

ABSTRACT

WHITEHOUSE, MARY HOLLINGS. Evaluation of the Assessment Component of the HELPS Program. (Under the direction of John C. Begeny.)

The purpose of the research was to examine the assessment component of the Helping Early Literacy with Practice Strategies (HELPS) Program which is similar to curriculum-based measurement (CBM). CBM has demonstrated usefulness with determining student growth and intervention effectiveness, as well as predicting student outcomes on state-mandated assessment. The current study utilized data from one academic year of implementation of the HELPS Program with second grade students at an elementary school. The purpose was to evaluate whether the assessment data from HELPS indicated gains in students' reading fluency and whether those gain scores predicted student outcomes on the Gray Oral Reading Test (GORT) for fluency and comprehension. Students' scores were evaluated in terms of generalization of passage gains (initial reading of a new passage), immediate passage gains (gains within one session with one passage), and retention of passage gains (gains from one session to the next). Scores were evaluated with both linear and quadratic multi-level models. Results indicated that students' generalization of passage gain scores were most meaningful in terms of determining growth over time and predicting students' outcome scores. Immediate passage gains, however, can also be useful in evaluating students' growth over time. Implications for these results are discussed.

© Copyright 2009 by Mary Hollings Whitehouse

All Rights Reserved

Evaluation of the Assessment Component of the HELPS Program

by
Mary Whitehouse

A thesis submitted to the Graduate Faculty of
North Carolina State University
in partial fulfillment of the
requirements for the Degree of
Master of Science

Psychology

Raleigh, North Carolina

2009

APPROVED BY:

Shevaun Neupert

Scott Stage

John Begeny
Chair of Advisory Committee

DEDICATION

For Mom: I hope this in some way continues the work for which you were so gifted and about which you were so passionate. Your gifts are greatly missed.

For the three most important men in my life: we are one step closer to the finish line.

Thanks to the teachers who participated in this study in the 2007-2008 school year. You contributed greatly to the reading ability of your students. Your contribution to the research is secondary.

BIOGRAPHY

Mary Hollings Whitehouse was born and raised in Fayetteville, North Carolina. She attended college at the University of North Carolina at Greensboro, where she earned her Bachelor's degree in psychology, with a double minor in history and Russian. From there, she entered the Divinity School at Campbell University in Buies Creek, North Carolina, where she earned a Master of Divinity (M.Div.). Upon completing her M.Div., she worked as a hospital chaplain at WakeMed Health and Hospitals in Raleigh, North Carolina for one year. During that year, she realized she greatly missed working in the mental health field, and made the decision to pursue a Ph.D. in psychology. At the time of this writing, she is in her third year in the School Psychology Ph.D. program at North Carolina State University.

TABLE OF CONTENTS

LIST OF TABLES	vii
LIST OF FIGURES	viii
INTRODUCTION	1
CHAPTER 1: REVIEW OF THE LITERATURE	3
Federal Education Legislation: Changes Over Time	3
A Closer Look at RTI	7
Historical Perspectives	7
Current Perspectives	8
The RTI Framework	8
Multitier Intervention.....	10
Screening and Assessment: What does it mean to respond to intervention?.....	11
Curriculum-Based Measurement	13
A Timeline of CBM Development	13
Technical Adequacy of CBM	16
CBM: Empirically Based Assessment	16
Progress monitoring	17
Identifying at-risk students	19
Predicting outcomes.....	20
Strengths and limitations.....	25
Evaluating Intervention Effectiveness/Use in RTI	30
Reading Fluency	31
Assessment of Reading Fluency	33
Empirically Based Reading Fluency Interventions.....	38
CHAPTER 2: EVALUATING ASSESSMENT WITH THE HELPS PROGRAM	41
The HELPS Program	41
Using HELPS as a CBM-R Assessment Tool	44
Purpose and Rationale.....	45

Research Questions and Hypotheses	46
CHAPTER 3: METHOD	51
Participants and Setting.....	51
Teacher Participants.....	51
Student Participants	52
Setting	52
Assessment Materials.....	53
Gray Oral Reading Test, Fourth Edition.....	54
Curriculum-Based Measurement program reading passages.....	54
Procedure	55
Analyses.....	55
CHAPTER 4: RESULTS	60
Generalization of Passage Gains.....	60
Immediate Passage Gains	61
Retention Passage Gains	62
Further Analyses	64
Generalization of Passage Gains.....	65
Immediate Passage Gains	66
Retention Passage Gains	67
Summary of Results.....	67
CHAPTER 5: DISCUSSION.....	69
Practical Implications.....	72
Limitations	76
Future Research Directions.....	77
REFERENCES	79
APPENDICES	91
Appendix A. HELPS Program Implementation Protocol.....	92

Appendix B. Tips and Reminders for Implementing the HELPS Program.....	94
Appendix C. Scripted Protocol—HELPS	97
Appendix D. Standard HELPS Graph for Grade Two.....	98
Appendix E. HELPS Program Prize Chart	99
Appendix F. Description of the HELPS Session	100

LIST OF TABLES

Table 1. Linear Multi-Level Models for Generalization of Passage Gains	103
Table 2. Linear Multi-Level Models for Immediate Passage Gains	104
Table 3. Linear Multi-Level Models for Retention Passage Gains	105
Table 4. Quadratic Multi-Level Models for Generalization of Passage Gains.....	106
Table 5. Quadratic Multi-Level Models for Immediate Passage Gains.....	107
Table 6. Quadratic Multi-Level Models for Retention Passage Gains	108
Table 7. Summary of Significant Relationships	109

LIST OF FIGURES

Figure 1. Growth for Generalization of Passage Gains	110
Figure 2. Quadratic Interactions for Immediate Passage Gains—Fluency.....	111
Figure 3. Quadratic Interactions for Immediate Passage Gains—Comprehension	112
Figure 4. Quadratic Interactions for Retention Passage Gains—Fluency	113
Figure 5. Quadratic Interactions for Retention Passage Gains—Comprehension.....	114

Introduction

Recent legislation, such as the Individuals with Disabilities Education Improvement Act (IDEIA, 2004) and No Child Left Behind (NCLB, 2001), has spurred our education system in a new direction. Amid the copious amounts of high-stakes testing and the pressure to achieve “adequate yearly progress” is the permission—and funding—to evaluate academic achievement differently. Previously, children were classified with a learning disability solely through a significant discrepancy between their IQ and their score on a standardized achievement test. While that is still a widely used option, there is now another alternative. Children may also be classified if they do not make academic progress after receiving documented, empirically based interventions. At that point, they may then be referred for special education services. This new approach, known in the educational field as Response to Intervention (RTI), helps professionals make decisions concerning students’ academic growth earlier in students’ education.

The most common referral concern for special education is difficulty in reading (Fuchs & Fuchs, 2006). Nearly 80% of children with a learning disability are reading disabled (Begeny & Martens, 2006; Fuchs & Fuchs, 2006; Lyon, 1995; Snow, Burns, & Griffin, 1998). Certainly the advantages of early reading success and the austere effects of reading failure have become more readily apparent in this nation. Schools are charged with the task of responding to the demands of a knowledge-based 21st century by providing quality reading instruction to all students. In order to achieve reading success, it is necessary to prevent reading difficulties and to ensure that all children become proficient readers early

in their educational careers (Good, Simmons, & Kame'enui, 2001). Reading First, a federal initiative tied to No Child Left Behind, was created for the purposes of providing empirically based reading instruction to all children in order for them to become proficient readers by the end of third grade. Key among the areas of reading development and instruction outlined in Reading First is reading fluency, or the ability to read text with speed and accuracy (U.S. Department of Education, 2002).

Since the recent changes in federal legislation have greatly influenced the field of education, the current discussion begins there. The subsequent section will describe RTI and its relevant components. Because assessment is a core piece of RTI, and because extensive research has focused on the use of Curriculum-Based Measurement (CBM) as the primary assessment tool for RTI models, a discussion of CBM will follow, including its technical adequacy and uses. Finally, reading fluency, its place in RTI, and its relationship to evaluating student reading outcomes is described. Discussion of each aforementioned topic will then lead to the primary purpose and rationale for the proposed study. In short, I will first argue that more immediate educational decision-making from brief assessment may improve individual outcomes, as such decisions may inform the frequency and duration of intervention required. I will then describe a study meant to evaluate whether progression through an intervention is associated with student outcomes on post-test measures.

Chapter 1:

Review of the Literature

Federal Education Legislation: Changes Over Time

The practice of school psychology is guided, in large part, by legal and ethical statutes. Since the mid-1960s, the United States federal government has passed numerous legislative acts pertaining to special education. Primary among those is Public Law 94-142, the Education for All Handicapped Children Act, passed in 1975. This piece of federal legislation was the first of its kind to guarantee children with disabilities the right to “free and appropriate public education services” (Merrell, Ervin & Gimpel, 2006). Since 1975, P.L. 94-142 has undergone several major revisions. In 1990, the Act was amended and the name was changed to the Individuals with Disabilities Education Act (IDEA). The 1997 version of IDEA outlined the process for classification of disabilities in children. Under this law, any child thought to benefit from special education services underwent a complete individual evaluation. The results of this evaluation were used to determine whether the child met disability criteria for one of thirteen categories. These categories included physical disabilities such as deafness, blindness, orthopedic impairment or other health impairments, developmental disabilities, and various cognitive disorders, including specific learning disabilities (Merrell et al., 2006). According to IDEA-97, a child was classified as learning disabled when the individual evaluation results indicated a significant discrepancy between the child’s intellectual ability and an area of academic achievement, such as reading or mathematics. Once it was determined that a child has a disability in one of thirteen areas, an

individualized educational plan (IEP) was established to outline the particular special education services the child should receive and the expected progress the child should make over the course of an academic year (Jacob & Hartshorne, 2007).

