Registered Holistic Scoring. Using a logistic regression method, two cut off scores were identified by the judges to create the three categories of proficiency at: less than proficient = 100-199, proficient = 200-249, advanced proficient = 250-300 (State of New Jersey Department of Education, 2007).

Reliability of scores is based on Cronbach's coefficient alpha measure of internal consistency. Content validity of the sections is based on the Pearson computed item means, response frequencies, and biserial correlations as analyzed by the content review

committee members. Pearson product-moment correlations between student scores are also used. Criterion validity is also possible with relationships of scores to the National Assessment of Education Progress scores (State of New Jersey Department of Education, 2007).

In addition, to measure initial status in reading before program participation, scores from the New Jersey Assessment of Skills and Knowledge (NJ ASK) test were used. The NJ ASK was developed to replace the prior Grade Eight Proficiency (GEPA) Assessment and was first used in 2008. The GEPA was also contracted by Measurement, Inc. and was developed concurrently with the HSPA. Due to this the GEPA shared psychometric properties of the HSPA (C. Blue, personal communication, October 16, 2013). The Language Arts Literacy section of the NJASK test contains short reading passages, two writing prompts, and multiple-choice questions. Questions are made to measure knowledge of the New Jersey Core Curriculum Content Standards (State of New Jersey Department of Education Assessment, 2009).

Reading questions for the NJ ASK attempt to determine interpreting, analyzing, and critiquing text; while writing questions focus on speculation and persuasion. Reading passages are either narrative or informational. Students are scored similarly to the state HSPA testing with either being partially proficient (100-199), proficient (200-249), or advanced proficient (250-300). (State of New Jersey Department of Education Assessment, 2009).

Test items are subjected to classical item analysis before item calibration, scaling, and equating is performed. The process consists of using “P-Value”, “r-biserial”, and

distractor analysis for questions. Differential Item Functioning is used with the Mantel-Haenszel procedure to ensure multiple choice question fairness (State of New Jersey Department of Education Assessment, 2009).

Graduation Rate. To measure rate of high school graduation at the project site, the researcher verified student status via the school information system, NJ SMART, to identify for each program and non-program participant in the sample whether or not high school graduation was achieved.

Data Collection Methods

SRI tests were administered each year by READ 180 program teachers or the reading specialist in a regular classroom setting in September and again in February. The exams were taken by students on a computer station through instructional software. Students were placed at a computer station and given up to 45 minutes to complete the assessment. Students were presented a series of short reading passages followed by multiple-choice questions. The HSPA was administered to students every March over the course of three days. Students answered multiple-choice, open-ended, and essay questions with pencils and testing booklets for two-and-a-half to three hours each day. The math section comprised one day while the language arts literacy section comprised two days (State of New Jersey Department of Education, 2006). NJ ASK tests had been previously administered to participants while in middle school over the course of three days. Students answered multiple choice, open-ended, and essay questions with pencils and testing booklets for two-and-a-half to three hours each day. The math section comprised

one day while the language arts literacy section comprised two days (State of New Jersey Department of Education Assessment, 2009).

Data was collected in several ways. The school reading specialist delivered SRI results to the researcher in printed-file format with years of student participation and class rosters from READ 180. The district's Information Technology Director delivered to the researcher in electronic format NJ ASK and HSPA scores. The researcher identified program and non-program participants by collating data from all three reports. The researcher identified graduation status by verifying student records via the school's electronic information system, NJ SMART. In this manner, cohorts were extracted for several years starting with the 2006-07 cohort through to the 2010-11 cohort. Students in this range were the most recent students to have taken the HSPA and graduated. Student data could only be taken after each cohort had completed the five-year cycle from 8-12th grades. Selecting this particular range of years was beneficial due to the potential lack of hard copy records for years before electronic reporting. However, due to incomplete records, only the 2007-2008 and 2010-2011 cohorts were able to be used for analysis.

Data Analysis Methods

In order to determine if the program was successful it was necessary to first examine internal effects before looking to external effects. Examining the READ 180 program, and the variable of reading achievement as defined by READ 180 itself, the study benefited from less potential noise from variables. The self-measurement created by the publisher (SRI) was originally intended to measure reading achievement as increased by the program. Utilizing the SRI added greater certainty to this specific

analysis. After evaluating internal measurements, it is then best to move to external measurements and their effects such as language arts scores and graduation rates. Each successive level of analysis is one more removed from the proximity to the statistics most directly attributed to the program itself. Analyzing in this manner provided a logical progression of results and helped with building a progressive interpretation and summation.

Since goals-based evaluation is designed to assess the alignment between program goals and program outcomes, this study analyzed data by comparing student achievement to three program goals aligned with several hypotheses. Data were displayed for analysis via the summary tables of each test that was performed in the statistical analysis program SPSS.

To evaluate the goal of improving reading achievement at the classroom level, scores from the SRI were entered into a paired difference *t*-test (Creswell, 2012), in which the READ 180 program was considered the independent variable, and student achievement on the SRI was considered the dependent variable. Statistical significance will suggest the null hypothesis should be rejected. The null hypothesis was tested at the .05 level of probability. $H_1$1: There will be an effect in student reading achievement, as measured by the SRI reading assessment, throughout the years of READ 180 implementation. $H_0$1: There will be no effect in student reading achievement, as measured by the SRI reading assessment, throughout the years of READ 180 implementation. The program participant's males and females also had the scaled data of SRI posttest scores analyzed by an independent samples *t*-test to determine if there was

an affect upon gender and proficiency. The hypotheses to be tested was $H_2 2$: There will be an effect upon gender and proficiency as measured by the SRI reading assessment. $H_0 2$: There will be no effect upon gender and proficiency as measured by the SRI reading assessment.

To evaluate the goal of improving performance on the language arts section of the HSPA, a multiple regression analysis was applied in which program participation was the independent variable, eighth-grade reading score on the NJASK was the control variable, and HSPA score was the dependent variable. Following a hierarchical procedure (Triola, 2012), the control variable was forced into regression in order to factor out the contribution of eighth-grade reading scores on HSPA scores. Such analysis yielded the incremental variance in HSPA scores contributed by program participation after the effects of eighth-grade reading score were already accounted for. A statistically significant incremental variance at the .05 level of probability was used as the standard for program success. Therefore, $H_3 3$: There will be significant linear correlation between program participation and HSPA language arts, after the contribution of initial status in reading is factored out. $H_0 3$: There will be no significant linear correlation between program participation and HSPA language arts, after the contribution of initial status in reading is factored out.

To evaluate the goal of increasing the rate of graduation, the ratio data of graduation rates for each year of program implementation were analyzed alongside the program participants and the general population graduation rates with a Chi-Square test to determine independence (Triola, 2012). The hypotheses tested was $H_4 4$: There will be

a dependency upon graduation rates for program participants, throughout the years of READ 180 implementation. $H_0$4: There will be no dependency upon graduation rates for program participants, throughout the years of READ 180 implementation.

Assumptions

I have made several assumptions regarding the implementation of READ 180 during the years being investigated. First, I have assumed that all program teachers were given the same or similar professional development administered by the program provider. Second, I have assumed that teachers in all classrooms were implementing the program with fidelity. Finally, I have assumed that test data and graduation data for program students were faithfully recorded and were accurately retrieved by district personnel before being submitted for data analysis.

Limitations

Potential limitations for the study are directly related to the aforementioned assumptions. To the extent that not all program teachers received the same level of professional development, or not all program classrooms implemented the program with fidelity, results may reflect variations in implementation rather than program outcomes. To the extent that test scores or graduation rates are not accurate, results also may not reflect program outcomes.

In addition, goals-oriented evaluation is subject to the limitation of predetermining data most likely tied to the programs goals. Such orientation raises the potential for evaluators to not see the true context or best connection between program inputs and outputs (Fitzpatrick et al., 2012). This potential weakness can be further

complicated by the fact that the evaluator is told in advance what the program goals are, and is not in a position to determine if the goals properly fit a program (Fitzpatrick et al., 2012). Thus, such a narrowing of research parameters can limit findings.

A related methodological limitation of this research study is that only archival quantitative data sources are available to the researcher. No qualitative data sources, such as classroom observations, teacher interviews, or student interviews, are available since the years being studied are past and the program has been discontinued. Often program evaluations rely upon qualitative data to provide context or explanation for results obtained from quantitative data.

Because the evaluation does not use an experimental or quasi-experimental design, statistical analysis is limited in terms of inferring potential causality. At best, the analyses provided evidence of relationships among critical variables such as program participation and program outcomes. Although regression analysis is common to the field of education, results of such analyses rely heavily on how well independent variables can account for the original variance (Creswell, 2012). If there are variables not being used that may benefit the analysis, the researcher may have a hard time finding any predictive value of worth.

Sample size suffered from a historically high attrition rate common to the evaluation study site. Cohorts used for sampling were not always intact and potentially useful data was missing for students who were not present for all phases of instrumentation. In addition, data sets were available for only two of the years of the

program and a complete data set was available for only one year. This limit to sample size may have potential impact on generalizability of results.

Finally, having only one evaluation study site also limits generalizability of findings. Performing a similar evaluation at multiple peer sites would add validity to the findings. Therefore any findings of this study may not easily be used for projection of potential outcomes at other schools.

Scope

The scope of this evaluation study was limited to the high school that served as the research site, an urban underperforming high school with approximately 2,400 students from ninth through 12th grade. Although the READ 180 program is nationally available, its specific implementation was particular to the project site and indeed changed from year to year. The purpose of this study is to provide valid quantitative evidence relating to the effectiveness of the READ 180 program at the project site and therefore no attempt was made to draw conclusions regarding the efficacy of READ 180 beyond the project site.

Summary of Methodology

To determine the effectiveness of the READ 180 intervention program in achieving three specific goals (reading achievement at the classroom level, performance on the state language arts test, and graduation rate), a goals-based evaluation design was used. Goals-based evaluation can accommodate post-hoc analysis of a defunct program, whereas other forms of evaluation cannot. Both program participants and non-participants were included in the analysis in order to contrast outcomes on all three goals.

Data analysis was conducted for two years of program implementation but complete data was available for only one year. Instrumentation reflected each of the three program goals, with SRI scores showing reading achievement at the classroom level, HSPA scores showing language arts achievement on the state test, and graduation status showing graduation rate. In addition, NJ ASK scores allowed the researcher to control for effects of initial status in reading as a potential confounding variable. Paired difference *t*-tests, which analyze repeated measures of a sample over time, examined SRI impact. Multiple regression measured the effect of program participation on HSPA scores after controlling for initial status in reading. Chi-square provided information regarding potential dependency of program participants and graduation rates.

Results

Based on availability of data sets, two cohorts (2007-2008 and 2010-2011) were used in the data analysis. The 2007 cohort comprised 714 students, of whom 619 made the graduate/dropout list as reported to the State of New Jersey Department of Education through NJSMART. The remainder was classified as either excluded or transferred. Within the cohort students were designated as Hispanic (559), Black (37), White (17), Asian (3), and Native Hawaiian or Pacific Islander (1). There were also 140 students designated as English Language Learners. Of these, 56 were placed into the READ 180 program. This cohort had no complete GEPA scores for READ 180 program and non-program participants that could be used for regression analysis; therefore, the regression analysis was not performed.

The 2010 cohort comprised 358 students who made the graduate/dropout list as reported to the State of New Jersey Department of Education, with none reported as excluded or transferred. The change in cohort size was due in part to new procedures in how the state computed each category for district reporting. Within the cohort, students were designated as Hispanic (329), Black (23), White (2), and Asian (1). There were also 140 students designated as English language learner. Of these, 30 were placed in the READ 180 program. This cohort had complete data available to run all tests required by this evaluation study.

2007 Cohort Analysis

Hypothesis 1: Reading achievement. In order to test the hypothesis that there will be an effect on student reading achievement throughout READ 180 implementation, a paired-difference *t*-test was used to determine if there was a statistically significant mean difference between the pre- and posttest SRI scores of READ 180 participants during 2007-2008. Score differences for the SRI pre- and posttest were normally distributed as assessed by the Shapiro-Wilk's test (p = .811). READ 180 participants scored higher on the posttest (M = 703.00, SD = 259.357) than on the pretest (M = 687.96, SD = 270.066) (see Table 3). The READ 180 program elicited a mean increase of 15.03 on the SRI test from September through May of the school year. Furthermore, an increase of 15 (M) on a test that ranges on such a wide scale cannot be considered a practically important difference (see Table 4). There was no significant difference in pretest and posttest scores, $t(55) = .591$, $p = .557$ with a confidence interval of 95% (CI_{95}: 66.0-35.9); therefore, the null hypothesis cannot be rejected (see Table 5).

Table 3

Paired Difference t Test Sample Statistics for 2007 Cohort SRI Scores

		Mean	N	Std. Deviation	Std. Error Mean
Pair 1	Posttest	703.0000	56	259.35788	34.65815
	Pretest	687.9643	56	270.06659	36.08917

Table 4

Paired Difference t Test for 2007 Cohort SRI Scores Mean

		Mean	Std. Deviation	Std. Error Mean	95% Confidence Level Lower
Pair 1	Posttest-Pretest	15.03571	190.44417	25.44917	-35.96557

Table 5

Paired Difference t Test for 2007 Cohort SRI Scores Significance

		t	df	Sig. (2-tailed)	95% Confidence Level Upper
Pair 1	Posttest-Pretest	.591	55	.557	66.03700

Next, to test where there was an effect of gender upon reading achievement as measured by the SRI reading assessment, SRI scores of male (N = 36) versus female (N = 20) participants were entered into an independent samples t-test to find out if females performed similarly to males on the SRI posttest. Assumption of homogeneity of variances was violated, as assessed by Levene's test for equality of variances (p = .04). Females scored 61, 95% CI [67.10 to 189.84] higher than males (see Table 6). However, there was not a statistically significant difference in mean scores between males and females, $t(52.90) = .958$, $p = .342$; therefore, the null hypothesis cannot be rejected (see Table 6).

Due to the violation of Levene's test for equality of variances it became necessary to use the *Equal variances not assumed* section of each table for the posttest (see Table 6). Variances were not equal due to unbalanced grouping (see Table 7). The difference in groups lowered the equality of means upper confidence interval (see Table 8). Females did show a greater gain than males, but only in a non-significant amount of 61 (see Figure 2).

Table 6

Equality of Means in Gender for 2007 Cohort Posttest

		Sig. (2. Tailed)	Mean Difference	Std. Error Difference	95% Confidence Interval Lower
Post test	Equal Variances Assumed	.401	-61.36667	72.51889	-206.75826
	Equal Variances not assumed	.342	-61.36667	64.05130	-189.84254

Table 7

Equality of Means p-Value for 2007 Cohort Posttest

		F	Sig.	*t*	df
Post test	Equal Variances Assumed	4.438	.040	-.846	54
	Equal Variances not assumed			-.958	52.908

Table 8

Equality of Means Confidence Interval for 2007 Cohort Posttest

		95% Confidence Interval Upper
Post test	Equal Variances Assumed	84.02492
	Equal Variances not assumed	67.10921

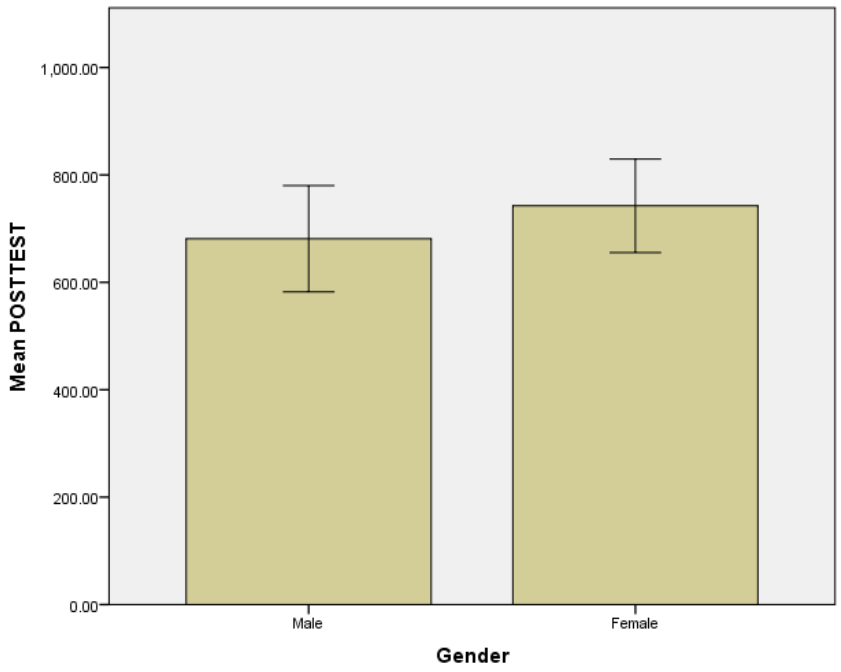

Figure 2. Equality of means gender for 2007 cohort posttest.

