AN EVALUATION OF A SCHOOL-BASED SUMMER LITERACY PROGRAM

by

MARK C. ROBERTSON

A DISSERTATION

Presented to the Department of Educational Methodology, Policy, and Leadership
and the Graduate School of the University of Oregon
in partial fulfillment of the requirements
for the degree of
Doctor of Education

December 2014

DISSERTATION APPROVAL PAGE

Student: Mark C. Robertson

Title: An Evaluation of a School-Based Summer Literacy Program

This dissertation has been accepted and approved in partial fulfillment of the requirements for the Doctor of Education degree in the Department of Educational Methodology, Policy, and Leadership by:

Keith Zvoch	Chair
Joe Stevens	Core Member
Gina Biancarosa	Core Member
Christopher Murray	Institutional Representative

and

J. Andrew Berglund	Dean of the Graduate School

Original approval signatures are on file with the University of Oregon Graduate School.

Degree awarded December 2014

DISSERTATION ABSTRACT

Mark C. Robertson

Doctor of Education

Department of Educational Methodology, Policy, and Leadership

December 2014

Title: An Evaluation of a School-Based Summer Literacy Program

The purpose of this study was to evaluate the change in oral reading fluency among a sample of students (N = 44) who were randomly assigned a summer school placement. A second goal was to identify relationships between student background characteristics, student learning engagement, and reading fluency outcomes among those students who participated in summer school. Results indicated that students who were assigned to or participated in summer school did not achieve statistically greater summer learning outcomes than students who did not participate. However, summer school participants showed substantial growth in fluency outcomes during the summer intervention period. Implications for summer programs are discussed.

CURRICULUM VITAE

NAME OF AUTHOR: Mark C. Robertson

GRADUATE AND UNDERGRADUATE SCHOOLS ATTENDED:

University of Oregon, Eugene, Oregon
Concordia University, Portland, Oregon

DEGREES AWARDED:

Doctor of Education, Education, Methodology, Policy, and Leadership, University of Oregon, 2014
Master of Arts, Education, 2005, Concordia University
Bachelor of Arts, Biology, 2004, Concordia University

AREAS OF SPECIAL INTEREST:

Use of Assessment in Classroom Teaching
Pedagogical Instruction in Science
Program Evaluation

PROFESSIONAL EXPERIENCE:

Concordia University, Assistant Professor, Portland, Oregon, 2010-Present

Concordia University, Adjunct Professor, Portland, Oregon, 2009

Beaverton School District, High School Science Teacher, Beaverton, Oregon, 2008-2010

Beaverton School District, Middle School Science Teacher, Beaverton, Oregon, 2005-2008

ACKNOWLEGEMENTS

I would like to thank my advisor Dr. Keith Zvoch for his time, advice and critical review during this project. I would also like to express appreciation to my committee members, Dr. Joe Stevens, Dr. Gina Biancarosa, and Dr. Christopher Murray for their time and review of this dissertation.

I would also like to thank Concordia University for the opportunity to attend the University of Oregon to further my education and for the financial support provided to complete this degree.

Additionally, I would like to thank my cohort members for their help and encouragement over the past three years.

DEDICATON

For my students, they demonstrated great courage and a willingness to learn. I hope they pursue their dreams with the same passion.

TABLE OF CONTENTS

Chapter | Page

LIST OF FIGURES

LIST OF TABLES

CHAPTER I

INTRODUCTION

Summer recess poses a challenge for teachers and administrators as students are disengaged from the scholastic environment for three months. In general, annual learning trajectories that are positive during the academic year flatten or become negative during summer months (Alexander, Entwisle, & Olson, 2001; Downey, von Hippel, & Hughes, 2008). For students who are at an economic disadvantage, the long summer break is often detrimental to achievement gains made the prior year. Over the summer, the achievement growth of students who are economically disadvantaged tends to decline, whereas the learning gains of more advantaged students only slows in comparison to in-school learning rates (Alexander, et al. 2001; Entwisle & Alexander, 1992; Schacter, 2001). For example, studies conducted in the Baltimore School District (Alexander et al., 2001) demonstrated that students from low socio-economic status (SES) backgrounds had achievement progress that stagnated or declined during the summer, whereas their higher SES peers' achievement improved during the same period. Moreover, the summer performance trend continued over the life of the five year study, resulting in increasing disparity between students with higher and lower SES backgrounds.

The observed drop in learning and increasing disparity in performance that emerges during the summer months increases the burden for school district personnel charged with reducing the achievement gap and ensuring that all students meet grade-level proficiencies required by federal legislation. The summer drop is particularly problematic as the No Child Left Behind Act (NCLB, 2001) requires school districts to demonstrate adequate yearly progress (AYP) and achieve 100% proficiency in reading and mathematics across disaggregated subgroups by 2014. However, in 2012, a Flexibility Waiver was adopted giving states the opportunity to supplement traditional status-based methods for measuring AYP, focusing instead on State created Annual Measurement Objectives (AMOs; U.S. Department of Education, 2012). Under the new waiver, states are now given flexibility to meet federal mandates by focusing on improving student learning and increasing the quality of teacher instruction. As part of the Flexibility Waiver, districts may use funding to implement 21st Century Community Learning Centers (CCLCs) in effort to extend student learning time. Districts may also implement programs to support students during non-school hours, intermittent school breaks, and during summer recess (U.S. Department of Education, 2012). Supplemental programs can be used to help districts mediate the negative effects associated with learning decline while school is not in session.

With the responsibility for learning that occurs both within and outside of the traditional nine month academic year, some school districts have attempted to offset the summer learning setback by either adopting a year round schedule (Cooper, Valentine, Charelton, & Melson, 2003; von Hippel, 2007) or by implementing targeted summer programs for students most at-risk of summer loss (McCombs, Augustine, & Schwartz, 2011). Proponents of a continuous school calendar advocate for shorter breaks (e.g. 15-20 days of intersession for every 45-60 instructional days) keeping the rhythm of instruction constant (Ballinger, 2000; Kneese, 2000). However, the benefits from shifting to a year-round calendar may not outweigh the costs of modifying other community infrastructures currently in place (Cooper, Nye, Charlton, Lindsay, & Greathouse, 1996; von Hippel, 2007). A rearrangement of the school calendar to a year round schedule affects factors proximal to students, teachers, and parents (e.g. time for teacher planning and professional development, complications with child care for parents (Sardo-Brown & Rooney, 1992). Moreover, changes made to the calendar to create a more constant instructional rhythm do not necessarily lead to improved learning outcomes. The 180 day calendar does not increase in year round schools as it would in an extended school year calendar. Instead, the same number of days is redistributed. When year round and nine month calendar schedules are compared, findings show that students in year round schools learn faster during summer months while students in nine month schools learn faster during the rest of the year. Overall, learning rates generally do not differ (von Hippel, 2007).

In contrast, summer programs serve as an extension to the school calendar and offer students who are most at-risk strategic supplemental instruction. However, estimating the causal impact of supplemental summer instruction is a challenge. The difficulty in separating instructional effects from the distinct background characteristics of summer school participants weakens the inferences drawn regarding summer program outcomes (Zvoch & Stevens, 2011, 2013). Yet, as summer programs are both time and resource intensive, it is essential that stakeholders have access to unbiased estimates of school-based summer learning outcomes. In response, this paper presents an evaluation of a district summer reading initiative. Findings will contribute to the literature on the causal effects of summer school by examining second grade fluency outcomes as a function of treatment assignment and participation. The following provides a (1) review of the summer learning/summer school literature, and (2) offers a rationale for the study described herein.

CHAPTER II

LITERATURE REVIEW

The following section is a review of both the summer learning summer school literatures. Also, Oral Reading Fluency (ORF) is discussed as an indicator of student reading progress and a predictor of reading competence.

Summer Learning Literature

Seasonal achievement growth rates established in the analysis of data from nationally representative sources (e.g. Beginning School Study, Early Childhood Elementary Study, Sustained Effects Study) demonstrate that when students from divergent economic backgrounds are compared, economically advantaged students maintain or increase their achievement performance over the summer recess while their peers from economically disadvantaged backgrounds tend to stagnate or experience declines in achievement over the same time frame (Alexander, et al., 2001; Borman & Dowling, 2006; Bryk & Raudenbush, 1988; Entwisle, & Alexander, 1992; Heyns, 1978; Lee & Burkam, 2002). More specifically, findings indicate that students from higher SES backgrounds tend to lose ground in mathematics and gain in reading. Students from mid-SES backgrounds tend to lose ground in mathematics and maintain reading achievement levels, whereas students from low SES backgrounds lose ground in both subject areas (Borman & D' Agostino, 1996; Cooper et al., 1996; Downey, von Hippel, & Broh, 2004). Overall, achievement decreases by of one tenth of a standard deviation or one month of grade level progress across all SES backgrounds and subject areas (Cooper et al., 1996).

Differences in learning between SES levels and content areas (i.e. reading and mathematics) may be explained to some extent by the amount of opportunity students have to practice academic skills outside the scholastic environment (Alexander et al., 2001; Bryk & Raudenbush, 1988). Home and community environments tend to provide more opportunity for continued development of reading (e.g. books, education media, libraries) than for development of factual and procedural skills needed to progress in mathematics. Children are therefore better able to maintain or increase proficiencies in reading than in mathematics over the summer (Byrk & Raudenbush, 1988; Cooper et al., 1996). A similar explanation for the decline in student achievement growth in both mathematics and reading over summer months is based on the "faucet theory," where the flow of educational resources slows or stops (e.g. books, library access, educational manipulative or tools) for students from low SES backgrounds (Alexander et al., 2001, Entwisle, Alexander, & Olson, 2002). Besides fewer resources, disadvantaged students also have fewer opportunities to engage in learning activities (e.g. trips to the library, reading books) that are conducive to building academic skills and academic patterns of thinking (Alexander, Entwisle, & Olson, 1997; Heyns 1978). Moreover, parents in low SES homes tend to lack the cognitive and time resources to instill a culture of learning (Barton & Coley; 2008; Lee & Burkman, 2002; Alexander et al., 2001). Because there are fewer resources and opportunities to engage in learning for students of low SES backgrounds compared to their peers, the difference in students' summer environments creates a learning inequity that results in poorer summer outcomes for disadvantaged students.

While the literature demonstrates that summer learning rates differ among students of divergent economic backgrounds, Cooper et al. (1996) suggest that these findings be interpreted as only suggestive as the summer learning estimate often includes a portion of the school year (Cooper et al, 1996; Entwisle & Alexander 1994; Zvoch & Stevens, 2013). Frequently, the tests on which summer learning estimates are based are not given on the last and first day of school year. Instead, students are often assessed with 3-5 weeks remaining in the school year. After the assessment has been administered, students are still in school progressing toward learning outcomes. The same is true in the fall of the following school year as students are often assessed with post summer assessments after 3-5 weeks of school instruction has been received. As a consequence, summer recess may be more detrimental than the research indicates.

Despite the caution Cooper et al. attaches to summer learning estimates, the effect of the summer slide on low SES students is non-trivial. When students of low SES backgrounds enter kindergarten, they are already behind their advantaged peers in both mathematics and reading (Barton & Coley, 2008; Coley 2002; Lee & Burkman, 2002). As students move forward in school, the gap in achievement widens (Downey et al. 2004). Downey et al. (2004) used national survey data on over 20,000 kindergarteners and first graders to assess changes in math and reading scores at the beginning and end of summer in kindergarten and first grade for two cohorts. Reported results indicate that student growth rates were similar between advantaged students and disadvantaged students during the school year despite large differences in initial status. Over summer months, student learning rates diverged. Findings suggest that during the academic year, schools increased learning rates and decreased inequity amongst students, but during the summer disadvantaged students lost ground relative to their more advantaged peers.

The achievement gap that is exacerbated by differential summer learning has been associated with higher dropout risk, lower college preparatory placement, graduation, and college attendance (Alexander, Entwisle, and Olson 2007). Alexander et al., (2007) conducted a longitudinal study of students in the Beginning School Study (BSS) through 9th grade and into post high school, vocational, and higher education. Among their findings, it was reported that disadvantaged students showed a 33% dropout rate compared to the 3% of their advantaged peers. Also, 13% of students of disadvantage backgrounds were placed in college preparatory course in comparison to 62% of students from advantaged backgrounds. Logically, college preparatory courses would then impact attendance in two- and four-year colleges. A substantive portion of variance between SES groups was explained by the summer effect. The authors note that the BSS data is sampled from a specific geographic, racial, and economic group of 20 districts and is not entirely generalizable, but Cooper et al., (1996)'s findings suggest that similar outcomes are likely for other districts as trends in learning trajectories are similar in other studies.

From the literature on summer learning outcomes, it is evident that summer brings challenges for lower SES students that set students back from their prior year's progress. Over time, gaps in performance increase and disadvantaged students find themselves less likely to take advanced courses and continue toward college preparation. Instead, school becomes a process that is left unsuccessful and unfinished. However, Downey et al. (2004) argue that schools are the mediating resource required to provide an equal opportunity to learn year round for all students. Following Cooper and colleagues' recommendation to increase the quality and volume of instruction over summer break for at-risk students (Cooper et al., 1996), school officials may be able to offset losses over the summer and decrease the need for remedial instruction during the academic year by providing targeted instructional programs in the summer.

Summer School Literature

The literature examining summer learning rates among students from divergent economic backgrounds shows the academic challenges students face during summer recess, and the student, school, and environmental factors that diminish student learning, widen the achievement gap, and increase the burden on school leaders working toward meeting NCLB mandates. In their effort to offset the unequal distribution of educational opportunities and resources available to students, education leaders have increasingly moved toward implementing supplemental summer instruction to keep their most at-risk students (i.e. students performing below grade-level benchmarks and/or economically disadvantaged students) progressing toward grade-level proficiency. These instructional programs or "summer schools" are strategic in that education resources can be used to target at-risk students in need of greater support who would otherwise not have access to such educational opportunities.

In general, students attending summer programs benefit both academically and non-academically (McCombs et al., 2011). Summer programs provide additional time and resources for students to practice academic tasks. As a result, students are more likely to increase academic achievement in the content areas in which supplemental instruction is delivered. Additionally, increased achievement may also lead to higher attendance the following school year, greater perseverance in future learning challenges, and stronger self-efficacy (McCombs et al., 2011). Summer programs may also serve to support school efforts in lowering drop-out rates as well as facilitating transition periods for students moving from middle school to high school. If summer assessments are administered, teachers and administrators also obtain learning outcome data that can be used to monitor students' progress toward meeting grade-level proficiencies (McCombs et al., 2011).

Although the provision of summer school increases the financial burden on school districts, summer programs can often be supported through several competitive initiatives. Because summer programs support students and families by providing resources beyond academic instruction (e.g. food provision, child care, transportation), support funds are available for district summer programs through Title I, Child Care and Development Funds (CCDF), Temporary Assistance for Needy Families (TANF), and grants issued for 21st Century Community Learning Centers (CCLC funds) (McCombs et al., 2011). However, it should be noted that these funding sources are only available for a designated time period and districts should make long terms plans to sustain summer programs after the initial funding is no longer available. Should districts choose to invest in long term summer programs, McCombs et al. (2011) observes that summer programs operate at a 53%-63% lower cost than that of a typical academic year school day. Moreover, summer programs are less disruptive to community infrastructures than alternatives such as year-round school and are thus becoming a more popular choice for school districts.

Although summer programs are argued to be a viable means to support at-risk students, the manner in which these programs are implemented vary, making evaluation of summer programs difficult (Cooper, Charlton, Balentine, Muhlenbruck, & Borman, 2000; McCombs et al, 2011). Summer school program characteristics tend to vary in provider type (whether the program is offered by the district or an external entity), whether the program is voluntary or mandatory, dosage (length of the program day and duration of the total program), setting (whether the program is site-based or home-based), and purpose (e.g. remediation, acceleration, recreation) (McCombs et al., 2011). More specifically, dosage and setting are closely associated with the intensity and purpose of the summer program. For example, studies examining summer literacy programs for elementary students in Baltimore lasted for seven weeks. Site-based daily instruction focusing on literacy accounted for half of the daily activities and the other half were reserved for student enrichment (e.g. mathematics, science exploration, physical activity, art projects). Also, intermittent enrichment and recreational field trips were provided (e.g. museums, swimming, bowling, research centers) (Borman, Benson, & Overman, 2005; Borman & Dowling, 2006). Analytic findings from randomized field trials revealed effect sizes for students attending summer school in vocabulary, reading comprehension, and overall reading as, respectively, d =.32, d = .28, and d = .30 compared to their students who chose not to attend or were assigned to the control group (Borman & Dowling, 2006).

In contrast, Schacter and Jo (2005) examined a summer program that focused only 1.5 hours of a 9 hour camp day on literacy instruction and the rest of the time was allotted for recreational activities not associated with academics. Seventy-two students were randomly selected from a pool of 162 first grade students. Compared to a control group of peers similar in age, minority status, and free or reduced lunch status, outcomes from the program for attendees showed effect sizes of .96 and .59 standard deviations for decoding levels one and two of the *Gates-MacGinite Word Decoding* assessment, and effects of 1.35 and 1.25 standard deviations for levels one and two of the *Gates-MacGinite Comprehension* assessment.

Other summer programs have used only half-day sessions focusing solely on basic literacy skills and outcomes in small group environments that facilitated individual instruction and feedback (Zvoch & Stevens, 2013). Here, program outcomes were measured using DIBELS' Nonsense Word Fluency (NWF) and DIBELS' Test of Oral Reading Fluency (TORF). DIBELS' NWF is a standardized one minute probe that measures the alphabetic principle and the ability to blend letters into words (Kaminski & Good, 1996). DIBELS' TORF is a standardized one minute probe that measures accuracy and fluency with passages connected to text (Shinn, 1998). The effect size for kindergarten students moving to first grade assigned to summer school using DIBELS' NWF was .6 relative to student peers with similar preprogram assessment scores assigned to the control group. First grade students moving to second grade, who were assigned to summer school, when compared to an equivalent control group, showed an effect size of .78 measured with the TORF. In each of the studies mentioned above, positive findings observed using randomly selected student samples indicated that site-based academic instruction in programs was beneficial in stemming summer learning loss and increasing summer gains. Furthermore, despite the varied amount of time allocated to academic instruction, most programs that target at-risk students are structured to provide intensive academic instruction over several weeks (Cooper et al., 2000; McCombs et al., 2011).