Although the IQ-achievement discrepancy model has been standard practice for many years, it has garnered much criticism. To begin with, the IQ-achievement discrepancy is not psychometrically valid. This approach lacks adequate reliability, especially for accurately diagnosing learning disabilities (Fletcher, Coulter, Reschly, & Vaughn, 2004; Fletcher, Denton, & Francis, 2005). The IQ-achievement discrepancy also delays intervention for children, preventing eligibility for services until the student's performance is so low the discrepancy is met. For instance, a child may clearly perform below grade level but his or her academic deficits are not great enough to warrant a diagnosis of a learning disability. Consequently, they are not identified as needing additional services (Fletcher et al., 2004; Fuchs & Fuchs, 2006; Speece, Case, & Malloy, 2003). Additionally, there is no standardized approach to determining whether a student has met the discrepancy criteria. The IQ score may be the full scale IQ, but it may also be the verbal or performance score by itself (Fletcher et al., 2005). Not only does the IQ-achievement discrepancy lack an evidentiary basis, it is also not an equitable method for determining learning disabilities. More specifically, minority students are overrepresented in special education based on this criterion (Fletcher et al., 2004; Speece & Case, 2001). There are additional concerns that classifying students for special education on the basis of the IQ-achievement discrepancy does not inform instructional decisions that could be useful for improving student outcomes (Fletcher

et al., 2004; Gresham, 2007; Speece & Case, 2001; Speece et al., 2003).

The Elementary and Secondary Education Act (ESA) was signed into law in 1965. Under this Act, school districts received federal funding, primarily for the education of children who were economically disadvantaged. There was no specific mention of children with handicaps. In 2001, the ESA was reauthorized and renamed the No Child Left Behind (NCLB) Act. At its core, NCLB is about greater effectiveness of and accountability by school districts (Merrell et al., 2006). School effectiveness was operationalized in terms of adequate yearly progress as indicated by statewide accountability testing (NCLB, 2001). Such testing occurs in grades 3-8 and covers reading and math. Science is assessed at least once in elementary, middle and high school. According to the law, all students must be proficient in these subject areas by the 2013-2014 school year. Failure to display adequate yearly progress results in specific consequences for the particular school and district. Consequences vary, but may include withholding of federal funds, internal restructuring, replacement of staff, and/or state takeover of school operations (Merrell et al., 2006).

The NCLB Act has been met with considerable controversy, particularly over the requirements for mandatory state testing and adequate yearly progress. Some of the controversy centers on lack of sufficient funding to implement assessment procedures. Concerns have also been voiced over the use of test scores for students with disabilities or limited English proficiency as part of the school's overall level of performance (Merrell et al., 2006). Because of the high-stakes nature of statewide testing, there is also concern that more students will be placed in special education or that more will be retained in early grades

in an attempt to inflate overall district scores. Additionally, the results of statewide testing are used to evaluate teacher and school performance. Schools with low test scores may experience negative publicity and greater outside scrutiny, while schools with high scores will likely receive public praise, greater autonomy and, possibly, financial rewards (Jacob & Hartshorne, 2007).

It has been justifiably argued that waiting until the end of 3rd grade to assess students and make decisions about their level of learning is not time-efficient and may not be instructionally relevant. While they can gauge progress and levels of performance at the school and district levels, the data obtained through end of year assessments do not provide information regarding specific areas of weakness for students (Good et al., 2001).

In December 2004, IDEA was reauthorized again (now called *Individuals with Disabilities Education Improvement Act*, IDEIA), though much of the legislation remained unchanged. One significant change, however, was the *optional* use of the intelligence-achievement discrepancy to determine the presence of a learning disability. That is, IDEIA now allows schools to “use a process that evaluates whether a child responds to a research-based intervention in determining whether the child has a learning disability” (Merrell et al., 2006, p. 125). In essence, this response-to-intervention (RTI) approach allows practitioners to identify children with learning disabilities, as well as areas of relative weakness, through the use of early interventions. IDEIA also allows school districts to use up to 15% of their special education monies to fund early intervention (Fuchs & Fuchs, 2006).

A Closer Look at RTI

Historical Perspectives. Although RTI was just recently outlined in U. S. federal regulations for education, the concept is well established in other fields, such as medicine (Jimerson, Burns, & VanDerHeyden, 2007). Gerald Caplan (1964) outlined a model of preventive psychiatry in *Principles of Preventive Psychiatry*. The overarching goal of preventive psychiatry is to lower the rate of incidence in mental disorders in a population over time. The intent, according to Caplan, is not to avert sickness in any one person. Instead, the aim of prevention is to reduce the risk of incidence for the whole population in order to reduce the overall number of individuals who will become ill (Caplan, 1964).

Caplan's program involved three levels, or tiers, of prevention. At the primary level, an individual is seen as representative of a larger group. The individual's treatment is developed in response to his or her individual needs as well as in response to the community problem the individual represents. The treatment also utilizes appropriate available resources. Additionally, the information obtained about the individual also provides information regarding the status of the larger community. Whatever the presenting problem of the individual at the primary level, basic supports are put in place to help the individual through his or her crisis (Caplan, 1964).

The goal of secondary prevention is to reduce the prevalence of a disorder in a community. This reduction can occur in one of two ways: (a) the rate of new cases is lowered by altering factors that lead to the disorder, or (b) the number of existing cases is lowered through early diagnosis and an effective treatment plan. In other words, this second tier

encompasses and builds on the first tier. Likewise, tertiary prevention encompasses and adds to secondary prevention. The goal at this level is to reduce the rate of deficient functioning due to mental disorder in the greater community. Tertiary prevention relies on large-scale treatment of mentally disordered patients so they can return to their “productive capacity” as soon as possible (Caplan, 1964).

Within the field of education, it has been suggested that the basis of RTI can be found in the 1980 National Research Council report on special education. That report outlined the way in which the validity of special education services was to be evaluated. Three criteria were stipulated: (a) the caliber of the general education curriculum, (b) whether the special education program is instrumental in producing significant outcomes for students, and (c) the reliability and validity of the assessment process in identifying a disability (Gresham, 2007).

Current Perspectives. RTI is most often conceptualized in one of two ways. The first is a problem-solving approach, or “a systematic analysis of instructional variables designed to isolate target skill/subskill deficits and shape targeted interventions” (Jimerson et al., 2007, p. 4). The second is a standard protocol approach, where a standard set of research based instructional methods are implemented to remediate academic difficulties. While this dichotomous conceptualization of RTI is quite common, many RTI models described in the literature combine the two views. Both approaches are, at their essence, problem-solving approaches and are most effective when combined and integrated into a three-tier model of service delivery (Jimerson et al., 2007).

The RTI Framework. The whole notion of RTI is determining whether there is

sufficient or insufficient change in academic progress as the result of an intervention. Within RTI, decisions about changing or intensifying a student's intervention are based on how well the student does or does not respond to an evidence-based intervention implemented with integrity. The child's response to an intervention, as determined by frequent assessment, provides information that "is used to select, change, or titrate interventions" (Gresham, 2007, p. 10). The assumption of RTI is that if a child does not adequately respond to the best feasible interventions in a given setting, then that child should be eligible for additional supports, including more intense intervention and special education services (Gresham, 2007).

Researchers have outlined a model for RTI based on general education (Fuchs, 2003; Fuchs & Fuchs, 1998; Jenkins, Hudson, & Johnson, 2007). A general education environment that has proven to be effective for the majority of students is the principal criteria used to test responsiveness. The primary assumption is that children in an effective general education curriculum or environment experience strong rates of academic growth. The smaller subset of students performing below their peers, then, stands out (Fuchs & Fuchs, 1998). Those students who succeed in a general education program will likely continue to do so and not require intervention. For those students who perform lower than their peers, intensive interventions are required to affect adequate learning (Fuchs 2003). RTI provides a multi-tiered framework that addresses the need for early and, when necessary, continued intervention for those students who experience difficulty in general education (Jenkins et al., 2007).

Multitier Intervention. The RTI framework is based on a three-tiered model of intervention implementation. While prevention is the overarching emphasis, students receive interventions at the various levels, based on their needs (Glover & DiPerna, 2007).

The first, or primary, tier of intervention is universal. Theoretically, a strong, empirically-based curriculum is in place class- and school-wide. Ideally, this helps to rule out inadequate instruction as a contributing factor to a student's academic difficulties. The second tier utilizes supplemental, research-based strategies to support those students who continue to struggle academically. These strategies may be implemented at the small group or individual level. Finally, the third, or tertiary, tier employs targeted, intensive interventions for those students who continue to lack academic progress after receiving supports at the first two levels (Daly, Martens, Barnett, Witt, & Olsen, 2007). Tier 3 interventions should still be feasible in the classroom. Failure to respond to such intensive intervention results in a special education referral because the potentially effective level of intervention then goes beyond what the teacher can do in the regular education setting (Ardoin, Witt, Connell, & Koenig, 2005).

As a child progresses through these tiers of intervention, the interventions increase in intensity. Fuchs and Fuchs (2006) suggested several ways that this may occur: "(a) using more teacher-centered, systematic, and explicit instruction; (b) conducting it more frequently; (c) adding to its duration; (d) creating smaller and more homogenous student groupings; or (e) relying on instructors with greater expertise" (p. 94).

Screening and Assessment: What does it mean to respond to intervention?

According to Gresham (2007), the aim for all interventions is to improve overall levels of performance. RTI is about detecting and producing changes in student responding over time (Daly et al., 2007). Thus, at each tier, regular assessment occurs to determine whether the intervention is effective, and decisions are then made as to how to modify the intervention as needed (Glover & DiPerna, 2007).

Assessment at Tier 1 involves universal screening for the purpose of accurately classifying students as at-risk for learning difficulties or poor academic outcomes. A screening process comprised of more than one measure appears to yield better classification accuracy than a process based solely on one measure. Additionally, dynamic assessment can potentially identify those students whose skills in an academic area are low but who also respond quickly to intervention. When used as part of the screening process, dynamic assessment may also increase classification accuracy (Jenkins et al., 2007).

Local data may be collected at different points in the academic year (e.g., fall, winter and spring). These data are then used to establish distributions of “typical” performance based on mean and standard deviation statistics (Christ & Hintze, 2007). From these results, screening at Tier 1 can utilize a dual discrepancy model to identify students with academic weaknesses. The dual discrepancy is based on an individual’s rate of growth *and* on their level of performance relative to their classmates. Students with scores one standard deviation or more below the mean level and slope of their classmates have fewer skills in a particular area than their classmates. They are also acquiring new skills at a rate slower than their peers.