Hypothesis 2: Scores on state language arts test. Due to a lack of data availability it was not possible to perform a regression analysis.

Hypothesis 3: Graduation rates. To test the hypothesis that there will be a dependency upon graduation rates for program participation, graduation rates of participants and non-participants were entered into a chi-square test for independence. To test whether proportions were different for each group, a χ^2 test of independence was used with $\alpha = .05$ as criterion for significance. There was not a statistically significant association between program participation and graduation for the cohort year of 2007, $\chi^2(1) = 3.532$, $p = .060$. Thus, there was little to no association between program participation and graduation (see Table 9). The expected counts and counts were very similar (see Table 10). The value of Phi was not significant (see Table 11). Upon graphing the analysis, it may be determined that the small size of the READ 180 participant pool had an impact on the accuracy of the analysis. The clustered bar chart illustrates from the existing data the program had little or no effect on student graduation rate (see Figure 3).

Table 9

Chi-Square Tests for Dependence Graduation for 2007 Cohort

	Value	df	Asymp Sig. (2-sided)	Exact Sig. (2-sided)	Exact Sig. (1-sided)
Pearson Chi-Square	3.532[a]	1	.060		
Continuity Correction[b]	2.459	1	.117		
Likelihood Ratio	2.843	1	.092		
Fisher's Exact Test				.108	.068
Linear-by-Linear Association	3.527	1	.060		
N of Valid Cases	619				

Table 10

Chi-Square Test for Dependence Graduation Cross-tabulation for 2007 Cohort

	Value		Graduated	Dropout	Total
READ 180	No	Count	539.0	39.0	578.0
		Expected Count	536.0	42.0	578.0
		% within READ 180	93.3	6.7	100.0
		% within Graduated	93.9	86.7	93.4
		% Total	87.1	6.3	93.4
		Count	35.0	6.0	41.0
		Expected Count	38.0	3.0	41.0
		% within READ 180	85.4	14.6	100.0
		% within Graduated	6.1	13.3	6.6
		% of Total	5.7	1.0	6.6
	Yes	Count	574.0	45.0	619.0
		Expected Count	574.0	45.0	619.0
		% within READ 180	92.7	7.3	100.0
		% within Graduated	100.0	100	100.0
		% of Total	92.7	7.3	100.0

Table 11

Chi-Square Test for Dependence Symmetric Measures for 2007 Cohort

		Value	Approx. Sig.
Nominal by Nominal	Phi	.076	.060
	Cramer's V	.076	.060
N of Valid Cases		619	

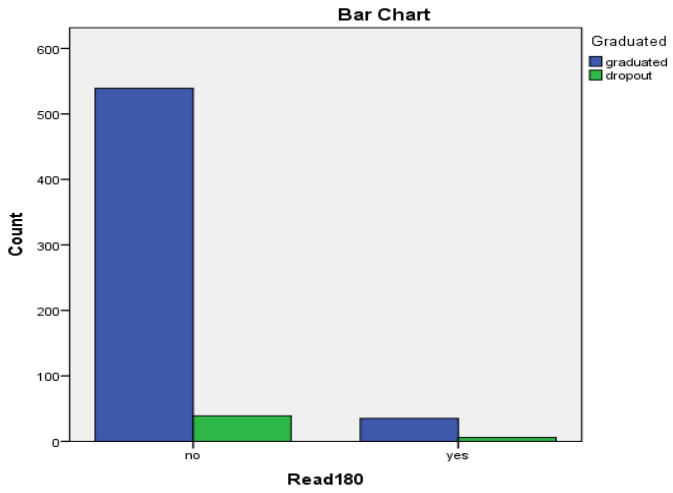

Figure 3. Chi-square test for dependence clustered bar chart for 2007 cohort.

2010 Cohort Analysis

Hypothesis 1: Reading achievement. In order to test the hypothesis that there will be an effect on student reading achievement throughout READ 180 implementation, a paired-difference *t*-test was used to determine if there was a statistically significant mean difference between the pre- and posttest SRI scores of READ 180 participants during 2010-2011. Score differences for the SRI pre- and posttest were normally distributed as assessed by the Shapiro-Wilk's test (p = .996). READ 180 participants

scored higher on the posttest (*M* = 895.200, *SD* = 140.197) than on the pretest (*M* = 856.950, *SD* = 134.185) (see Table 12). The READ 180 program elicited a mean increase of 38.25 (*SE* = 14.602) on the SRI test from September through May of the school year (see Table 13). The posttest mean was greater than the pretest (see Figure 4). There was a significant increase in scores from pretest to posttest, t (19) = .2.619, $p < .001$ with a confidence interval of 95% (CI_{95}: 68.81-7.68); therefore, the null hypothesis can be rejected (see Table 14). Cohen's effect size value (d = .28) suggested a small effect size as shown in Table 14.

Table 12

Paired Difference t Test Sample Statistics for 2010 Cohort SRI Scores

		Mean	N	Std. Deviation	Std. Error Mean
Pair 1	Posttest	895.2000	20	140.1978	31.3491
	Pretest	856.9500	20	134.1854	30.0047

Table 13

Paired Difference t Test for 2010 Cohort SRI Scores Mean

		Mean	Std. Deviation	Std. Error Mean	95% Confidence Level Lower	95% Confidence Level Upper
Pair 1	Posttest-Pretest	38.25000	65.30566	14.60279	7.68601	68.81399

Table 14

Paired Difference t Test for 2010 Cohort SRI Scores Significance

		t	d	df	Sig. (2-tailed)
Pair 1	Posttest-Pretest	2.619	.28	19	.017

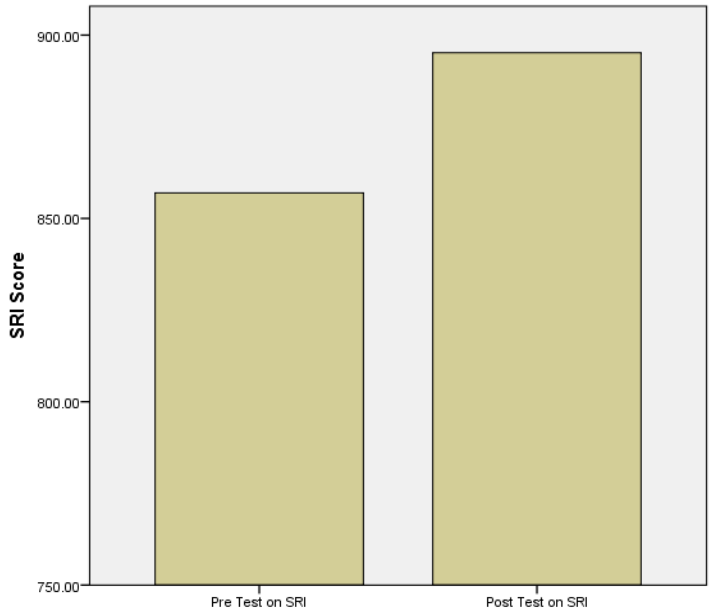

Figure 4. Pretest and posttest mean for 2010 cohort SRI scores.

To test whether there was an effect of gender upon reading achievement as measured by the SRI reading assessment, SRI scores of male (N = 12) versus female (N = 18) participants were entered into an independent samples *t*-test to find out if females performed similarly to males on the SRI posttest. Score differences between genders on the SRI pre- and posttest were normally distributed as assessed by the Shapiro-Wilk's test (p = .996). There was homogeneity of variances for posttest scores for males and females, as assessed by Levene's test for equality of variances (p = .650) (see Table 16). With an

equality of variances assumed the upper confidence interval was 109.348 (see Table 18). Females (*M* = 907.90, *SD* = 131.824) scored higher than males (*M* = 882.50, *SD* = 154.30) (see Figure 5). Yet, females only scored slightly higher than males (see Table 15). There was not a statistically significant difference in mean scores between females and males, $t(18) = .396$, $p = .697$; therefore, the null hypothesis cannot be rejected (see Table 17).

Table 15

Group Statistics in Gender for 2010 Cohort Posttest

	Gender	Mean	N	Std. Deviation	Std. Error Mean
Post	Male	882.5000	10	154.30939	48.79691
	Female	907.9000	10	131.62442	41.62330

Table 16

Equality of Variances p-Value for 2007 Cohort Posttest

		F	Sig.	*t*	df
Post test	Equal Variances Assumed	.213	.650	-.396	18.000
	Equal Variances not assumed			-.396	17.563

Table 17

Equality of Means Confidence Interval for 2010 Cohort Posttest

		Sig. (2. Tailed)	Mean Difference	Std. Error Difference	95% Confidence Interval Lower
Post test	Equal Variances Assumed	.697	-25.40000	64.13765	-160.14820
	Equal Variances not assumed	.697	-25.40000	64.13765	-160.38874

Table 18

Equality of Means Confidence Interval for 2010 Cohort Posttest Continued

		95% Confidence Interval Upper
Post test	Equal Variances Assumed	109.34820
	Equal Variances not assumed	109.58874

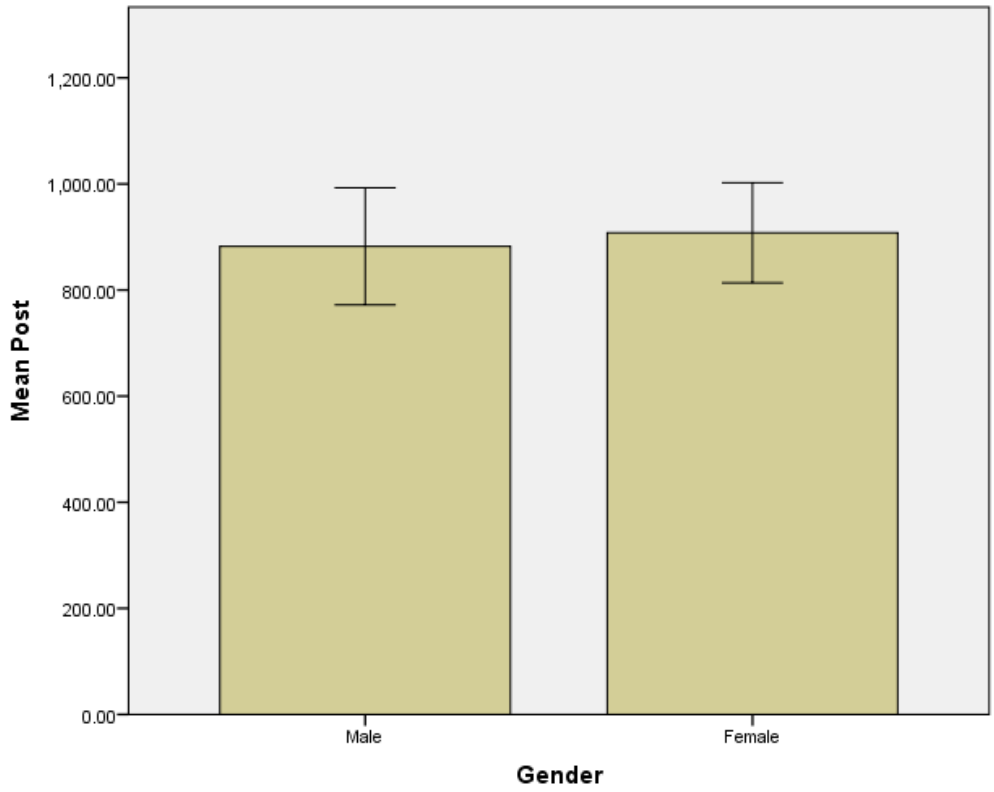

Figure 5. Equality of means gender for 2010 cohort posttest.

Hypothesis 2: Scores on state language arts test. Next, to test the hypothesis that program participation will impact HSPA performance after controlling for prior reading ability, a hierarchal multiple regression test was run. Since only HSPA scores and READ 180 scores were available for the 2007-2008 cohort, regression analysis was performed on the 2010-2011 cohort. Program participation was placed as a nominal variable in SPSS where "0" stood for the value label of students who did not take part in the READ 180 program and "1" stood for the value label of students who did take part in the READ 180 program. NJASK 8 and HSPA scores were both scaled variables. NJASK 8 scores were entered first into the hierarchical multiple regression test as a control variable. Next, program participation was entered to determine the amount of variance in

HSPA scores contributed by the READ 180 program after NJASK 8 variance had been accounted for. A statistically significant incremental variance at the .05 level of probability was used as the standard for program success. There was independence of residuals, as assessed by a Durbin-Watson statistic of 1.541. The assumptions of linearity, independence of errors, homoscedasticity, unusual points, and normality of residuals were met after removing three outliers and seven cases, respectively, which held high leverage values. HSPA mean scores were higher than NJASK 8. The NJASK 8 had a strong correlation to HSPA scores, while READ 180 participation had a negative correlation. The model of NJASK 8 scores to predict HSPA performance alone was statistically significant with $R^2 = .568$ $F(1, 237) = 311.681$, $p < .0005$, adj. $R^2 = .566$ (see Tables 19 and 20). The addition of READ 180 program participation to the prediction of HSPA scores added a statistically insignificant increase in R^2 of .003, $F(1, 236) = 1.677$, $p = .197$, adj. $R^2 = .567$ (see Table 19). Since only one variable, NJASK 8, proved to be statistically significant in regards to the prediction, $p < .0005$; READ 180 participation did not factor in any statistically significant way, $p = .197$; therefore, the null hypothesis cannot be ruled out (see Table 20).

Table 19

Multiple Regression Analysis Model Summary for Cohort 2010

Model	Predictors (constant)	R	R Square	Adjusted R Square	Std. Error of the Estimate	Change Statistics R Square Change	F Change	df1
1	NJASK8	.754[a]	.568	.566	14.256	.568	311.681	1
2	NJASK8, READ 180	.756[b]	.571	.567	14.236	.003	1.677	1

Note. Dependent Variable: HSPA

Table 20

Multiple Regression Analysis Change Statistics for Cohort 2010

Model	Predictors (constant)	df2	Sig. F Change	Durbin-Watson
1	NJASK8	237[a]	.000	
2	NJASK8, READ 180	236[b]	.197	1.541

Note. Dependent Variable: HSPA

Additional analyses were conducted to assess the differential effects of English language status on program success. First, a series of descriptive statistics were run. Students in the 2010 cohort who were not English language learner and participated in the READ 180 program (N = 16) had mean scores of 188.00 and 211.08 respectively on the NJASK 8 and HSPA, while students who were not English language learners and did not in the READ 180 program (N = 158) had mean scores of 213.22 and 227.79 respectively. English language learner students in READ 180 therefore exhibited a lower

initial status on the NJASK 8 (see Table 21). Although both groups showed growth in reading as measured by the HSPA, the READ 180 students showed more growth (see Table 22). This observation suggested a series of inferential tests to test if the differences observed in the descriptive statistics were statistically significant. The researcher postulated that by removing nominal data (*program participation)* and instead focusing on scaled data (*NJASK, HSPA)* a more accurate result would be determined by the analysis.

Table 21

Comparison of Means Non-ELL Students

READ 180		NJASK8	HSPA
	Mean	188.00	211.08
Yes	N	16.00	13.00
	Std. Deviation	15.04	17.81
	Mean	213.22	227.79
No	N	158.00	163.00
	Std. Deviation	20.66	20.83
	Mean	210.90	226.55
Total	N	174.00	176.00
	Std. Deviation	21.46	21.04

Table 22

Comparison of Means ELL Students

READ 180		NJASK8	HSPA
	Mean	195.00	219.58
Yes	N	14.00	12.00
	Std. Deviation	11.28	16.73
	Mean	204.68	198.42
No	N	72.00	132.00
	Std. Deviation	17.88	28.99
		203.10	200.19
Total	N	86.00	144.00
	Std. Deviation	17.30	28.73

Therefore, to test the hypothesis that there will be an effect of READ 180 participation upon reading achievement among English language learner students, NJASK 8 scores of the 2010 cohort were first entered into an independent samples *t*-test to find out if program English language learners ($N = 14$) performed similarly to non-program English language learner ($N = 72$). A statistically significant result at the .05 level of probability was used as the standard for success (see Tables 24 and 25). Scores for the English language learner based NJASK 8 and HSPA assessments of students within the READ 180 program and not within the program were normally distributed as assessed by the Shapiro-Wilk's test ($p > .05$). There was homogeneity of variances for NJASK 8 scores of READ 180 English language learner and non-READ 180 English

language learner students, as assessed by Levene's test for equality of variances (p = .074). READ 180 English language learner for NJASK (M = 193.50, SD = 11.261) participants scored lower than non-READ 180 English language learner (M = 204.68, SD = 17.883) students, indicating that program participants had a lower initial status in reading than non-program participants. There was a statistically significant difference in mean scores between program participants and non-program participants, $t(82) = -2.092$, $p = .040$; therefore, the null hypothesis has sufficient cause to be rejected. Homogeneity of variances was violated due to unbalanced group sizes for HSPA scores of READ 180 English language learner and non-READ 180 English language learner students, as assessed by Levene's test for equality of variances ($p = .001$) (see Table 23). However, despite their lower initial status, English language learner program participants (M = 225.10, SD = 9.792) scored higher than English language learner non-program participants (M = 198.42, SD = 28.991) on the HSPA. Indeed, READ 180 English language learner students scored higher scores on the HSPA than on the NJASK 8 (see Figures 6 and 7). There was a statistically significant difference in mean scores between English language learner program participants and English language learner non-program participants, $t(24.190) = 6.678$, $p < .001$; therefore, the null hypothesis has sufficient cause to be rejected (see Table 23).