Other summer learning research suggests that students with high economic needs may be partially supported through the provision of book regimens over the summer months (Allington, et al., 2010; Kim, 2006, 2007; Kim & Guryan, 2010; Kim & White, 2008). In a series of studies examining voluntary summer reading programs where books were randomly assigned to students during the summer months, Kim (2006) reported that students assigned a book provision had higher reading achievement scores on the *Iowa Test of Basic Skills* (ITBS) in September of grade 5. More specifically, compared to students in the control group, effects for Black students were reported as .22 of a standard deviation, .14 of a standard deviation for Latino students, .17 of a standard deviation for students identified as having low fluency skills, and .13 of a standard deviation for students who owned less than 50 books at home. Kim (2007) employed a continuation study for 331 students in grades 1-5. All students were given a pretest/posttest in June and September using the SAT10 reading test. Students were randomly assigned to a regiment of 10 books over the summer. Results indicated that despite an increase in books read and participation in reading activities for students assigned to receive books, there were no statistically significant differences between the treatment group and the control regarding posttest outcomes. Kim and White (2008) further examined voluntary book provision programs by conducting a randomized field trial for 400 students between third grade and fifth grade. In the study, students were assigned to the control, a book regiment, a book regiment with oral reading fluency scaffolding, or a book regiment with oral reading fluency and comprehension scaffolding. Participants of the study were measured on oral reading fluency using DIBELS, and silent reading ability using the ITBS. Authors concluded that book provisions were a positive, but insufficient support for students from low SES backgrounds. As a result, they recommended that book provisions be provided simultaneously with instructional scaffolding.

The literature on summer school identifies several options for school administrators regarding the manner in which summer instruction and resources can be provided to support at-risk students. Because educational funding is limited, program cost is a large factor in determining a summer program's feasibility, sustainability, and overall merit (McCombs et al., 2011). McCombs et al. (2011)'s analysis of seven summer programs' financial costs, including administrative costs, transportation, food, and school facilities, indicates that large externally-led programs require the most financial support ($2,058-$2,801 per slot, per summer). In contrast, district-led programs are more cost effective ($1,109-$2,601 per slot, per summer), whereas programs using only book provisions without site-based instruction are substantially less costly ($245 per slot, per summer). However, the authors note that although book provision programs without site-based instruction are more cost effective, these programs do not yield the additional benefits that site-based instruction programs provide (e.g. mathematical achievement, improved safety, social and behavioral outcomes, and recreational opportunities). Overall, district-led summer programs are the most cost effective option for offering site-based core instruction to targeted at-risk students.

In comparison to costs of academic programs run during the school year, district-led summer programs are more cost effective (McCombs et al., 2011). The reduction in program cost is due to less academic support for special education and English language learners. Also, district summer school programs spend less time and financial resources on curriculum planning and accountability reporting. Additionally, certified teachers receive lower wages working for summer programs relative to wages earned during the academic year. As a result, summer programs operate at costs 40% to 50% less than that of programs run during the academic year. Considering the different methods to support student learning over the summer, the greatest benefits may be provided from district-led, site-based academic programs that target students who are academically and economically at-risk with intensive teacher-directed instruction.

Methodological Challenges. Although the literature on summer school indicates that students who attend summer programs generally benefit from the experience (Allington et al., 2010; Borman & Dowling, 2006; Entwisle & Alexander, 1992; Kim, 2006, 2007; Zvoch & Stevens, 2013), Cooper et al. (2000) advises caution when interpreting summer program effects. Cooper and colleagues' comprehensive review of summer programs identified several methodological issues that arise from measuring program success with single group pre/posttest designs. For example, if the program's effectiveness is measured by pre/posttest gains, then statistical regression to the mean may inflate scores and lead to an overestimation of program effects. Also, if a comparison group is not available, then it is not possible to measure the effect the program had on students relative to students not attending the program. To begin to evaluate the efficacy of summer school, it is requisite that a control condition or a comparison group be available to contrast achievement outcomes. However, researchers also need to be aware of the manner in which students are assigned to participate in summer school as selection bias may also lead to distorted inferences regarding program efficacy.

Cooper et al. (2000) addresses the selection-based validity threats that commonly arise in the summer school literature. The strength of inferences made regarding summer program efficacy are negatively impacted by how students are invited and whether students choose to attend (Bloom, 1985; Borman & Dowling; 2006; Zvoch & Stevens, 2013). Programs using specific criteria to select students consequently separate participants into groups with characteristic differences. For example, summer reading programs that select students based on low test scores, may also be selecting students with a home environment that is less conducive to academic success. In contrast, a comparison group not receiving treatment because of high test scores may be benefitting from extra instruction at home or access to other educational resources. The presence of extraneous factors can distort treatment and control group comparisons leading to biased inferences regarding program outcomes. Cooper et al. (2000) recommend using stronger methodological approaches such as experimental designs using random assignment. In situations where complete randomization is not feasible alternative quasi-experimental designs can be employed such as the regression discontinuity design (RDD), where cut scores establish a counterfactual condition. In summer programs, approaches like RDD increase the strength of inferences made from data gathered using non-equivalent group pre/posttest designs (for an example of RDD used in summer school evaluation, see Zvoch & Stevens, 2011).

As attendance at a summer program is often voluntary, researchers who study these programs using random assignment or other techniques should be prepared to examine outcomes as a function of the participation status of students (i.e., students who (a) fully participate, (b) partially participate, (c) refuse treatment, and (d) are ineligible for programs) (Cooper et al., 2000). For example, Borman and Dowling (2006) randomly assigned over 300 early elementary students to a need-based summer program in which reported results indicated that the experimental effect of assigning to treatment was insufficient as students assigned to summer programs showed no differences in summer reading gains when compared relative to those not assigned to summer school. However, further analysis showed that the frequency of attendance was predictive of positive achievement outcomes the following fall. Participation for the full six weeks was associated with an effect size of .27 of a standard deviation compared to students who volunteered for the study and were assigned to the control group. Each additional week students participated was associated with a .05 of a standard deviation increase in fall achievement scores. These results suggest that students who are assigned to treatment and do not participate (i.e. refusers or non-compliers) in summer programs lead to an underestimation of the program effect. When examining outcomes associated with treatment assignment, statistical adjustment is therefore requisite to more accurately estimate treatment effectiveness by removing the effects of non-compliers. In current studies, statistical adjustments are made for non-compliant students in the assignment group by using either the Bloom's Adjustment (Zvoch and Stevens, 2013), or Complier Average Adjustment Effects (CACE) analysis (Borman & Dowling, 2006).

In addition to selection and dosage factors, programs that deviate from the academic year curriculum and rigor of instruction (programs that focus on remedial skills below grade level) make it difficult to compare and generalize summer program outcomes. Construct-irrelevant factors (Messick, 1994) that do not align with what is taught during the school year cloud inferences regarding the authenticity and size of program outcomes. More specifically, Cooper et al. (2000) discuss the congruence of curriculum with the dependent measures used to measure changes in program outcomes. For example, summer literacy programs that align the skill-based curriculum that is used during the summer program to the curriculum used for instruction during the academic year increase the generalizability of inferences made from score changes relative to growth rates observed when school is in session (Borman & Dowling, 2006; Borman et al., 2009; Borman et al., 2005; Zvoch & Stevens, 2011, 2013). In other words, summer programs that align outcome measures with program curriculum increase the accuracy and comparability of score changes during summer programs to student progress made during the academic year.

Regarding the type of outcome measures used for evaluation of summer programs, programs that measure student outcomes with norm-referenced tests may be measuring concepts that are more distal and less sensitive to changes in procedural skill gains made during the duration of the program leading to smaller measured outcomes. Conversely, when a program uses measures proximal to treatment (Zvoch & Stevens, 2013) by examining procedural reading skills with fluency measures, (i.e., DIBELS and the Test of Oral Reading Fluency) the magnitude of effects may be greater than what would be expected using a larger norm-referenced test that measured components of reading such as vocabulary and reading comprehension. Without consideration of measure sensitivity, programs that use outcome measures proximal to treatment may be deemed highly successful when, in actuality, large effect sizes may be partially due to practicing effects or other extraneous factors. Given that the magnitude of program effects may be partially due to the type of outcome measure used, program evaluators should be mindful of a measure's sensitivity when interpreting outcome data regarding program efficacy.

With consideration of all the methodological challenges present in summer program studies, Cooper et al. (2000) concluded that economically disadvantaged students were able to maintain or slightly improve upon academic gains from the prior year in reading and mathematics. Additionally, middle-class students were able to accelerate acquisition of reading and mathematics skills. Other findings revealed a u-shaped relationship across grade levels as early-elementary and late-secondary students were the largest benefactors from program participation. The overall size of the summer school effect across all students and samples examined was approximately one quarter of a standard deviation. However, the middle class students' average effect (.45-.56) was larger than low income students (.20-.24).

Cooper et al. (2000)'s findings also indicate differences in outcomes depending on the targeted grade-level and provider type. Students who are older may be enrolled in accelerated programs which attract more motivated students producing greater positive program outputs. Using the same logic, younger students in summer programs may be more apt to want to please instructors within programs, consequentially increasing student engagement yielding greater program outcomes. In addition, differences between middle and lower income students can be explained partially as a function of differences in motivation based on whether the program is a remedial or an accelerated program (Cooper et al., 2000). Remedial instruction programs may also be serving students with lower motivation to engage academically during the program hours and in a home environment, whereas students in accelerated programs may be more likely to engage academically during program hours and in home environments. As a result of differences in student motivation, program outcomes may indicate greater program efficacy due to treatment, when in actuality, differences in outcomes are reflective of levels of engagement with treatment. As student motivation is a factor in program outcomes, it is requisite for investigators to consider how and what factors motivate learning outcomes.

Participation and Engagement. Differences in outcomes within need-based voluntary summer school programs are partially based on the level of student interaction with treatment. Heyns (1978) suggests that individual student interest, motivation, and engagement may influence summer program outcomes. Engagement factors are thought to affect student/teacher, student/peer, and curriculum interactions both during and outside of the program environment. Furthermore, comparisons of students who attend summer school and engage and those who attend and do not engage can be misleading. Student engagement requires effort, interest, and motivation that is expressed as either cognitive (e.g. motivated when using strategic thinking skills) or behavioral (motivation for completing active tasks) (Bodovski & Farkas, 2007; Culatta, Setzer, Wilson, & Aslett, 2004; Greenwood, Horton, & Utley, 2002 Guthrie, Wigfield, Barbosa, Perencevich, & Taboada, et al., 2004). For example, in the context of a summer program that focuses on strategic instruction for reading literacy outcomes (e.g. proficiency in procedural skills, reading comprehension), students that attend the program but cognitively disengage or fail to persevere or attempt to think critically when asked to understand a challenging text will most likely fail to make progress despite receiving treatment (Guthrie et al., 2004; Taylor, Pearson, Peterson, & Rodriguez, 2003; Wigfield et al., 2008). With the same logic, students that disengage with antisocial behaviors when asked to sound out words, read aloud, or to participate in social interactions with peers will most likely make minimal progress (Bodovski & Farkas, 2007; Guthrie et al., 2004). Conversely, students demonstrating high interest in cognitive and behavioral tasks consistency, but who have lower attendance rates, may make greater academic progress than their disengaged peers. In other words, the treatment effect is relative to the quality of interaction and engagement the student has with the program instruction. Engagement and compliance is thus a prerequisite for positive program outcomes.

The literature examining engagement factors and reading behavior varies in how engagement is defined. Earlier research suggests that cognitive engagement is expressed primarily when a child reads to comprehend text, or reads to learn. To facilitate this process, both self-regulation and multiple strategies for higher-order thinking are required (Dole, Duffy, Rochler, & Pearson, 1991; Guthrie, Von Meter, McCann, Whitfield, & Bennet, et al., 1996; Pintrich & DeGroot, 1990). Pintrich and DeGroot (1990)'s framework identifies three components of student motivation: a) belief that the child is able to complete the task (self-efficacy), b) belief that the task is important to complete, and c) emotional responses felt by the child when completing the task. Also, factors that influence motivation are dichotomized into intrinsic (when the individual is personally motivated) and extrinsic factors (when the individual is motivated by outside source) (Pintrich & Schrauben, 1992). Furthermore, intrinsic motivation factors (e.g. wanting to know more about a subject, increased pleasure or satisfaction) tend to result in greater cognitive engagement behaviors during reading activities whereas external motivation factors (e.g. grades, stickers, verbal affirmation, avoiding punitive consequences) result in compliance related behaviors. Effects resulting from intrinsic motivation were found to be greater than those resulting from extrinsic factors.

Other research suggests that reading engagement is inadequately explained by cognitive engagement alone (Guthrie, Hoa, Wigfield, Tonks, Humenick, & Little, 2007). Instead engagement is driven by a more comprehensive and multidimensional construct of motivation (Guthrie & Wigfield, 2000, Guthrie et al., 2006) comprised of motivation when reading different text genres, general reading motivation, and students and teachers perception of the student's motivation. Regardless of the definitions used to explain engagement, research indicates that motivation for reading and reading comprehension are linked (Baker & Wigfield, 1999; Wang & Guthrie, 2004) as well as motivation and self-regulation (i.e. management and control of cognitive and active behaviors) (Pintrich & DeGroot, 1990). In contrast, negative or *avoidant* engagement behaviors are negatively correlated with reading achievement (Guthrie & Coddington, 2009). As program outcomes are a function of both the quality of treatment provided and the level of engagement and motivation offered by participants, supplemental measurements of student motivation and engagement levels may better inform school personnel of program outcomes regarding evaluation and decision-making for summer reading programs.

In summary, the literature on engagement and motivation associated with reading identifies several motivational and engagement factors (e.g. whether or not the child is motivated intrinsically or extrinsically, preference of reading text, the length of time student engages in the reading activity) that should be considered when evaluating student interactions during instructional reading activities or during autonomous reading.

The relationship between motivation to read and reading proficiently, though intuitive, systemically contributes to success or failure later in education and beyond school (Cunningham & Stanovich, 1997, Stanovich, 1986). Stanovich argues that motivation to read and reading proficiency is a reciprocal partnership, which build off each other. Strong readers gain pleasure and curiosity from reading which increases motivation to read further. Struggling readers become unmotivated and avoid reading which results in a greater challenge to improve as a reader. Stanovich calls this phenomenon the *Matthew Effect*. Because reading is a developmental process, programs aiming to support struggling readers at early ages should examine indicators of reading competence using metrics that are technically sound, efficient, and are predictive of reading goals at higher difficulties.

Oral Reading Fluency Predictive of Reading Competence

Reading is an essential skill that enables learners to acquire general knowledge requisite to interact socially, successfully pursue professional goals, and engage in civic participation (Chall, 1983; Gillet, Temple, Temple, & Crawford, 2012). In addition to an overall increase in the quality of life, adults who are better readers are found to have a larger vocabulary and a greater general knowledge of the surrounding world (Stanovich, 1992). For those who struggle to read proficiently, research indicates poor readers tend to disengage in activities that practice and develop reading skills. As a result, reading deficiencies increase as the reader matures (Stanovich, 1986). Over time, difficulty reading leads to long term negative effects that extend beyond academic years into adulthood.

Longitudinal research demonstrates that early reading outcomes are predictive of academic outcomes several years later (Cunningham & Stanovich, 1997; Juel, 1988; Slavin, 1991). For example, reading scores of twenty-six first grade students were correlated with scores on academic measures administered ten years later. Findings indicated that reading outcomes (for the students sampled), were predictive of cognitive abilities and overall general knowledge (Cunningham & Stanovich, 1997). Similarly, first grade students who were identified as poor readers were shown to have an 88% likelihood of remaining a poor reader by fourth grade (Juel, 1988). Furthermore, students identified as poor readers early in school are observed to be at a greater risk of struggling in later grades and ultimately dropping out of high school (Alexander Entwisle & Horsey, 1997; Annie E. Casey Foundation, 2010). In addition to persistent negative academic outcomes, struggling to read at an early age is associated with negative self-perceptions that can lead to a lower self-esteem and depression (Arnold, Goldston, Walsh, Reboussin, Daniel, et al., 2005; Boetsch, Green, & Pennington, 1996; Maughan, Rowe, Loeber, & Stouthamer-Loeber, 2003). Negative reading outcomes for learners early on in development often lead to more severe negative outcomes in adulthood.

Data collected from national surveys provides evidence that adults who are not proficient readers tend to have lower academic credentials and suffer greater economic hardship (NCES, 2001:534; NAAL 2006). Furthermore, adults who are not proficient readers have a greater likelihood of incarceration (NAAL, 2006), and lack opportunities for acquiring a job or advancing within a vocation. Moreover, poor literacy and lower education status are associated with higher recidivism rates for individuals who have been previously incarcerated (Hrabowski, III, & Robbi, 2002). Because proficient reading is a requisite skill for all individuals to function in society, and the negative outcomes for those who do not become proficient readers are costly to both the individual and to society in general, policy makers have concentrated efforts to make sure that all students can read by the end of third grade (Annie E. Casey Foundation, 2010; National Research Board, 1998). As reading is a complex process, formative assessment of specific skills that are predictive of proficient reading may provide teachers and school leaders with timely information to make decisions regarding curriculum, instructional support, and overall student progress toward reading proficiency.