Without intervention, the achievement gap between an individual student and his or her classmates will likely continue to widen and, potentially, pass the point at which the student will succeed in the general education curriculum (Ardoin et al., 2005).

As a result of the universal screening, students who evidence need are matched with appropriate services (Glover & DiPerna, 2007). Measurement, then, becomes the means for determining what adjustments need to be made in intervention implementation. As measurements indicate inadequate levels of student response to the intervention, the intensity of the intervention is strengthened (Daly et al., 2007).

According to Fuchs (2003), intervention responsiveness requires the following: (a) determining the timing of measurement in a student's response to intervention; (b) establishing criterion for responsiveness (i.e., a cut point); and (c) consideration of the nature of the intervention (i.e., appropriate for the child's situation). These three factors should be considered when addressing how to assess intervention responsiveness.

Assessment may be post-intervention, periodically throughout the intervention, or based on dual-discrepancy, which can be applied at any time. Perhaps the most feasible approach of measuring RTI is to assess students following the intervention and then "apply a standard" for distinguishing children who responded to the intervention with those who did not. However, basing the decision on a final status measurement does not directly address responsiveness. In other words, the final score may fall below the standard, but the student may have experienced more growth when compared to others (Fuchs, 2003).

To more directly assess responsiveness to an intervention, pre- to post-treatment

growth (e.g., based on pre-/post-testing) or slope (e.g., based on more frequent testing) should be considered. These data, in addition to local norms from universal screening, can be combined to determine an appropriate cut point for responsiveness. The cut point may also be based on a specific benchmark. However it is determined, the cut point, along with local norms, can be useful for generating a normative profile with a range of response to the intervention based on the normal (i.e., local or peer) population. It should be noted, however, that intensive interventions (i.e., highly individualized, small group or lasting an extended period of time) preclude such a normative framework for identifying disability. One alternative may be to compare growth scores to a limited norm (Fuchs, 2003).

According to Busch and Reschly (2007), within RTI, there is a need for “measures that are technically adequate, can be administered frequently, and are sensitive to student growth” (p. 224). The overall goal of prevention-focused assessment is to furnish schools with a system of measurement that provides early and reliable predictions concerning performance on critical outcomes, and that does so in a way relevant to instruction (Good et al., 2001). Research to date has focused primarily on the use of Curriculum-Based Measurement (CBM) to monitor student progress and to differentiate between students performing at typical and atypical levels (Glover & DiPerna, 2007).

Curriculum Based Measurement (CBM)

A Timeline of CBM Development

Researchers have long recognized the need for measures of academic progress outside state achievement testing for several reasons. First, the state-mandated testing

procedures are difficult for classroom teachers to replicate. Testing is done in several-hour blocks of time over the course of a few days and the test materials are developed by a group of people at the state, rather than local, level. Second, decisions being made based on those test scores are so important that other confirming information is needed to complement the data (Crawford, Tindal, & Steiber, 2001). Achievement tests administered once every spring are not relevant for day-to-day decision-making in the classroom (Crawford et al., 2001; Deno, 1985). In general, statewide assessments do not begin until the end of third grade. By then, the information provided by student performance comes at the expense of valuable time that could have been spent providing instructional support (McGlinchey & Hixson, 2004). Third, teachers need the information gathered from more frequent performance indicators in order to improve instructional programs in a timely fashion (Crawford et al., 2001; Deno, 1985). Finally, it is possible that statewide tests are insensitive to change in academic progress for low-performing students (Crawford et al., 2001).

Current federal initiatives such as No Child Left Behind and Reading First have increased the demand not only for state and district accountability, but also for early identification of students needing additional support and intervention provision for those students (McGlinchey & Hixson, 2004). Consequently, there is the need for a method of assessment that can bridge the gap between school entry and annual assessment beginning at the end of third grade. Curriculum-based measurement (CBM) is proving to be an effective method.

The development of CBM began in the late 1970s and early 1980s. By design, the measures are tied to curricula, simple and efficient to administer, capable of having alternate forms, are reliable and valid, inexpensive to produce, and sensitive to change over time (Crawford et al., 2001; Deno, 1985; Marston, 1989). Stanley Deno and Phyllis Mirkin focused their research in the early 1980s on field testing the technical adequacy of potential curriculum-based measures. They identified the following as validated “curriculum-based measures of student achievement:

1. In the area of reading, counting the number of words a student reads correctly from either a basal reader or word list in a 1-minute interval.
2. In the area of written expression, counting the number of words or letters written during a 3-minute interval, given either a story starter or topic sentence.
3. In spelling, counting the number of correct letter sequences (White & Haring, 1980, as cited in Marston, 1989) or words spelled correctly during a 2-minute interval, given words dictated every 7 seconds.
4. In the area of math, counting the number of correctly written digits during a 2-minute interval from grade-level computational problems” (Marston, 1989).

By the late 1980s, the scope of CBM expanded from evaluating instructional effectiveness in special education to developing local norms at the class, school and district levels. From there, CBM outcomes became “associated with systematic screening, eligibility, and diagnostic decisions” (Christ & Silberglitt, 2007, p. 130). More recently, CBM has found its home in RTI as the primary means to evaluate intervention effectiveness (Busch & Reschly, 2007; Christ & Silberglitt, 2007; Fuchs, 2003; Glover & DiPerna, 2007; Shinn, 2007). In addition, CBM is useful for identifying children for special education, establishing Individualized Educational Plan (IEP) goals and monitoring progress towards meeting them (McGlinchey & Hixson, 2004; Shinn, 2007), progress monitoring for instructional

programming, developing local norms for purposes of instructional decision-making, and improving assessment for minority students (Stage, 2001).

Technical Adequacy of CBM

During the early development period of CBM, researchers understood the importance of the technical adequacy of the measures (Christ & Silberglitt, 2007). The results of the early psychometric research on CBM were summarized in a chapter by Marston (1989), which has become one of the most widely cited sources on the subject (Christ & Silberglitt, 2007). Marston (1989) discussed the research as it pertains to the four main subject areas of CBM: reading, spelling, writing and math.

In the area of reading, Marston (1989) cited studies that examined the relationship between several forms of CBM measures (e.g., oral passage reading, reading in context, Cloze comprehension procedures and word meaning) to the Stanford Diagnostic Test, the Woodcock Reading Mastery Test and the Reading Comprehension subtest from the Peabody Individual Achievement Test. Correlation coefficients ranged from .63 to .90, with most coefficients above .80. Reliability estimates were also strong, with test-retest reliability coefficients ranging from .82-.97, parallel form estimates from .84-.96 and interrater reliability at .99.

CBM: Empirically-based Assessment

Research on the use of CBM as a tool for assessment and decision-making has grown extensively since its early development. The focus has primarily been on the measurement of academic skills to monitor student progress, identify at-risk and low performing students,

predict student outcomes, and evaluate intervention effectiveness (Deno, 1985; Silbergitt & Hintze, 2005; Glover & DiPerna, 2007; Shinn, 2007). Each of these will be discussed in turn.

Progress Monitoring. Progress monitoring with CBM is a dynamic approach that involves regular, systematic assessment based on curricular goals. These assessment outcomes provide data that inform instructional decision-making (Fuchs, 1989). In addition, the use of CBM progress monitoring results in increased instructional quality and student achievement (Fuchs, 1989; Hintze & Christ, 2004). For instance, Fuchs, Deno, and Mirkin (1984) found that teachers who utilize CBM to monitor students' reading growth (a) tend to employ more measurable and appropriate achievement goals, (b) are more realistic about goal attainment, (c) are able to cite more data for determining sufficient progress and for deciding whether instructional modifications are necessary, and (d) modify individual instructional programs more frequently. Additionally, students experienced better outcomes on global achievement tests assessing decoding and reading comprehension with CBM progress monitoring compared to students who were assessed less systematically.

CBM-Reading (CBM-R) originated as a true measure of children's performance within their specific curriculum. For assessing reading fluency, children read passages from the basal reader, or the text used as part of the curriculum used in the classroom (Deno, 1985; Marston, 1989). As CBM-R evolved, however, "non-curriculum" passages were created by outside sources (e.g., Dynamic Indicators of Basic Early Literacy Skills (DIBELS); AIMSweb) with the same goal of assessing students' oral reading fluency. The procedure of this newer "version" of CBM-R is the same, and the text, when chosen appropriately, is the

same difficulty level as what the student encounters in the regular classroom. Both types of CBM-R reading probes (i.e., materials from the student's reading curriculum and materials the student otherwise does not come in contact with) are appropriate for measuring students' progress over time (Powell-Smith & Bradley-Klug, 2001; Riley-Heller, Kelly-Vance, & Shriver, 2005).

Currently, the most common procedures for progress monitoring in reading incorporate periodic assessments with a tool such as CBM-R, using material that is unfamiliar to the student. These assessments typically occur at three different points during the school year: fall, winter, and spring (Ardoin & Christ, 2008; Good et al., 2001; Hintze & Silbergitt, 2005; McGlinchey & Hixson, 2004; Stage & Jacobsen, 2001). The two most commonly used sets of standardized reading materials for these tri-annual assessments comes from the Dynamic Indicators of Basic Early Literacy Skills (DIBELS; Good & Kaminski, 2002) and AIMSweb (AIMSweb, 2008).

For each of these systems of CBM-R, the standard tri-annual assessment involves administering a set of three passages to students at each assessment period. The passages are presumably of equivalent difficulty, based on a readability formula and/or field testing with the passages. Each student's median score from the three readings at each assessment period is used to estimate the student's level of performance at the time of assessment. Scores from all three assessment points (fall, winter, spring) are used to estimate each student's reading growth over the course of the school year (Ardoin & Christ, 2008; Shinn, 2002).

Additionally, benchmarks, or cut-scores, for these timeframes have been established

in the literature, specifically in the area of CBM-R (DIBELS; Good & Kaminski, 2002; Good et al., 2001; Hasbrouck & Tindal, 1992; Silbergitt & Hintze, 2005). The use of these cut-scores as targets for student performance and progress allows educators to evaluate students at each benchmark period and determine whether individual students are “on track.”