Table 23

Equality of Means in ELL p-Value for 2010 Cohort t Test

		F	Sig.	*t*	df
NJASK8	Equal Variances Assumed	3.284	.074	-2.092	82.000
	Equal Variances not assumed			-2.886	21.598
HSPA	Equal Variances Assumed	10.627	.001	2.889	140.00
	Equal Variances not assumed			6.678	24.190

Table 24

Equality of Means Confidence Interval for ELL 2010 Cohort t Test

		Sig. (2. Tailed)	Mean Difference	Std. Error Difference	95% Confidence Interval Lower
NJASK8	Equal Variances Assumed	.040	-11.181	5.346	-21.815
	Equal Variances not assumed	.009	-11.181	3.874	-19.224
HSPA	Equal Variances Assumed	.004	26.676	9.234	8.420
	Equal Variances not assumed	.000	26.676	3.994	18.435

Table 25

Equality of Means Confidence Interval for ELL 2010 Cohort t Test Continued

		95% Confidence Interval Upper
NJASK8	Equal Variances Assumed	-0.547
	Equal Variances not assumed	-3.137
HSPA	Equal Variances Assumed	44.932
	Equal Variances not assumed	34.916

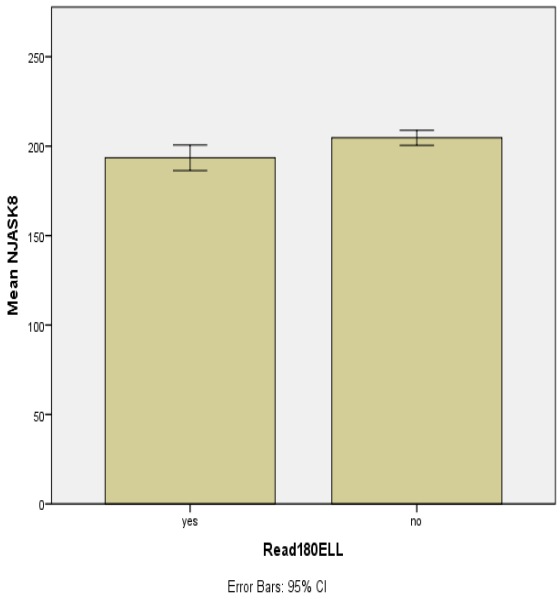

Figure 6. Equality of means NJASK 8 for ELL 2010 cohort *t* test.

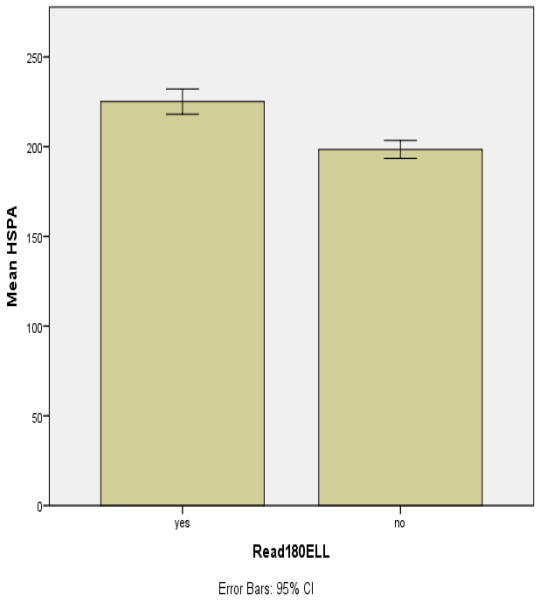

Figure 7. Equality of means HSPA for ELL 2010 cohort *t* test.

Next, to test the hypothesis that there will be an effect of READ 180 participation upon reading achievement among non-English language learner students, NJASK 8 scores of the 2010 cohort were first entered into an independent samples *t*-test to find out if program non-English language learners (N = 16) performed similarly to non-program non-English language learner (N = 161). A statistically significant result at the .05 level of probability was used as the standard for success (see Tables 27 and 28). Scores for the non-English language learner non-program participants for the NJASK and HSPA assessments were not normally distributed as assessed by the Shapiro-Wilk's test ($p <$ .05). There was homogeneity of variances for NJASK 8 scores of non-English language learner program participants and non-English language learner non-program participants, as assessed by Levene's test for equality of variances respectively (p = .881, p = .417). Non-English language learner program participants (M = 188.00, SD = 15.047)

scored lower than non-English language learner non-program participants (*M* = 212.29, *SD* = 19.018) on NJASK, indicating lower initial status in reading. There was a statistically significant difference in mean scores between program participants and non-program participants, $t(170) = -4.947$, $p < .001$; therefore, the null hypothesis has sufficient cause to be rejected. Homogeneity of variances was violated due to unbalanced group sizes for HSPA scores only of program participants and non-program participants, as assessed by Levene's test for equality of variances ($p = .001$). Contrary to findings among English language learner students, non-English language learner program participants (*M* = 211.08, *SD* = 17.816) scored lower than non-English language learner non-program participants (*M* = 227.38, *SD* = 20.603) on the HSPA. Program participants averaged passing scores but not as high as non-program participants (see Figures 8 and 9). There was a statistically significant difference in mean scores between non-English language learner program participants and non-English language learner non-program participants, $t(172) = -2.769$, $p = .006$; therefore, the null hypothesis has sufficient cause to be rejected (see Table 26).

Table 26

Equality of Means in Non-ELL p-Value for 2010 Cohort t Test

		F	Sig.	*t*	df
HSPA	Equal Variances Assumed	.023	.881	-2.769	172.000
	Equal Variances not assumed			-3.134	14.719
NJASK8	Equal Variances Assumed	.661	.417	-4.947	170.000

	Equal Variances not assumed	-5.985	20.265

Equality of variances was violated for HSPA scores, but still were significant (see Table 27).

Table 27

Equality of Means Confidence Interval for Non-ELL 2010 Cohort t Test

		Sig. (2. Tailed)	Mean Difference	Std. Error Difference	95% Confidence Interval Lower
HSPA	Equal Variances Assumed	.006	-16.302	5.888	-27.924
	Equal Variances not assumed	.007	-16.302	5.201	-27.407
NJASK8	Equal Variances Assumed	.000	-24.288	4.909	-33.979
	Equal Variances not assumed	.000	-24.288	4.058	-32.746

Table 28

Equality of Means Confidence Interval for Non-ELL 2010 Cohort t Test Continued

		95% Confidence Interval Upper
HSPA	Equal Variances Assumed	-4.680
	Equal Variances not assumed	-5.197
NJASK8	Equal Variances Assumed	-14.598
	Equal Variances not assumed	-15.830

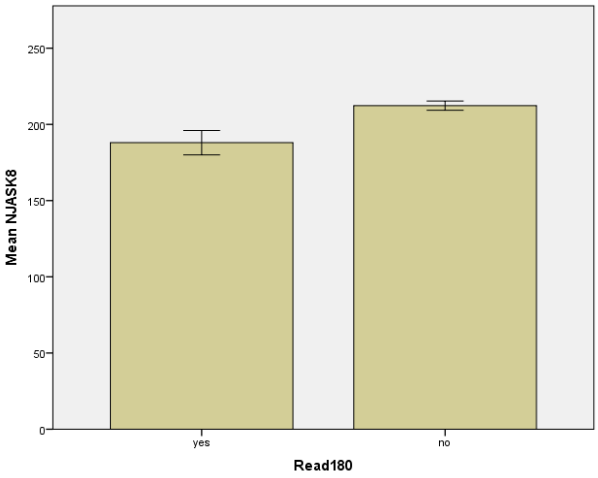

Figure 8. Equality of means NJASK 8 for non-ELL 2010 cohort *t* test.

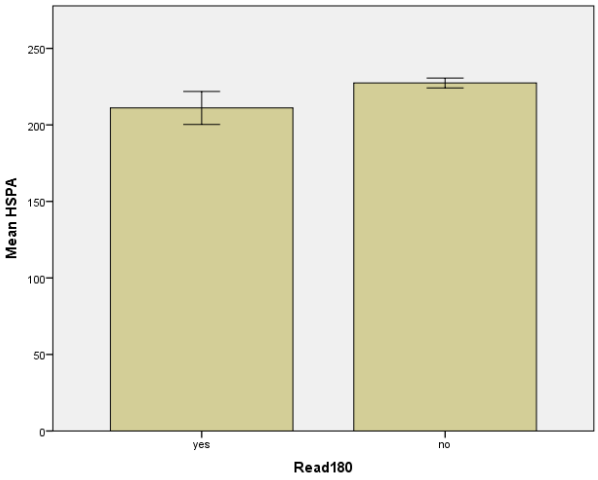

Figure 9. Equality of means HSPA for non-ELL 2010 cohort *t* test.

Hypothesis 3: Graduate rates. Next, to test the hypothesis that there will be a dependency upon graduation rates for program participation, graduation rates of participants and non-participants were entered into a chi-square test for independence. To test whether proportions were different for each group, a χ^2 test of independence was used with $\alpha = .05$ as criterion for significance. READ 180 students graduated at a slightly lesser rate (83.3%) then non-READ 180 students (85.1%) (Table 29). There was not a statistically significant association between program participation and graduation for the cohort year of 2010, $\chi^2(1) = .064$, $p = .800$ (see Table 30). There was little to no association between program participation and graduation (see Figure 10). Phi was also found to show no significant strength of association (see Table 31).

Table 29

Chi-Square Test for Dependence Graduation Cross-tabulation for 2010 Cohort

	Value		Dropout	Graduated	Total
READ 180	Yes	Count	5.0	25.0	30.0
		Expected Count	4.5	25.5	30.0
		% within READ 180	16.7	83.3	100.0
		% within Graduated	9.3	8.2	8.4
		% Total	1.4	7.0	8.4
	No	Count	49.0	279.0	328.0
		Expected Count	49.5	278.5	328.0
		% within READ 180	14.9	85.1	100.0
		% within Graduated	90.7	91.8	91.6
		% of Total	13.7	77.9	91.6
		Count	54.0	304.0	358.0
		Expected Count	54.0	304.0	358.0
		% within READ 180	15.1	84.9	100.0
		% within Graduated	100.0	100.0	100.0
		% of Total	15.1	84.9	100.0

Table 30

Chi-Square Tests for Dependence for 2010 Cohort

	Value	df	Asymp Sig. (2-sided)	Exact Sig. (2-sided)	Exact Sig. (1-sided)
Pearson Chi-Square	.064[a]	1	.800		
Continuity Correction[b]	.000	1	1.000		
Likelihood Ratio	.062	1	.803		
Fisher's Exact Test				.790	.484
Linear-by-Linear Association	.064	1	.800		
N of Valid Cases	358				

Table 31

Chi-Square Test for Dependence Symmetric Measures for 2010 Cohort

		Value	Approx. Sig.
Nominal by Nominal	Phi	.013	.800
	Cramer's V	.013	.800
N of Valid Cases		358	

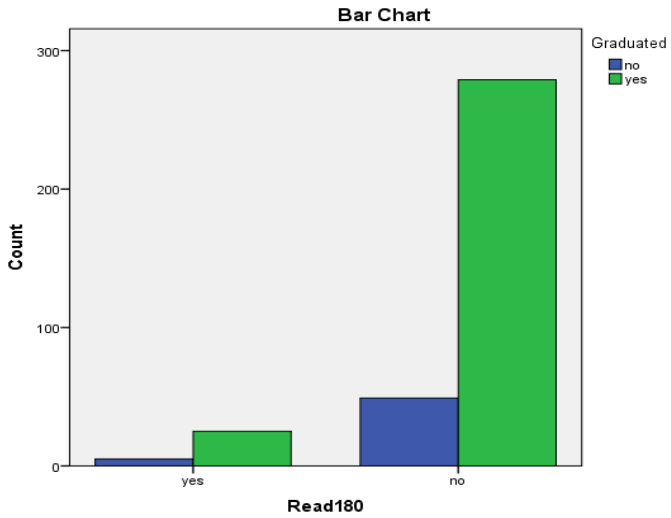

Figure 10. Chi-square test for dependence clustered bar chart for 2010 cohort.

Summary of Analyses

This evaluation study was guided by three research questions based upon locally-developed performance goals for the READ 180 reading intervention: (a) improve reading achievement, (b) improve language arts scores on the state HSPA (HSPA), and (c) increase graduation rates.

The first goal was evaluated by determining if program participation affected reading achievement at the classroom level, as measured by the Scholastic Reading Inventory. Although the participants showed growth at the classroom level in both the 2007 and 2010 cohorts as analyzed by paired difference *t*-tests, the growth of 15.03 (p = .557) was not significant for the 2007 cohort. On the other hand, the 2010 cohort did achieve significant growth ($p < .001$). Performance differences by gender showed no significant change in either the 2007 ($p = .060$) or 2010 cohort ($p = .697$). Such results

may imply that the program was equally successful at benefiting students regardless of gender.

The second goal was evaluated by performing a multiple regression analysis of program participation upon HSPA scores using NJASK 8 as a control variable. The regression analysis could be performed for only the 2010 cohort due to lack of complete archival data for the 2007 cohort. Only one variable, NJASK 8, added statistical significance to the prediction, $p < .05$. READ 180 participation did not factor in any statistically significant way, $p = .197$.

The third goal of increasing graduation rates was evaluated by performing a chi-square test for independence of program participants and non-participants. There was not a statistically significant association between program participation and graduation for the 2007 cohort, $\chi^2(1) = 3.532$, $p = .060$. There was little to no association between program participation and graduation. There was not a statistically significant association between program participation and graduation for the 2010 cohort, $\chi^2(1) = .064$, $p = .800$. There was little to no association between program participation and graduation.

Although the initial analyses did not provide confirmation for program effectiveness, results of the regression analysis did not yield information in a way that would rule out program effectiveness for all populations served by the program. In an effort to analyze the more at-risk populations served by the program, the researcher reviewed descriptive statistics to understand why READ 180 participation did not provide a predictive factor in HSPA scores. A tentative hypothesis was developed that program effects may have been differentiated by English language ability. Specifically,

English language learners (ELL's) may have benefitted from the program differently than English-proficient students (non-ELL's).

By comparing mean scores on the NJASK and HSPA, first for English language learners and then for non-English language learners, a pattern emerged indicating a differential effect of READ 180 participation based on language ability. Due to incomplete archival data, only the 2010 cohort could be analyzed in this manner. There was a statistically significant difference in NJASK 8 mean scores between READ 180 English language learner (M = 193.50) and non-READ 180 English language learner (M = 204.68) students, $t(82) = -2.092$, $p = .040$. Such a difference showed the READ 180 English language learner students started at a lower achievement level than English language learner students not in the program and were more at risk of failing the HSPA. After participating in the READ 180 intervention, there was a statistically significant difference in HSPA mean scores between READ 180 English language learner (M = 225.10) and non-READ 180 English language learner (M = 198.42) students, $t(24.190) = 6.678$, $p < .001$. READ 180 students not only grew at a greater rate, but safely entered the HSPA passing score range (200) required for graduation, while non-READ 180 English language learner students actually regressed to a lower score range that would not allow them to graduate. This loss of passing score range for non-READ 180 students also could help to account for any maturation threats.

The score pattern for English proficient students (non-ELL's) was very different. There was a statistically significant difference in NJASK 8 mean scores between READ 180 non-English language learner (M = 188.00) and non-READ 180 non-English

language learner (M = 212.29) students, $t(170) = -4.947$, $p < .001$. READ 180 participants were clearly, once again, in a greater position to fail the HSPA and therefore not achieve graduation status. After program participation, it was found that there was a statistically significant difference in HSPA mean scores, but this time in favor of non-program participants. READ 180 participants (M = 211.08) once again had an increase, but not as much as the non-program group (M = 227.38), $t(172) = -2.769$, $p = .006$. Thus, both READ 180 English learners and English proficient students were lower in initial reading status than non-program participants, as would be expected since reading status was a factor in program selection. However, whereas the English proficient students remained lower in reading achievement than their non-program peers even after the READ 180 intervention, the English learners not only improved but also outperformed the non-program English learners who were initially superior to them prior to the READ 180 intervention.