In conjunction with efforts to ensure all students are proficient readers by the third grade, the National Reading Panel (2000) identified five components necessary for proficient reading: a) alphabetic principle, b) phonemic awareness, c) reading fluency, d) vocabulary, and e) comprehension. Within the spectrum of reading development, an important focus of instruction for second and third grading reading goals and a predictor of overall reading competence is oral reading fluency (ORF) (National Reading Panel, 2000; Good III, Simmons, & Kame'enui, 2001). Moreover, adequate progress on ORF measured in the spring of second grade and the spring of third grade is necessary for students to be on track for attaining high-stakes achievement outcomes (Kame'enui & Simmons, 2001). As oral reading fluency is an important component of reading development, the following section will discuss oral reading fluency as a predictor of overall reading competence, examine theoretical frameworks for reading, and present evidence regarding the validity of interpreting reading progress using oral reading fluency measures.

A large body of literature identifies ORF as an early predictor of future reading proficiency (Adams, 1990; Fuchs, Fuchs, Hosp & Jenkins, 2001; Shinn, 1998; Yovanoff, Duesbery, Alonzo, & Tindal, 2005). Reading fluency, defined as speed and accuracy with appropriate expression (prosody), is comprised of phonological segmentation (the ability to break a word into smaller sounds), phonological recoding skills (sequencing letters sounds to create a string of sounds in order to sound out a word), and word recognition (recognizing a word correctly from memory) (Fuchs et al., 2001). As children become able to effortlessly translate text into spoken language with speed and accuracy, they are able to access lexical representation to derive meaning from text (Adams, 1990). Multiple models within the literature depict varying processes of language acquisition that increases translation of text.

Models explaining language acquisition follow two directions of processing: bottom-up or top-down. In general, bottom-up models for reading are based on the work of Laberge and Samuels (1974). *Automaticity theory*, as a basis for bottom up models, operates from the premise that comprehending text requires an orchestration of many components. If each component requires attention, the simultaneous execution of all components in a short time would be impossible. With repetition, a reader's lower level processing skills (decoding, word recognition) become automatic thus freeing up more attention for the reader to make semantic and contextual decisions (contextual facilitation) that derive meaning from text. Essentially, higher order skills (passage fluency and comprehension) are dependent on proficiency of lower order skills (sublexical skills) (Laberge & Samuels, 1974, Fuchs et al., 2001). In contrast, top-down reading models suggest that students interpret text through a psycholinguistic interaction between thought and language (Goodman, 1967). Such models operate with the assumption that reading is a holistic process where the reader uses minimal cues from prior semantic (meaning of the word), syntactic (formation of grammatical sentences), and general knowledge to guess or predict expected meaning of proceeding text from partial comprehension of text already read. This process is inverse to bottom-up models as logic or comprehension skills are required in order for the reader to encode (provide an oral output) text. In other words, recoding text is not dependent on a reader's ability to decode. Instead, the child uses iterative recoding cycles to predict overall meaning or comprehension. The *signal*, or oral output, is partially due to decoding and partially due to prediction within the recoding process, "any matching or coded signal which results is a kind of byproduct" (Goodman, 1967, pg. 6).

An intermediate approach relative to the other two models presented, interactive-compensatory models, operate from a premise that higher order skills work concurrently with lower order skills. Contextualization facilitates compensation for inefficient decoding processes (Stanovich, 2000). Overall, bottom-up models and inter-compensatory models share the assumption that proficiency in low level word identification skills free up processing attention required for comprehension of meaning within text (Fuchs et al., 2001). The notion that speed and accuracy through text frees up attention for comprehension provides a logical argument that ORF is theoretically predictive of reading competence. Oral reading fluency, as an indicator of reading competence, is frequently measured using Curriculum-Based Measures for reading (R-CBMs). R-CBMs are reading probes that measure reading accuracy within a given timeframe. Some researchers, critiquing the interpretation of fluency measures which result in words correct per minute (WCPM) outcomes, argue that fluency measures do not assess all subcomponents that encompass the construct of oral reading fluency (speed, accuracy, and prosody) (Valencia, Smith, Reece, Li, Wixson, & Newman, 2010). As rate and accuracy are directly assessed in CBM measures, expression and phrasing (prosody) are not. Moreover, research suggests that prosody may be linked to oral reading fluency and comprehension in that expression provides evidence that the reader is comprehending text (Khun & Stahl, 2003). However, Good and Jefferson (1998)'s extensive review of reading fluency provides other evidence suggesting that CBM probes are in fact sound indicators of reading comprehension and fully measure the construct of oral reading fluency in that, across all concurrent and predictive validity studies correlation data was reported as ranging from .6-.8.

In conjunction with the notion that CBM ORF metrics theoretically and empirically are demonstrated to have adequate construct validity, R-CBM probes are criterion-referenced and are considered efficient, valid, and reliable when monitoring growth and making data driven decisions regarding basic reading competence (Fuchs & Fuchs, 1992; Hintz, Owens, Shapiro, & Daly, 2000). Multiple studies demonstrate that ORF measures have strong criterion-related validity regarding reading comprehension measures and state tests. More specifically, a cut score for the WCPM on fluency measures was highly predictive of student success rates on state accountability tests. For example, McGlinchey and Hixon (2004) report a 74% pass rate for student over 100 WCPM on the Michigan State accountability test for fourth grade. Also, Good, et al., (2001), reported a 96% pass rate on the third grade Oregon Assessment of Knowledge and Skills (OAKS) for students scoring above a 110 WCPM cut scores on fluency measures administered in the spring of third grade.

In addition to R-CBMs being empirically sound measures for predicting success on state tests, several studies, noted by Shapiro, Keller, Lutz, Sontoro, and Hintz (2006), indicate moderate (.44) to moderately-strong (.79) concurrent validity data for R-CBM ORF measures and multiple state assessments at the third and fourth grade year: Colorado (Shaw & Shaw, 2002), Florida (Buck & Torgenson, 2003), Illinois (Sibley, Biwer, & Hesch 2001), Michigan (McGlinchey & Hixson, 2004), Minnesota (Hintz & Silberglitt, 2005), North Carolina (Barger, 2003), Oregon (Crawford, Tindal, & Stieber, 2001), and Washington (Stage & Jacobson, 2001). Other concurrent validity studies show that fluency measures correlated with the Iowa Tests of Basic Skills (ITBS) comprehension in grades 1, 2, and 3. Correlations were reported as .69-.74, .68-.75, and .63-.65, respectively (Shilling, Carlise, Scott & Zeng, 2007). Moreover, concurrent data between oral reading fluency and the SAT-9 total reading score reported the strength of the relationships in first grade as .80-.84, second grade .74-.77, and third grade .77-.81 (Klein & Jimerson, 2005).

Other research has shown strong internal reliability and validity data on specific fluency measures. The Test of Oral Reading Fluency (TORF) is a one minute probe used to test speed and accuracy with connected text (Children's Education's Services, 1987). Tindal, Marston, and Deno (1983) report TORF alternate form reliability coefficient as .89-.94. Test retest reliability was reported as .92-.97. Also, criterion-related validity was reported in eight separate studies ranging from .52-.91 (Good & Jefferson, 1998). Likewise, the DIBELS' Oral Reading Fluency (DORF), fluency passages designed to align with the validity and reliability of TORF, were examined for technical adequacy for the second grade DORF passages (Good, Kaminski, Smith, & Bratten, 2001). Median alternate form reliability was reported as .94, and concurrent validity correlation coefficients with the TORF ranged from .92-.96, with a median concurrent reliability coefficient of .95.

The evidence presented on fluency measures demonstrates that R-CBM probes, as indicators, are not only internally sound, but they are strong predictors of high stakes tests and overall reading competence. Despite concerns raised regarding the brevity of fluency probes and making a determination about a child's reading proficiency (see Hamilton & Shinn, 2003), CBM's are widely used in school districts and possess high utility in that they can be used for several purposes: screening, diagnostics, progress-monitoring, and outcome measures (Hasbrouck & Tindal, 2006). Because fluency measures are instrumentally sound and have high utility, districts who utilize fluency probes in summer reading programs can strategically apply the data collected to not only monitor students reading progress over the summer months in congruence with progress during the academic year, but also use fluency outcomes as program data to evaluate the program's efficacy.

Proposed Research

The literature surrounding summer learning and summer programs provides administrators and teachers with a logical basis for offering low performing students, and students at an economic disadvantage, supplemental-strategic support over the summer. The literature also suggests that districts seeking increased opportunity for disadvantaged students will benefit from supplementary literacy-based programs (Alexander et al., 1997). However, given the challenges associated with program evaluation, strong methodological designs are needed to distinguish treatment effects from external factors (e.g. sampling bias, history) (Cooper et al., 2000).

In response to the challenge that stems from non-equivalent control group designs, Zvoch & Stevens (2013) implemented a randomized field experiment to assess the efficacy of an intensive school-based summer literacy program. The study is distinct from other studies because of the summer program's focus on fundamental reading instruction (i.e. phonemic awareness, alphabetic understanding, and fluency/automaticity) and the large dose of reading specific instruction (Cooper et al., 2000, Zvoch & Stevens, 2013). Additionally, the fluency-based measures used (i.e. DIBLES NWF; TORF) were contiguous to the curriculum provided to students during the academic school year and within the program.

Based on the prior research of Zvoch and Stevens (2011, 2013), the proposed study will estimate the change in literacy performance associated with assignment to, and participation in summer school among an older sample of struggling readers from Zvoch and Stevens' database. The sample of students is comprised of those who completed second grade during the previous academic year and were randomly assigned a summer school placement. An Intent-to-Treat (ITT) and Treatment-on-Treated (TOT) analysis will distinguish between effects of assignment to treatment and participation in treatment, as well as the predictive relationship between student engagement and summer oral reading fluency outcomes. The following questions will be examined:

1. How do reading fluency outcomes for students who participated in a district summer school programs compare with students who refused a summer school placement and those who were not invited?

2. Does student participation (i.e. attendance, engagement, homework completion) relate to summer reading fluency outcomes?

3. What is the relationship between student background characteristics and summer reading fluency outcomes?

CHAPTER III

METHODS

Experimental and non-experimental design components were used to examine the summer reading fluency outcomes of students who participated in a summer reading intervention and their peers who did not receive or declined a summer school placement. Students eligible for the program were identified by scoring within a cut score interval on a formative literacy assessment the prior spring (Zvoch & Stevens 2011, 2013). Sections below will (a) describe the demographics of the district, (b) describe the procedures used to conduct a randomized control field experiment, (c) describe the treatment sample, (d) describe the treatment, (e) identify measurement tools and empirical evidence of their reliability and validity , and (f) explain the statistical methods used to analyze the data.

Population, Location, and Treatment

The study was conducted in a moderately- sized Pacific Northwest school district that serves approximately 6,000 students (Zvoch & Stevens, 2011, 2013). In recent years, the student population has been approximately 74% White, 14% Latino, 3% African American, 3% Asian American, 3% Native American, and 2% other. Also, 44% of the district population receives free or reduced lunch services, and site-based instruction is provided to English language learners at three of the elementary schools.

Treatment Assignment. Scores on the Test of Oral Reading Fluency (see below) were used to determine students' eligibility for a summer school placement. Summer school placements were randomly assigned to second grade students who scored within a fluency performance interval (i.e., 70-90 WCPM) on the spring of second grade TORF assessment [1]. Three groups of students emerged from the assignment process: (a) those that were invited and attended, (b) those that were not invited (i.e., the control group) and (c) those that were invited but chose not to attend (i.e., “refusers”).

Analytic Sample. Forty-four second grade students scored within the proficiency interval and were randomly assigned an invitation to attend summer school. Sample demographics were as follows: fifty percent were female (n = 22), and seventy-five percent (n = 33) received free or reduced lunch services the previous year. Other demographic data indicated that twenty-seven percent of the population (n = 12) were ethnic minorities, and no students were classified as English language learners. Of the forty-four students in the assignement interval, thirty students were randomly assigned to summer school. However, only fourteen of the assigned students actually attended summer school resulting in a fifty-three percent refusal rate. Overall, the analytic sample was constituted by participants (n = 14), refusers (n = 16), and control students (n = 13).

Treatment Offered. Summer school was offered to students for five weeks during the months of July and August for 3.5 hours a day. Summer school ran from 8:30 am until 12:00 pm four days a week (Monday through Thursday). Class sizes were small (n < 20), and all classes were held in a single location at a designated school site. Two hours of the total time spent in the program per day were devoted to teacher directed instruction on three key reading components identified by the National Reading Panel (2000): (a) alphabetic understanding, (b) phonemic awareness, and (c) fluency. The rationale for focusing on these areas is that procedural skills in reading have a greater tendency to be retained over time (Cooper et al., 1996). Providing students with

[1] All students below 70 WCPM on the spring TORF assessment were invited to summer school but were not included in the current study.

activities where teachers can model skills and students have opportunities to practice with formative feedback was an integral part of the intervention (Zvoch & Stevens, 2011).

Daily lesson procedures began with all students meeting as a group for the purpose of taking attendance, reviewing homework, and engaging in a warm-up activity. Students then worked in homogeneous groups of 3-5 based on skill level so that instruction and support could be maximized to meet each student's needs (Zvoch & Stevens, 2013). Upon completion of small group work, students were given a short recess, continued by small group work and individual work within literacy stations. Stations made up of curriculum-identified tasks which focus on specific readings skills allowed teachers a greater ability to address individual student needs through differentiation. In the last section of the lesson, students reconvened as a whole group to review concepts addressed during the lesson. The daily class schedule for the five week intervention is presented below.

Summer Reading Program Schedule:

8:30-8:45: Opening

8:45-9:45: Reading Groups

9:45-10:00: Snack and Story

10:00-10:15: Recess

10:15-11:45: Centers – Rotate Every 30 Minutes

11:45-12:00: Closing – Whole Group Activity

Measures

Students' reading performance was measured by the Test of Oral Reading Fluency (TORF). TORF is an individually administered, curriculum based-measure (CBM) (Children's Education's Services, 1987). The test measures accuracy and fluency with written text. TORF scores may be used for multiple purposes: (a) screening-measurement of skills that predict future reading in order to provide information as to who will need additional support at the beginning of the school year, (b) diagnostic-measurement taken anytime during the year to provide information regarding a student's strengths and weaknesses, (c) progress monitoring- assessments conducted at minimum of three times a year to measure progress, identify students who do not demonstrate adequate progress, and evaluate effectiveness of instruction, and (d) outcome measure- to determine if students meet grade level performance (Hasbrouck & Tindal, 2006).

The TORF is calibrated so that each administration is grade specific. Test administration requires students to read a passage aloud for one minute. Each word omitted, substituted, or proceeded by a hesitation lasting for longer than three seconds in identified as an error. Words that are self-corrected are not considered errors. The correct total words per minute is the reading fluency score (Children's Education's Services, 1987). Test-retest reliability ranged from .92-.97, and the alternate form reliability ranged from .89-.94 in one study (Tindal, Marston, & Deno, 1983). Other research shows concurrent validity with DIBELS' Oral Reading Fluency ranging from .92-.96 (Good, Kaminski, Smith, & Bratten, 2001).

The spring of second grade TORF assessment served as the preprogram literacy assessment. Students invited to the summer programs received three test administrations of the second grade TORF form. The initial test administration was given early in week one of summer school. The second administration was given the third week of summer school, and the final administration was given at the end of week five. A post program literacy assessment was given to all students in the fall of third grade using the third grade TORF form. In addition to summer TORF administrations, three administrations of the TORF were given to all students each academic year in September, January, and May.

The change in TORF scores between the spring of second grade and the fall of third grade served as the primary outcome measure for all students. The outcome measure for the second set of analyses on the subset of students who participated in the summer program was the change in TORF scores between the first and last assessment during summer school.

Predictor Variables

Dummy codes were used to form two summer status indicators to estimate and compare student literacy outcomes. The first indicator distinguished between students who were randomly assigned to summer school (whether they attended or not) and their peers who were in the assignment pool, but did not receive an invitation to attend (i.e. the control group). A second set of dummy codes were used to distinguish between treatment participants, treatment refusers, and students in the control group.

Demographic variables were also coded to indicate the students' gender, ethnicity (White or ethnic minority status), and free or reduced lunch status. Age (in months) was also used as a predictor. The rationale for including student background variables was to address possible extraneous characteristics students bring when volunteering for a summer program (Cooper et al., 2000), and to examine if student background characteristics were predictive of changes in fluency outcomes.

Because engagement with treatment is likely to affect program outcomes, treatment engagement was examined as a predictor of literacy outcomes for the subset of students who attended summer school. Student engagement was measured by teacher observation using a five-point scale ranging from zero to four (see Appendix A). Teachers rated each child's engagement with the lesson in each of two daily 45 minute small group instructional periods. A composite score was calculated by taking the mean rating of thirty-two lessons across sixteen days. In addition to engagement, accuracy scores were computed on the same five point scale (see Appendix A). A composite score for accuracy was computed by averaging the twice daily scores over the total of thirty-two lessons across sixteen days. As a second indicator of treatment engagement, attendance was monitored for all students who attended summer school.

Analytic Procedures

Two multiple regression models were used to evaluate summer program performance. The first multiple regression model was specified to derive an ITT estimate that contrasted the summer change in fluency for students assigned to treatment (includes all students assigned to summer school whether they participated or not) and those assigned to the control group. The ITT analysis provides an estimate of the effect of assigning students to summer school. The second regression model (i.e. TOT) estimated the effect of summer school participation by distinguishing between treatment participants, treatment refusers, and those assigned to the control group. The regression models were as follows:

ITT: $Y_{i} = \beta_0 + \beta_1$(assigned to treatment) + e_i

TOT: $Y_{i} = \beta_0 + \beta_1$(treatment participants) + β_2(refusers) + e_i

The ITT model represents a comparison between students assigned an invitation to summer school and students assigned to the control. However, participation was voluntary and a percentage of students who were invited to summer school did not attend. The students who were assigned treatment and did not attend (refusers) represent a proportion of the treatment group who did not receive treatment. Refusers create a bias in the interpretation of the ITT analysis and potentially lead to an underestimation of treatment effects (Bloom, 1985). To account for the refuser effect, a supplemental analysis was conducted using Bloom's non-compliance adjustment (i.e. ITT= $(M_{assigned} - M_{control})/P_{c,}$ Pc is the proportion of the compliers).