Individual progress monitoring may also take place much more frequently—as many as several times per week (Fuchs, 2003). Frequent data points, as provided by frequent assessment, provide information to educators in order to develop successful academic strategies. The cut-scores for the class can be a useful tool for comparison at the individual level. Likewise, individual goal lines serve as the basis for evaluating progress. These goal lines, according to Deno, Marston and Tindal (1986) “are established by plotting a line on a graph from the median baseline performance to the level of performance expected (goal rate) at the program termination (goal date)” (p. 11). The slope of the student’s improvement over time represents the average amount of anticipated unit gain from one week to the next. Consequently, the slope serves as an indicator of the student’s growth and skill development (Silbergitt & Hintze, 2007). The goal line establishes the required rate of growth to meet the program goal (Deno et al., 1986). Regular progress monitoring provides information about the individual student’s rate of growth and whether modifications need to be made to increase the student’s potential for goal attainment.

Identifying at-risk students. As previously discussed, the conventional model for identifying students as at-risk or in need of special education services has been the IQ-achievement discrepancy. With the 2004 reauthorization of IDEIA, however, a new option is

available for determining eligibility. According to the new regulations, “in determining whether a child has a specific learning disability, a LEA [local education agency] may use a process that determines if the child responds to a scientific, research-based intervention as part of the evaluation procedures” (IDEIA, 2004).

In order to determine whether a child is responding to an intervention, frequent assessment must be employed. Repeated measures, such as CBM, provide time-series data, which reveal actual individual growth over time and can be graphically displayed (Deno et al., 1986). Published achievement tests do not provide the same kind of information because they are given only once or twice per year. Practitioners can then use slope and variability of individual performance (data which can only be derived from time-series data) to make educational decisions (Deno et al., 1986; Fuchs & Fuchs, 2006; Shinn, 2007; Silberglitt & Hintze, 2007). These data can estimate responsiveness within as little as two-month’s time (Fuchs, 2003).

Under this new model, the discrepancy now is between the student’s grade placement and the student’s performance level in the curricula. The student’s individual performance is relative to the local norms (i.e., classroom or school) established through the screening process. The task at that point is to determine whether the student’s level of performance is significantly different from his or her peers such that additional supports (e.g., intervention or special education) are warranted (Fuchs & Fuchs, 1998; Glover & DiPerna, 2007; Shinn, 1989). As previously discussed, dual discrepancy is an appropriate model for determining differences in levels of performance and making subsequent decisions concerning the need

for additional interventions and/or special education eligibility (Busch & Reschly, 2007; Fuchs, 2003; Glover & DiPerna, 2007).

Predicting outcomes. With the issuance of legislation such as No Child Left Behind, the educational vernacular now includes words such as “standards-based reform,” “accountability,” and “high-stakes testing.” The most prominent of these is the use of assessment for purposes of accountability. This high-stakes accountability movement “calls for an assessment system that produces trustworthy and reliable results that are instructionally relevant and capable of forecasting educational change that positively impacts and sustains student learning” (Good et al., 2001, p. 258).

Given that high-stakes testing does not typically occur until the end of third grade, a system of assessment during primary grades would provide useful information concerning student progress. Such a system can verify whether students are progressing at an appropriate pace before the end of third grade and before some learning problems become too great. This type of assessment, therefore, needs to allow tenable predictions as to how likely children who achieve gains on one measure or set of measures one year will perform at specified benchmark levels in successive years (Good et al., 2001).

Previous findings, as discussed by Glover and DiPerna (2007), suggest variability and uncertainty concerning the predictive validity of academic screening measures and their use in eligibility decision-making. For example, Hintze, Ryan and Stoner (2003) examined the concurrent and predictive validity of the Dynamic Indicators of Basic Early Literacy Skills (DIBELS) and the Comprehensive Test of Phonological Processing (CTOPP). Results

indicated moderate to strong correlations between the two, suggesting they measure the same construct. The authors also utilized suggested and alternative cut scores in their examination of whether DIBELS predicted CTOPP performance. Although the cut scores resulted in high sensitivity, there were a large number of false positive results, resulting in false classifications for eligibility. With follow-up analyses and adjusted cut scores, however, the authors were able to reduce false positives and increase specificity and predictive power.

Silberglitt and Hintze (2005) conducted a study using CBM in the area of reading (CBM-R). Their purposes were twofold. The first was to “examine the extent to which CBM-R probes could predict success on state-mandated achievement tests” (p. 310). The second was to “compare various methods of standard setting and examine the accuracy and appropriateness of each of these approaches in determining cut scores” (p. 310). Data were compiled from the 1996-97 through the 2001-02 school years. Multiple cohorts were included in the sample, with a total of 2,191 students from one large school district. Students completed at least one (though most completed several or all) CBM-R probe in the fall, winter or spring of their 1st, 2nd, or 3rd grade year. In addition, the students all completed the Minnesota Comprehensive Assessment (MCA) in reading in the spring of 3rd grade. The data analysis plan was to examine the correlation between CBM-R and the state test (across benchmark periods) to determine whether CBM-R demonstrated predictive validity for these purposes. All correlations among CBM-R scores and scores on the MCA from winter Grade 1 through spring Grade 3 were significant. However, the strength of the relationship decreased with an increase in the amount of time between test administrations (1st grade

Winter: $r = .41$; 3rd grade Spring: $r = .71$). The researchers do point out that there was a nonlinear relationship between CBM-R and MCA at earlier administrations of CBM-R. This may have impacted the correlations to show a less substantial relationship between the two (Silberglitt & Hintze, 2005).

In addition, the researchers examined the number of false positives and false negatives as measured by predictive power. The overarching goal in this type of assessment is to minimize the number of false negatives in determining eligibility for academic services. In this instance, “high values of negative predictive power indicate that, of the students who reached the CBM-R cut score, a high percentage of these students also passed the MCA” (Silberglitt & Hintze, 2005, p.320). In this particular study, 81-83% of students whose scores were above the appropriate cuts scores for their grade level also passed the MCA. The cut scores established for this study were also consistent with previous research (Silberglitt & Hintze, 2005).

Stage and Jacobsen (2001) examined the predictive validity of CBM-R as related to student outcomes on the Washington Assessment of Student Learning (WASL), a state-mandated achievement test. Students were assessed at three different points during the school year (September, January and May) using oral reading probes taken from the reading curriculum for the appropriate grade level. The researchers used hierarchical linear modeling (HLM) analyses to model individual oral reading fluency slopes across all three assessments. An overall average slope was determined for the participants. HLM analyses examined whether the difference of the y-intercept from zero and whether the slope across the three

measurement periods were statistically significant. The researchers also evaluated the relationship between students' WASL reading test scores with the students' slope and y-intercept from the fluency measures. Results indicated that students' oral reading fluency scores as measured at each assessment point better predicted reading performance on the WASL when compared to growth in oral reading fluency across the entire school year.

McGlinchey and Hixson (2004) replicated Stage and Jacobsen's (2001) study in Michigan. The purpose was similar, but was extended to a larger, more diverse sample over a longer period of time. Students were administered reading passages two weeks before the Michigan Educational Assessment Program (MEAP) reading assessments. Researchers examined the relationship between reading rate (with a cut score of 100 words correct per minute (WCPM) and scores on the MEAP. A higher cut score, such as 100 WCPM, decreased the probability of predicting a failing score while it increased the probability of predicting a satisfactory score. Results indicated a moderately strong relationship between oral reading rate and performance on the MEAP. The correlation here was higher than that of the Stage and Jacobsen (2001) study (.64 and .44, respectively) (McGlinchey & Hixson, 2004).

In an unpublished study cited by Fuchs (2003), Fuchs, Fuchs, Hosp, and Hamlett utilized a dual discrepancy normative approach, with the normative sample limited to other children in the same class. They administered a battery of reading measures as a pre-test in September and again as a post-test in the Spring. Over the course of the first 15 weeks of school, the researchers collected weekly CBM-R data on all children in 14 different second

grade classrooms. They used the CBM-R data to examine end-of-year performance and beginning-to-end-of-year growth in the area of reading. Students' level and slope (as generated by CBM-R data) were compared to other students in their own classroom or to the entire sample. For predicting student growth from the beginning to the end of the school year, CBM-R level and CBM-R dual discrepancy yielded moderate to large effect sizes for both comparison to the classroom sample and the entire sample (.60-.67 and .84-.87, respectively). Use of CBM-R level and CBM-R dual discrepancy to predict post-test performance yielded very large effect sizes (1.77-2.82 and 1.60-2.27, respectively). The use of CBM-R slope also yielded a moderate effect size for post-performance (.57), and large effect sizes for predictions of growth (.99-1.16). These effect sizes indicate that the use of CBM-R level and dual discrepancy are solid tools for predicting growth and outcomes.

As this was not the original intent behind the development of CBM-R, more research is needed to determine whether CBM-R is a valid tool for predicting high-stakes outcomes, as well as for guiding high-stakes decisions (Christ & Silbergitt, 2007). Preliminary research, however, does provide a promising future for the expanded use of CBM-R.

CBM-R strengths and limitations. The strengths of CBM-R have contributed to it being the standard method of choice for assessing reading fluency. In addition to its established reliability and validity (Christ & Silbergitt, 2007; Marston, 1989), CBM-R is useful for a variety of purposes including predicting important student outcomes (McGlinchey & Hixson, 2004; Silbergitt & Hintze, 2005; Stage & Jacobsen, 2001), aiding in educational decision-making (Deno et al., 1986; Fuchs & Fuchs, 2006; Shinn, 2007;

Silberglitt & Hintze, 2007), and monitoring student growth (Ardoin & Christ, 2008; Good et al., 2001; Hintze & Silberglitt, 2005; McGlinchey & Hixson, 2004; Stage & Jacobsen, 2001).