One can infer from the results that the English language learner students benefit from the intervention. One can also infer that English language learner students not in the intervention actually regress in achievement performance and actually lead to harmful results. Both sets of means comparison demonstrate the READ 180 students growing more than non-program participants do. Holdaway's theory can explain English language learner students lowered *literacy set* becoming a hindrance when students are just entering the study of literature in middle school, but over time such a lower set can become a detriment as more advanced literature studies are practiced in high school. The lack of growth could be due to the fact that inference and connotation are the

predominant modes of analysis by instructors at this level. A balanced literacy program can only be effective if it is geared towards the student individual needs. It is not possible for one program to engage all levels of all students at all times. Kim et al. (2011) reported this when only students who scored at the 40-45th percentile on the SAT 9 or SAT 10 comprehension test showed significant gains in reading comprehension in balanced literacy programs. Further support is shown by the smaller increases made by READ 180 non-English language learner students. Non-English language learner students entered the program with a higher *literacy set* did not grow at a rate similar to their English language learner counterparts in the program. It may be concluded that the READ 180 program is more geared for students in need of a bottom-up heavy approach while offering top-down skills at a lower rate. Thus, students entering the program of English language learner status and failing the NJASK 8 present as the most receptive to benefit from the program's offerings.

Overall, the evaluation study site can find consolation in the results pointing to the average score for an intervention READ 180 student being safely in the proficient category of the HSPA. Even if the program was not fully effective for all participants, it was still capable of providing the means to an end, chiefly, increasing state test language arts scores. A balanced literacy program should be reinstituted based on such results.

One may point to the inconsistency in growth for both cohorts due to their stark differences in success. Growth in gains through 2010 only supported my initial concerns about program logistics and fidelity through the early days of the intervention. The differing effects from the early cohort of 2007 through the cohort of 2010 seemed to

support a concern about early program adaptation. The 2007 cohort had only been implemented for less than 2 years, with much of the resources and technology not in place until closer to 2007. A lack of technology components was a regular occurrence in the early years of the program. After contemplating this point, I also began to ponder what could be learned if complete archive records were kept and made available in order to determine and pinpoint where in time the *turn* happened from insignificant to significant scores. Knowing what the *critical mass* for implementation is could help in properly resourcing future programs and help them to start more effectively earlier.

Ultimately, the evaluation study was a success since it was able to determine to what degree the three goals of the READ 180 intervention were effective and what sub-groups benefited most. Section 3 will describe the deliverable project developed as a result of the evaluation study, project goals, project rationale, review of relevant literature, and how the project will address community needs locally and wider.

Section 3: The Project

Introduction

This evaluation study provided statistical evidence regarding the effectiveness of the READ 180 intervention program. In Section 3 I explain the deliverable project I developed as a result of this evaluation. Detailed in Section 3 are a description of the project and its goals, a rationale for the project, and a review of the literature pertaining to the appropriateness of the project to address the local problem. The project itself is included in the appendix.

Description and Goals

The evaluation study undertook a goals-based program evaluation that addressed the problem of low reading achievement in urban public high schools by providing a quantitative analysis of the efficacy of the READ 180 intervention program. The immediate goal of the evaluation was to determine whether the intervention program achieved the three measures of success as defined by the school: (a) improving reading achievement at the classroom level; (b) improving scores on state language arts tests; and (c) increasing graduation rates. A long-term goal of such an analysis provides direction and contributes to the knowledge policy makers need in order to make informed decisions before proceeding with future interventions. To accomplish this long-term goal, I developed a white paper to inform policy makers in a format most accessible and familiar to public school stakeholders. Furthermore, it is the social goal that educators will look upon such a project as a template to carry out future evaluations.

Rationale

A white paper was selected for this project in order to clearly and directly inform policy makers of the results and implications of the evaluation. White papers are reports written authoritatively and offer facts alongside proposals regarding pressing issues (Juricek, 2009). As opposed to technical reports, which may contain dense theoretical language and unfamiliar statistics, a white paper can ease transmission of ideas to an audience with limited research background. Reporting findings of a program evaluation to a school district that has never performed such program evaluation before can be challenging, as ease of understanding is necessary for policy makers and practicing educators when being presented with a research report (Creswell, 2012). This ease of understanding may increase the opportunity to affect permanent change, as the statistical analysis of hypothesis testing and regression may pose difficulties to comprehension. Further, the school district where the evaluation took place has been beset by administrative turnover in recent years. There have been four superintendents in 5 years and three principals in that same time. The vice principals have fully turned over their numbers twice and are close to a third. The board of education has also had a similar turnover rate. Within the confines of the English department, there have been six supervisors in that time, plus 1 year with no supervisor at all. With such turnover, curricular decisions can become the whim of the current person in charge. Instead, to best serve the needs of the student constituents, there is great need not only for a system of evaluation, but also for a culture of evaluation that allows quantitative evaluation results rather than personal preference to influence curricular decisions. In fact, findings

discussed in the white paper suggested that a successful remedial program was terminated without benefit of any evaluative data. Perhaps the project can provide a template that will far outlast the persons in place to make curricular decisions.

Review of the Literature

Disseminating Results of Program Evaluation

In order to research the literature on relevance of a white paper to disseminate results of a program evaluation study, I conducted an online search of the Walden University Library using terms with Boolean algorithms such as *program evaluation, evidence based policy, evidence-based practice, organizational knowledge distribution, white papers, grey literature, constructivism, collaborative inquiry,* and *professional learning communities*. To perform this search I used Education Research Complete, ProQuest Central Database, and Google Scholar. Often times I would select the "more articles like this" option to further find materials when I discovered a useful article. When researching the inclusion of white papers as a genre, I could find only a few articles within the allowable period of 5 years. ProQuest returned five results inputting *white paper* and *genre,* none of which were suitable for use. Modifying the field to remove *genre* opened the search too all white papers (regardless of topic) written within that time range, returning over 2,000 results. Typing *white papers* into Education Search Complete returned 311 results, of which only two were directly addressing the issue of analyzing white papers as a genre. This necessitated using Google Scholar's *related articles* feature and utilizing the reference lists from the articles that could be found. In addition, recent articles on the specific style of program evaluation used were scarce as it has not been the

most popular trend in recent program evaluation literature. Overall, 28 articles from peer-reviewed sources published within the past 5 years are included in this review.

Program evaluations that can become institutionalized may provide for a continuous cycle of adaptation and change, supporting the need for adaptation and change required of any institution or society to survive (Shaw, 2002). It is only logical to suppose that if an organization is failing, it may need to increase its ability to change. Determining the best way to use a program evaluation would necessitate a state of ongoing change and adaptation.

In order to facilitate movement through organizational barriers and bureaucracy, the ACE Star model of knowledge transformation may provide a useful template to promote such adaptation and change (Stevens, 2013). Each step of the five-part "star" is designed as part of a continuous cycle, similar to the logic model of program evaluations, in order to transform evidence and research into readily applicable information and action (Stevens, 2013). The ACE Star model cycle contains the following in order: *discovery research, evidence summary, translation into guidelines, practice integration,* and *outcome evaluation* (Stevens, 2013). Stevens (2013) explained each point as follows: *discovery research,* primary research studies; *evidence summary,* knowledge synthesis into a simple statement; *translation into guidelines,* a combination of evidence and expertise to create recommendations; *practice integration,* when practice and evidence align; *outcome evaluation,* a view of the outcome and effects of the practice utilized. Utilizing the ACE Star logic model, program evaluations are a critical component of two of the five components for an organization to undergo successful ongoing change through

knowledge transformation (Stevens, 2013). The READ 180 program evaluation specifically satisfies both the *discovery research* and *outcome evaluation* steps in the ACE Star knowledge transformation process.

Program evaluation is an appropriate selection as an avenue of research due to its ability to fit into an organization that is going through reform, such as at the evaluation study site. As a reform tool, program evaluations could help decision-makers at the evaluation study site by focusing on institutional problems with data in real-time (Mohamadi, 2013). A goals-based program evaluation would help further by aligning outcomes with predetermined goals (Spaulding, 2008). Such evaluations are useful at isolating specific program elements to be studied (Fitzpatrick et al., 2012). Recommendations from program evaluations have grown over time to now be considered of great usefulness to practitioners seeking evidence (Imbens & Wooldridge, 2008). The ACE Star model is useful at removing barriers when moving evidence into practice, which would add rapidity to the ongoing evaluation process. Research suggested using such a model for a program evaluation is a logical choice. Primarily, the ACE Star model achieves the removal of barriers by transforming one form of knowledge into a form that suits the needs of an organization more aptly (Stevens, 2013). White papers are likewise known for their speed, primarily in how fast they can be produced, as well as how flexible they are in form (Okoroma, 2011). Therefore, the adoption of a white paper into the ACE Star model, as previously shown through research, is also natural fit. All three methods (program evaluation, ACE Star model, and white papers) work well together and serve to use the strengths of each as a whole.

Project Genre

The local school site is currently undergoing a reform model including many aspects of constructivist theory including professional learning communities, teacher leaders, and collaborative inquiry. A program evaluation report presented in a white paper may benefit schools undergoing such reforms due to the requirements of successful collaborative inquiry in a model of collaborative inquiry (Barth, 2001). What the research suggests is that such a paper will quickly demonstrate methods to be utilized by stakeholders in future research while simultaneously providing needed data for ongoing intervention practices in a manner that satisfies current reform doctrine. A white paper will also complete the *evidence summary* component of the ACE Star model.

A white paper can be a valuable tool for researchers to disseminate research results, especially to a non-technical audience. Often the white paper is prepared for a group of decision-makers who can act upon the recommendations of the white paper. White papers can be the first, and often only, resource on a topic lending them great value (Childress, 2003). Medical researchers have moved towards using white papers for research in recent years (Pappas & Williams, 2011). A study performed by Juricek (2009) found white papers as a simple method for delivering original research as a means of reporting on relevant issues directly to an audience. Juricek went on to define white papers as a category of *grey matter* or *grey literature*, and noted a key difference in such papers being that they are not produced by the publishing industry and therefore are not commercially available. He explains how common examples can range from evaluation reports to conference papers. He points out that white papers have become so commonly

accepted they can be found on ProQuest's *Historical Annual Reports* and scholarly journals such as *The Grey Journal*.

Grey literature as a term was originally used in Germany at a conference in York in 1978 as 'Graue literatur' (Lawrence, 2012). With the advent of online publishing, its use has since grown over time. Such growth has made it difficult to fit into a specific category as a genre (Smith, 2009). Thusly, g*rey matter* can often be difficult to define by librarians (Frater, Myohanen, Taylor, & Keith, 2007). Olson (2013) describes *grey matter* as generally not indexed by major databases and not published with the same standards as peer review articles, yet it also can include theses and dissertations of high quality. Some even believe *grey matter* will one day be equal to peer reviewed articles in abundance of use by scholars (Banks, 2006). Yet, there are several traits that are desirable for *grey matter*: it is the first source of research on a topic, it contains major innovations by practitioners not able to publish in peer journals, it can focus the view of a topic with great flexibility since there are no publisher restraints, and it can be completed at a fast rate (Olson, 2013). The last trait is also one of the greatest characteristics of *grey matter* in that it can fill the void left between professional journals and timely research (Happe & Walker, 2013). This trait would be of particular importance to the evaluation study site since limited independent research on the READ 180 program is available, especially with demographics similar to the evaluation study site.

Recently the use of *grey matter* is increasing in historically quantitative fields such as health care. An example of this is a recent medical journal meta-analysis by Van Cauwenberghe, Maes, Spittaels, Van Lenthe, Brug, Oppert, & De Bourdeaudhuij (2010)

where seven *grey* research reports were included along with traditional articles in an effort to get the most up to date and accurate view of child obesity across Europe. One of the main reasons for the inclusion of the seven pieces of research was that they were the only existing places for information from various schools and countries that did not have capabilities to have large amounts of published material on the topic. Focusing on English language learner students with low literacy rates entering the program, this program evaluation will also fill a void in a similar manner.

Another example of *grey matter* filling a research void is a mental health meta-analysis by Stickley and Wright (2011), which had a similar problems researching mental health recovery literature in Britain. The study was done to evaluate the efficacy of the British mental health evidence base. After reviewing the traditional sources of articles, they found it necessary to include a comparison of *grey* research. Which was done in an effort to acquire the most complete review possible. By attempting to understand the READ 180 relationship to graduation rates, this project will also fill a void for the evaluation study site.

Many empirically sound initiatives do not make it to implementation (Odom, 2009). Simply generating good research is not enough, good research needs to be integrated into practice (Reigle, Stevens, Belcher, Huth, McGuire, Mals, & Volz, 2008). Influencing stakeholders will require more than simply using elements of constructivism in the project and product. With such a wide birth, the shadow of constructivism is often referenced. Many rarely understand it as a school of reform, so it can become a catch-all phrase with little specificity. The evaluation study school also uses this term and elements

of the more popular tactics of it similarly. As Philips (1995) indicated, such a diverse dominion over seemingly endless sets of branches can cause constructivism to be confusing for newcomers. Philips went on to distill this theory down to its most fundamental aspects and declared that at its' root, constructivism is merely the assertion that human knowledge and by extension all public knowledge is constructed (1995). It will then become the responsibility of the researcher to produce a product in a way that is most easily constructed. A *white paper*, due to its flexibility, could produce such an effect.

In order to explain complicated research to a school with such a diversity in problems, *grey matter* is a natural fit. Since plans need to be modified and evaluated during implementation, an adaptable product needs to be produced. The genre of a white paper will satisfy the ability to inform policy makers how to construct interventions in a way for evaluation, while still utilizing components the site has chosen for implementation. To best plan and ultimately influence policy based on evidence there are three types of knowledge, which need to be included, as stated by Head (2008): political judgment, rigorous scientific analysis, and practical organizational knowledge.

A careful selection of evidence can be used best for influencing political judgment when seen as in the political perspective. Utilizing information in a way that presents a persuasive approach more than being purely objective satisfies this requirement (Head, 2008). If the policy crafter wishes to demonstrate why something may cause a positive or negative impact, utilizing appropriate evidence in large areas would be effective. In writing the project product (*white paper*) and performing the project evaluation

appropriate evidence could be utilized with results from the statistical analysis, such as the multiple regression analysis, independent *t*-test, and paired difference *t*-test. Of particular importance, would be to repeat in various ways the results of the English language learner population and effects upon HSPA scores. It is often considered a necessary practice to use the best available evidence, which includes white papers, when policy makers are assessing this type of information (Carter, 2010). Results that coincide with the overall findings should also be presented in several varieties to ensure they are fully effective. This would be possible with the various bar charts used to explain the growth of means between various sub-groups that were analyzed and easy to read tables to show population type of samples. To accomplish this the English language learner READ 180 and English language learner non-READ bar chart, which demonstrated the gains students made from the NJASK 8 to the HSPA would be very successful, as opposed to a more dry and difficult to understand *p* value discussion. This would also be useful for the SRI growth for the 2010 cohort of the READ 180 analysis. Showing the clustered-bar chart of the 2010 cohort graduation chi-square results would similarly help in this manner. Also, adding color to the English language learner table indicating state test score gains could help with this.

In order to politically judge information to use as persuasive evidence one must have a strong coalition of support (Flitcroft, Gillespie, Salkeld, Carter, & Trevena, 2011). Such support will be possible if care and attention are made to the organizational climate and culture of the site at the time when analysis and advice is given (Head, 2010). White papers can provide a learning experience that can strengthen ties for researchers and

communities (Willerton, 2013). Having partnered with several key stakeholders throughout the research study and project process, it was possible for the project genre to accomplish this as their preferences in consuming information are known to me. Of particular concern is the Spanish speaking population being brought up to a passing level on state testing, which was a centerpiece of this project as it was analyzed and presented with multiple inferential and descriptive analyses.

A potential downfall to the level political judgment can be influenced will come from the current crisis mode of the evaluation study site. When organizations are in crisis response mode often a desire for rapid responses trumps the ability to perform long-term research and analysis. This was demonstrated by constant relegation to *low* priority of information requests made during this study, which were never responded to. Similarly, risk-management may cause policy managers to be more cautious than they would otherwise be (Head, 2008). Any potential stakeholders, in such an atmosphere, that appear less than favorable from past decisions regarding the READ 180 program may be skeptical or reluctant to believe the findings.