For students who attended summer school, additional analyses were run to examine if students who were more engaged in summer school demonstrated greater literacy outcomes, and if student background characteristics were associated with positive changes in literacy outcomes and engagement levels. Descriptive data and a correlation analysis were used to examine relationships between student demographics, student engagement and accuracy levels, and summer school fluency outcomes. Demographic variables used in the correlational analysis were gender, ethnic minority status, free or reduced lunch status, and age.

Sequential regression models were run examining relationships between student predictor variables and changes in summer school literacy outcomes. In the first block, student demographic variables (gender, ethnic minority status, free or reduced lunch status, and age) were specified. In the second block, engagement and accuracy factors were entered. The two regression models were as follows:

$Y_i = \beta_0 + \beta_1(\text{female}) + \beta_2(\text{ethnic minority}) + \beta_3(\text{free or reduced lunch recipient}) + \beta_4(\text{age}) + e_i$

$Y_i = \beta_0 + \beta_1(\text{female}) + \beta_2(\text{ethnic minority}) + \beta_3(\text{free or reduced lunch recipient}) + \beta_4(\text{age}) + \beta_5(\text{engagement}) + \beta_6(\text{accuracy}) + e_i$

CHAPTER IV

RESULTS

Background characteristics for all students as a function of their assignment status are presented in Table 1. As can be seen in Table 1, for students assigned to summer school, 57% percent were identified as male (n = 17), 70% were identified as White (n = 21), and 80% were identified as a free or reduced lunch recipient (n = 24). For students assigned to the control group, 36% were identified as male (n = 5), 79% percent were identified as White (n = 11), and 64% were identified as a free or reduced lunch recipient (n = 9)[2]. Comparisons between summer school assignment status groups revealed that a relatively greater percentage of females, ethnic minority, and free and reduced lunch recipients were assigned to summer school. The control group was also older (98.15 months) relative to the treatment group (97.37 months). However, an independent t-test revealed that the mean age difference between status groups was not statistically significant, $t(42) = 0.92$, $p = 0.36$.

Table 2 presents the scores on the preprogram TORF assessment given in the spring of second grade and scores from the post program TORF assessment given to students in the fall of third grade. As can be seen in Table 2, the range of scores on the preprogram assessment were almost identical between the treatment group (70-88) and the control group (70-89). This is to be expected as students were selected for the study based on an interval scale of 70-90 WCPM on the preprogram assessment in spring of second grade. Also, mean scores on the TORF administration in spring of second grade

[2] One student was dropped from the control group due to missing data on the fall TORF assessment.

Table 1
Demographic statistics of all students in the study by assignment

	Treatment	%	Control	%
Female	13	43	5	38
Ethnic Minority	9	30	3	23
Free or Reduced Lunch Recipient	24	80	9	69

	M	*SD*	*M*	*SD*
Mean age (in Months)	97.37	2.81	98.15	3.76

Notes. The treatment group represents students who were randomly assigned to summer school. The control group represents students who did not receive an invitation.

show that the treatment group (*M* = 80.73, *SD* = 5.45) scored below, but similar to, the control group (*M* = 82.85, *SD* = 6.77). An independent t-test was conducted to examine if mean scores between assignment groups were statistically different during the spring TORF administration. Findings showed that there was no statistical difference between groups on the spring TORF administration, $t(42) = .71$, $p = 0.48$. In contrast, scores on the post program TORF administration showed that the variance of scores between groups increased. Although the treatment group (*M* = 66.27, *SD* = 10.50) had similar mean scores compared to the control group (*M* = 69.08, *SD* = 17.88), $t(41) = 0.65$, $p = 0.52$, the difference in standard deviations was noticeably larger during the fall assessment.

Table 2
Descriptive statistics of TORF scores given in spring of second grade and fall of third grade for students in the treatment and control group.

TORF Scores	Treatment					Control				
	N	*Min*	*Max*	*M*	*SD*	*N*	*Min*	*Max*	*M*	*SD*
Grade 2 Spring	30	70	89	80.73	5.45	13	70	88	82.85	6.61
Grade 3 Fall	30	51	81	66.27	10.50	13	28	98	69.08	17.88

Notes. Students in the treatment group represent those who were randomly assigned to summer school. The control group represents students who were not invited.

Mean TORF scores for the treatment group and the control group at the end of grade 2 and beginning of grade 3 are presented in Figure 1. As can be seen in Figure 1, both groups performed similarly on the spring TORF administration. The change in TORF scores from spring of Grade 2 to fall of Grade 3 was negative for both groups. A paired *t*-test was conducted to examine if group means, for both assignment status groups, were statistically different between spring and fall TORF administrations. For the treatment group, mean scores on the spring and fall TORF administration were statistically different, $t(29) = 8.06$, $p < .001$. Control group mean scores for spring and fall were also statistically different, $t(12) = 3.44$, $p < 0.01$. Furthermore, a comparison of the change in mean scores from spring of second grade to the fall of third grade indicate that the treatment group showed an average loss of 14.46 words-correct-per-minute (WCPM), which was greater than the loss of 13.77 WCPM of the control group.

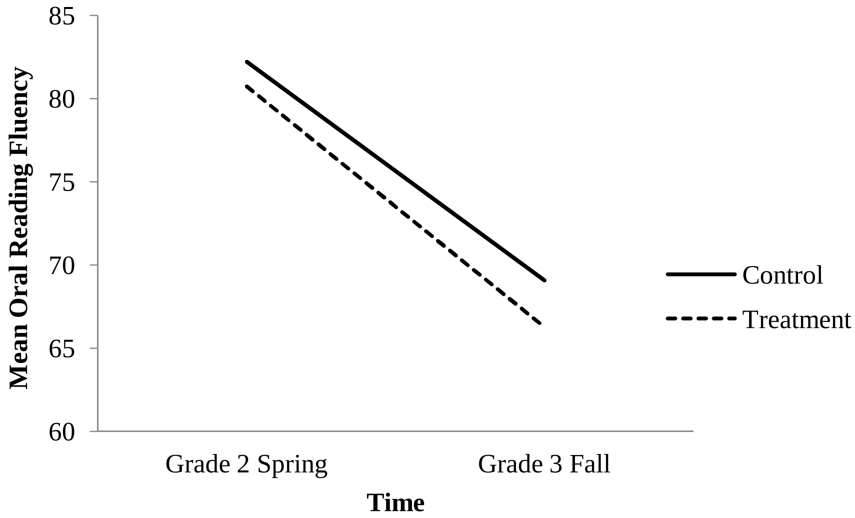

Figure 1. Mean oral reading fluency as a function of summer treatment status and time.

To identify whether student background characteristics were related to the change in TORF scores over the summer, a series of correlations were computed. The correlations between student background characteristics and changes in TORF scores are presented in Table 3. Coefficients were generally low to moderate in size, and only ethnic minority status was statistically related to the change in TORF scores from spring of second grade to fall of third grade (ethnic minority status, $r = .30$, $p < .05$).

Intent to Treat

In order to determine if the descriptive group differences were statistically significant, the ITT model specified in equation 1 was estimated. Results for the Intent-to-Treat model are presented in Table 4**.** The ITT as a predictive model of summer TORF outcomes was not statistically significant, $R^2 = .001$; F (1, 42) = .034, $p = .89$. For students in the control group, the average change in WCPM was a statistically significant

Table 3
Correlations for student background characteristics and the change in fluency outcomes

	Female	Ethnic Minority	Free or Reduced Lunch Recipient	Age (in months)	TORF Change Grade 2 to Grade 3
Female		0.20	0.05	0.06	0.23
Ethnic Minority			0.35*	-0.04	0.30*
Free or Reduced Lunch Recipient				-0.12	-0.05
Age (in months)					0.28
TORF Change Grade 2 to Grade 3					

Notes. * Correlation is significant at the 0.05 level (2-tailed).

loss of 13.77 WCPM. For those assigned to treatment, including refusers, there was a loss of 14.47 WCPM (i.e. -0.70 – 13.77 = 14.47). The unstandardized partial regression coefficient associated with the between group contrast was not statistically significant (b_1 = -0.70, SE = 3.77, t = .185, p = .85). Computation of Cohen's *d* for the -0.70 WCPM between group differences represented an Intent-to-Treat effect size of -0.06 standard deviation (i.e. $g = b_1/SD_p$ = -0.70/11.23 = -0.06).

Table 4
The Intent-to-Treat (ITT) model

	ITT			
	b	*SE*	*ES*	*squared semi-partials*
Intercept	-13.77*	.15		
Assigned to Tx	-0.70	3.77	- 0.06	0.0008
R^2		0.001		

Note. The referent group in the ITT model consisted of students who were randomly assigned to the control group. Standard errors are in parentheses; *ES* = effect size. * $p < .05$.

Results associated with a model examining the unique relationship between student background characteristics and the change in summer TORF performance are presented in Table 5. The ITT model that included student background characteristics was not statistically significant, $R^2 = 0.23$; F (5, 42) = 2.20, $p = .08$. The intercept in this model represents the change in TORF scores from spring of second grade to the fall of third grade for a white, economically advantaged male of average age who was assigned to the control group. For a representative control group student, the average change in summer WCPM was a statistically significant loss of 11.88 WCPM ($p < .05$). For a representative student in the treatment group, an average 12.02 WCPM loss was estimated (i.e. -0.14 – 11.88 = -12.02). The beta coefficient associated with the between group contrast was not statistically significant ($b_1 = -0.14$, $p = .97$). Computation of Cohen's *d* for the -0.14 WCPM group performance contrast revealed a difference of -0.01 of a standard deviation (i.e. $g = b_1/SD_p = -0.14/11.23 = -0.01$).

Table 5
The Intent-to-Treat (ITT) model with background characteristics included

	ITT			
	b	*SE*	*ES*	*squared semi-partial*
Intercept	-11.88*	4.44		
Assigned to Tx	-0.14	4.13	-0.01	0.00004
Female	-4.17	3.31	-0.32	0.03
Ethnic Minority	8.16*	3.90	0.62	0.09
Free or Reduced lunch Recipient	-3.62	4.03	-0.28	0.02
Age (in months)	1.02	0.54	0.08	0.07
R^2		0.23		

Note. The referent group in the ITT model consisted of students who were randomly assigned to the control group. Standard errors are in parentheses; ES = effect size.
* $p < .05$.

As can be seen in Table 5, with the exception of ethnic background, student demographic characteristics (i.e. gender, ethnic minority status, free or reduced lunch recipient, age) were not statistically related to the change in WCPM from spring to fall. The unstandardized partial regression coefficient associated with ethnicity was statistically significant (b_2 = 8.16, SE = 3.90, t = 2.09, $p < .05$). Computation of Cohen's *d* showed that the 8.16 WCPM difference represented a 0.62 of a standard deviation group difference (minority group advantage). Results indicate that students who were of ethnic minority loss less ground over summer than students of the ethnic majority. Overall, the control group did not differ statistically from the treatment group during the summer. The ITT model, with background characteristics, explained 23% of the variation in the change in spring to fall TORF scores.

To further examine summer school participation and to correct for potential bias associated with the effects of non-compliance within the assignment group (Bloom, 1985), Bloom's adjustment was calculated for both of the ITT models (i.e. ITT= ($M_{assigned}$ – $M_{control}$)/ 0.47). Bloom's adjustment in the ITT model without background characteristics revealed a relative loss of 1.49 WCPM for students in the treatment group. Computation of Cohen's *d on the adjusted group difference* estimate resulted in -0.13 of a standard deviation contrast. In the ITT model including background characteristics, adjustment for refusers indicated a loss of 0.30 WCPM for students in the treatment group. Computation of Cohen's *d* resulted in a -0.03 of a standard deviation contrast between groups. Overall, the Bloom's adjustment resulted in an increased group difference in both of the ITT models. However, the change in group differences was not statistically significant.

Treatment on Treated

Background characteristics for all students as a function of their assignment status are presented in Table 6**.** As can be seen in Table 6**,** the participant group was 35% female (n = 5), 28% ethnic minority status (n = 3), and 79% free or reduced lunch status (n = 10). Students in the refuser group were 50% female, 31% ethnic minority, and 81% free or reduced lunch. The statistics reported for the control group were the same as the statistics reported in Table 1. Comparisons using descriptive statistics between participants, refusers, and the control group revealed a greater relative percentage of males, ethnic minority, and free or reduced lunch recipients participating in summer school. A series of Chi-square tests were run to statistically analyze the differences in background characteristic frequencies (i.e. gender, ethnic minority status, and free or reduced lunch recipients) between participants, refusers, and the control group. Findings revealed no statistical differences between status groups in any of the comparisons. Regarding age, the control group (*M* = 98.29, *SD* = 3.65) was older than both the participants (*M* = 97.57, *SD* = 3.09) and the refusers (*M* =97.86, *SD* = 3.18). A one-way ANOVA was run to compare the age of summer status groups. Findings revealed no statistical differences in age between groups, F (2, 42) = 0.86, p = 0.43.

Table 6
Descriptive statistics for summer status groups

	Participants	%	Refusers	%	Control	%
Female	5	35	8	50	9	69
Ethnic Minority	4	28	5	31	3	23
Free or Reduced Lunch Recipient	11	79	13	81	9	69
	M	*SD*	*M*	*SD*	*M*	*SD*
Mean age (*in Months)*	97.57	3.09	97.86	3.18	98.29	3.65

Notes. Participants and refusers represent students randomly assigned to summer school. Students who were not randomly assigned to summer school were controls.

Table **7** presents the mean scores on the preprogram TORF assessment given in the spring of second grade and scores from the post program TORF assessment given to students in the fall of third grade. As can be seen in Table 7, participants (*M* = 78.57, *SD* = 5.96) scored below both refusers (*M* = 82.63, *SD* = 4.53) and the control group (*M* = 82.21, *SD* = 6.61) on the spring administration of the TORF. In general, status groups showed similar group means, standard deviations, and range of scores. An ANOVA was conducted to determine if mean scores between status groups were different on the preprogram spring TORF administration. Findings revealed that there was no statistical difference between the group means, F (2, 43) = 2.16, p = 0.13. On the third grade TORF administration in the fall, group mean comparisons showed that the treatment group (M = 65.21, SD = 10.67) scored below both the refusers (M = 67.19, SD = 10.63) and the control group (M = 69.07, SD = 17.88). Also, there was an increase in disparity between participants, refusers, and the control regarding the size of score range and standard deviation. Both the treatment group and refusers were relatively similar in score distribution. In contrast, the control group showed a much larger disparity in score range and standard deviation relative to the other two treatment groups. An ANOVA was conducted to determine if mean scores between status groups were different on the post program TORF assessment given in fall of third grade. Findings revealed no statistical difference between summer status group mean scores, F (2, 42) = 0.28, p = .75.

Mean TORF change scores from second grade spring TORF scores to fall of third grade TORF scores, for all three groups, are presented in Figure 2. As noted above, group mean scores were not statistically different on the preprogram TORF administration in spring of second grade. Over the summer months all groups lost ground. Paired t-tests were conducted for each group to examine if changes in TORF scores from spring of second grade to fall of third grade were statistically different within groups.

Table 7
Descriptive statistics for the program fluency assessments given in the spring of second grade and the post program assessment given in the fall of third grade

	Participants					Refusers					Control				
TORF Scores	*N*	Min	*Max*	*M*	*SD*	*N*	*Min*	*Max*	*M*	*SD*	*N*	*Min*	*Max*	*M*	*SD*
Grade 2 Spring	14	70	88	78.57	5.96	16	74	89	82.63	4.53	13	70	88	82.85	6.61
Grade 3 Fall	14	51	81	65.21	10.67	16	48	79	67.19	10.63	13	28	98	69.08	17.88

Notes. Participants and refusers represent students randomly assigned to summer school. Students who were not randomly assigned to summer school were controls.

Findings revealed that for participants, *t*(13) = -5.31, $p < .001$; refusers, *t*(15) = -6.02, $p < .001$; and the control group *t*(12) = -3.44, $p < .01$, mean scores were statistically different between the two TORF administrations. Additionally, a comparison of the change in mean TORF scores from spring to fall indicated that the refuser group lost the most ground (15.44 WCPM loss) compared to the control group (13.77 WCPM loss) and participants (13.36 WCPM loss).