CBM-R is an inexpensive measure that is easy for teachers and paraprofessionals to administer in a brief period of time (Deno, 1986; Francis et al., 2008; Marsten, 1989). Research has demonstrated that assessment based on “non-curricular” CBM-R passages, such as those published by AIMSweb and DIBELS, are just as effective (if not more so) as passages derived from students’ classroom curriculum (Fuchs & Deno, 1994; Riley-Heller et al., 2005; Powell-Smith & Bradley-Klug, 2001). CBM-R passages, such as those included in DIBELS and AIMSweb, also come pre-packaged, making them a more convenient choice for educators.

However, there are several limitations of CBM-R. Measurement error and researchers’ lack of reporting measurement error are among the primary limitations (Ardoin & Christ, 2008; Poncy, Skinner, & Axtell, 2005). More recently, research has addressed the issue of measurement error and its relationship to interpreting CBM-R outcome scores (Christ & Silberglitt, 2007).

Christ and Silberglitt (2007) set out to estimate likely magnitudes of the standard error of measurement (*SEM*) for CBM-R. They also examined the magnitudes of the *SEM* to evaluate and identify a range of *SEM* values to help with CBM-R interpretation. The overall median estimate of *SEM* across all grades (first through fifth) and conditions was 10 words correct per minute (WCPM) (range = 5-15). This estimate of *SEM* is rather high, especially when one considers that most elementary-aged students are only expected to increase oral

reading fluency scores by no more than 1 to 2 words per week over the course of a school year (AIMSweb, 2008; Good & Kaminski, 2002; Hasbrouck & Tindal, 2006). Christ and Silbergitt (2007) also pointed out that the *SEM* can be used to estimate the magnitude of the standard error of the estimate (*SEE*) and the standard error of the slope (*SEb*), the latter being most relevant for decision making purposes (e.g., diagnostic and/or eligibility decisions). Though they did not provide values for the *SEE* or *SEb*, Christ and Silbergitt suggested that the variability (*SEE*) around the trend line is likely to approximate the *SEM*. Although they recommended that estimates of the *SEM*, *SEE*, and *SEb* should be included with standardized progress monitoring materials, they also stated that measurement error has not yet been sufficiently addressed in research literature.

Researchers agree that one way to increase reliability and decrease measurement error is to obtain more data with a specific student. In other words, students should be assessed on multiple (e.g., 10 to 40) passages that are of similar difficulty levels (Ardoin, 2006; Ardoin & Christ, 2008; Christ & Silbergitt, 2007; Poncy et al., 2005). Researchers have also found that the accuracy of CBM-R scores improved with standardized administration and scoring procedures (Colon & Kranzler, 2006; Hintze & Silbergitt, 2005) as well as consistent measurement conditions (Ardoin & Christ, 2008).

One common means for obtaining multiple data points for students is to assess all students at three time points during the school year (i.e., fall, winter and spring). Although these tri-annual assessments have provided sufficient data to predict end of year outcomes across students (Stage & Jacobsen, 2001; McGlinchey & Hixson, 2004), the tri-annual

assessment system does not provide sufficient data for teachers to make monthly, weekly, or daily instructional decisions. This fact is particularly important for students who may be struggling with reading and receiving some form of reading intervention. In other words, teachers need to know on a more frequent basis (e.g., weekly rather than every three to four months) whether such students are responding to the reading intervention so that timely instructional decisions can be made to maximize each student's learning. This logic therefore fits as part of a RTI framework, in that students receiving Tier 2 or Tier 3 support ought to be assessed more than once per three to four months.

However, the need for more frequent progress monitoring of students receiving supplemental instruction or intervention (such as at a Tier 2 or Tier 3 level) brings about the obvious logistical problem of ensuring that teachers have the time and resources to administer more frequent assessments. At the present time, no known research has specifically investigated this possible logistical problem, but clinical experience suggests that many teachers would find weekly or even by-weekly oral reading fluency assessments for all students receiving supplemental reading support a time-consuming and unwanted task. A small amount of evidence from one recent study supports this observation. Specifically, as a small component of a study investigating teachers' perceptions of students reading ability, Begeny, Groce, Krouse, and Mann (2008) asked 27 first through fifth grade teachers about their willingness to administer (and/or facilitate) CBM-R oral reading fluency assessments for struggling readers three times per year versus two times per month. Using a Likert-type scale from 1 to 5 (1 = Strongly Disagree, 3 = Agree; 5 = Strongly Agree), teachers' average

rating of willingness to administer CBM-R assessments three times per year was 4.6 ($SD = 0.7$), whereas their willingness to administer such assessments two times per month was 3.9 ($SD = 1.1$). This difference was statistically significant ($p = .002$). Also important to note, this same sample of teachers may have valued the importance of reading fluency more than a “typical” sample of teachers. When asked, “How important do you feel reading fluency is in teaching students to read?” teachers’ average rating was 4.5 ($SD = 0.5$) on a 5-point scale (1 = Not Very Important, 3 = Important; 5 = Extremely Important). Collectively then, these data suggest that with a sample of teachers who clearly value the importance of reading fluency, their willingness to administer or facilitate CBM-R assessments with struggling readers two times per month was significantly lower than their willingness to administer these assessment three times per year. Although this topic is in need of further investigation, these data, coupled with clinical experience in schools, suggest that teachers may have at least some reservation about administering CBM-R assessments with students they also assist regularly with reading intervention.

As another limitation of CBM-R assessment, the difficulty levels of passages, such as those included in DIBELS, are still viewed as an important problem when using progress monitoring procedures. For example, Francis and colleagues (2008) found that the typical method for determining passage difficulty (e.g., the use of readability formulas such as the Spache (1953) readability formula) may not be an accurate determination of the actual level of difficulty for each passage. In other words, students’ WCPM scores may substantially vary from passage to passage and, therefore, may not be the most accurate reflection of their oral

reading abilities.

In sum, although CBM-R assessment procedures in reading have a number of associated strengths (e.g., brevity of administration, strong indicators of reliability and validity), important limitations (e.g., measurement error, varying difficulty levels of passages, and logistical problems for teachers to achieve regular progress monitoring) still exist. To address these limitations, research suggests that CBM-R be administered frequently enough to reduce measurement error (thereby increasing the amount of data obtained and increasing the consistency in measurement conditions). However, finding ways to do this that are acceptable and feasible for teachers have yet to be thoroughly explored.

Evaluating intervention effectiveness/Use in RTI

According to Hintze and Christ (2004), the primary purposes of CBM-R, when used within a problem solving model, are twofold. The first is to obtain sufficient data related to basic skills performance to detect and confirm academic weaknesses. The second is to monitor students' responsiveness to intervention or modified instruction over time in a decisive manner.

Busch and Reschly (2007) outlined the role of CBM-R when integrated within each of the three tiers of RTI. CBM-R at Tier 1 serves two functions: (1) to assess whether the instruction provided in the classroom is sufficient to expect students' growth in reading (or another relevant subject) and (2) to provide classes, schools and/or school districts with the means to "collect normative data on all students' levels and rates of reading growth" (p. 226). The instruction provided at Tier 2 is usually done by teachers in the general education

setting. The intent is to modify the general learning curriculum or educational environment in an attempt to increase the performance of students otherwise not making adequate progress in a particular academic area. Progress monitoring with CBM-R occurs several times per week. Decisions are based on tracking a student's CBM-R data and comparing them to locally established norms or to those norms suggested in the literature. After 10-15 weeks of intervention, the student is moved to either Tiers 1 or 3 based on the change in slope of performance as demonstrated by the CBM-R data. Busch and Reschly (2007) equate Tier 3 with special education. Here, CBM-R is used to set and monitor progress toward performance and/or IEP goals. The use of CBM-R at this tier also helps teachers meet several of the requirements outlined as part of IDEIA—specifically, setting measurable annual goals and monitoring student progress. CBM-R also allows special education teachers to give timely feedback to parents and others regarding students' progress.

Much of the discussion related to CBM-R to this point has centered on using it as an assessment tool, with emphasis on the area of reading. It is worthwhile at this point to deviate from the discussion on CBM-R and look at reading fluency in general, as well as interventions that aim to improve reading fluency for emerging readers. A more thorough description of CBM-R as a means for assessing reading abilities will be included within that discussion.

Reading Fluency

The ability to read is a necessary skill for functioning in our society, yet many students still struggle to acquire basic reading skills. In 2007, the National Center for

Education Statistics (NCES) reported that 33% of children in Grade 4 and 26% of those in Grade 8 read below the basic level (Lee, Grigg, & Donahue, 2007). To address this critical issue, five key areas of reading of reading development and instruction have been identified as primary areas of focus. They are: phonemic awareness, phonics, vocabulary development, reading fluency (including oral reading skills) and reading comprehension (National Institute of Child Health and Human Development [NICHD], 2000; U.S. Department of Education, 2002). Reading fluency serves as a critical bridge among these five skills. A fluent reader has a solid command of phonemic awareness, phonics and vocabulary development, and they are better able to comprehend the meaning of the text (LaBerge & Samuels, 1974; Martens, Eckert, Begeny, Lewandowski, DiGennaro, Montarello, et al., 2007; O'Connor, White, & Swanson, 2007). Additionally, reading fluency arguably serves as the best overall predictor of overall reading ability, which includes reading comprehension and performance on state-mandated achievement tests (Fuchs, Hosp, & Jenkins, 2001; Jenkins et al., 2005; McGlinchey & Hixson, 2004; NICHD, 2000).

Reading fluency has been the subject of much research since the beginning of the 19th century. However, it was not until the work of LaBerge and Samuels in the early 1970s that “reading fluency took a more prominent role in our understanding of reading development” (Chard, Vaughn, & Tyler, 2002, p. 386).

LaBerge and Samuels (1974) proposed a theory of automatic information processing as it relates to reading. This theory proposes that the more efficient an individual is at processing word units into recognizable words and connecting those words while reading

text, the more cognitive energy is available to comprehend the meaning of the text. Put another way, when each component of a complex skill requires attention, the completion of the complex skill would be impossible because it has exceeded attentional capacity. When enough components are completed automatically, however, the “attentional load” falls within “tolerable limits,” and the skill is successfully executed. This theory also assumes that lower level cognitive processes must be completed before higher level processes (Fuchs et al., 2001).