Utilizing political incentives may overcome this obstacle (Head, 2008). During politically volatile times, it is often common for policy managers to be concerned with external support (Head, 2010). Educational leaders during such a period often must think strategically about how they are *outwardly* performing (Brewer, 2011). The evaluation study school will be accepting of a project evaluation since it demonstrates to the external stakeholders of the community that decisions are being made based on evidence more than opinion. Demonstrating a culture of evaluation can also be seen as a culture of

learning (Head, 2010). Such a statement would help soften otherwise politically opposed external forces. Using the results to bolster an outreach of community support for recent immigrant status students, who struggle with language (ELL's), could give credibility to administrators attempting to show more inclusive reform efforts.

In order to achieve a rigorous scientific analysis, a precise and full analysis of data is needed. As defined by Head (2008) this is possible when a researcher produces policy change research efforts after a careful and systematic analysis of trends and the past conditions of those trends as well as taking into account the current conditions. Initially, there may not be a large contingency of evidence, so adherence to strong scientific principles should be used (Hughes and Titler, 2008). Policy managers need to see an entire picture of how the research fits into the institutions structure. Having contextualization of possible effects is necessary for evidence to be effective (Crouch, 2010). Presented information must also be based on topic areas of high relevance to those involved (Head, 2010). Instead of a simple snapshot the research needs to also explain how the history and potential future of the research coincide. By having an evidence-based evaluation analysis, it is more likely that it will be credited with being synonymous with *what works* (Bell & Dolainski, 2005). Since the project and product are explaining data of students taken over a multi-year period, and data is linked to understanding a very relevant trend in a lack of student achievement that is attached to funding and state oversight, it will easily fit this requirement. An in depth analysis was evidenced by taking the time to perform additional descriptive statistical analysis after the completion of the initial testing was performed, allowing the evaluation to uncover the sub-group of

English language learner students that benefitted greatly from the READ 180 program. Also, using the newfound information from the descriptive statistics to then perform a hierarchical regression analysis lends support to the academic diligence used in the test procedure.

Schools in crisis mode are also in need of showing that rigorous scientific analysis is only used if it provides good information that is democratically accountable (Head, 2010). In order for a decision to be taken with full merit in such an institution as one in crisis, stakeholders are aware that when action is taken there will be plaudits or criticism. During a crisis, consequences matter. By drawing on data from a variety of sources policy managers are more assured that proper analysis has taken place (Flitcroft et al., 2011). The unique nature of a program evaluation is that it is by its very nature a highly systematic approach to data gathering (Metcalfe, Aitken, & Gaff, 2008). Since the project and product were a program evaluation analyzing multiple and various statistical measures taken from many sources, this requirement was satisfied.

From the standpoint of a stakeholder using evidence to inform policy, it is necessary for the researcher to have practical organizational knowledge. Knowing the internal workings of an organization, which will make it more likely the research is seen as an infrastructure investment (Shirey, Hauck, Embree, Kinner, Schaar, Phillips, Ashby, Swenty, & McCool, 2011). Knowledge of this type is strongly centered on the current context of program effectiveness (Head, 2008). Head (2008) goes on to state that this is precluded by a mutual understanding of the different roles played by a researcher and the policy makers. If research is not balanced by practicality in its ability to be used it will

not be used effectively (Head, 2010). Of equal importance is having clarity of the report (Cresswell, 2012). The current project and product is centered upon evaluating program efficacy and is performed by a teacher-leader as defined in the current reform model. As long as priorities and activities are adapted to meet the need of research utilization the project and product will be successful (Head, 2010). With the policy makers at the highest position already involved in the project and product, the final result was one they are confident in. Utilizing this organizational knowledge made it possible to gather information to be analyzed that was otherwise unavailable.

Implementation of Project

First, the researcher will contact via email all key stakeholders, establishing the preferred method for dissemination of the white paper. Potential methods of dissemination include regular post, email, PowerPoint, and personal presentation. Next, the researcher will deliver the white paper to each of the following stakeholders:

- After the white paper has been delivered to all stakeholders, the researcher will follow up to establish a time to discuss findings and potential concerns regarding the content of the white paper with each constituent group. (Intended start date: March 2015. Intended complete date: March 2015.)
- The researcher will make himself available for a face-to-face meeting with the superintendent of schools and the high school principal. (Intended start date: April 2015. Intended complete date: April 2015.)

- The researcher will also offer to meet with the school board, either in open or closed session. (Intended start date: April 2015. Intended complete date: April 2015.)
- Such dialogue with stakeholders is intended to establish which essential elements of the white paper are most capable of being acted upon, to satisfy the *translation into guidelines* component of the Ace Star model to be completed within the coming school year. (Intended start date: May 2015. Intended complete date: May 2015.)

Roles and Responsibilities

It will be the responsibility of the researcher to manage dissemination of the project by contacting recipients and providing information in the form (electronic, paper, oral) that is most suited to their needs. Timely responses and being flexible will be necessary. Following up on whether the project was received and establishing a time that is most convenient for the recipients to meet and discuss the project will also be necessary. In order to establish a culture of evaluation, demonstrating how such a project could be turn-keyed will also be beneficial. Also, by organizing the desires of each stakeholder, and acting as the facilitator of such information, it will be possible to stay involved in the process more fully in order to ensure action and results are forthcoming.

The high school principal and vice-principals have responsibility to suggest curricular programs for approval at the district level. In addition, they are responsible for oversight of implementing short-term curricular or pilot programs. The principal and assistant principal then need to justify the continuation or termination of such programs.

Thus, the white paper could set a precedent for the use of ongoing evaluations to measure success of interventions that address literacy. Whereas teachers have the responsibility of implementing curricular programs, high school administrators are responsible for assuring that long-term goals are met across the spectrum of grades and ability levels.

In addition to their role as implementers of classroom programs, English teachers also support curricular decision-making in several ways. They can bring forward ideas for new literacy development programs and they can write new courses of study and unit plans for existing courses of study. In addition, they write lesson plans that constitute the actual instructional activities students experience on a biweekly basis. Teachers can also be included in decisions regarding the future status of new programs. Teachers will need to fully assume the role of scholar-practitioner by working with the department chair to ensure they are promoting an environment rich in evidence useful for evaluation. Taking on this role will help teachers to feel as co-researchers, not just agents of the administration. They will also be utilizing the *practice integration* component of the Ace Star model. When teachers feel they are co-researchers and are part of the formulation and solution in an evaluation, they are more likely to be strong supporters (Henke, 2001).

In order for decisions to be enacted, district-level administrators must give approval of all long-term curricular decisions. Decisions of this nature must be balanced between resources that are available and personnel capable of enacting the decision. Ultimately, school board members are the final voice and stage to the process of curricular decisions. Any long-term decisions must be passed with a vote to either approve funding of programs and resources that support those programs or to approve

curriculum that may be implemented. As the voice of the community most impacted by the results of the school system, they must look at all of their decisions by establishing what is best for students and the community as a whole in the short and long-term.

Resources, Supports, Barriers

Many initial resources required to evaluate white papers are already in place such as time, structure, and political will. Professional development is a requirement for any eco-system of evidence-based practice (Odom, 2009). The current contract at the evaluation study site between the teachers and district already includes several hours per month of mandatory meeting time. Utilizing this time to read and discuss the project would be a realistic use of this time.

Politically the theme of reform is strong within the district. Reorganization of the structure from buildings to administration is already taking place. Having the *will* to support such initiatives makes white papers much more realistic.

Support for teachers producing white papers would be in the form of monthly professional development meetings. During this time meeting with stakeholders is possible for many areas of the English department due to the myriad of pseudo-administrative staff such as reading specialists, literacy coaches, department chairs and supervisors. Having support to implementation is needed if organizational practices are to be changed (Odom, 2009).

A barrier to the implementation of the white paper could be administrative commitment. Administrators must be convinced such a project effort is worthwhile and practical. With such a high turnover rate at key positions of leadership, administrators are

already overworked and behind schedule in many areas of their duties. If they deem the project a low priority, time may never be granted to review and discuss its findings. Also, they may consider the ramifications of the project too far beyond the scope of what can realistically be done in the current environment. Not receiving meeting time to discuss or present findings could lead to other teachers deciding not to engage in future white papers due to a lack of interest and response, ultimately preventing the spread of similar teacher-scholar led projects.

Project Evaluation Plan

Performing an evaluation of the impact the project has had towards creating a more evidenced-based system of reading interventions (utilizing balanced literacy) requires a summative design. Summative designs are best when determining the final outcome of an event that is not developmental (Fitzpatrick et al., 2012). In opposition to formative evaluations, which operate under the motive of program improvement (often incremental), summative evaluations will explain the end result of a program and its ultimate merit (Fitzpatrick et al., 2012). A summative evaluation of the impact of the project would best serve the purposes of the evaluation study site.

Participant-oriented approaches empower stakeholders and often seek social justice (Fitzpatrick et al., 2012). A summative evaluation using a participant-oriented approach fits the needs of the evaluation site and the social justice underpinning this study. Following the Stake Responsive Approach a full description of effects produced by the project could be used to provide judgment of merit by a stakeholder (Fitzpatrick et al., 2012). Utilizing benefits of being an internal evaluator such as: familiarity with the

evaluation site, access to information in a timely manner, understanding of political nuances, and the ability to follow-up with stakeholders in a rapid manner will enhance the evaluator's ability to ascertain the final worth of the project (Spaulding, 2008).

Focusing on Stake's responsive evaluation model, it will be possible to engage on-site in a pluralistic manner consistent with current reform methods already undertaken by the evaluation site (Fitzpatrick et al., 2012). Such a model will ensure that the evaluator (a key stakeholder) is focused on adapting to changing knowledge, how programs are seen in different ways by different participants, and emphasizing local knowledge to understand nuances and sensitivities (Fitzpatrick et al., 2012).

In order to gather information for ascertaining final worth of the project the following stakeholders will be participants: principal, English supervisors, and reading intervention specialists (both reading and literacy). To gather data Likert-style surveys will question how they feel the progress of the following has been since the project was provided: balanced literacy interventions; program evaluations as a decision making tool; data-based policy decisions; effects of interventions on student sub-populations (i.e. ELL). A space on the back of each survey will be provided to leave additional comments of perceptions. The results of this information will be placed into a bar graph with numeric tallies and percentages for each response category as well as summaries of comments.

Project Implications

A critical aspect of social change and impact on local stakeholders is how the project and genre fit this study. One can only see this impact fully after understanding

how important the efforts of the professional learning communities set up by the school are to reform. Learning communities also satisfy any number of Ace Star model components including: *discovery research, outcome evaluation,* and *practice integration*. Many themes relevant to the project such as democracy are reflected in this aspect.

A strong reason for performing this project is the democratic equality called for by Dewey (2013), when he reasoned democracy cannot exist if all are not treated equally by institutions. Habermas felt similarly when he stated that schools exist to promote democratic ideals (Coulter, 2001). In order to fit this maxim into professional learning communities (PLC's), it is necessary to understand how they at their very foundation echo such a call. Such a link is supported by many. More support is garnered by Merriam (2007) when he explains that democracy is nothing more than an opportunity for the most power to be wielded by citizens directly. A statement that coincides with lack of local control by failing schools in crisis. To regain control locally, and demonstrate democracy, fitting the using the genre of a white paper will ensure an ease of assimilation into the existing PLC's because of how it fits easily into those characteristics.

PLC's embody what Chance (2002) called a change from within. The aspect of collaborating in this model represents how all levels of stakeholders have a chance to enact change, as may happen with this project product. Professional learning communities at the evaluation study site meet daily in small PLC groups to engage in professional inquiry and collaboration. Sincere collaborations of this type are more inclined to be sincere and candid, they are more likely to be accepted by organization members (Beatty, 2007). Weinbaum (2004) demonstrated how groups instill a feeling of

leadership in teachers once they have control over their professional development, which also is key in creating a successful community of leaders. This is pointed out by Barth (2001) when he explains that teachers assuming responsibility for issues they care about have the most opportunity for growth. Barth (2001) goes on to say that teachers who inquire collaboratively make schools more democratic. Growth of this type can exist in the collaborative inquiry component of PLC's. In such a setting, due to time constraints and small groupings, a white paper would be most fitting to convey such an inquiry.

School communities devoted to collaboration must also include allowing local community members access to the decision making process (Epstein, 2008). Posting an easy to comprehend white paper, such as the project product of this evaluation, will enable this type of involvement. With such easy access to relevant data by community and school officials new directions in how their democracy is run may be attained (Trachtman, 2007). These collaborative relationships can lead to directions empowering minority stakeholders, which would increase democratic equality (Carnmarota, 2009). It is of great importance to lower the distance from decision making and power to bolster democratic growth (Carnmarota, 2009). The project product being a white paper will give more teachers the impetus to follow such a path of empowerment.

Section 4: Reflections and Conclusions

Project Strengths in Addressing the Problem

Goal-based program evaluation provides data to permit informed decision-making regarding the future direction of the program being evaluated. For this evaluation study, access to a large set of archival data was conducive to rich statistical analysis. Having three separate statistical analyses helped the evaluation achieve analysis triangulation and provided a more robust result (Hussein, 2009). In addition, clearly articulated program goals permitted me to effectively measure program effectiveness. Given the extreme need of target students in the areas of reading and language arts, the evaluation study was able to shed important light on the efficacy of the READ 180 program for English language learners at the New Jersey high school under study. The post hoc design necessitated by the termination of the READ 180 program was not optional but proved useful when ongoing collection of new data was not feasible. The large store of archival data available to me as the researcher for this study permitted savings in both time and resources. Indeed, the many archives of data maintained by the school district represent opportunities for considerable future research. The excellent record-keeping by the school district will allow continuing evaluation of the READ 180 program, and future researchers could pursue a multiyear longitudinal study. Public school systems, like many modern institutions, are filled with a wealth of collected data (Jones, 2010), much of which may be untapped to inform decision-making. Utilizing a white paper genre to relate findings is a very effective and easy way to replicate method that will transfer

easily to other scholarly practitioners. White papers are also easier for nonacademics to understand, which will increase knowledge transmission.

Project Limitations in Addressing the Problem

Program evaluation at its best has the ability to impact an ongoing process. However, the program I evaluated was no longer in operation at the evaluation study site. Of course, that decision was out of my control, and I would have much preferred that ongoing evaluation of the program had occurred before an administrative decision was made to terminate the program. Therefore, in the sense that the present evaluation cannot impact directly an ongoing program, one might conclude that the evaluation study is limited. On the other hand, to the extent that this evaluation raises awareness and concerns regarding the overall direction of reading intervention in the school district, it may prove a success. As it is, there have been two cohorts of students moving through the school system who received no reading intervention for these last few years.

Several limitations impacted the content of the data used, and thereby the analyses conducted. Of the available data, I was only able to use two cohorts out of seven. Of those, only one had complete data. The only way to fully examine how the students' attitudes and skills were impacted in this cohort would be over a longer period of time. Achieving this would only be possible with longitudinal tracking, which was not within the parameters of this study. Paper-and-pencil tests of reading, which one could argue are irregular and performed rarely by students in their everyday literacy activities, are not always valid in how they simulate student experiences in school. Finally, perhaps one of the largest limitations would be not including how home environment, students'

predisposition to recreational reading, or study habits of individual students influenced results.

As much as archival data have potential, they also have drawbacks. Tracking a cohort of students for 5 years would be a Herculean feat for a doctoral candidate of limited resources if not for having access to an archive. Yet, archival data posed another drawback to the project in the sense that a richness to the context is lost when data are not as current as possible (Shultz, Hofman, & Reiter-Palmon, 2005). Archival data would work best if merged with more of a prospective research design. I also feel having no qualitative data limited the conclusions I could draw because educators deal with the whole student, yet in this study I only analyzed test scores. Interviews and surveys may have helped me draw different or more detailed conclusions. Using a white paper as the deliverable project for the study may not be seen by all academics as the most traditional format and may incur skepticism merely due to genre selection. In addition, a white paper, in its attempt to provide a diverse audience with technical information in an easy-to-understand format, may oversimplify dense content or omit certain detailed information that provides nuance to the technical information. Thus, while administrators, school board members, and parents may appreciate its brevity, teachers and curriculum specialists may prefer more statistical reporting. Furthermore, it remains to be seen if recommendations of any reform model will be implemented long-term due to the constant turnover at key levels of policy and implementation positions.

Recommendations

Program evaluation should become formalized and operationalized as a normal and expected aspect of curricular oversight in the research site's school district. Program evaluation results, over several years or student cohorts, should guide curricular decision-making, including textbook and software adoption.