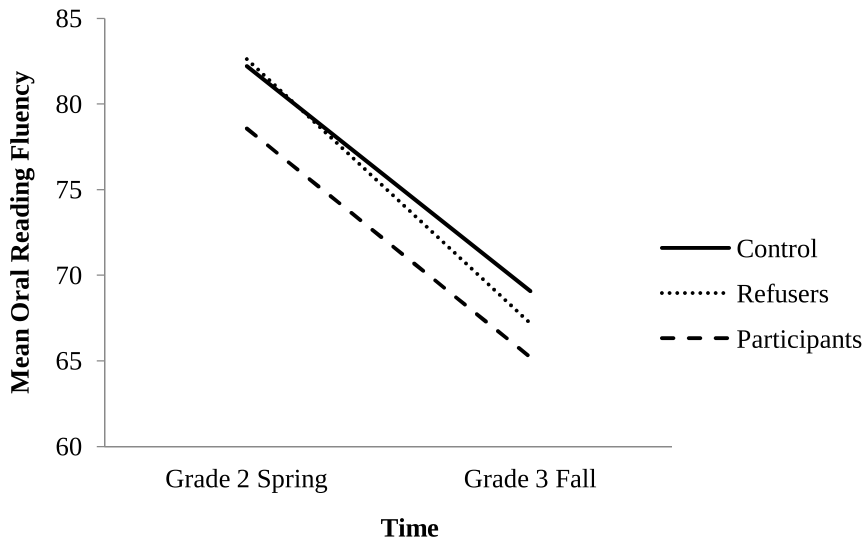

Figure 2. Mean oral reading fluency as function of participants, refusers, the control group and time

In order to determine if the descriptive group differences were statistically significant, the TOT model specified in equation 2 was estimated. Results for the Treatment-on-Treated model are presented in Table 8**.** The TOT as a predictive model of summer TORF outcomes was not statistically significant, $R^2 = .007$; $F(1, 42) = .014$, $p = .87$. For students in the control group, the average change in WCPM was a statistically significant loss of 13.77 WCPM. For those who participated in summer school, there was a non-statistically significant loss of 13.36 WCPM (i.e., 0.41 – 13.77 = 13.36). The beta coefficient associated with the between group contrast was not statistically significant (b_1 = 0.41, p = .92). Computation of Cohen's *d* for the 0.41 WCPM between group differences represented a Treatment-on-Treated effect size of a 0.03 standard deviation (i.e. $g = b_1/SD_p = 0.41/13.01 = 0.03$).

Table 8
The Treatment-on-Treated (TOT) model

	TOT			
	b	*SE*	*ES*	*squared semi-partials*
Intercept	-13.77	3.18		
Participants	0.41	4.42	0.03	0.004
Refusers	-1.67	4.23	-0.15	0.0004
R^2		.007		

Note. The referent group in the TOT model consisted of students who were randomly assigned to the control group. Standard errors are in parentheses; ES = effect size. * p < .05.

Results associated with a model examining the unique relationship between student background characteristics and the change in summer TORF performance are presented in Table 9. The TOT model that included student background characteristics was not statistically significant, R^2 = 0.23; F (6, 42) = 1.80, p = .13. The intercept in this model represents the change in TORF scores from spring of second grade to the fall of third grade for a White, economically advantaged male of average age who was assigned to the control group. For a representative control group student, the average change in summer reading fluency was a statistically significant loss of 11.94 WCPM (p < .05). For a representative student who attended summer school, an average 11.06 WCPM loss was estimated (i.e. 0.34 – 11.94 = 11.06). Also, refusers lost an estimated average of 12.5 WCPM. The unstandardized partial regression coefficient associated with the between group contrast was not statistically significant (b_1 = 0.34, SE = 4.42, t = .093, *p* = .94). Computation of Cohen's *d* for the 0.34 WCPM group performance contrast revealed a difference of 0.03 of a standard deviation (i.e. $g = b_1/SD_p$ = 0.34/11.23 = 0.03).

Table 9
The Treatment-on-Treated (TOT) model with student background characteristics included

	TOT			
	b	*SE*	*ES*	*squared semi-partials*
Intercept	-11.94	4.41		
Participants	0.34	4.25	0.03	0.0004
Refusers	-0.56	4.13	-0.04	0.0001
Female	-4.07	3.38	-0.31	0.03
Ethnic Minority	8.20*	3.96	0.63	0.09
Free or Reduced Lunch	-3.62	4.09	-0.28	0.02
Age (in months)	1.01	0.54	0.08	0.07
R^2	.23			

Note. The referent group in the TOT model represents students who were randomly assigned to the control group. Standard errors are in parentheses; ES = effect size. * p < .05.

Except for minority status, student background characteristics were not predictive of changes in TORF scores from spring to fall. The beta coefficient for ethnic minority status indicated that students identified as ethnic minorities showed a statistically significant 8.20 WCPM relative difference in performance (Cohen's *d = .63*). Results indicate that ethnic minority students showed less summer loss than students who of the ethnic majority. Overall, summer school participants outperformed the refuser group by a difference of 0.90 WCPM and the control group by a difference 0.34 WCPM, during the summer. The TOT model, with background characteristics, explained 23% of the variation in the change in spring to fall TORF scores.

Summer School Participants

For students who were assigned to summer school and participated, three additional fluency assessments were administered during the summer intervention period. Table 10 presents the scores on the preprogram TORF assessment given in the spring of second grade, three summer school TORF administrations, and scores from the post program TORF assessment given to students at the in fall of third grade. As can be seen in Table 10, participants lost 13.36 WCPM from the initial post program assessment given in the spring of second grade to the first administration of the summer TORF assessment week one of summer school. Comparisons of the change in mean scores between beginning, middle, and end TORF administrations show that mean scores increased an average of 15.93 WCPM over the duration of summer school. Also, the distribution of scores widened as twelve of the fourteen students improved on fluency measures and two students remained relatively constant. Comparisons between the ending summer TORF assessment and the TORF assessment given in the fall of third grade showed an average loss of 15.72 WCPM. Trends in fluency scores from spring of second grade to fall of third grade indicate that, despite large gains made during summer school, participants lost 13.36 WCPM overall.

Table 10
TORF scores of participants for preprogram spring assessment, three summer school TORF assessments, and the post programs fall TORF assessment

TORF Scores	*N*	*Min*	*Max*	*M*	*SD*
Grade 2 Spring	14	70	88	78.57	5.96
Summer Beginning	13	52	83	65.00	11.53
Summer Middle	13	56	101	73.77	12.83
Summer Final	14	52	109	80.93	18.25
Grade 3 Fall	14	51	81	65.21	10.67

Notes. Participants represent students assigned to summer school that attended.

Changes in student learning, using group mean TORF scores for spring of second grade, three summer TORF administrations, and fall of third grade are presented in Figure 3. As can be seen, participants lost ground between the end of second grade and start of summer school in July. Participants made substantial gains during the duration of summer school, but showed an overall loss on fluency measures when assessed again in the fall of third grade. A series of paired t-test indicate that participants showed a statistically significant loss in reading fluency during both summer breaks before and after summer school. Changes in mean score on the spring TORF assessment in grade 2 to the beginning summer school assessment in week 1 indicated a statistically significant 13.57 WCPM loss, $t(13) = -5.01$, $p < .001$. Similarly, changes in mean scores from the last summer assessment in week 5 to the fall TORF assessment in grade 3 indicated a statistically significant loss of 15.72 WCPM, $t(13) = -3.78$, $p < .01$. In contrast, while students attended summer school a statistically significant average gain of 15.93 WCPM was observed, $t(13) = 4.07$, $p < .01$. The magnitude of change in fluency scores during summer school was an estimated 1.3 standard deviations. Despite large gains during summer school, overall mean changes indicated that participants lost 13.36 WCPM from spring of second grade to fall of third grade.

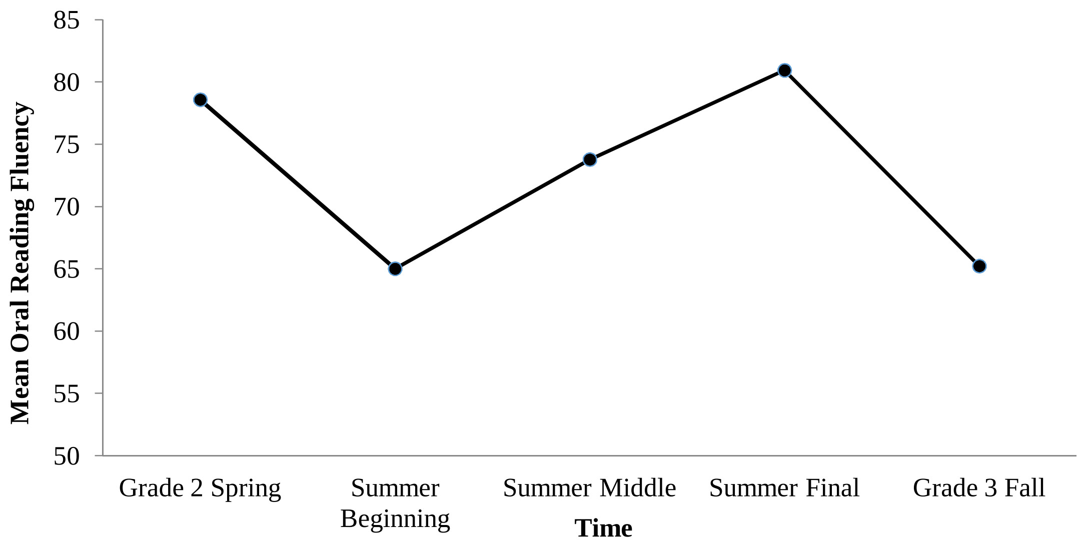

Figure 3. Mean oral reading fluency a as function of participation in summer school and time

To identify whether student background characteristics and student engagement levels during summer school were related to positive changes in fluency outcomes, a series of correlations were computed. Of additional note, attendance was considered as a variable to be included in the combination of background characteristics. However, attendance turned out to be a constant among summer participants (i.e., all students had perfect attendance), and was not analyzed further. The correlations between student background characteristics, engagement and accuracy levels, and changes in TORF scores are presented in Table 11. Coefficients were generally low to moderate in size, and only age was statistically related to the change in TORF scores from the beginning of summer school to the end of summer school (age, *r* = .51, *p* = .07). Other correlation data revealed that ethnic minority status was statistically linked to accuracy scores during summer school (ethnic minority status, *r* = .55, *p* = .04). Results indicate that students of ethnic minority status showed high levels accuracy and engagement during summer school. Also, students who were older tended to outperform students who were younger.

Table 11
Correlations for background characteristics, engagement and accuracy, and changes in TORF outcomes

	Female	Ethnic Minority	Free or Reduced Lunch	Age	Accuracy	Engagement	TORF Change
Female		-0.14	-0.34	-0.01	0.18	-0.05	0.27
Ethnic Minority			0.33	0.15	0.55*	0.49	0.47
Free or Reduced Lunch				0.15	-0.19	-0.03	0.31
Age (in months)					-0.08	-0.23	0.51*
Accuracy						0.47	0.28
Engagement							0.04
TORF Change							

Notes. * Correlation is significant at the 0.05 level (2-tailed).

In order to determine if background characteristics were predictive of changes in TORF scores from the beginning of summer school to the end of summer school, the regression model specified in equation 3 was estimated. Results for the regression model are presented in Table 12**.** The combination of background characteristics (i.e. gender, ethnicity, free or reduced lunch status, and age) as a predictive model of summer school

TORF outcomes was not statistically significant, $R^2 = .58$; F (4, 12) = 2.76, $p = .10$. The intercept in this model represents the change in TORF scores from the beginning to end of summer school for a White, economically advantaged male of average age attending summer school. Although several of the beta coefficients were relatively large in absolute size, no student background characteristics were statistically significant predictors of the change in TORF during summer school. For further reference, the proportion of variance for the change in TORF scores uniquely explained by each of the predictors was $sr^2_{female} = .14$; $sr^2_{ethnic\ minority} = .12$; $sr^2_{free\ or\ reduced\ lunch\ recipient} = .05$; $sr^2_{age} = .19$. Overall the regression model explained 58% of the variance in changes in TORF scores from the beginning to the end of summer school.

Table 12
Regression model predicting change in summer fluency scores with demographic characteristics for students attending summer school

	b	*SE*	*squared semi-partial*
Intercept	4.18	7.43	
Female	10.62	6.34	0.14
Ethnic Minority	10.76	6.31	0.12
Free or Reduced Lunch Recipient	7.32	7.76	0.05
Age (in months)	1.90	1.01	0.19
R^2		0.58	

Note. Participants were randomly assigned to summer school and attended. SE = Standard errors; * p < .05.

Results associated with a model examining the relationship between student background characteristics, engagement and accuracy scores, and the change in TORF performance during summer school are presented in Table 13. The regression model that included student background characteristics and engagement and accuracy scores was not statistically significant, R^2 = 0.60; F (6, 12) = 1.48, p = .32. The intercept in this model represents the change in TORF scores from the beginning to the end of summer school for a White, economically advantaged male of average age with average engagement and accuracy scores.

Table 13
Regression model predicting change in summer fluency scores with demographic characteristics and engagement factors for students attending summer school

	b	*SE*	*squared semi-partial*
Intercept	3.35	8.55	
Female	10.00	7.30	0.13
Ethnic Minority	8.61	11.31	0.04
Free or Reduced Lunch Recipient	8.57	9.68	0.05
Age (in months)	1.89	1.15	0.19
Engagement	-2.57	8.34	0.006
Accuracy	5.60	12.29	0.01
R^2		0.60	

Note. Participants were randomly assigned to summer school and attended. SE = Standard errors; ES = effect size. * p < .05.

Despite a relatively large R^2 value, no student background characteristic was statistically significant predictors of the changes in fluency scores during summer school. The proportion of variance for the change is fluency scores uniquely explained by each of the predictors was sr^2_{female} = .13; $sr^2_{ethnic\ minority}$ = .04; $sr^2_{free\ or\ reduced\ lunch\ recipient}$ = .05; sr^2_{age} = .19; $sr^2_{engagement}$ = .0006; $sr^2_{accuracy}$ = .01. The overall regression model explained 60% of the variance for changes in TORF scores from the beginning to the end of summer school.

Overall Study Comparisons

In order to examine student performance on fluency outcomes across the entire study, descriptive statistics were analyzed by summer status group during second grade, summer school, and third grade. Comparisons of in-school learning rates and summer school learning rates were examined to determine if there were differences in performance between students in a particular status group throughout the study relative to their peers in other status groups. Table 14 presents the scores on the preprogram TORF assessments given in the second grade, summer school TORF scores for students who attended, and scores from the post program TORF assessments given to students in the third grade. As can be seen in Table 14, summer school participants averaged a 41.34 WCPM gain from the beginning of second grade to end of second grade compared to summer school refusers who gained 45.56 WCPM, and a gain of 39.29 WCPM for students in the control group. Summer school participants averaged 15.93 WCPM while attending. A comparison of second grade TORF mean scores and third grade TORF mean scores showed all three groups made less progress in third grade. Refusers made the most progress (37.94 WCPM) compared to participants (35.22 WCPM) and the control (34.38 WCPM). A series of paired t-tests revealed that status group differences in learning rates between second grade and third grade were not statistically significant, participants $t(13) = 1.98$, $p = .07$; refusers, $t(15) = 1.53$, $p = .14$; control, $t(13) = 1.12$, $p = .28$.

Table 14

Average TORF scores for second grade, summer school, and third grade for participants, refusers, and the control group.

	Participants					Refusers					Control				
TORF Scores	*N*	*Min*	*Max*	*M*	*SD*	*N*	*Min*	*Max*	*M*	*SD*	*N*	*Min*	*Max*	*M*	*SD*
Grade 2 Fall	13	15	55	37.23	11.81	16	11	66	37.07	13.43	13	18	60	42.92	11.30
Grade 2 Winter	13	51	73	62.00	7.46	15	47	84	69.40	12.67	14	36	85	71.08	13.45
Grade 2 Spring	14	70	88	78.57	5.96	16	74	89	82.63	4.53	14	70	88	82.85	6.61
Summer Beginning	13	45	83	65.00	11.53										
Summer Middle	13	56	101	73.77	12.83										
Summer Final	14	52	109	80.93	18.25										
Grade 3 Fall	14	48	81	65.21	10.67	16	48	89	67.19	10.63	13	28	98	69.08	17.88
Grade 3 Winter	14	72	100	84.57	7.51	16	59	111	89.00	13.51	13	68	108	86.62	13.09
Grade 3 Spring	14	84	131	100.43	11.88	16	80	136	105.13	17.55	13	66	132	103.46	18.62

Note. Participants and refusers were students randomly assigned to summer school. Students who were not randomly assigned to summer school were controls. The TORF assessment given in Grade 3 increased in passage difficulty relative to the assessment given in Grade 2.

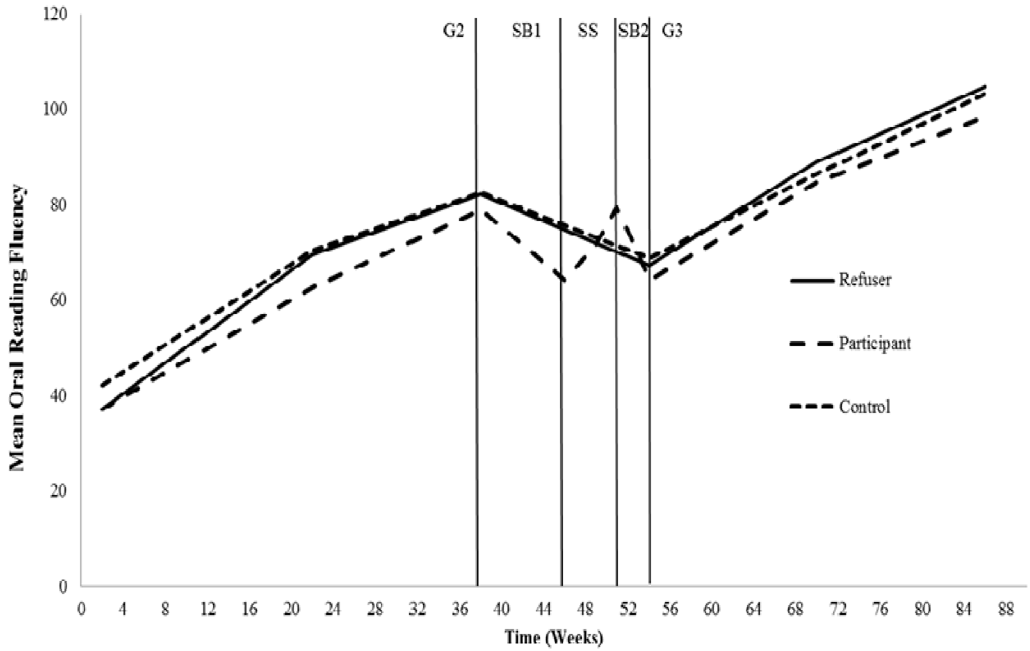

*Figure 4***.** Mean oral reading fluency as function of summer treatment status and time

Student learning trajectories over the entire study are presented in Figure 4**.** As can be seen in Figure 4, student learning rates over the entire study show that participants in summer school averaged a gain of 1.09 WCPM per week during second grade compared to 1.04 WCPM per week during third grade. Over the entire duration of summer, participants lost an average of 0.84 WCPM per week. However, participants averaged a gain of 3.3 WCPM per week while attending summer school. In comparison to summer school participants, refusers of summer school gained 1.14 WCPM per week in second grade and 1.04 WCPM in during third grade. During the summer break, refusers lost an average of .96 WCPM. For students not assigned an invitation to summer school, TORF scores indicate a gain of 1.03 WCPM during second grade, compared to a gain of 1.07 WCPM during third grade. Over the summer, students in the control lost .82 WCPM per week.