Reading fluency, on the whole, is a complex and multifaceted process that requires the synthesis of a variety of skills (Fuchs et al., 2001). The skills required for fluent reading include the identification of letters, phonological awareness (i.e., sound-symbol relationships), sight-word vocabulary and phonics skills (Snow, Burns, & Griffin, 1998). These skills are acquired gradually, with the greatest growth occurring between kindergarten and second grade (Fuchs et al., 2001).

Assessment of Reading Fluency

As Fuchs and colleagues (2001) suggest, “the most salient characteristic of skillful reading is the speed with which text is reproduced into spoken language” (p. 239). It makes sense, then, to utilize assessment measures that depend on oral reading rate. The validated CBM approach to measuring reading ability (Marston, 1989) is the most widely researched measure for oral reading fluency (ORF) (Busch & Reschly, 2007). In the CBM-R assessment, a student reads a text orally for one minute. The WCPM score is calculated by the examiner (Deno, 1985; Marston, 1989). The CBM-R oral reading procedure has strong

alternate form reliability, making it appropriate for monitoring student progress over time. It can also differentiate growth patterns for students across varying levels of achievement (e.g., high, average, low), making it a valid indicator of overall student progress (Busch & Reschly, 2007). Additionally, CBM-R has a standard protocol, making it easy to administer and time efficient (Busch & Reschly, 2007; Stage & Jacobsen, 2001).

As previously mentioned, a cut score is typically employed to determine whether a student is progressing. In the case of CBM-R, the cut score establishes the “critical number” of WCPM and serves as a benchmark for reading proficiency. Students scoring at or above the cut score are considered proficient readers, and students scoring below the cut score are considered to be in need of additional supports (Silberglitt & Hintze, 2005).

CBM-R cut scores have traditionally been based on local norms, and the percentiles were used to make decisions about placement in special education or to receive instructional intervention (Deno, 1985). However, it has been argued more recently that a norming sample comprised of more students from more diverse settings would offer a more meaningful foundation for data-based decision-making (Silberglitt & Hintze, 2005).

Hasbrouck and Tindal (1992) conducted a large study using data from thousands of students over the course of nine years in various districts around the country. Their study is widely cited and used in educational decision-making (Silberglitt & Hintze, 2005). More recently, Hasbrouck and Tindal (2006) updated their work and identified the following as cut scores for CBM-R performance at the end of each grade level: 53 WCPM (1st grade); 89 WCPM (2nd grade); 107 WCPM (3rd grade).

Silberglitt and Hintze (2005) examined the relationship between CBM-R and the Minnesota Comprehensive Assessment (MCA), a state-mandated achievement test administered annually beginning in the spring of Grade 3. They also sought to establish cut scores for CBM-R that best predicted students' outcomes on the MCA. While their cut scores were determined specifically for one test, they determined cut scores for two to three benchmark points during each grade level for first through third grades (i.e., fall, winter, spring). Results indicated that CBM-R was a strong tool for predicting outcomes on the MCA. The cut scores established in this study closely resembled those established by Hasbrouck and Tindal (1992). Additionally, the majority of students in the Silberglitt and Hintze (2005) study who read at or above the cut score at each benchmark period passed the MCA (83.5% in Grade 3 Spring). Likewise, the majority of those who read below the cut score failed the MCA (68.5% in Grade 3 Spring). The advantage of the cut scores at the benchmark periods as established in the Silberglitt and Hintze (2005) study is that use of these cut scores will help teachers better determine whether their students are "on track." These scores can also serve as criteria for intervention effectiveness within an RTI framework.

The primary assumption in using oral reading fluency (ORF) to monitor reading progress is that, when controlling for levels of passage difficulty, changes in performance between measurement points one and two reflect changes in a student's reading performance. However, given the inherent variability of scores between passages, the use of multiple passages at a single measurement point can reduce measurement error. Measurement error

due to variability in passage difficulty can be reduced by using multiple passages (e.g., three or four) at each test point (Jenkins, Zumeta, Dupree, & Johnson, 2005).

Another factor to consider when using assessment of ORF using CBM-R is the frequency and long-term duration of measurement. Teachers can make more reliable estimates of reading growth when they collect ORF data regularly over longer periods of time (e.g., 1-2 times per week for 1 year) (Fuchs, Fuchs, Hamlett, Walz, & Germann, 1993). When ORF is assessed over shorter intervals (e.g., 8 weeks), informative growth estimates can be obtained (Fuchs, 2003), but may include significant measurement error (Hintze & Shapiro, 1997). Evaluators should be cognizant of this fact as it can depict a less accurate picture of growth.

Assessment using CBM-R relies on the measurement of slope, or the average amount of gain expected from one week to the next. More specifically, slope is the average weekly expected increase in the number of words read correctly per minute (Silbergliitt & Hintze, 2007). With the repeated measurements inherent in CBM-R, time-series data are produced that indicate individual growth over days, weeks and months (Deno et al., 1986). These types of data make CBM-R more sensitive to change (Deno et al., 1986; Hintze & Christ, 2004; Stage & Jacobsen, 2001). Slope and variability of individual performance are then used to make educational decisions (Deno et al., 1986).

As previously mentioned, dual discrepancy is an effective way to determine a student's level of performance individually and as it compares to his or her peers. If the student's level of performance is more than one standard deviation below that of the mean

for the class, an achievement gaps exists that will likely only widen without intervention (Ardoin et al., 2005). An additional method for assessing performance level is median split. Here, the slope values for the group of students being evaluated are split equally into two groups: top and bottom or higher and lower. The level of performance of each group is compared to the other, which can provide more discriminate information about students' performance when other methods do not seem to differentiate the level of performance (Fuchs, 2003; Fuchs, Fuchs, & Compton, 2004). Median split is more useful for first grade students, though both dual discrepancy and median split have both emerged as the overall best methods for determining level of performance or responsiveness to intervention (Fuchs, 2003).

Overall reading achievement is usually not assessed until the end of third grade. Reading First was established as part of No Child Left Behind with the goal of all children reading proficiently by the end of third grade (U.S. Department of Education, 2002). With state assessments not occurring until the time children are expected to read proficiently, there are very little data to judge ongoing progress or guide educational decisions (e.g., the need for intervention) prior to the end of third grade (Crawford et al., 2001). Fortunately, the body of research concerning CBM-R provides the field with a valid tool to accurately assess students' level of performance and their ongoing progress in the area of reading fluency. Additionally, more recent research suggests that CBM-R is useful for predicting outcomes on state-mandated achievement tests (e.g., Crawford et al., 2001; Good et al., 2001; McGlinchey & Hixson, 2004; Silberglitt & Hintze, 2005; Stage & Jacobsen, 2001). Assessment is a vital

component of the education process in that it helps education professionals to make informed decisions regarding students' educational trajectories. Likewise, assessment informs educators about specific areas appropriate for intervention.

Empirically-based Reading Fluency Interventions

Approaches to improving reading fluency have been at the center of much research within the past ten years. A number of intervention studies have been published utilizing a variety of intervention programs and/or techniques such as model reading and guided oral reading. Additionally, several syntheses of reading interventions have been published to determine the most effective reading fluency interventions.

Therrien (2004) conducted a meta-analysis of repeated reading (RR) studies to determine whether RR was effective for improving reading fluency and comprehension. He was also interested in the components within a repeated reading program that are crucial to the success of the program. Effect sizes (ES) were based on either nontransfer measures (i.e., utilizing the same passage for multiple readings) or transfer measures (i.e., students' ability to read and comprehend a new passage after reading a different passage several times).

On the whole, Therrien (2004) found that, as a result of the fluency interventions evaluated, all subjects obtained a moderate increase in fluency with repeated reading. However, the best situation (i.e., yields the greatest increase in fluency) is one where the student repeatedly reads a passage to an adult who cues the reader to focus on speed and comprehension, and where performance criterion and charting are utilized. Kuhn and Stahl (2003) agreed that model reading, with the student following along, produced gains in

fluency and comprehension. In addition, they noted that repeated reading has measureable effects on fluency as well as speech pauses and intonation.

A meta-analysis by Chard and colleagues (2002) also examined various components of reading fluency interventions. They found that oral repeated reading by itself resulted in “significantly better scores for fluency and comprehension” (Chard, Vaughn, & Tyler, 2002, p. 389). When an adult modeled the passage for the student as he or she followed along, the student performed better on comprehension measures and had faster reading rates than those who only read a passage repeatedly. Several studies in this synthesis resulted in gains in reading fluency when the student read a text multiple times. Scores on a measure of reading fluency were significantly better when subjects read the same passage three times than when they read it only once. Similar differences were evident on a measure of comprehension as well.

Another intervention component, though it receives less attention in literature, is known as phrase-drill error correction. The phrase drill requires students to repeatedly read a phrase (i.e., 5-6 words) that contains a word read incorrectly during a previous reading (Begeny & Martens, 2006), which provides the student the opportunity for additional reading practice. The phrase drill can be considered part of an error correction or feedback procedure.

While each of the interventions described produce gains in fluency, a combination of intervention components has oftentimes produced larger, more immediate gains in oral reading fluency (Begeny & Martens, 2006; Chard et al., 2002; Therrien, 2004). In sum, the most effective interventions for reading fluency include the following components: adult

model reading with students following along, repeatedly reading a passage at least three times to an adult, the student being cued to focus on speed and comprehension, the use of performance criteria and charting, and the use of a phrase drill for error correction. One recently developed instructional program that utilizes each of these components is the Helping Early Literacy with Practice Strategies (HELPS) Program.

Chapter 2:

Evaluating Assessment with the HELPS Program

The HELPS Program

The Helping Early Literacy with Practice Strategies Program (henceforth referred to as HELPS) was originally designed as a reading fluency program that teachers, teacher-assistants, or even community volunteers can use as a time-efficient supplement to all early readers' (e.g., second grade students') core reading curriculum (Begeny, 2007; Begeny, Laugle, Krouse, Lynn, Parker, & Stage, 2008). As such, HELPS does not require extensive training for its successful implementation and it can be implemented in approximately 10 minutes per session (with implementation suggested two to three times per week). Since its development, HELPS has also been implemented (and shown effective) specifically as an intervention program for low-performing readers (Begeny, Mitchell, Whitehouse, & Harris, 2009).