Several avenues could be traversed in an effort to enhance program evaluations of future reading and literacy interventions. First, establishing program goals should take into account parental, as well as district, expectations, promoting a more democratic process of curriculum evaluation. In addition, evaluation results could be expanded to include qualitative data in the form of interviews and surveys with students and teachers. Such qualitative data might provide insights into why certain results are, or are not, achieved. Further, future evaluations should not be conducted post hoc, but rather should become part of the ongoing implementation of any reading or literacy intervention. Ongoing and current evaluation can provide for tracking of fidelity of implementation as well as student performance.

Program evaluations should always consider performance of student subgroups, such as English language learners, in addition to the group as a whole.

Dissemination of evaluation results should take into account the needs and abilities of multiple constituencies. Making evaluation results understandable to a larger group of constituents would further democratize the curriculum reform process. For example, the qualities of a white paper that make it accessible to many may provide

stakeholders who have not previously had voices in the reform process an avenue to take on more active roles.

A commitment by administrative and academic staff to more willingly engage in scholarly research, combined with adequate professional development, will allow future project evaluations to be communicated in a more technical format, to enhance the dissemination of results being provided in the less-technical white papers.

What Was Learned About Scholarship

Throughout the course of conducting this project, I learned a great deal about the process of scholarship. Scholarship is something people assume comes easily to those that are born with a high IQ, and as such only those same people are capable of performing scholarly tasks. I learned this could not be farther from the truth. Scholarship involves being able to align all aspects of a process to work in synergy. It is only at that time when a researcher appears to be intelligent. The process is what is smart, while the researcher is merely a mouthpiece to that process. The careful coordinating of research and analysis is vital. It would not be possible to write in a clear and cogent fashion if research was not on target for the proposed hypothesis and if the analysis was not likewise suitable. I have increased my respect for professional researchers and realize now it is not that they are just smarter, but that they have goals aligned to a process that ensures what they are writing is of the highest caliber.

What Was Learned About Project Development

Project development may start as one person's vision, but I learned it takes a shared vision by many to see it through to a successful end result. Without the belief from

my district that research was necessary, I would have never been given permission to view the archival data so critical to this evaluation. I think it speaks highly of those involved that they were willing to allow me to look into the cessation of a program that may actually end up showing the district made a poor decision. In the current climate of political volatility, I could easily see an approach of leaving things the way they are. Without my district's support, this project could not have been completed.

Having support of my district was necessary, but having a team of veteran scholar researchers as committee members was critical. Project development through the dissertation process is not about being told what to do. Many times my committee could have just told me what to do and I would have done it. It was only after very long and winding routes that I would end up at the best solution to the various stages of project development. Early on this made me realize I had help, but more in the sense of guidance and course correction. All of the actual work had to be done by me. This was scary at first, but over time my confidence grew enormously. I feel I could now develop my own project successfully, but also recognize the value of having input and guidance of veteran researchers.

What was Learned About Leadership and Change

I learned that leaders are under pressure from many sides and that for change to take place information needs to fit a certain criteria. If I research a problem and find a potential solution it may not be actionable unless I as the researcher can place the research into a model that fits the school politically, scientifically, and organizationally (Head, 2008). The explication of that research is as important as the research itself. This

was something I had never considered prior to this project. Understanding that research is of and by itself *not* the single most important factor was something I had to reconcile with at first. Using Head's model of knowledge transmission was something I had never done before, but I now fully expect to use it for any future evaluations.

Of particular importance is that fact that I am working at an urban school in a poor district, schools in these settings often have overwhelming issues to deal with (Jackson, 2005). Due to this, the effectiveness of utilizing existing reform models is also very important to having a project used for change by leaders. Leaders have selected various reforms due to a belief in how they work. By using a project that fits well into that existing model, I think I have made it easier for my school leaders to allow us to all take one step closer to being teacher leaders.

What was Learned About Being a Scholar

I have learned a great deal about being a scholar. In the past I would have assumed great writing is the key to great research, but I see that there is a process that must be respected. There are no short cuts to great research writing. It does not matter how many words I can type per minute, what matters is how many minutes I read other researchers words. The culmination of theory and research into a process that can analyze both is how great research writing happens. This can only happen if I spend the time properly exploring all possible avenues of theory and data before I even begin writing one sentence of my own. The actual research writing process is dwarfed by the preparation the process of research requires.

I have also learned that without a community to share your scholarly work, it is useless. By always keeping my mind's eye on the goal of having a final product I could share with parents and administrators alike, I was able to keep my perspective and adjust my goals accordingly. Ultimately, this is how I was able to distill what was at times a very wide direction for research into a specific path with individuals I knew I could help personally.

What I Learned About Being a Practitioner

By becoming a practitioner of research I was able to grow as an educator in many ways, most notably at being a careful consumer. My research skills at finding articles and evaluating them for inclusion to a study increased greatly. Along with the increase of that skill my ability to see how researchers may try to steer audiences with use of language and selective use of data grew. Many articles are written by people with agendas and that should not get in the way of useful data. As a producer of research I am now able to identify those moments where I am being led by an author and simply form my own conclusions based on the data and results. Such a skill will help me to continue to grow as a practitioner.

Performing a literature review was something I had never done before and other than the aforementioned skill at detecting loaded language, I have grown as a consumer of quantity. In the past I may have found an article or two to use as support for a paper I was writing for a course, but never before did I need to find 25-40. When first faced with this task I thought it was merely overdone and superfluous. Then as I made my way through, I started to see how small turns in the research of the exact same concept could

be seen when done with such a mass consumption of research. I was then able to break what I thought was the only theme into several sub-themes, which drove me to more research. There were points when I actually had to cut articles out and not use pages I had written for the sake of brevity. A far cry from my original perception of the literature review process.

What was Learned About Being a Project Developer

Being a project developer requires patience and flexibility. At first I thought time was my enemy and that the more time I used the worse I may be doing. After some point in the evaluation study process I realized time can be a benefit. Having more time to revise a plan, or contact a stakeholder to verify project site needs is indeed beneficial. Once I was able to understand this, my patience increased in regard to development.

Having a plan to reach a goal is important when developing a project, but being able to alter that plan in the midst of the development process requires flexibility. Flexibility enabled me to open myself up to many modes of project evaluation, even after I thought I had the one selected that fit my project best. Also, when selecting the various types of statistical analysis I was not sure just what was possible by using three diverse tests. After being pointed in a direction by my committee and learning more about the potential of regression analysis, for example, I realized it was a great fit for the project. I will now be willing to listen to input of others, even if I thought I already had a sufficient development plan in place.

Overall, I would state that project development is a highly iterative process with no room for complacency. The more a developer is willing to devote to the development

of a project the more that developer will get out of the final product. Resources exist to be used and frugality with any part of the process may unintentionally hamper the best efforts of a researcher. I will take these lessons with me as I move into the next phase of my career as a scholarly practitioner.

Importance of the Work

In a democracy reading is not reserved for the privileged, it is a fundamental right of all. At the heart of any democracy is the will of the people. Democracies only function properly when the will of all people is represented. Citizens can be under-represented when they do not have the same access to advance as others do, and can be left out of the democratic process. One tenant in the American Dream is that hard work will allow for economic mobility. In places where citizens do not have access to equal resources from public institutions, often the hardest workers are relegated to low paying jobs and become a permanent underclass. Education has been the mobility ladder for generations of Americans, yet in underperforming schools this too becomes just another reminder of how different its community members are.

The philosophical and social underpinning of this work is to provide democracy as it was truly intended, and so clearly explained by Dewey, as only being possible when institutions treat all members equally. By its very nature focusing on a single group of individuals would exclude all other groups. Yet, focusing on a single group because they can be brought up to where others already are is altruistic. Democracies flourish when the less fortunate are made fortunate by those with the means to do so.

A school system that is not serving all members of the community is failing at this standard. A school system that only is able to help those who need very little while the very needy devolve is failing even more. The large proportion of Hispanics populating the evaluation study school are generally broken into two groups, those that were born speaking English and Spanish and those that learned English only after years of speaking Spanish exclusively. The latter are at a great disadvantage when moving into literature studies focused primarily on connotative meaning, which is predominantly where secondary education classrooms focus lessons. Identifying a working intervention for this group would be a boon to their ability to be equal citizens in a democracy. The READ 180 program did this for those children as this evaluation study proved quantitatively, yet due to a lack of program evaluation it was taken away. Now at the conclusion of this study, even if the prior program is not brought back, at the very least new programs can be properly vetted for efficacy based on this model of program evaluation. In doing so future programs, for future populations in need, will be able to retain interventions that work and discard those that do not with more certainty. Thusly, helping the institution meet the demands of democracy.

Schools are one of the last bastions of equality and fairness left to our society. Within the walls of a school every student is special, every question asked worthy, and each is the master of a boundless future. To not fully support equality of resources for students in our schools would demonstrate a lack of caring about the most basic of our freedoms, the freedom to learn. And with that, the freedom to grow. A school with transparent walls is a school that can be trusted by its community. When community

members are informed and involved, schools and students succeed. The path to that success is a widely implemented, systemic, and open to the public program evaluation system. No student should feel they are locked into a life of few choices and must accept a low standard of living. Americans can take back the greatest ladder to upward mobility ever invented, a quality education.

Implications, Applications, and Directions for Future Research

Using data to make decisions is at the core of this evaluation study. I hope the implications would first be to realize decisions need to be made based on specific correlative data. Simply looking at the entire school HSPA or graduation rate will not tell if a program servicing a small percentage of the population has had any effect, as was shown in this evaluation study. The group must be analyzed after being statistically removed from the general sample population. Even after that, there may be a sub-group that benefits or is hindered to a greater degree than the whole of that group. Using descriptive statistics can enlighten a researcher to this situation as it did with my study and the English language learner sub-group population.

Applications of this study will be that project evaluations can become a regular part of how intervention programs are evaluated before decisions to continue or discontinue them are made. Applying this model of evaluation is a good start to creating a proper learning community. Even if people in administrative positions continue to come and go, the community will remain. The community will be responsible for shaping the education of its members and that will the greatest lasting application this study could ever have achieved.

Future research should entail longitudinal grouping to more accurately track trends and effects of what works with literacy development alongside brain development. Too often we address this issue after many years of schooling have already passed and students are already in high school. By identifying, in a much earlier way, how the children in this community learn both of their languages while studying only one academically, will create new paths and interventions. Initial intervention groups should be small, but they should not remain small. Future research could benefit from homogeneity in grouping to bolster statistical analyses, unlike the issues I ran across with this study. The very lifeblood of evaluation is a vast and accurate data pool. Time and again this was a detriment to this study as only one full cohort could be fully accounted for on what should be three very common standards of data for a school to maintain. As schools move forward into an age of digitization they should also look to the marvelous potential that lays beneath their fingertips if only they could maintain a constant vigil over data. There is no limit to what future researchers could produce if the required reports to the state were made easy to access and complete in their validity.

References

ACT, Inc. (2013). The condition of college & career readiness 2010. Retrieved from http://www.act.org/research/policymakers/cccr10/page_8.html

American Diploma Project. (2014). Ready or not: Creating a high school diploma that counts. *Achieve.org* Retrieved from http://www.achieve.org/files/ADPsummary_5.pdf

Applebee, A. N., Langer, J. A., Nystrand, M., & Gamoran, A. (2003). Discussion-based approaches to developing understanding: Classroom instruction and student performance in middle and high school English. *American Educational Research Journal, 40*(3), 685-730. Retrieved from

Au, K. H. (2009). Providing powerful comprehension instruction. *Reading Today, 27*(2), 17.

Bailey, T. (2009). Challenge and opportunity: Rethinking the role and function of developmental education in community college. *New Directions For Community Colleges, 2009*(145), 11-30. doi:10.1002/cc.352

Barth, R. S. (2001). Teacher leader. *Phi Delta Kappan, 82*(6), 443-449.

Basaran, M. (2013). Reading fluency as an Indicator of reading comprehension. *Educational Sciences: Theory & Practice, 13*(4), 2287-2290. doi:10.12738/estp.2013.4.1922

Beatty, B. (2007). Going through the emotions: leadership that gets to the heart of school renewal. *Australian Journal of Education, 51*(3), 328-340.

Bell, K., & Dolainski, S., (2005). What is evidence based reading instruction and how do

you know it when you see it? Retrieved from Literacy Information and Communication System https://lincs.ed.gov/publications/pdf/EDVAE09C0042EBRILAUSD.pdf

National Archives and Records Administration. (2013). Bill of Rights: The charters of freedom. Retrieved from http://www.archives.gov/exhibits/charters/bill_of_rights.html

Banks, M. A. (2006). Towards a continuum of scholarship: the eventual collapse of the distinction between grey and non-grey literature. *Publishing Research Quarterly*, *22*(1), 4-11.

Bishop-Kallmeyer, N. (2008). *A critical examination of the READ 180 program among a sample of english as a second language students and special education students.* (Doctoral dissertation). Retrieved from ProQuest Dissertations and Theses. (304558021)

Boltzmann, M., Rüsseler, J., Zheng, Y., & Münte, T. F. (2013). Learning to read in adulthood: An evaluation of a literacy program for functionally illiterate adults in Germany. *Problems Of Education In The 21St Century*, *51*, 33-46.

Brewer, C. (2011). School leaders as political strategists: William Boyd's contributions to our understanding of the politics of leadership. *Peabody Journal of Education*, *86*(4), 450-463.

Carnegie Corporation of New York. (2011). Time to act: An agenda for advancing adolescent literacy for college and career success. Retrieved from Carnegie Corporation of New York

http://carnegie.org/fileadmin/Media/Publications/PDF/tta_Main.pdf

Carnmarota, J. (2009). The cultural organizing of formal praxis-based pedagogies: a socio-historical approach to participatory action research. *Social Justice, 36*(4), 6-13.

Carmichael, M. (2012, September 20). Many are unprepared for college, study says. *Boston Globe*. Retrieved from http://search.proquest.com/docview/1041121025?accountid=14872

Carter, Bronwyn J, (2010). Evidence-based decision-making: Practical issues in the appraisal of evidence to inform policy and practice. *Australian Health Review, 34*(4), 435-40.

Chance, P. L. & Chance, E.W. (2002). Organizational structure: Fundamental constructs that define schools. In *Introduction to educational leadership and organizational behavior: Theory into practice* (pp. 11-33). Larchmont, NY: Eye on Education.

Childress, E. & Jul, E. (2003) Going grey. *Journal of Internet Cataloging, 6*(3), 3-6.

Coulter, D. (2001). Teaching as communicative action: Habermas and education In V. Richardson (Ed.), *Handbook of research on teaching* (4th ed., pp. 90-98). Washington, DC: American Education Research Association

Crawford, A. (2010, March 14). Pennsylvania's college drop-out rate traced to high schools. *Tribune - Review / Pittsburgh Tribune - Review.* Retrieved from

Creswell, J.W. (2012). *Educational research: Planning, conducting, and evaluating quantitative and qualitative research* (4th ed.). Boston, MA: Pearson Education

Crouch, Simon R., (2012). Opportunistic evidence: Evidence-based policy in the setting

of the australian government's chlamydia screening pilot. *Australian Health Review, 36*(1), 57-60. Retrieved from

Denson, K. (2008). Passport reading journeys™ effectiveness with ninth grade students identified for reading improvement instruction in an urban high school. Retrieved from http://www.voyagerlearning.com

Dewey, J. (2013). *Democracy and education* Retrieved from http://www.gutenberg.org/ebooks/852

Dole, J. A., Duffy, G. G., Roehler, L. R., & Pearson, P. D. (1991). Moving from the old to the new: Research on reading comprehension instruction. *Review of Educational Research, 61*(2), 239-264. Retrieved from

Duff, F., Hayiou-Thomas, M., & Hulme, C. (2012). Evaluating the effectiveness of a phonologically based reading intervention for struggling readers with varying language profiles. *Reading & Writing, 25*(3), 621-640. doi:10.1007/s11145- 010-9291-6

Duke, N. K., & Pearson, P. (2008). Effective practices for developing reading comprehension. *Journal Of Education, 189*(1/2), 107-122.

Education Law Center. (2013) The history of Abbot v. Burke. Retrieved from http://www.edlawcenter.org/cases/abbott-v-burke/abbott-history.html

Epstein, J. (2008). Improving Family and Community Involvement in Secondary Schools. *Education Digest, 73*(6), 9-12.

Fitzpatrick, J.L., Sanders, J.R., & Worthen, B.R., (2012). *Program evaluation: Alternative approaches and practical guidelines* (4th ed.). Upper Saddle River,

NJ: Pearson Education, Inc.

Flitcroft, K., Gillespie, J., Salkeld, G., Carter, S., & Trevena, L. (2011). Getting evidence into policy: The need for deliberative strategies?. *Social Science & Medicine, 72*(7), 1039-1046.