Overall, inferential models show non-significant differences between summer status groups. However, trajectories of summer status groups during second grade, over summer, and third grade indicate that, although not statistically different, participants are consistently lower achieving at both preprogram and post program measurement occasions compared to refusers and the control.

Summary of Learning Outcomes

In general, descriptive comparisons of seasonal learning gains indicate that students gain roughly one WCPM per week while in school. Over the summer months, students lost.82 to .96 WCPM. Trajectories for the change in mean fluency scores from fall of grade 2 through spring of grade 2 are similar for all three summer status groups. In comparison with the change in TORF scores during second grade, 3^{rd} grade TORF score gains tended to be lower. In addition, mean comparisons between status groups (i.e. participants, refusers, and controls) on the preprogram spring literacy assessment (second grade form) revealed no statistically significant group differences, F (2, 44) = 2.17, p = .13. Similarly, mean comparisons between student groups on the post program summer literacy assessment in the fall of third grade show no statistically significant group differences, F (2, 44) = .29, p = .75, despite the 15.93 WCPM mean gain for those attending summer school.

The ITT and TOT models indicated that contrasts between groups with and without background student characteristics were not predictive of changes in TORF scores from the end of second grade to beginning of third grade (summer duration). However, ethnic minority status was related to summer fluency change, with students identified as ethnic minorities outperforming their peers in both the ITT and TOT models. Age was also near statistical significance as a predictor of the change in TORF scores from spring of second grade to fall of third grade in both the ITT and TOT models. Older children had less of a decline in summer reading fluency than younger children. All other background characteristics were not predictive of changes in TORF scores.

For participants receiving treatment during the five week summer literacy program, the regression model representing a combination of background variables predicting change in TORF scores during summer school was not statistically significant. Additionally, no background variables were statistically predictive of, or linked to, changes in TORF scores during summer school. Furthermore, the regression model representing the combination of engagement and accuracy and background characteristics was not statistically significant predictor of change in TORF scores during summer school. No variable in the regression model was identified as a statistical predictor of changes in TORF scores during summer school.

CHAPTER V

DISCUSSION

For students at an economic or academic disadvantage, summer school offers an opportunity to engage with academic resources and instructional support during the summer vacation period. Because summer school programs are strategic in the manner they are delivered, these programs are often used by districts as a cost-effective means to support struggling students (McCombs et. al., 2011). However, due to the selective nature in which summer school programs recruit students for placement, it is often challenging to obtain unbiased program effect estimates (Cooper, et al., 2000). In the current study, a collaborative partnership with district personnel enabled the implementation of a field-based randomized trial that allowed for an examination of the effects of assignment to, and participation in, summer school. An exploratory examination of factors associated with summer learning outcomes for students who attended summer school was also performed.

An Intent-to-Treat (ITT) analysis was first conducted to compare students that were randomly assigned to summer school, regardless of participation, and those assigned to the control group. Results indicated that assignment to treatment was not statistically associated with summer reading fluency outcomes and the standardized group difference was small, -.06 of a standard deviation. Due to large refuser rates in the assignment group, Bloom's adjustment was applied to the ITT analysis. However, the adjustment for treatment non-compliance resulted in little change in the treatment effect estimate. Next, a TOT comparison of group means for those who actually participated (i.e. those assigned to summer school and attended) and those assigned to the control condition indicated a slightly more positive effect size of .04 of a standard deviation. However, the standardized difference was again small and the test of the associated beta coefficient was not statistically significant. Overall, the change in mean scores for students assigned to the control revealed a loss of 13.77 WCPM from the second grade TORF administration to the third grade TORF administration. Similarly, the group of treatment refusers lost an average of 15.44 WCPM from the end of second grade TORF to the beginning of third grade. In comparison, summer school participants lost 13.36 WCPM between spring of grade 2 and fall of grade 3 TORF administrations.

The absolute decline in performance for all groups, including summer school participants is somewhat inconsistent with the summer school literature, but was not completely unexpected as the third grade TORF form, by design, is more difficult than the second grade form (Shinn, Gleason, & Tindal, 1989). In particular, studies that track student attendance in summer reading programs tend to report that students with high levels of program attendance show an increase in reading achievement (Borman & Dowling, 2006; Schacter & Jo, 2005; Zvoch & Stevens, 2013). In the current study, students who participated in summer school had perfect attendance, and yet changes in reading scores declined from spring of second grade to fall of third grade. However, findings in the current study were consistent with other studies in that refusers of summer school and the control group lost ground over the summer months. The inconsistency of findings for summer school participants is most likely due to the discrepancy in form difficulty between pretest and posttest measures.

Comparisons with Other Research

The summer literacy performance of those who participated in summer school compared to students who refused treatment or were assigned to the control were somewhat distinct relative to the meta-analytic findings of Cooper et al. (2000). More specifically, participants assigned to summer school who participate tend to outperform refusers and students assigned to the control. In contrast, summer school participants in the current study showed no statistical differences relative to refusers and the control in both the ITT and TOT models. Furthermore, effects estimated in comparable studies calculated scores from pre/post assessments that were equated at the same grade level, whereas the current study measured preprogram scores and summer treatment scores with TORF assessment forms aligned for second grade, and posttest assessment forms aligned for third grade. Because forms were not equated at the same grade level, it is reasonable the effects estimated in the current study do not closely compare to other similar studies.

In addition to comparisons with refusers and the control group, the pre/posttest summer school gains of students participating in summer school were much larger in magnitude, $d = 1.3$ relative to that of Cooper et al., (2000)'s findings of .25 of a standard deviation. Cooper and colleagues caution however that the single group pre and posttest design is the weakest approach for studying summer school effects. With no comparison group, the change in summer fluency scores for summer school participants does not directly inform school personnel of the program's efficacy. In isolation, the 1.3 standard deviation change in summer school outcomes only provides suggestive evidence of the effect of offering summer school. However, when in-school learning rates were compared with the summer learning rates for students attending summer school, an increase of 2.2 WCPM per week was observed during summer school. Increased learning rates for summer attendees provides some additional evidence of the benefit summer programs may provide for students who actually attend and engage in summer school.

Current results are also distinct from other randomized field trials conducted on early-elementary school student samples. Results from the literature generally demonstrate positive literacy gains for students who participated in summer school, relative to students in a control group (Borman et al., 2009, Schacter & Jo, 2005; Zvoch & Stevens, 2013). In contrast, the current study indicates that students who attended summer school did not show gains in literacy skills over the summer. In addition, two randomized field investigations with students reported as economically-disadvantaged from large urban environments and one study partially comprised of economically-disadvantaged students from a medium-size city in the Pacific Northwest (Borman et al., 2009; Schacter & Jo, 2005; Zvoch & Stevens, 2013) reported that student background characteristics were generally not predictive of summer literacy outcomes. However, minority status was identified as a statistically significant predictor in the current study, suggesting that different relationships may be found in different samples. When examined relative to other randomized field experiments that adjusted for non-compliance among the treatment group, current results were somewhat distinct. For example, , Borman and Dowling (2006) examined effects of summer school on children attending the Teach for Baltimore program, a seven week academically intensive program that selected students from ten low-income school districts. Borman and Dowling (2006) applied Complier Adjustment Causal Effect (CACE) analysis to treatment literacy outcome data and found summer treatment effects when accounting for non-compliance. However, in the current study, findings indicated that even with Bloom's non-compliance adjustment, (i.e., the adjusted ITT model) group differences remained small and not statistically significant.

It is important to note that the CACE analysis is generally considered superior to the Bloom's non-compliance adjustment. Bloom's adjustment compares the treatment group with the control group by dividing the percentage of students who complied into the difference in means scores between the treatment assignment group and the control group. Alternatively, CACE analysis uses an index of student characteristics to predict the likelihood of compliance based on the students who attend. For example, Borman and Dowling (2006) used in-school student attendance rates as their complier index. The probability of compliance is used to weight students in the control group based on their probability of compliance. As a result, the CACE analysis more accurately compares program participants and students in the control who would be program participants if they were assigned.

Discrepancies between compliance adjusted studies (Borman & Dowling, 2006; Zvoch & Stevens, 2013) and the current study may also be due partially to the grade level of students sampled. The previous studies sampled students moving from kindergarten to first grade and first grade to second grade. The current study examined second grade students transitioning to third grade. Grade level may be a distinct factor in program outcomes. For example, students in the current study represent students who have shown a continuous sign of struggling to read in kindergarten, first grade, and second grade. Because these are older students with more history, two possible circumstances could influence students' performance or attendance. First, grade-level reading skills and text difficulty is different for students moving from second to third grade than for students progressing grade levels from kindergarten or first grade. Changes in scores on assessments at the second and third grade level may not generalize to changes in scores for assessments at earlier grade levels. Second, in consideration of Borman, Overman, and Benson (2005)'s conclusion that parental effort to support attendance partially explains achievement difference in summer school learning, parental values and expectation may differ across grade levels. As a result, studies that compare program outcomes affected by self-selection factors associated with attendance may lead to diverse program estimates across grade levels. Although effect sizes in the current study were not comparable to results reported in the other studies, the ITT and TOT results support Zvoch and Stevens (2013)'s observation that assignment to treatment in and of itself is insufficient for positive learning outcomes.

The discrepancy in results may also be due in part to practice effects associated with the measure of oral reading fluency. Practice effects could influence fluency scores for students of different backgrounds and with different home opportunities. Increased practice using grade-specific vocabulary would greatly increase the student's ability to recognize sight words resulting in improved speed or a WCPM response. Changes in

TORF scores from fall of grade 2 to spring of grade 3 demonstrated that time away from the academic environment lead to decreased fluency outcomes. Furthermore, positive student gains in both summer school and during three sequential grade 3 TORF assessment points indicated that when students engaged in practice and were provided teacher instruction in an academic environment, fluency outcomes increased. However, it is unknown how much time each student spent practicing oral reading in their at-home environment during the academic school year or during summer. Differences in the amount of time students spend practicing on their own may result in smaller or greater changes in fluency scores for individual students. Thus, programs that use assessments sensitive to procedural skills to measure programs outcomes should be mindful of the frequency and intensity students practice those skills outside of program.

Treatment Refusers

Refuser rates were comparable to those reported in other randomized trials. The refusal rate in Borman and Dowling (2006) ranged from 21%-43% across 10 sites. Zvoch and Stevens (2013) reported refuser rates for a kindergarten sample at 45% and 21% for a first grade sample. In comparison, the refuser rate for the current study was 53% of the second grade sample invited to summer school. Closer analysis of preprogram TORF assessment data during grade 2 suggests the relatively large refusal rates could be partially due to second grade and final spring fluency scores on the TORF. Raw scores on the second grade TORF given in the spring of second grade show that of the sixteen students who refused treatment only four of them scored below 80 WCPM, whereas ten of the fourteen students who participated in summer school scored below 80 WCPM. The proportion of students refusing summer school who scored above 80

WCPM and the proportion of students attending summer school who scored below 80 WCPM may indicate that students (and parents) based participation decisions on cues of academic performance.

In addition, only four students in the control group scored below 80 WCPM on the spring TORF assessment. The discrepancy between participants and the control group further explains why the control group outperformed the participant group on the grade 3 TORF assessment. Moreover, the mean TORF score on the grade 3 TORF assessment for students in the control group who scored below 80 WCPM on the grade 2 TORF assessment was 46.67 WCPM. Students in the participant group who scored below 80 WCPM scored much higher on the grade 3 TORF assessment with an average score of 61.60 WCPM. Comparisons between participants and control students demonstrate that for students of similar achievement levels, self-selection may impact outcomes associated with inferential models. As a result, data provided to educational administrators and stakeholders regarding the causal impact of summer school treatment that does not account for the manner and magnitude of refusal may be misleading.

A possible explanation for why the grade 2 spring assessment scores influenced the self-selection of students invited to summer school is that preprogram assessment may have influenced teachers' perceptions and beliefs regarding the necessity of each child to attend summer school. The five weeks remaining in school could provide an opportunity for teachers to recruit or dissuade students to attend summer school by making recommendations to parents and students based on results from the spring second grade fluency assessment. In addition to teacher influences, parents greatly contribute to whether a child attends summer school or not. Several factors may affect a parent's decision to enroll a student. As parents are often provide transportation to and from summer school, the duration of time per day the student attends summer school and the distance the parent is required to drive to transport their child could be considered too inconvenient relative to learning gains a student might make. Also, summer school is often seen as a form of daycare for some parents (Cooper et al., 2000; McCombs et al., 2011). Whether the program supports parental responsibilities of providing meals, an opportunity to recreate, and the length of supervision per day may make a program more or less desirable. Furthermore, because site-based summer school requires seat time for students, family time and other summer activities (e.g. sports, vacations, and religious events) could also be competing factors in the decision for the student to attend. Ultimately, parents must weigh their perceptions regarding the opportunity costs of attending summer school with student achievement data and possible teacher recommendation when deciding whether or not to enroll their child.

All factors considered, the greatest contributor to refuser rates in the current study was likely students' scores on the spring TORF exam in second grade. However, the study did not collect other home and environmental data such as educational resources available in the home, parent's level of education, whether or not one or both parents were working during day hours, or an index of other non-educational interests being pursued during summer. Additional survey measures may better inform educators as to why students (and parents) refuse summer school. It is recommended that voluntary need-based summer programs use multiple measures to gather background data that will better inform administrators and teachers of environment-specific factors influencing refuser rates.

Benefits of Summer Programs

Despite not identifying a causal effect of summer school, results demonstrated learning gains during summer school for students that actually attended. For students attending summer school, changes in mean scores from the beginning of summer school and the last week of summer school showed a 15.93 words-per-minute increase ($d = 1.3$). However, the observed gains should be considered with respect to the environmental context and the fidelity with which the summer program was delivered to students. The program was designed to provide an intense dose of daily direct reading instruction in a small group environment. The explicit and scripted lessons were aligned with academic school year curriculum and were carried out by highly trained and experienced teachers that utilized best practice instructional strategies. Students were provided modeled instruction, opportunities to perform guided practice of skills, and individualized student feedback. Furthermore, minimal deviation from protocol was observed between teachers during the five week program (Zvoch, 2012). Fidelity data from 26 ratings within Zvoch's study revealed that provider instruction was computed at 84.9% with a standard deviation of 14.7% demonstrating high fidelity and low variance instruction. Moreover, summer school fidelity data revealed that ecological factors related to students' adherence to treatment may have been optimal in the current study and thus would not be expected for other summer programs. More specifically, because the program adhered to strict protocols with strong fidelity and students showed high levels of attendance and adherence to treatment, it is likely that effects reported from this study are at the upper bound for what can be expected relative to similar summer programs.

In addition to high fidelity instruction from teachers and high engagement from students, the large fluency gains for participants during summer school should be interpreted with respect to the focus of instruction that was provided. More specifically, this summer program delivered an intense dose of instruction that supports basic reading skills (i.e. oral blending and segmentation, decoding and phonic analysis, and speed and accuracy in reading connected text). The combination of strong instructional fidelity, an instructional focus on procedural reading skills, and proximal fluency outcome measures likely resulted in estimates that over represent summer school gains. Also, due to the nature of fluency measures, which are sensitive to the amount of time in which students engage in practice, the time students practiced reading at home may have acted as a moderator between the treatment delivered and changes in TORF scores. Furthermore, the gains made by summer participants are based on procedural skills and thus these results may not generalize to higher order reading skills such as reading comprehension.

A summer school treatment focused on fluency scores that aligned with second grade passage difficulty (summer program curriculum, pretest, and posttest measures were based on second grade TORF outcomes), effects from summer school should be interpreted more as remediation gains rather than acceleration gains (McCombs et al., 2011).

Engagement in Summer School

In addition to well-planned fidelity that ensured the treatment provided an intense dose of reading instruction in a small group environment by trained teachers, student engagement data was analyzed to examine fluency outcomes associated with student engagement and accuracy. In general, engagement and accuracy scores were not statistically predictive of changes in fluency scores for students attending summer school. However, descriptive statistics and correlation data suggest that for students of ethnic minority status, increased engagement and demonstrated accuracy was associated with greater fluency gains. More specifically, three of the four students identified as ethnic minority status were observed to have above average engagement relative to attending students, as well as greater than average demonstrated accuracy during treatment. However, because the participant sample was small, findings in the current study may not generalize to other studies or different samples. Of additional note, the engagement rubrics used to score student engagement and accuracy were scored on a five point scale. Due to the small range of scores possible, composite scores for both engagement and accuracy may not be sensitive enough to distinguish differences in overall student adherence to treatment. As a result of the tool's relative sensitivity, inferential models within the current study were limited in their ability to detect significant effects associated with student engagement and accuracy regarding changes in fluency scores during summer school.