Procedurally, HELPS is a systematic series of instructional strategies aimed to improve early readers' basic reading skills, with a particular focus on improving students' reading fluency. HELPS instructional procedures were selected from all of the available evidence supporting strategies that improve students' reading fluency, and as a result, students' reading comprehension (see Chard et al., 2002; NICHD, 2000; Therrien 2004). Thus, HELPS incorporates the following instructional components and strategies: (a) repeated reading, (b) model reading, (c) phrase-drill error correction, (d) goal-setting, (e) graphical displays of student progress, (f) performance feedback, (g) use of systematic praise

and a token economy for student reading behaviors and accomplishments, and (h) verbal cues for students to read with fluency and for comprehension.

In the initial studies evaluating HELPS, instructional materials comprised of 88 passages that ranged in difficulty from the beginning of first grade to the end of fourth grade. The difficulty levels of the passages were calculated using the Spache readability formula (Spache, 1953). Eight passages, representing readability levels from 1.17 to 2.31, were selected from the first grade reading series from Silver, Burdett, and Ginn (Pearson et al., 1989). These passages were used as a more appropriate starting point for lower performing second grade students. The remaining passages were selected from the Dynamic Indicators of Basic Early Literacy Skills, 6th Edition progress monitoring materials (DIBELS; Good & Kaminski, 2002). Readability levels of these passages ranged from 2.27 to 4.99. All passages were sequenced by level of difficulty (Begeny, 2007, in press).

In addition to the instructional passages, HELPS materials include a general implementation protocol (which describes the strategies to implement and the sequenced order of procedures), as well as a scripted protocol for the HELPS implementer (henceforth referred to as “teacher”) to read at transition points during HELPS implementation. HELPS materials also include a customized, standard-interval graphing chart and a “Star Chart,” which is used as part of the token economy. A customized tracking sheet is also included as part of the HELPS materials so that teachers can record students’ reading scores across sessions and easily communicate students’ progress to all teachers using HELPS with a specific student. Appendices A through F show the above materials in the order they are mentioned.

Thus far, three separate studies have supported the effectiveness of HELPS. The initial study found that when used as a supplement to second grade students' core reading program (from February until April), students receiving HELPS significantly improved their basic reading skills (including their reading fluency) when compared to a control group (Begeny et al., 2008a). In a follow up study that again evaluated HELPS as a supplement to second grade students' core reading curriculum, Begeny (2009) compared the effects of HELPS when implemented at different frequencies. Specifically, one group of students (N = 29) received HELPS three times every two weeks, another group of students (N = 29) received HELPS six times every two weeks, and a third groups of students (N = 30) served as a control group. Also, in contrast to the more brief study described above, the duration of the Begeny (2009) study lasted from October to April. Preliminary findings from this study confirmed that all students receiving HELPS significantly outperformed students in the control group across various measures of reading skills. Furthermore, students benefitted significantly more on some measures of reading when receiving HELPS six times every two weeks (i.e., approximately every Monday, Wednesday, and Friday) versus three times every two weeks.

In the most recent study (Begeny et al., 2009), HELPS was implemented as an intervention program for struggling second grade readers. Also, in contrast to the university volunteers who implemented HELPS in the first two studies, this study utilized classroom teachers and teacher assistants as the implementation agents. Similar to earlier findings, preliminary findings from this study demonstrate that students receiving HELPS significantly

outperformed the control group across several measures of early reading.

Using HELPS as a CBM-R Assessment Tool

HELPS was initially developed as an instructional program for early readers and has been shown to improve reading fluency and comprehension skills for early readers of all ability levels. It was designed to supplement regular classroom literacy instruction, providing additional support to students in a way that requires relatively little instructional time (Begeny et al., 2008b). Subsequent research demonstrated that HELPS significantly improved reading fluency and comprehension for second grade struggling readers, and was successfully implemented by teachers, teaching assistants, and community volunteers (Begeny et al., 2009).

In addition to the instructional components integrated within the HELPS Program, a unique characteristic of this program is that it also integrates three separate forms of CBM-R progress monitoring: (a) student progress made within a session from practicing the same passage (i.e., immediate passage gains, IPGs); (b) student retention of progress made across 2 to 3 days without practicing the passage (i.e., retention of passage gains, RPGs), and (c) student performance on novel passages, presumably resulting from the overall effect of HELPS (i.e., generalization of passage gains to new materials, GPGs). The latter of these three forms of CBM-R progress monitoring exemplifies the traditional form of CBM-R progress monitoring (i.e., using novel passages over time to assess students' oral reading fluency). The first two progress monitoring procedures (IPGs and RPGs) have been used in research evaluating the effects of reading fluency interventions (e.g., Begeny & Silber, 2006;

Martens et al., 2007), but unlike GPGs, they have yet to be explored as predictors of more general reading outcomes.

Purpose and Rationale

The primary purpose of this study is to explore the psychometric characteristics and predictive validity of the three forms of CBM-R progress monitoring that are built within the HELPS Program. The main reasons for this investigation are as follows, all of which speak to limitations of traditional CBM-R progress monitoring: (a) HELPS naturally produces multiple assessments over time (i.e., usually up to 3 assessment per week), yielding a large number of data points and presumably decreasing measurement error associated with CBM-R, (b) passages developed independently of the student's curriculum are used, which leads to a more systematic control of passage difficulty, (c) measurement conditions are highly consistent because students experience the same "conditions" up to three times per week across multiple weeks, thereby reducing possible measurement error due to inconsistent measurement conditions, and (d) assessment procedures are built directly into instructional procedures, thereby possibly increasing teachers' use (and acceptability of) frequent progress monitoring with students getting supplemental reading support. Collectively, if the HELPS assessment procedures demonstrate similar or better psychometric characteristics compared to traditional CBM-R procedures, the advantages of the HELPS assessment procedures may prove to be a preferable alternative from traditional CBM-R.

Research Questions and Hypotheses

Based upon this purpose and rationale, the specific research questions and associated hypotheses for this study are as follows:

Research question 1.

Are there statistically significant changes over time in students' generalization of passage gains to new passages (GPG), as measured by students' slope of GPGs?

Hypothesis 1.

The bulk of CBM-R research has focused on progress monitoring using students' WCPM scores for the readings of unfamiliar passages (Fuchs, 1989; 2003; Fuchs et al., 2004; Hintze & Christ, 2004; McGlinchey & Hixson, 2004; Silberglitt & Hintze, 2005; 2007), or evaluating intervention effectiveness based on WCPM scores for either new or practiced passages (Begeny & Martens, 2006; Busch & Reschly, 2007; Fuchs & Fuchs, 2006; Glover & DiPerna, 2007). Over time, particularly when students participate in fluency interventions, students' WCPM scores tended to increase (Begeny & Martens, 2006; Jones & Wickstrom, 2002; Martens et al., 2007). Thus, it is hypothesized that, as a result of participating in HELPS, students' slopes for GPGs will be positive over time.

Research question 2.

Are the changes in GPGs over time associated with the differences in students' pre- to post-test scores for measures of reading fluency and comprehension as assessed by the *Gray Oral Reading Test, Fourth Edition* (GORT; Wiederholt & Bryant, 2001)?

Hypothesis 2.

Increasingly, researchers have evaluated the extent to which CBM-R scores will predict students' outcomes on end of year assessments such as state-mandated achievement tests (Crawford et al., 2001; Glover & DiPerna, 2007; Good et al., 2001; McGlinchey & Hixson, 2004; Silberglitt & Hintze, 2005; Stage & Jacobsen, 2001). In general, CBM-R scores are useful for predicting broad outcomes of reading achievement, such as the GORT, which can be considered a broad assessment of reading fluency and comprehension. To date, research that shows that students receiving fluency interventions show increases in GORT scores is limited to previous HELPS studies (Begeny, 2009; Begeny et al., 2009). However, research has not evaluated whether CBM-R predicts GORT outcomes. Given that participation in HELPS results in fluency gains (Begeny, 2009; Begeny et al., 2008b; Begeny et al., 2009) and that CBM-R is useful for predicting broad outcomes of reading achievement, it is hypothesized that there is a relationship between students' CBM-R scores and the differences in students' pre- to post-test scores for the GORT Fluency and Comprehension subtests.

Research question 3.

Are there statistically significant changes over time in students' immediate passage gains (IPG), as measured by students' slope of IPGs?

Hypothesis 3.

Previous research examining IPG has demonstrated positive results such that students participating in interventions with various components (e.g., repeated reading, listening

passage preview, etc.) experienced immediate increases in WCPM (Begeny & Silber, 2006; Martens et al., 2007). Martens and colleagues (2007) noted that students' gains in fluency within each intervention session actually decreased over time. They suggest that this decrease over time indicated that as students performed the first reading at a higher level of fluency, they benefitted less from the fluency training. In other words, the students were reaching their upper fluency limits as the intervention progressed.

Given the paucity of research regarding IPGs over time, the question regarding the changes in students' slope over time is more exploratory in nature. It has been demonstrated that HELPS, an intervention that includes the various components described above, results in increases in students' WCPM scores (Begeny, 2009; Begeny et al., 2009). It is likely that, just as with the study conducted by Martens and colleagues (2007), as students' levels of fluency increase over the course of the intervention their IPGs may actually decrease over time, resulting in a negative slope.

Research question 4.

Are the changes over time in students' immediate passage gains (IPG), as measured by students' slope of IPGs, associated with the differences in students' pre- to post-test scores for measures of reading fluency and comprehension as assessed by the GORT?

Hypothesis 4.

Although students' IPGs have been evaluated to determine intervention effectiveness (Begeny & Silber, 2006; Daly, Martens, Hamler, Dool, Eckert, 1999; Martens et al., 2007), students' IPGs have not been used to predict outcomes. Thus, the question of whether IPGs

are associated with pre- to post-test differences in the GORT is exploratory. If, however, a decrease in students' slope for IPGs over time indicates that students are reading at a higher level of fluency (Martens et al., 2007), then it is reasonable to hypothesize that these negative changes in performance over time will be associated with changes on outcome measures such as the GORT.

Research question 5.