Forbes, J.B., Piercy, J.E., (1991). *Corporate mobility and paths to the top: Studies for human resource management development specialists* Retrieved from http://books.google.com/books?id=vyRZvzF64FUC&printsec=frontcover#v=onepage&q&f=false

Frater, J., Myohanen, L., Taylor, E., & Keith, L. (2007). What would you tell me if I said grey literature? The New York Academy of Medicine Grey Literature report. *Journal of Electronic Resources in Medical Libraries, 4*(1-2), 145-153.

Frattura , E., Capper, C. A., (2006). Segregated programs versus integrated comprehensive service delivery for all learners: assessing the differences *Remedial and Special Education, 27,* 355-364. doi:10.1177/07419325060270060501

Frey, B. B., Lee, S. W., Tollefson, N., Pass, L., & Massengill, D. (2005). Balanced literacy in an urban school district. *Journal Of Educational Research, 98*(5), 272-280.

Gruenbaum, E. A. (2012). Common Literacy Struggles with College Students: Using the Reciprocal Teaching Technique. *Journal Of College Reading And Learning, 42*(2), 110-116.

Happe, L. E., & Walker, D. (2013). Using Grey Literature to Prepare Pharmacy Students

for an Evolving Healthcare Delivery System. *American Journal Of Pharmaceutical Education, 77*(4), 1-7.

Hargrave, A. C., & Sénéchal, M. (2000). A book reading intervention with preschool children who have limited vocabularies: The benefits of regular reading and dialogic reading. *Early Childhood Research Quarterly, 15*(1), 75-90.

Hasselbring, T. S., & Goin, L. I. (2004). Literacy instruction for older struggling readers: What is the role of technology?. *Reading & Writing Quarterly, 20*(2), 123-144

Hausheer, R., Hansen, A., & Doumas, D. M. (2011). Improving Reading Fluency and Comprehension Among Elementary Students: Evaluation of a School Remedial Reading Program. *Journal of School Counseling*, *9*1-20.

Head, B. W. (2008). Three lenses of Evidence-Based policy. *Australian Journal of Public Administration*, *67*(1), 1-11.

Head, B. W. (2010). Reconsidering evidence-based policy: Key issues and challenges. *Policy and Society, 29*(2), 77-94.

Henke, R. (2001). A framework for understanding cross-national differences in the relationship between research and policy. *Journal of International Migration and Integration, 2*(4), 581-611. doi:10.1007/s12134-001-1013-0

Hicks, T., Jacobs, A. M., and Matthews, J. S. (2013). "Are the rich better off than they were four years ago? class-biased economic voting in comparative perspective."

Holdaway, D. (1979). *The Foundations of Literacy.* Toronto, Canada: Ashton Scholastic

Holdaway, D. (1982). Shared book experience: teaching reading using favorite books. *Theory Into Practice, 21*(4), 293.

Hughes, R. G., & Titler, M. G. (2008). The evidence for evidence-based practice implementation. In: Hughes RG, editor. Patient safety and quality: an evidence-based handbook for nurses. Rockville (MD): Agency for Healthcare Research and Quality (US); 2008 Apr. Chapter 7.

Hunsberger, P. (2007). "Where am I?" A call for "connectedness" in literacy. *Reading Research Quarterly, 42*(3), 420-424.

Hussein, A. (2009). The use of triangulation in social sciences research: Can qualitative and quantitative methods be combined. *Journal of Comparative Social Work, 1*(8), 1-12.

Imbens, G. M., & Wooldridge, J. M. (2008). *Recent developments in the econometrics of program evaluation* (No. w14251). National Bureau of Economic Research.

Jackson, J. (2005). Leadership for Urban Public Schools. Essays. *The Educational Forum, v69*(n2), p192.

Jones, C. (2010). Archival data: Advantages and disadvantages for research in psychology. *Social and Personality Psychology Compass*, *4*(11), 1008-1017.

Juricek, J. (2009). Access to Grey Literature in Business: An Exploration of Commercial White Papers. *Journal Of Business & Finance Librarianship, 14*(4), 318-332.

Kessinger, T. A. (2011). Efforts toward education reform in the united states since 1958. *American Educational History Journal, 38*(1/2), 263-276.

Kim, J., Samson, J., Fitzgerald, R., & Hartry, A. (2010). A randomized experiment of a balanced literacy literacy intervention for struggling readers in grades 4–6: effects on word reading efficiency, reading comprehension and vocabulary, and oral

reading fluency. *Reading & Writing*, *23*(9), 1109-1129. doi:10.1007/s11145-009-9198-2

Kim, J.S., Capotosto, L., Hartry, A., & Fitzgerald, R. (2011). Can a balanced literacy literacy intervention improve the reading achievement of low-performing elementary school students in an after-school program?: results from a randomized controlled trial of read 180 enterprise *Educational Evaluation and Policy Analysis*, 33 (2) , pp. 183-201. doi:10.3102/0162373711399148

Kostewicz, D. E., Lemons, C. J., Mrachko, A. A., & Paterra, M. F. (2012). Effectiveness of decoding and phonological awareness interventions for children with down syndrome. *Exceptional Children*, *79*(1), 67-90.

Lawrence, A. (2012). Electronic documents in a print world: grey literature and the internet. *Media International Australia, (143),* 122–131. Retrieved from http://www.uq.edu.au/mia/2012-issues#143

Lawrence, S. A., McNeal, K., & Yildiz, M. N. (2009). Summer program helps adolescents merge technology, popular culture, reading, and writing for academic purposes. *Journal Of Adolescent & Adult Literacy*, *52*(6), 483-494.

Liu, F. (2010). A short analysis of the nature of reading. *Journal English Language Teaching*, *3*(3), p152.

Lodico, M.G., Spaulding, D.T., Voegtle, K.H. (2010). *Methods in educational research: From theory to practice* (Vol. 28). John Wiley & Sons, Inc.

Metcalfe, S. A., Aitken, M., & Gaff, C. L. (2008). The importance of program evaluation: How can it be applied to diverse genetics education settings? *Journal of Genetic*

Counseling, *17*(2), 170-9. doi:10.1007/s10897-007-9138-8

McWhorter, H. (2009). *Facilitating high school student success through READ 180: Analysis of program impact using measures of academic progress (MAP).* (Order No. 3379873, Walden University). *ProQuest Dissertations and Theses*, 116-n/a. Retrieved from http://search.proquest.com/docview/305068939?accountid=14872. (305068939).

Melekoglu M., (2011). Impact of motivation to read on reading gains for struggling readers with and without learning disabilities *Learning Disability Quarterly, November 2011;vol. 34: pp. 248-261*

Metametrics, Inc. (2004). Lexiles: a system for measuring reader ability and text difficulty, a guide for educators Retrieved from http://teacher.scholastic.com/products/sri_reading_assessment/pdfs/SRI_LexileProfesionalPaper.pdf

Meyer, R. J. (2013). The truth behind manufactured malpractice: The impacts of NCLB upon literacy teaching and learning. *New England Reading Association Journal, 49*(1), 1-6.

Mohamadi, Z. (2013). Program Evaluation on General English Course: A Case Study at Tabriz University. *Journal Of Language Teaching & Research, 4*(6), 1285-1297. doi:10.4304/jltr.4.6.1285-1297

Morrow, L. M., Tracey, D. H., (2012) *Lenses on reading: An introduction to theories and models* New York, NY: Guilford Press

National Center for Education Statistics. (2013) NAEP main nde data explorer Retrieved

from http://nces.ed.gov/nationsreportcard/naepdata/report.aspx

Nave, J. (2007). "An assessment of *READ 180* regarding its association with the academic achievement of at-risk students in sevier county schools." *Electronic Theses and Dissertations.* Paper 2107. Retrieved from: http://dc.etsu.edu/etd/2107

Nelson, T. (2008). *Predictive factors in student gains in reading comprehension using a reading intervention program.* (Order No. 3318825, University of South Dakota). *ProQuest Dissertations and Theses,* , 157-n/a. Retrieved from http://search.proquest.com/docview/304460113?accountid=14872. (304460113).

Odom, S. L. (2009). The tie that binds. *Topics in Early Childhood Special Education, 29*(1), 53-61. doi:http://dx.doi.org/10.1177/0271121408329171

Okoroma, F. (2011) Towards effective management of grey literature for higher education, research and national development, *Library Review*, Vol. 60 Iss: 9, pp.789 - 802

Olson, C. A. (2013). Using the Grey Literature to Enhance Research and Practice in Continuing Education for Health Professionals. *Journal of Continuing Education in the Health Professions.* pp. 1-3. doi:10.1002/chp.21159.

Olson, L. (2006). Skills for work, college readiness are found comparable. *Education Week, 25*(36), 1-1,19. Retrieved from http://search.proquest.com/docview/202714535?accountid=14872

Onosko, J. (2011). Race to the top leaves children and future citizens behind. *Democracy & Education, 19*(2), 1-11.

Papalewis, R. (2004). Struggling middle school readers: Successful, accelerating

intervention. *Reading Improvement, 41*(1), 24-37. Retrieved from http://search.proquest.com/docview/215797754?accountid=14872

Pappas, C., & Williams, I. (2011). Grey literature: its emerging importance. *Journal of Hospital Librarianship, 11*(3), 228-234.

Parker, C. A., Holland, G., & Jones, D. (2013). The Effectiveness of Two Reading Intervention Programs in a South Texas Urban School District. *National Forum of Applied Education Research Journal, vol. 26* (3) Retrieved from http://www.nationalforum.com/Journals/National%20Forum%20of%20Applied%20Educational

Parker, C. (2011). *The effectiveness of READ 180 and journeys III in a south texas urban school district.* (Order No. 3534090, Texas A&M University - Kingsville). *ProQuest Dissertations and Theses,* , 80. Retrieved from http://search.proquest.com/docview/1238003107?accountid=14872. (1238003107).

Phillips, D. C. (1995). The good, the bad, and the ugly: The many faces of constructivism. *Educational researcher,* 5-12.

Pressley, M., Roehrig, A., Bogner, K., Raphael, L. M., & Dolezal, S. (2002). Balanced literacy instruction. *Focus on Exceptional Children, 34*(5), 1. Retrieved from http://search.proquest.com/docview/224050218?accountid=14872

READ 180:adolescent literacy. (2009). *What Works Clearinghouse* U.S. Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance retrieved from

http://ies.ed.gov/ncee/wwc/pdf/intervention_reports/wwc_read180_102009.pdf

Reigle, B. S., Stevens, K. R., Belcher, J. V., Huth, M. M., McGuire, E., Mals, D., & Volz, T. (2008). Evidence-based practice and the road to magnet status. *Journal of Nursing Administration*, *38*(2), 97-102.

Robby, M. A. (2008). *Evaluation of the academic effectiveness of the READ 180 program: Educational software intervention in reading for at risk high school students in riverside county, california.* (Order No. 3323282, Argosy University/Orange County). *ProQuest Dissertations and Theses*, 194-n/a. Retrieved from http://search.proquest.com/docview/304825309?accountid=14872. (304825309).

Roth, L. (2010, Nov 27). Many not prepared to cut it in college. *Orlando Sentinel*. Retrieved from http://search.proquest.com/docview/814423735?accountid=14872

Rumelhart, D. E. (2004). Toward an interactive model of reading. In Robert B. Ruddell et al. (Eds.). *Theoretical Models And Processes of Reading*, (Fifth Edition): International Reading Association, Newark, DE.

Ryder, J., Tunmer, W., & Greaney, K. (2008). Explicit instruction in phonemic awareness and phonemically based decoding skills as an intervention strategy for struggling readers in whole language classrooms. *Reading & Writing*, *21*(4), 349-369. doi:10.1007/s11145-007-9080-z

Salinger, T., Moorthy, S., Toplitz, M., Jones, W., & Rosenthal, E. (2010). *Implementation matters: Systems for success. A descriptive study of READ 180 in urban middle schools.* Washington, DC: American Institutes for Research.

Samuels, S. J., & Flor, R. F. (1997). The importance of automaticity for developing expertise in reading. *Reading & Writing Quarterly: Overcoming Learning Difficulties, 13*(2), 107-121.

Savage, R., Piquette, N., Deleveaux, G., Abrami, P. C., Wood, E., Sanghera-Sidhu, S., & Burgos, G. (2013). A (Pan-Canadian) Cluster Randomized Control Effectiveness Trial of the ABRACADABRA Web-Based Literacy Program. *Journal Of Educational Psychology, 105*(2), 310-328. doi:10.1037/a0031025

Schacter, J. (1999). Reading programs that work: A review of programs for pre-kindergarten to 4th grade. *Santa Monica, CA: Milken Family Foundation, jschacter@mff.org*.

Scholastic (2007). Scholastic reading inventory technical guide Retrieved from http://teacher.scholastic.com/products/sri_reading_assessment/pdfs/SRI_TechGuide.pdf

Scholastic (2007a). Scholastic reading inventory & scholastic achievement manager training model Retrieved from http://www.scholastic.com/dodea/pdfs/SRI_SAM_handouts.pdf

Scholastic (2011). *2011 Compendium of READ 180 research.* Retrieved from http://read180.scholastic.com/pdf/research/1_2011_EfficacyStudy_ResearchCompendium2011_READ180.pdf

Scholastic (2013). *About READ 180 Next Generation*. Retrieved from http://read180.scholastic.com/reading-intervention-program/about

Scholastic (2013a). *Scholastic Education: READ 180 Next Generation.* Retrieved from

http://education.scholastic.ca/category/READ_180

Schlager, K,. Staab, A., (2012) Education 2012: New Jersey's Top High Schools *New Jersey Monthly* Retrieved from http://njmonthly.com/articles/towns_and_schools/top-new-jersey-high-schools-2012-.html

Shaw, T. M. (2002). *Innovation and social learning: institutional adaptation in an era of technological change*. M. S. Gertler, & D. A. Wolfe (Eds.). New York: Palgrave Macmillan.

Shaw, M., & Davidson, M. (2009). Using the Phono-Graphix reading program as a literacy support intervention strategy. *Support For Learning*, *24*(1), 42-48. doi:10.1111/j.1467-9604.2009.01396.x

Shirey, M. R., Hauck, S. L., Embree, J. L., Kinner, T. J., Schaar, G. L., Phillips, L. A., . . . McCool, I. A. (2011). Showcasing differences between quality improvement, evidence-based practice, and research. *Journal of continuing education in nursing*, *42*(2), 57.

Shultz, K. S., Hofman, C. C, & Reiter-Palmon, R. (2005). Using archival data for I/O research: Advantages, pitfalls, sources, and examples. *The Industrial/Organizational Psychologist, 42(3),* 31-37.

Sigears, K. A. (2008). *The impact of the implementation of the scholastic read 180 model on reading skills development of middle school students with learning disabilities as compared to those using the traditional resource reading model.* (Order No. 3320196, Tennessee State University). *ProQuest Dissertations and Theses,* 114-

n/a. Retrieved from http://search.proquest.com/docview/89257262?accountid=14872. (89257262).

Sladky, L., (2010) One-third of students need remedial college math, reading *USA TODAY* retrieved from http://usatoday30.usatoday.com/news/education/2010-05-11-remedial-college_N.htm

Slavin, R. E., Cheung, A., Groff, C., & Lake, C. (2008). Effective reading programs for middle and high schools: A best-evidence synthesis. *Reading Research Quarterly, 43*(3), 290-322. Retrieved from http://search.proquest.com/docview/212134541?accountid=14872

Slavin, R. E., Lake, C., Chambers, B., Cheung, A., & Davis, S. (2009). Effective reading programs for the elementary grades: A best-evidence synthesis. *Review of Educational Research, 79*(4), 1391-1466. Retrieved from http://search.proquest.com/docview/214120445?accountid=14872

Smith, J. (2009). The evidence base: where is it? *The Australian Library Journal, 58*:1, 28-38, DOI:10.1080/00049670.2009.10735833

Spaulding, D.T, (2008). *Program evaluation in practice: Core concepts and examples for discussion and analysis* San Francisco, CA: Jossey-Bass

Stanovich, K. E., West, R. F., & Feeman, D. J. (1981). A longitudinal study of sentence context effects in second-grade children: Tests of an interactive-compensatory model. *Journal of Experimental Child Psychology, 32*(2), 185-199.