Study Limitations

It is requisite to consider the limitations in the current study to further contextualize the results discussed prior. First, because the summer school program was administered within the context of a moderately-sized school district in the Pacific Northwest, estimates between summer participants, refusers, and students in the control, may not generalize to summer programs with dissimilar population demographics, instructional focus and dosage, or environmental settings. Second, due to the voluntary nature in which students were sampled for the study, students in the treatment group who received an invitation to summer school but refused treatment led to a large refuser bias (53%) in the ITT analysis. As a result, comparative estimates representative of the causal impact for assigning summer school were confounded as over half the students estimated to receive treatment did not attend summer school. Without correction for refusers in the ITT model, between-group contrasts underrepresent effects associated with assignment to summer school.

Other study limitations created challenges when interpreting results from study outcome measures. The scores in second grade and summer school represent changes on the second grade TORF form, and third grade changes represent performance on the third grade TORF form. Therefore, form effects may have influenced the summer fluency outcome estimates. Other studies indicate that curriculum-based measures that are not vertically and horizontally articulated within or between grade-levels may lead to differences in passage difficulty between forms. As a result, changes in means and variances within score distributions may be confounded by differential form difficulty (Betts, Pickart, & Heistad, 2009). As noted by Zvoch and Stevens (in press), the variance in passage difficulty that would undermine estimates derived at different assessment points of fluency measures is improved by using a median score across three separate TORF forms given in the same test administration (Children's Education Services, 1987). However, using a median score across three forms of the same difficulty would not necessarily improve estimates from forms across grade levels.

To further contextualize the difference in form difficulty between grade levels, oral reading fluency norms (Hasbrouk & Tindal, 2006) indicate that a student in the spring of second grade reading with speed and accuracy in the 50th percentile has a

WCPM score of 89 WCPM. In contrast, a student in the fall of third grade reading with speed and accuracy at the 50th percentile has a WCPM score of 71 WCPM. The 18 WCPM difference between spring of Grades 2 and fall of Grade 3 provides reasonable evidence and a rationale as to why TORF score changes between the preprogram and post program assessments were negative in direction.

In consideration of the positive changes in TORF scores for summer participants, in addition to high fidelity of instruction and possible practice effects, statistical regression to the mean may have inflated summer gains. Participants attending summer school demonstrated low initial summer school TORF scores (65 WCPM). As initial summer school scores were well below participant preprogram TORF scores in the spring of Grade 2 (78.57 WCPM), it would be expected that ending summer school TORF scores would shift toward prior mean scores.

Additionally, statistical power was a major limitation when detecting effects and interpreting study results. A power analysis for the current study indicated that, for the regression models run, a sample size of 68-98 is required to detect a moderate effect sizes (f^2= .15). The actual study sample analyzed was much smaller (N = 43). Given the sample size of the current study, the sensitivity of the regression models run could only detect effects sizes that were larger than f^2 = .23-.37. In addition, an analysis of statistical power for correlations indicated that the sensitivity for ITT and TOT models (n = 30) was r = .56, and for models including only summer participants (n = 14) sensitivity was r = .72.

Given the low statistical power within the current study, factors previously interpreted in the ITT, TOT, and summer school analyses could tentatively have implications for practical application. For example, although age was not statistically significant as a predictor of changes in TORF scores in the ITT and TOT analyses, the unique variance explained by age (sr^2 = .07) suggests that developmental differences between students may account for some of the variance in TORF score changes between summer status groups.

Furthermore, ethnic minority status was a statistically significant predictor of changes in TORF scores in both the ITT and TOT analyses. Findings indicate that ethnic minority students lost less ground compared to students who were identified as ethnic majority status. Moreover, correlational data revealed that ethnic minority status was statistically linked to free or reduced lunch status (r = .35) and also statistically linked to TORF Score Change (r = .30). Taken together, results from the correlational data and the ITT and TOT models suggests that student background factors may be independent predictors of summer learning outcomes.

For students who attended summer school, similar inferences are supported in regression models that analyzed student background characteristics as a predictor of change in TORF scores during summer school which revealed large R^2 values ranging from .58-.60. Within the predictive models, the background factors age (sr^2 = .19); ethnic minority status (sr^2 = .12), and gender (sr^2 = .14) were the greatest contributors to the unique variance explained for changes in TORF scores during summer school. Additionally, moderately-sized correlation coefficients indicate that engagement may be linked to ethnic minority status (r^2 = .49), and ethnic minority status may be linked to the change in TORF scores during summer.

Tentative interpretations of the data reported suggest that developmental differences, background factors, and students' levels of engagement with treatment are possible contributors to treatment outcomes during summer school. As accurate estimates of treatment impact on students learning is essential when making decisions regarding the initial design or changes to existing summer programs, administrators should pay close attention to the environmental context (e.g. student population, type of instruction provided, the students' level of engagement) in which results are reported.

Recommendations for Summer School

The summer school literature for voluntary site-based programs demonstrates that additional support over summer months provides assistance to education leaders in two capacities. The first being that students showing signs of academic stagnation or a lack of progress toward grade-level targets may receive additional instruction in content-specific areas at a portion of the cost of in-school instruction (McCombs et al., 2011; Schacter & Jo, 2005; Zvoch & Stevens, 2013). Second, divergent learning trajectories for students of different economic backgrounds may be minimized through summer school attendance as the faucet of educational resources, shut off to many students during summer, continues to flow through internal district programs and external state and national programs (Alexander et al., 2001; Borman, Benson, & Overman, 2005; Borman & Dowling; 2006; McCombs et al., 2011). Because selection methods, strength of treatment dose, treatment fidelity, and outcome measures vary between programs, greater examination of program-specific environmental factors is requisite to further contextualize outcomes within summer programs.

The literature on summer learning and summer school that examines the causal effects of providing treatment to targeted student populations identifies several methodological challenges that undermine the validity of inferences drawn from summer school outcomes (Cooper et al., 2000). In consideration for such challenges, more research on the effect of summer school is needed using diverse study samples, explicit and transparent documentation of program fidelity protocols (e.g. instructor training, instructional observations, tracking of assigned homework , measurement of student engagement during treatment), and the use of both formative and larger norm-referenced achievement measures. As summer school is often provided to students based on need (academic or economic) and attendance is voluntary, increased student background data on the amount and types of educational resources students have access to in the home and community will better inform researchers in describing a more complete picture of student opportunities and learning over the summer. Also, data on student opportunities and resources may further serve to provide researchers and education leaders a more complete context regarding the reasons students (and parents) accept or decline an invitation to summer school as well as a more in-depth profile of attending summer school students.

The examination of summer school effects also requires attention to the manner students are assigned a summer school placement. Voluntary programs that target economically and academically disadvantaged students should plan research designs that utilize random assignment when possible (Borman et al., 2005; Borman & Dowling, 2006). When not feasible, ethical alternative methods such as the regression discontinuity and interrupted time series design (Zvoch & Stevens, 2013) can be used to derive unbiased estimates. It is also important that researchers pay close attention to treatment non-compliance as large refuser rates will often lead to an underestimation of treatment effects. Furthermore, inferences should be contextualized within the ecological framework in which the study is situated. Although explicit fidelity data may serve to better inform administrators of a program's operation, the context in which an effect is observed may evolve over time. Given the notion that community and student factors may change over time, multi-year studies may better inform educators and stakeholders regarding need and effectiveness of a particular program. Such data is beneficial when applied to inform decisions regarding instructional practice, curricular needs, and resource allocation.

Because education leaders must make decisions about summer programs with limited financial resources (Borman & Dowling, 2006; Zvoch & Stevens, 2013; Schacter & Jo, 2005), programs that provide data on the treatment's cost-effectiveness will benefit administrators and stakeholders regarding opportunity costs and achievement criteria for program success (Copper et al., 2000; McCombs et al., 2011). Programs that report the student cost-per-slot in addition to unbiased effects of treatment creates the opportunity to compare program expenditures with program learning gains. As a result, education leaders and stakeholders could more easily determine the most cost effective method to provide additional resources and instructional support for disadvantaged students at a greater economic and academic need. For example, programs that provide many recreational activities increase the cost of the program (McCombs et al., 2011) but may not increase overall learning gains. Conversely, recreational activities may increase attendance and compliance rates which serve to increase the effectiveness of instructional time within a program. Similarly, a smaller teacher to student ratio creates greater instructor expenditures making the cost-per-slot of a program more expensive, but the smaller teacher to student ratio may also increase treatment effects. In other words, program outcomes as a function of treatment type and the manner in which treatment is provided is relative to costs of providing said treatment. Furthermore, because not all summer programs have the same program goals, additional measures (e.g. delinquency data for the following year, in-school attendance rates, reported school and community activity involvement) may also be used to compare program cost effectiveness.

In consideration of programs reporting summer learning gains and program expenditures as a means of assessing program effectiveness, McCombs et al. (2011) notes that initial startup cost of summer programs would overestimate the student cost-per-slot during in the first one or two years of the program. Evaluations that compare learning gains to program expenditures should consider adjusting expenditures reported by subtracting expenditures required for starting the program, or report the comparison after the program has been running for more than one year. Also, because program outcomes are often the result of a complex environment-specific experience, program administrators should take into account population demographics, quality of instruction, and home factors relative to in-school learning when planning program goals and success criteria (Cooper et al., 2000, pg. 104).

CHAPTER VI

CONCLUSION

Summer holds the potential to be a time for students to explore new interests, to engage in activities that nurture creativity, and to spend additional time with family and friends. But for some students, summer adds additional challenges as resources available during the academic school year dry up, limiting opportunities to maintain or improve academic gains made the prior school year. Disadvantaged children in early grades often find themselves behind when they reengage the academic environment in the fall relative to their more advantaged peers. Students who lack proficiency in academic skills are greeted with increasingly more difficult challenges as the educational process progresses. As a result, struggling students tend to disengage from activities or behaviors of learning that improve reading outcomes. The cyclical process of poor reading performance and further avoidance of academic behaviors leads to greater deficiencies over time. Deficiencies cumulate and eventually result in higher dropout rates, fewer vocational opportunities beyond high school, and a greater burden to society overall.

Education leaders have taken measures to stem the summer slide and increase education opportunities and resources for disadvantaged and at-risk students with the goal that all children can read by the third grade. In light of such efforts, summer school has become a popular choice for school administrators as resources can be strategically allocated to target said students. Results from the current study indicate that despite the ineffectiveness of assignment to summer school, students who attended showed substantial gains in fluency scores during the summer school period. However, because no control group was available for direct comparison, it is difficult to gauge the true impact of the summer literacy program here. Alternatively, the ITT and TOT models revealed no statistically significant difference between-group effects for fluency outcomes over the entire summer. Consideration of possible form effects and the fact that the study was statistically underpowered may explain in part the inability to detect effects associated with treatment assignment. Additionally, the high refuser rate further highlights the need to pay close attention to the proportion of summer school refusers in future studies.

In general, the provision of a strong dose of summer instruction by trained teachers in combination with high levels of student engagement may serve to offset the summer slide. However, different summer programs implemented in different contexts may not produce results that are consistent with those reported elsewhere. Replication of the current study using larger and more diverse samples, vertically-equated and comprehensive-reading measures, and diverse geographic locations will better serve to gauge the size and direction of the summer school effect.

Summer school has been a viable and highly utilizable means for providing additional support for disadvantaged students for over a quarter century. With increasing policy demands, greater student need, and limited educational resources, education leaders require accurate and unbiased data to make difficult decisions regarding resource allocation. To the extent that researchers can provide education leaders with more descriptive measures that explain students' summer learning experiences in greater depth and provide unbiased treatment impact estimates , district personnel and stakeholders will be better equipped to make decisions regarding the implementation of cost-effective programs that provide the greatest benefit to students with the greatest need.

APPENDIX

STUDENT DAILY MONITORING FORM

Interventionist_____________________________

Student Daily Monitoring Form

At the end of each session, please rate each student's ***accuracy*** and ***engagement*** during the instructional time you spent with the students. Also indicate if they were in **attendance** as well as completed their **homework** (if appropriate) for that day.

Date: **Reading Session #1**

✓ Accuracy of Student Responses During the Lesson					
Student	**Highly Accurate**	**Very Accurate**	**Somewhat Accurate**	**Not Very Accurate**	**Not Accurate**
	4	**3**	**2**	**1**	**0**
	4	**3**	**2**	**1**	**0**
	4	**3**	**2**	**1**	**0**
	4	**3**	**2**	**1**	**0**
	4	**3**	**2**	**1**	**0**
	4	**3**	**2**	**1**	**0**
	4	**3**	**2**	**1**	**0**
	4	**3**	**2**	**1**	**0**

Engagement of Student Attention During the Lesson					
Student	Highly Engaged	Very Engaged	Somewhat Engaged	Not Very Engaged	Not Engaged
	4	3	2	1	0
	4	3	2	1	0
	4	3	2	1	0
	4	3	2	1	0
	4	3	2	1	0
	4	3	2	1	0
	4	3	2	1	0
	4	3	2	1	0

Attendance & Homework Completion

Please indicate if students were in attendance and completed their homework (if appropriate) for today.

Student	Absent?	Homework**		
	Y	Y	N	N A
	Y	Y	N	N A
	Y	Y	N	N A
	Y	Y	N	N A
	Y	Y	N	N A
	Y	Y	N	N A
	Y	Y	N	N A
	Y	Y	N	N A

**NA=not applicable (No homework expected today)

REFERENCES CITED

Adams, M. J. (1990). Beginning to read: Learning and thinking about print. *Beginning to read: Learning and thinking about print*. Cambridge, MA: MIT Press

Alexander, K. L., Entwisle, D. R., & Olson, L. S. (2001). Schools, achievement, and inequality: A seasonal perspective. *Educational Evaluation and Policy Analysis, 23*(2), 171-191. doi: 10.3102/01623737023002171

Alexander, K. L., Entwisle, D. R., & Olson, L. S. (2007). Lasting consequences of the summer learning gap. *American Sociological Review, 72*(2), 167-180. doi:10.1177/000312240707200202

Alexander, K. L., Entwisle, D. R., & Horsey, C. S. (1997). From first grade forward: Early foundations of high school dropout. *Sociology of Education, 70*(2) 87-107.

Allington, R. L., McGill-Franzen, A., Camilli, G., Williams, L., Graff, J., Zeig, J., ... & Nowak, R. (2010). Addressing summer reading setback among economically disadvantaged elementary students. *Reading Psychology*, *31*(5), 411-427. doi: 10.1080/02702711.2010.505165

Arnold, E. M., Goldston, D. B., Walsh, A. K., Reboussin, B. A., Daniel, S. S., Hickman, E., & Wood, F. B. (2005). Severity of emotional and behavioral problems among poor and typical readers. *Journal of Abnormal Child Psychology*, *33*(2), 205-217.

Angrist, J. D., Imbens, G. W., & Rubin, D. B. (1996). Identification of causal effects using instrumental variables. *Journal of the American Statistical Association*, *91*(434), 444-455. doi**:**10.1080/01621459.1996.10476902

Baker, L., & Wigfield, A. (1999). Dimensions of children's motivation for reading and their relations to reading activity and reading achievement. *Reading Research Quarterly*, *34*(4), 452-477.

Ballinger, C. (2000).Changing time: Improving learning. *High School Magazine*, *7*(9), 5-8.

Barger, J. (2003). Comparing the DIBELS oral reading fluency indicator and the North Carolina end of grade reading assessment. Asheville: North Carolina Teacher Academy. Retrieved October, 31, 2004.

Barton, P. E., & Coley, R. J. (2010). The Black-White achievement gap: When progress stopped. Policy Information Report. *Educational Testing Service*.

Betts, J., Pickart, M., & Heistad, D. (2009). An investigation of the psychometric evidence of CBM-R passage equivalence: Utility of readability statistics and equating for alternate forms. *Journal of School Psychology, 47*(1), 1-17. doi: 10.1037/a0022987

Bloom, H. S. (1984). Accounting for no-shows in experimental evaluation designs. *Evaluation Review, 8*(2), 225-246. doi: 10.1177/0193841X8400800205

Bodovski, K., & Farkas, G. (2007). Mathematics growth in early elementary school: The roles of beginning knowledge, student engagement, and instruction. *The Elementary School Journal, 108*(2), 115-130. doi 10.1086/525550

Boetsch, E. A., Green, P. A., & Pennington, B. F. (1996). Psychosocial correlates of dyslexia across the life span. *Development and Psychopathology*, *8*(03), 539-562. doi.org/10.1017/S0954579400007264

Borman, G. D., Goetz, M. E., & Dowling, N. M. (2009). Halting the summer achievement slide: A randomized field trial of the KindergARTen Summer Camp. *Journal of Education for Students Placed at Risk*, *14*(2), 133-147.

Borman, G. D., Benson, J., & Overman, L. T. (2005). Families, schools, and summer learning. *The Elementary School Journal, 106*(2), 131-150. doi: 10.1086/499195

Borman, G. D., & D'Agostino, J. V. (1996). I and student achievement: a meta-analysis of federal evaluation results. *Educational Evaluation and Policy Analysis, 18*(4), 309-326. doi: 10.3102/01623737018004309

Borman, G. D., & Dowling, N. M. (2006). Longitudinal achievement effects of multiyear summer school: Evidence from the Teach Baltimore randomized field trial. *Education Evaluation and Policy Analysis*, *28*(1), 25-48. doi: 10.3102/01623737028001025

Bryk, A. S., & Raudenbush, S. W. (1988). Toward a more appropriate conceptualization of research on school effects: A three-level hierarchical linear model. *American Journal of Education*, *97*(1), 65-108. http://www.jstor.org/stable/1084940

Buck, J., & Torgesen, J. (2003). The relationship between performance on a measure of oral reading fluency and performance on the Florida Comprehensive Assessment Test. Tallahassee, FL: Florida Center for Reading Research.

Casey, A. E. (2010). Early Warning. *Why Reading by the End of Third Grade Matters.* A KIDS COUNT Special Report for the Annie E. Casey Foundation.

Chall, J. S. (1983). Literacy: Trends and explanations. *Educational Researcher*, *12*(9), 3-8. American Educational Research Association

Children's Educational Services (1987). *Test of Reading Fluency*. Minneapolis, MN. Author.

Cooper, H., Charlton, K., Valentine, J. C., Muhlenbruck, L., & Borman, G. D. (2000). Making the most of summer school: A meta-analytic and narrative review. *Monographs of the society for research in child development*, i-127. doi: http://www.jstor.org/stable/3181549

Cooper, H., Nye, B., Charlton, K., Lindsay, J., & Greathouse, S. (1996). The effects of summer vacation on achievement test scores: A narrative and meta-analytic review. *Review of Educational Research*, *66*(3), 227-268. doi: 10.3102/00346543066003227

Cooper, H., Valentine, J. C., Charlton, K., & Melson, A. (2003). The effects of modified school calendars on student achievement and on school and community attitudes. *Review of Educational Research*, *73*(1), 1-52. doi: 10.3102/00346543073001001

Crawford, L., Tindal, G., & Stieber, S. (2001). Using oral reading rate to predict student performance on statewide achievement tests. *Educational Assessment*, *7*(4), 303-323. doi: 10.1207/S15326977EA0704_04

Culatta, B., Setzer, L. A., Wilson, C., & Aslett, R. (2004). Project SEEL Part III. Children's engagement and progress attainments. *Communication Disorders Quarterly*, *25*(3), 127-139. Retrieved from http://cdq.sagepub.com/

Cunningham, A. E., & Stanovich, K. E. (1997). Early reading acquisition and its relation to reading experience and ability 10 years later. *Developmental Psychology*, *33*(6), 934. doi: 10.1037/0012-1649.33.6.934

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

Downey, D. B., von Hippel, P. T., & Broh, B. A. (2004). Are schools the great equalizer? Cognitive inequality during the summer months and the school year. *American Sociological Review*, *69*(5), 613-635. doi: 10.1177/000312240406900501

Downey, D. B., von Hippel, P. T., & Hughes, M. (2008). Are "failing" schools really failing? Using seasonal comparison to evaluate school effectiveness. *Sociology of Education*, *81*(3), 242-270. doi: 10.1177/000312240406900501

Entwisle, D. R., & Alexander, K. L. (1992). Summer setback: race, poverty, school composition, and mathematics achievement in the first two years of school. *American Sociological Review*, *57*(1), 72-84. doi: 10.2307/2096145

Entwisle, D. R., Alexander, K. L., & Olson, L. S. (1997). *Children, Schools, and Inequality*. Boulder, CO. Westview Press.

Entwisle, D. R., & Alexander, K. L. (1994). Winter setback: The racial composition of schools and learning to read. *American Sociological Review*, *59*(3), 446-460. doi:http://www.jstor.org/stable/2095943

Entwisle, D. R., Alexander, K. L., & Olson, L. S. (2002). Baltimore Beginning School study in perspective. In E. Phelps, F. F. Fustenberg, A. Colbey (Eds.) *Looking at lives: American longitudinal studies of the twentieth century*, (pp. 167-193). New York, NY: Russel Sage Foundation.

Fuchs, L. S., & Fuchs, D. (1992). Identifying a measure for monitoring student reading progress. *School Psychology Review*.

Fuchs, L. S., Fuchs, D., Hosp, M. K., & Jenkins, J. R. (2001). Oral reading fluency as an indicator of reading competence: A theoretical, empirical, and historical analysis. *Scientific Studies of Reading*, *5*(3), 239-256. Retrieved from http://www.triplesr.org/

Gillet.,W. J., Temple.,C. Temple.,C., & Crawford.,A.(2012). *Understanding reading problems: Assessment and Instruction* (8th ed.). Boston: Allyn and Bacon.

Good III, R. H., Simmons, D. C., & Kame'enui, E. J. (2001). The importance and decision-making utility of a continuum of fluency-based indicators of foundational reading skills for third-grade high-stakes outcomes. *Scientific Studies of Reading*, *5*(3), 257-288. doi:10.1207/S1532799XSSR0503_4

Good, R. H., & Jefferson, G. (1998). Contemporary perspectives on curriculum-based measurement validity. *Advanced applications of curriculum-based measurement*, (pp. 61-88) Guilford Press.

Good, R. H., Kaminski, R. A., Smith, S., & Bratten, J. (2001). Technical adequacy of second grade DIBELS oral reading fluency passages (Technical Report No. 8). Eugene, OR: University of Oregon.

Goodman, K. S. (1967). Reading: A psycholinguistic guessing game. *Literacy Research and Instruction*, *6*(4), 126-135.

Greenwood, C. R., Horton, B. T., & Utley, C. A. (2002). Academic engagement: Current perspectives on research and practice. *School Psychology Review*. *31*(3), 328-349. Retrieved from http://www.nasponline.org/publications/spr/index-list.aspx

Guthrie, J. T., Wigfield, A., Barbosa, P., Perencevich, K. C., Taboada, A., Davis, M. H., … & Tonks, S. (2004). Increasing reading comprehension and engagement through concept-oriented reading instruction. *Journal of Educational Psychology, 96*(3), 403-423. doi: 10.1037/0022-0663.96.3.403

Guthrie, J. T., Hoa, A. L. W., Wigfield, A., Tonks, S. M., Humenick, N. M., & Littles, E. (2007). Reading motivation and reading comprehension growth in the later elementary years. *Contemporary Educational Psychology, 32*(3), 282-313.

Guthrie, J. T., & Wigfield, A. (2000). *Engagement and motivation in reading*. Mahwah, NJ. Lawrence Erlbaum Associates Publishers.

Guthrie, J. T., Meter, P., McCann, A. D., Wigfield, A., Bennett, L., Poundstone, C. C., ... & Mitchell, A. M. (1996). Growth of literacy engagement: Changes in motivations and strategies during concept-oriented reading instruction. *Reading Research Quarterly, 31*(3), 306-332.

Guthrie, J. T., Coddington, C. S., & Wigfield, A. (2009). Profiles of reading motivation among African American and Caucasian students. *Journal of Literacy Research, 41*(3), 317-353. doi: 10.1080/10862960903129196

Hamilton, C., & Shinn, M.R. (2003). Characteristics of word callers: An investigation of the accuracy of teachers' judgments of reading comprehension and oral reading skills. *School Psychology Review*, *32*, 228–240.

Hasbrouck, J., & Tindal, G. A. (2006). Oral reading fluency norms: A valuable assessment tool for reading teachers. *The Reading Teacher*, *59*(7), 636-644. doi: 10.1598/RT.59.7.3

Heyns, B. (1978). *Summer learning and the effects of schooling*. New York, NY: Academic Press.

Hintze, J. M., Owen, S. V., Shapiro, E. S., & Daly III, E. J. (2000). Generalizability of oral reading fluency measures: Application of G theory to curriculum-based measurement. *School Psychology Quarterly*, *15*(1), 52. doi: 10.1037/h0088778

Hintze, J. M., & Silberglitt, B. (2005). A longitudinal examination of the diagnostic accuracy and predictive validity of R-CBM and high-stakes testing. *School Psychology Review*, *34*(3), 372.

Hrabowski III, F. A., & Robbi, J. (2002). The benefits of correctional education. *Journal of Correctional Education*, *53*(3).

Jo, B. (2002). Estimation of intervention effects with noncompliance: Alternative model specifications. *Journal of Educational and Behavioral Statistics*, *27*(4), 385-409. doi: 10.3102/10769986027004385

Juel, C. (1988). Learning to read and write: A longitudinal study of 54 children from first through fourth grades. *Journal of Educational Psychology*, *80*(4), 437. doi: 10.1037/0022-0663.80.4.437

Kame'enui, E. J., & Simmons, D. C. (2001). Introduction to this special issue: The DNA of reading fluency. *Scientific Studies of Reading*, *5*(3), 203-210.

Kaminski, R. A., & Good III, R. H. (1996). Toward a technology for assessing basic early literacy skills. *School Psychology Review*. *11*(4), 326-336.

Kim, J. S. (2006). Effects of a voluntary summer reading intervention on reading achievement: Results from a randomized field trial. *Educational Evaluation and Policy Analysis*, *28*(4), 335-355. doi: 10.3102/01623737028004335

Kim, J. S. (2007). The effects of a voluntary summer reading intervention on reading activities and reading achievement. *Journal of Educational Psychology*, *99*(3), 505. doi: 10.1037/0022-0663.99.3.505

Kim, J. S., & Guryan, J. (2010). The efficacy of a voluntary summer book reading intervention for low-income Latino children from language minority families. *Journal of Educational Psychology*, *102*(1), 20. doi: 10.1037/a0017270

Kim, J. S., & White, T. G. (2008). Scaffolding voluntary summer reading for children in grades 3 to 5: An experimental study. *Scientific Studies of Reading*, *12*(1), 1-23. doi: 10.1080/10888430701746849

Klein, J. R., & Jimerson, S. R. (2005). Examining ethnic, gender, language, and socioeconomic bias in oral reading fluency scores among Caucasian and Hispanic students. *School Psychology Quarterly*, *20*(1), 23. doi: 10.1521/scpq.20.1.23.64196

Kneese, C. C. (2000). The impact of year-round education on student learning: A study of six elementary schools. *ERS Spectrum*, *18*(1), 20-26.

Kuhn, M. R., & Stahl, S. A. (2003). Fluency: A review of developmental and remedial practices. *Journal of Educational Psychology*, *95*(1), 3. doi: 10.1037/0022-0663.95.1.3

Kutner, M., Greenburg, E., Jin, Y., & Paulsen, C. (2006). The health literacy of America's adults: Results from the 2003 National Assessment of Adult Literacy. NCES 2006-483. *National Center for Education Statistics*.

LaBerge, D., & Samuels, S. J. (1974). Toward a theory of automatic information processing in reading. *Cognitive Psychology*, *6*(2), 293-323. doi: 10.1016/0010-0285(74)90015-2

Lee, V. E., & Burkam, D. T. (2002). *Inequality at the starting gate: Social background differences in achievement as children begin school*. Washington, DC: Economic Policy Institute.

Little, R. J., & Yau, L. H. (1998). Statistical techniques for analyzing data from prevention trials: treatment of no-shows using Rubin's causal model. *Psychological Methods*, *3*(2), 147. doi: 10.1037/1082-989X.3.2.147

Maughan, B., Rowe, R., Loeber, R., & Stouthamer-Loeber, M. (2003). Reading problems and depressed mood. *Journal of Abnormal Child Psychology*, *31*(2), 219-229.

Messick, S. (1994). The interplay of evidence and consequences in the validation of performance assessments. *Educational researcher, 23*(2), 13-23. doi: 10.3102/0013189X023002013

McCombs, J. S., Augustine, C. H., & Schwartz, H. L. (2011). *Making summer count: How summer programs can boost children's learning*. Rand Corporation.

McGlinchey, M. T., & Hixson, M. D. (2004). Using curriculum-based measurement to predict performance on state assessments in reading. *School Psychology Review*, *33*, 193-203.

National Assessment of Adult Literacy. (2006). *The Health Literacy of America's Adults*. NCES 2006-483. U.S. Department of Education.

National Center for Educational Statistics. (2001). *Adult literacy and education in America: Four studies based on the National Adult Literacy Survey*. NCES 2001-534. U.S. Department of Education.

National Reading Panel & National Institute of Child Health and Human Development. (2000). *Teaching children to read: An evidence-based assessment of the scientific research literature on reading and its implications for reading instruction.* Retrieved from: http://www.nichd.nih.gov/publications/pubs/nrp/documents/report.pdf

National Research Council. (1998). *Preventing reading difficulties in young children*. U.S. Department of Education.

No Child Left Behind (NCLB) Act of 2001, Pub. L. No. 107-110, § 115, Stat. 1425 (2002).

Pintrich, P. R., & De Groot, E. V. (1990). Motivational and self-regulated learning components of classroom academic performance. *Journal of Educational Psychology*, *82*(1), 33.

Pintrich, P. R., & Schrauben, B. (1992). Students' motivational beliefs and their cognitive engagement in classroom academic tasks. *Student Perceptions in the Classroom*, (pp. 149-183). Routledge.

Sardo-Brown, D., & Rooney, M. (1992). The vote on all-year schools. *American School Board Journal*, *179*(7), 25-27.

Schacter, J. (2003). Preventing summer reading declines in children who are disadvantaged. *Journal of Early Intervention, 26*(1), 47-58. doi: 10.1177/105381510302600104

Schacter, J. (2001). *Reducing social inequality in elementary school reading achievement: Establishing summer literacy day camps for disadvantaged children*. Santa Monica, CA.

Schacter, J., & Jo, B. (2005). Learning when school is not in session: a reading summer day-camp intervention to improve the achievement of exiting First-Grade students who are economically disadvantaged. *Journal of Research in Reading, 28*(2), 158-169. doi: 10.1111/j.1467-9817.2005.00260.x

Schilling, S. G., Carlisle, J. F., Scott, S. E., & Zeng, J. (2007). Are fluency measures accurate predictors of reading achievement?. *The Elementary School Journal, 107*(5), 429-448.

Shapiro, E. S., Keller, M. A., Lutz, J. G., Santoro, L. E., & Hintze, J. M. (2006). Curriculum-based measures and performance on state assessment and standardized tests reading and math performance in Pennsylvania. *Journal of Psychoeducational Assessment*, *24*(1), 19-35. doi: 10.1177/0734282905285237

Shaw, R., & Shaw, D. (2002). DIBELS oral reading fluency-based indicators of third grade reading skills for Colorado State Assessment Program (CSAP). *University of Oregon Technical Report).*

Shinn, M. R., Gleason, M. M., & Tindal, G. (1989). Varying the difficulty of testing materials implications for curriculum-based measurement. *The Journal of Special Education*, *23*(2), 223-233. doi: 10.1177/002246698902300208

Shinn, M. R. (1998). *Advanced applications of curriculum-based measurement*. Guilford Press.

Sibley, D., Biwer, D., & Hesch, A. (2001). *Establishing curriculum-based measurement oral reading fluency performance standards to predict success on local and state tests of reading achievement*. ERIC Clearinghouse.

Slavin, R. E., Madden, N. A., Karweit, N. L., Dolan, L. J., & Wasik, B. A. (1991). Research directions: Success for all: Ending reading failure from the beginning. *Language Arts*, *68*(5), 404-409.

Stage, S. A., & Jacobsen, M. D. (2001). Predicting student success on a state-mandated performance-based assessment using oral reading fluency. *School Psychology Review, 30*(3), 407-419.

Stanovich, K. E. (1986). "Matthew Effects in reading: Some consequences of individual differences in the acquisition of literacy." *Reading Research Quarterly 21*, 360-406

Stanovich, K. E. (1992). *Speculations on the causes and consequences of individual differences in early reading acquisition*. Hillsdale, NJ. Lawrence Erlbaum Associates, Inc.

Stanovich, K. E. (2000). *Progress in understanding reading: Scientific foundations and new frontiers*. New York. Guilford Publications.

Taylor, B. M., Pearson, P. D., Peterson, D. S., & Rodriguez, M. C. (2003). Reading growth in high-poverty classrooms: The influence of teacher practices that encourage cognitive engagement in literacy learning. *The Elementary School Journal*, *104*(1), 3-28. doi: 10.1086/499740

Tindal, G., Marston, D., & Deno, S. L. (1983). The reliability of direct and repeated measurement. *Journal of Special Education Leadership, 12*(2), 3-10. Retrieved from http://www.sagepub.com/journals/Journal201881/manuscriptSubmission

U.S. Department of Education (2012). *ESEA Flexibility*. Retrieved from http://www2.ed.gov/policy/elsec/guid/esea-flexibility/index.html

Valencia, S. W., Smith, A. T., Reece, A. M., Li, M., Wixson, K. K., & Newman, H. (2010). Oral reading fluency assessment: Issues of construct, criterion, and consequential validity. *Reading Research Quarterly*, *45*(3), 270-291. doi: 10.1598/RRQ.45.3.1

von Hippel, P. T. (2007). Regression with missing Ys: An improved strategy for analyzing multiply imputed data. *Sociological Methodology*, *37*(1), 83-117. doi: 10.1111/j.1467-9531.2007.00180.x

Wang, J. H. Y., & Guthrie, J. T. (2004). Modeling the effects of intrinsic motivation, extrinsic motivation, amount of reading, and past reading achievement on text comprehension between US and Chinese students. *Reading Research Quarterly*, *39*(2), 162-186. DOI: 10.1598/RRQ.39.2.2

Wigfield, A., Guthrie, J. T., Perencevich, K. C., Taboada, A., Klauda, S. L., McRae, A., & Barbosa, P. (2008). Role of reading engagement in mediating effects of reading comprehension instruction on reading outcomes. *Psychology in the Schools*, *45*(5), 432-445. doi: 10.1002/pits.20307

Yovanoff, P., Duesbery, L., Alonzo, J., & Tindal, G. (2005). Grade-level invariance of a theoretical causal structure predicting reading comprehension with vocabulary and oral reading fluency. *Educational Measurement: Issues and Practice*, *24*(3), 4-12. doi: 10.1111/j.1745-3992.2005.00014.x

Zvoch, K., & Stevens, J. J. (2011). Summer school and summer learning: An examination of the short-and longer term changes in student literacy. *Early Education & Development*, *22*(4), 649-675. doi 10.1080/10409289.2010.489891

Zvoch, K. (2012). How does fidelity of implementation matter? Using multilevel models to detect relationships between participant outcomes and the delivery and receipt of treatment. *American Journal of Evaluation*, *33*(4), 547-565. doi: 10.1177/1098214012452715

Zvoch, K., & Stevens, J. J. (2013). Summer school effects in a randomized field trial. *Early Childhood Research Quarterly*. *28*(1), 24-32. doi: 10.1016/j.ecresq.2012.05.002