State of New Jersey. (2012). New Jersey leads the way with call for statewide accountability system to turn around failing schools. Retrieved from

http://www.state.nj.us/governor/news/news/552012/approved/20120209b.html

State of New Jersey Department of Education. (2006). Your guide to the hspa Retrieved from http://www.nj.gov/education/assessment/hs/hspa_guide_english.pdf

State of New Jersey Department of Education. (2007). Assessment: new jersey assessment of skills and knowledge. Retrieved from http://www.state.nj.us/education/assessment/ms/gepa_technical_report.pdf

State of New Jersey Department of Education. (2008). Middle school statewide assessments Retrieved from www.state.nj.us/education/assessment/ms/5-8/tech/2008TechReport.pdf

State of New Jersey Department of Education. (2010a). 2010-2011 School Report Card retrieved from http://education.state.nj.us/rc/rc11/rcreport.php?c=23;d=4090;s=050

State of New Jersey Department of Education. (2010b). Bureau of bilingual education Retrieved from http://www.state.nj.us/education/bilingual/

State of New Jersey Department of Education. (2010c). High school proficiency assessment additional information retrieved from http://www.state.nj.us/education/assessment/hs/hspa/info/

State of New Jersey Department of Education. (2012). NJ school performance report Retrieved from http://www.state.nj.us/education/pr/2013/23/234090050.pdf

State of New Jersey Department of Education. (2013a). Final list of priority and focus schools retrieved from http://www.state.nj.us/education/reform/PFRschools/PriorityFocusSchools.pdf

State of New Jersey Department of Education. (2013b). NJ school performance report retrieved from http://www.state.nj.us/education/pr/1213/23/234090050.pdf

State of New Jersey Department of Education. (2014a). Educator evaluation. Retrieved from http://www.nj.gov/education/genfo/faq/faq_eval.shtml#_AchieveNJ_for_Teachers_1

State of New Jersey Department of Education. (2014b). NJ school performance reports interpretive guide. Retrieved from http://education.state.nj.us/pr/Interpretive%20Guide%202014.pdf

State of New Jersey Department of Education. (2014c). Taxpayers Guide to Education Spending 2013. Retrieved from http://www.state.nj.us/cgi-bin/education/csg/13/csg.pl?string=dist_code4090&maxhits=650

Sterzik, A. M., & Fraser, C. (2012). RC-MAPS: Bridging the comprehension gap in EAP reading. *TESL Canada Journal*, *29*(2), 103.

Stenner, A. J., Smith, D. R., Horabin, I., & Smith, M. (1987). Fit of the Lexile Theory to Sequenced Units from Eleven Basal Series. MetaMetrics, Inc. Retrieved from http://www.lexile.com/research/25/

Stevens, K. (2013). The impact of evidence-based practice in nursing and the next big ideas. *OJIN: The Online Journal of Issues in Nursing*, *18*(2).

Stickley, T. T., & Wright, N. N. (2011). The British research evidence for recovery, papers published between 2006 and 2009 (inclusive). Part Two: a review of the grey literature including book chapters and policy documents. Journal Of

Psychiatric & Mental Health Nursing, 18(4), 297-307. doi:10.1111/j.1365-2850.2010.01663.x

Teale, W. H., Paciga, K. A., & Hoffman, J. L. (2007). Beginning Reading Instruction in Urban Schools: The Curriculum Gap Ensures a Continuing Achievement Gap. *Reading Teacher, 61*(4), 344-348. doi:10.1598/RT.61.4.8

Trachtman, R. (2007, March). Inquiry and Accountability in Professional Development Schools. *Journal of Educational Research*, *100*(4), 197-203.

Triola, M. F. (2012). *Elementary statistics technology update.* (11th ed.). Boston, MA: Pearson Education, Inc.

U. S. Census Bureau. (2013). State and county quickfacts retrieved from http://quickfacts.census.gov/qfd/states/34/3458200.html

U.S. Department of Education. (1993). Understanding evaluation: the way to better prevention programs Retrieved from http://www2.ed.gov/PDFDocs/handbook.pdf

U.S. Department of Education. (2013). Statement from U.S. secretary of education secretary duncan on house ESEA reauthorization bill H.R. 5 Retrieved from http://www.ed.gov/news/press-releases/statement-us-secretary-education-secretary-duncan-house-esea-reauthorization-bil

U.S. Department of Labor. (2013a). Employment projections Retrieved from http://www.bls.gov/emp/ep_chart_001.htm

U.S. Department of Labor. (2013b). The 30 occupations with the largest projected employment declines, 2010-20 Retrieved from http://www.bls.gov/news.release/ecopro.t08.htm

U.S. Department of Labor. (2013c). The 30 occupations with the largest projected employment growth, 2010-20 Retrieved from http://www.bls.gov/news.release/ecopro.t06.htm

Uzuner, Y., Girgin, Ü., Kaya, Z., Karasu, G., Girgin, M., Erdiken, B., & Tanridiler, A. (2011). An examination of balanced literacy instructional model implemented to youths with hearing loss. *Educational Sciences: Theory & Practice, 11*(4), 2126-2133.

Van Cauwenberghe, E., Maes, L., Spittaels, H., van Lenthe, F. J., Brug, J., Oppert, J. M., & De Bourdeaudhuij, I. (2010). Effectiveness of school-based interventions in Europe to promote healthy nutrition in children and adolescents: systematic review of published and 'grey'literature. *British journal of nutrition,103*(06), 781-797.

Vaughn, S., Wexler, J., Leroux, A., Roberts, G., Denton, C., Barth, A., Fletcher, J., (2012). Effects of intensive reading intervention for eighth- grade students with persistently inadequate response to intervention *The Journal of Learning Disabilities, Nov 2012; vol.45: pp. 515-525*

Vogel, J. T. (2013). *A case study on the impact of the READ 180 reading intervention program on affective and cognitive reading skills for at-risk secondary level students* (Doctoral dissertation, Liberty University). Retrieved from http://digitalcommons.liberty.edu/doctoral/654/

Weimbaum et al., (2004) Foundations for inquiry: Reviewing the research. In *Teaching as inquiry*: *Asking hard questions to improve practice and student achievement*

(pp. 13-30). New York: Teachers College Press; Oxford, OH: National Staff Development Council. Reprinted by permission of the publisher. 2004 Teachers College, Columbia University.

Whitehurst, G. J., Arnold, D. S., Epstein, J. N., Angell, A. L., Smith, M., & Fischel, J. E. (1994). A picture book reading intervention in day care and home for children from low-income families. *Developmental Psychology, 30*(5), 679-689. doi:10.1037/0012-1649.30.5.679

Whitehurst, G. J., Zevenbergen, A. A., Crone, D. A., Schultz, M. D., Velting, O. N., & Fischel, J. E. (1999). Outcomes of an emergent literacy intervention from Head Start through second grade. *Journal Of Educational Psychology, 91*(2), 261-272. doi:10.1037/0022-0663.91.2.261

Willerton, R. (2013). Teaching white Papers through client projects. *Business Communication Quarterly, 76*(1), 105-113. doi:10.1177/1080569912454713

Wonder-McDowell, C. (2010). The hidden peril of differentiation: fragmented instruction. *College Reading Association Yearbook*, (31), 45-59.

Wong, K. K., & Langevin, W. E. (2005). Reconstructing local governance in american public education: politics, policy, and process. *Conference Papers -- Southern Political Science Association*, 1-20.

Wu, C. H. V. (2009). *Adolescent english language learners in the classroom: students' perceptions of using READ180* (Doctoral dissertation, University of Florida). Retrieved from: http://etd.fcla.edu/UF/UFE0025001/wu_c.pdf

Appendix: White Paper

An Evaluation of the READ 180 Program:
A White Paper Submitted to the XXXXX School District

Submitted by

Daniel Lombardi

Date of Report:
April 2015

Executive Summary

This paper addresses the problem of low reading achievement in the XXXXX School District. It reports results of an evaluation of READ 180, a balanced literacy intervention that was used at the high school until 2012. For the only year in which complete data is available, it was found that the balanced literacy program was highly successful for English language learners (ELL's). It is therefore recommended that the district reconsider use of balanced reading interventions for at-risk students. Further, it is recommended that the district consider a more formal policy of ongoing evaluation of literacy programs.

Need for the Study

XXXXX School District has long striven to provide academic programs to meet the needs of its student population. The challenges faced in a low-income, urban school district are not unique to XXXXX; nonetheless, curricular decision-making remains the prerogative of the individual school district, within the parameters established by the State Department of Education and regulations governing federal programs such as Title 1. As a long-time reading intervention teacher, the author is particularly concerned with helping our high school students attain the reading and literacy abilities necessary for postsecondary and workplace success. To better understand the effectiveness of our reading interventions, I chose to perform, as my evaluation study in graduate school, a program evaluation of the READ 180 literacy program. This program was used for seven years at the XXXXX High School to give an equal chance to students who struggled with literacy. After seven years this program was stopped, and now three years since its

cessation I have had an opportunity to gather program data to analyze how effective it was. During this process I also wished to understand if a form of continual program evaluation could be used in the future to help the district decide whether interventions should continue or not. Therefore, I performed a goals-based program evaluation of the now-defunct READ 180 intervention program based on its three explicit goals: 1) improving reading achievement at the classroom level 2) improving scores on state language arts tests and 3) increasing graduation rates.

The READ 180 reading intervention program is a balanced literacy model that focuses on reading through student-centered holistic reading experiences, but also develops basic reading skills through teacher-directed instruction. In a typical class session, a student would move through three stations involving a computer program with video and text, individual reading of student-selected text, and group reading of current event articles (see Figure A1). It differs from programs such as Reading Horizon in that it focuses on both fundamental reading skills such as phonics as well as next-level reading skills such as interpreting main ideas and inference. Unlike other programs, it incorporates equal time for group work, individual work, and computer based learning. Maintaining this balance while allowing students to grow holistically separates READ 180 from many other intervention programs.

Figure A1. READ 180 class cycle.

Over the course of seven years, the program was used at different times for freshman or sophomores. The majority of data on the program was put together in-house by various coordinators. To determine if these three goals were achieved I gathered all available data necessary to perform several robust and advanced statistical analyses.

The first stated goal of the READ 180 program was to increase reading achievement at the classroom level. To analyze this I gathered student scores from a test given within the program itself called the Scholastic Reading Inventory (SRI). Questions from this test are given to students through class computers, and they increase or decrease in difficulty to match the student's answers until the proper student achievement level is found. An example question is displayed below (see Figure A2).

At Lunch, Gilberto was hoping he'd see Maddie. He wanted to tell her his good news and also see how she was doing. But the lunch period passed, and he saw no sign of her. Gilberto decided that after school he would job over to her house and check on her.

He was _____for her.

- Cooking
- Studying
- Reading
- Looking

Figure A2. SRI sample question.

Results of the Study

It was possible to obtain data for only two cohorts of students who had gone through the program (2007-2008 and 2010-2011). A cohort comprises those students who began the 9th grade and were tracked until senior year. Both cohorts showed an increase in SRI scores from the beginning (September) to the end (February) of the intervention. The 2010 cohort had enough increase to be statistically significant and rule out the possibility of the scores having gone by up chance. In other words, there is a 95% confidence level that SRI reading scores were increased after participation in the READ 180 intervention (see Figure A3 below).

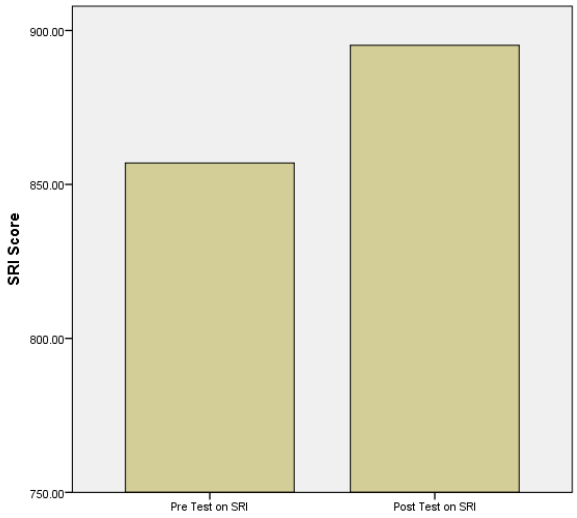

Figure A3. 2011 SRI score increase.

The second goal of the program was to improve scores on state language arts tests. It was only possible to obtain enough data to perform this test on the 2010 cohort. A comparison of average scores of each student demonstrated interesting results for regular education students. Of those students, the NJASK 8 score of READ 180 students (prior to entering the program) routinely scored below the passing range (200) of the test. After the intervention those same students scored above the passing range of the HSPA (also 200). Students not within the READ 180 program also grew but not as much over the same period of time between the two tests. This result was statistically sound to the point of 99.99% confidence (see Figure A4 below).

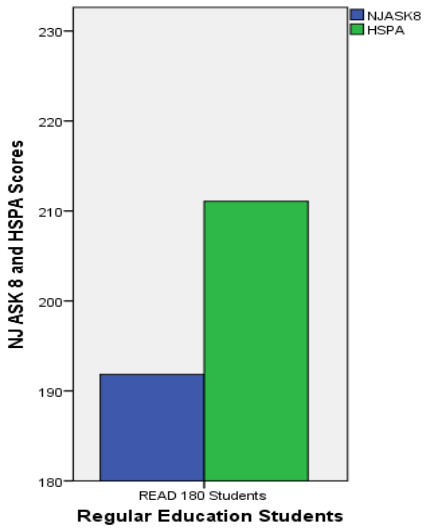

Figure A4. NJ ASK 8 to HSPA student growth.

Looking further into the data by subdividing the students by language proficiency, it became apparent that English language learners (ELL's) were most profoundly affected by participation in READ 180. English language learner students within the program went from a failing state test score of 193.50 to a passing score of 225.10, while English language learner students not in the program went from a passing score of 204.68 to a failing score of 198.42 (see Table A).

Table A

English Language Learner Growth

	ELL Students	Average score	Score Change Percentage
NJASK 8 Test	READ 180	193.50	
	Not READ 180	204.68	
HSPA Test	READ 180	225.10	READ 180 Students = +16%
	Not READ 180	198.42	Non READ 180 Students = -3%

It is clear that English language learner students both benefitted the most from being in the READ 180 program and suffered the most for not being in it. It is at present unknown, but might be important to learn, if HSPA scores for English language learner students have similarly dropped in the past three years.

To answer the question of whether or not participation in READ 180 affected graduation rates was difficult due to lack of archival data. A statistical test returned little to no association. On the other hand, deduction would suggest that if READ 180 participants' HSPA scores were higher than they would have been without the program, there may be an association between program participation and graduation, since students cannot graduate unless they pass the HSPA. Indeed, regular education students went from failing state test scores to passing scores when taking part in the program. The strongest data to support this tentative conclusion regarding graduation would be from the English language learner sub-group, as English language learners went from failing HSPA to

passing HSPA, while English language learners *not* within READ 180 went from passing HSPA to failing HSPA. In other words, we may deduce that English language learners graduate at a higher rate when they are in the READ 180 program and graduate at a lower rate when they are not in the READ 180 program.

Conclusions

Preliminary analyses revealed that the READ 180 program was only partially effective in meeting the stated program goals. However, follow-up analyses indicated that READ 180 was highly effective for English language learner students. Given the high population of English language learner students at the project site, it would appear that a balanced literacy program similar to READ 180 (if not READ 180 itself) should be re-instituted to serve the reading needs of the underperforming English language learner students. In addition, given the absence of any formal evaluation of reading programs at the project site, it may be beneficial to formalize an ongoing program evaluation system to determine program success, before decisions are made to continue or discontinue an existing program. A formalized system of program evaluation would fit well into the already-existing professional learning communities. Some training of administrators and teacher leaders would likely facilitate the evaluation system. Results of such evaluations should be made readily available to the school community at large. Such transparency of information could enable community leaders outside the school to take a more active and more educated role in curricular reform initiatives.

Implications for Action in XXXXX

Because XXXXX High School is designated as a low-performing school, it remains under intense scrutiny at both the local and state levels. Indeed, continued low academic performance may have ramifications for school oversight, funding, and staffing. Thus there exists a pressing need to better understand whether current curricular programs have an impact upon improving student performance. Curricular decisions based on administrative or teacher preferences, instead of proven results, do not serve the best interests of the students. One of the important lessons taken from the present evaluation study is that performance of all student sub-groups should be analyzed to determine if any program being evaluated may have differential effects based on student characteristics. In the present study, a balanced literacy intervention was shown to be effective in improving reading test scores of one of the most at-risk populations, the English language learner students.

At the present time there is no balanced literacy program to serve the English language learner students at XXXXX High School. Moving forward, an ongoing evaluation system could aid students who demonstrate literacy shortfalls by matching struggling readers with an intervention that meets their needs. In such a system, students would ultimately be able to not only improve academic performance, but also hopefully come closer to attaining their longer-term career and personal goals.

Specific Recommendations

Reconsider balanced literacy for at-risk students particularly English language learners. Consider formalizing evaluation process to support curricular decision-making. Improve archiving of student achievement data. Consider longitudinal studies that would

provide a better picture of student literacy achievement over all four years of high school. Intervention programs should utilize all available data for students to decide if any sub-group is benefiting and adjust enrollment accordingly. Ongoing program evaluations should take place to provide information to decision makers.