Are there statistically significant changes over time in students' retention passage gains (RPG), as measured by students' slope of RPGs?

Hypothesis 5.

Research evaluating intervention effectiveness typically focuses on students' IPGs as opposed to RPGs (Daly et al., 1999; Jones & Wickstrom, 2002). However, some research has examined students' RPGs and shown that, again, with interventions including multiple components, students' tended to retain their IPGs over the two day period between intervention sessions (Begeny & Silber, 2006; Martens et al., 2007). To date, there is little research to examine whether RPGs are associated with overall increases in fluency growth over time. Thus, this question is also exploratory. It is possible that students will experience fluency gains over time (Begeny & Silber, 2006; Martens et al., 2007), and that they will retain these gains, though their WCPM scores for retention passages during an intervention session may not be as high as they were for the readings in the session two days prior. However, similar to students' IPGs decreasing over time as fluency increases (Martens et al., 2007), students' may experience a parallel process with their RPGs over time. It is

hypothesized that there will be significant changes in students' slope for RPGs over time, such that RPGs will decrease over time.

Research question 6.

Are the changes over time in students' retention passage gains (RPGs), as measured by students' slope of RPGs, associated with the differences in students' pre- to post-test scores for measures of reading fluency and comprehension as assessed by the GORT?

Hypothesis 6.

As with students' IPGs, students' RPGs have not been used to predict outcomes. Thus, the question of whether RPGs are associated with pre- to post-test differences in the GORT is exploratory. If students' RPGs do, in fact, decrease over time as a result of a parallel process to the decreases in IPGs as suggested by Martens and colleagues (2007), then such decreases may indicate gains in reading fluency. Thus, it is hypothesized that there will be a negative association between students' RPGs and the differences in students' pre- to post-test scores for the GORT Fluency and Comprehension subtests.

Chapter 3:

Method

The data analyzed in the present study come from a previous study evaluating the implementation of the HELPS Program by teachers in the regular education setting (Begeny et al., 2009).

Participants and Setting

Teacher participants. A total of nine second grade classrooms from one rural school in the southeastern United States participated in the study. Four teachers and each of their teaching assistants were randomly selected to be trained to implement HELPS with low-performing students in their classrooms. Students and teachers from five other classrooms made up the control group. For the purposes of this study, only the teachers who implemented HELPS (and the students receiving HELPS) will be described. Information about the control group teachers and students can be found in a separate report (Begeny et al., 2009).

Teachers in the HELPS condition had an average of 11.75 years of teaching experience (range = 4-26 years), and an average of 10 years teaching second grade (range = 3-20 years). HELPS teacher assistants had an average of 2.25 years of teaching experience (range = 1-4 years), and an average of two years' experience teaching second grade (range = 1-3 years).

Prior to implementing HELPS with students, the HELPS teachers and teacher assistants (henceforth referred to as teachers) were trained how to use HELPS procedures and program

components during two, four-hour training workshops held within one week of each other in October. Following the training workshops, each teacher was given a take-home HELPS implementation quiz consisting of 12 True/False, 12 multiple choice, and three short-answer questions. To better ensure the teachers' understanding and correct implementation of the program, the quizzes were reviewed by research assistants trained on HELPS procedures and the research assistants provided teachers with feedback regarding any incorrect responses.

Student participants. Thirty students were selected by their teachers from four separate classrooms to receive HELPS. One student moved during the study, bringing the total number of participants to 29 second grade students. The average age of student participants was 7 years, 6 months (range = 6 years, 9 months-8 years, 10 months). Of the 29 participants, 14 were female, 17 were Caucasian, five African-American, five Hispanic, and two were identified as "Other ethnicity." None of the students received services for English as a Second Language. Additional student demographic information (e.g., presence of an education and/or psychological disability, eligibility for free or reduced lunch) could not be obtained for specific students due to state and county regulations.

Reading instruction across all second grade teachers in the participating school was similar. All teachers incorporated language arts into their curriculum for approximately 90 minutes each day. All utilized the *Houghton Mifflin* basal reading series, daily reading groups (e.g., small groups determined by student reading abilities), independent reading, phonics and vocabulary lessons, and writing activities.

Setting. All intervention sessions were implemented one-on-one (teacher-student)

either in the classroom or just outside the classroom in a quiet area in the hall, as determined by each teacher.

Assessment Materials

In the original investigation (Begeny et al., 2009), several measures of reading fluency were used to evaluate students' reading performance. For the purposes of this investigation, the Fluency and Comprehension subtests from the *Gray Oral Reading Test, Fourth Edition* (GORT; Weiderholt & Bryant, 2001) will be analyzed. Because the GORT provides alternate test forms, Form A was used during pre-test and Form B was used during post-test.

In their *Analysis of Reading Assessment Instruments for K-3* (Big Ideas in Beginning Reading—Assessment Domain, 2005), the Institute for the Development of Educational Achievement indicated the GORT and Curriculum Based Measurement (CBM-R) as having sufficient evidence for assessing reading fluency (and comprehension, in the case of the GORT Comprehension measure). These measures also met important standards for reliability and validity as assessed by expert committees associated with national reading centers such as the Institute for the Development of Educational Achievement (Big Ideas in Beginning Reading—Assessment Domain, 2005), and the Florida Center for Reading Research (Florida Department of Education, 2005). In addition, the GORT measures reading comprehension, which is a skill that has been linked to reading fluency (Chard et al., 2002; LaBerge & Samuels, 1974). Thus, the GORT measure was selected as the measure of overall student achievement because of its measurement of reading skills considered essential to early

reading instruction (e.g., fluency, comprehension). None of the other measures used in the Begeny et al., 2009 study evaluated both fluency and comprehension.

Gray Oral Reading Test, Fourth Edition (GORT-4). The GORT-4 was used to evaluate students' reading accuracy, rate, and comprehension on 14 developmentally sequenced passages of 50 to 200 words in length. Students also answered five multiple choice comprehension questions following each passage. Five scores were obtained from the test: Accuracy; Rate; Fluency (Rate + Accuracy); Comprehension; and the Oral Reading Quotient, comprised of the sum of the Fluency and Comprehension standard score. Coefficient alphas for the Fluency and Comprehension components, as well as for the Oral Reading Quotient, are all at or above .90 for early primary students. Test-retest reliability ranged from .85 to .95 for the four subtest scores and .95 for the Oral Reading Quotient (Wiederholt & Bryant, 2001).

Curriculum Based Measurement (CBM-R) program reading passages. The previously described HELPS reading passages were used according to the traditional CBM-R method. The passages provided to students during the HELPS sessions were typed and did not contain pictures. Examiner copies of the passage contained a cumulative word count at the end of each line of text. Participants had not read any of these passages prior to their participation in the HELPS Program. The scores for the number of words read correctly per minute (WCPM) were used to determine the WCPM score for the first reading of each new passage, for students' gains per session (the difference between the WCPM scores for the first and last readings of a passage within the session), and for students' retention gains (the

difference between the WCPM score for the last passage reading of one session and the first reading of the same passage the next session).

Procedure

All students were assessed with a series of pre- and post-test measures, including the GORT (see Begeny et al., 2009 for a complete list of measures). Students participating in the HELPS Program received the intervention two to three times per week from mid-October to early April. Students in the HELPS Program received an average of 50.43 sessions (range = 44 to 58) throughout the study. Most HELPS sessions were implemented by the teacher, who followed a procedural checklist of steps and scripted directions to ensure implementation integrity. Each session lasted approximately 12.5 minutes. For a more detailed description of a HELPS session, see Appendix F.

Analyses

Multilevel modeling (MLM; Raudenbush & Bryk, 2002) was selected for the analyses of these data for its advantages over statistical techniques that evaluate change between two time points (Neupert, Miller, & Lachman, 2006). MLM models growth over repeated measures, allowing the researcher to analyze individual subjects' fluctuation, or slope, over time—a fluctuation that is lost with analyses such as multiple regression (Stage, 2001; Stage & Jacobsen, 2001). Additionally, MLM provides a measure of the percentage of the total variability around each parameter estimated in the model (Stage, 2001). Such an estimate allows for conclusions regarding the variability within subjects across occasions as well as the differences between subjects (Neupert et al., 2006).

In the current study, within-subject changes may provide information regarding students' individual growth in reading fluency. The between-subject differences may be indicative of how students progress through HELPS relative to their peers. These differences may have implications for local norms for reading fluency as measured by CBM-R, as well as how much progress teachers should expect of the students participating in the intervention.

Analytical techniques such as the repeated measures analysis of variance (ANOVA) require subjects to be measured at each time point, prohibiting the inclusion of subjects with missing data in the final analyses (Stage, 2001). For example, as teachers were initially learning to correctly implement HELPS with their students, the protocol was not followed completely. Consequently, students may not have completed the three to four passage readings per session, thereby having an impact on the IPG and RPG scores. MLM uses all available data from all subjects, providing an estimated trajectory for each individual participant (Neupert et al., 2006).

The number of participants for the present study was relatively small ($n = 29$). However, each student received a total of three (Track B) or four (Track A) one-minute timed readings during each session. The WCPM scores from each reading provided multiple data points that provide three separate measures of progress monitoring: (a) student performance on novel passages (i.e., generalized passage gains; GPG) (b) student progress made within a session from practicing the same passage (i.e., immediate passage gains; IPG), and (c) student retention of progress made across 2 to 3 days without practicing the passage (i.e., retention passage gains; RPG). These assessments at multiple time points provided

sufficient data to obtain a reliable estimate of change over time. Time was marked by session as opposed to a unit of time such as day. The standard scores for the GORT Fluency and GORT Comprehension subtests served as the reading fluency outcome measures.

A 2-level hierarchical model was used for analyses. At Level 1, each student's variability (e.g., fluctuation in WCPM over time) is represented by an intercept and slope which become the outcome variables at Level 2. These Level 2 outcome variables may depend on person-level characteristics (Neupert et al., 2006) or, as in the present study, the student's GORT score. A correlation was conducted to determine the relationship between GORT Fluency and Comprehension change scores, and to determine whether these scores could be included simultaneously in each of the MLM models. Results from this correlation indicate that the two scores are not significantly related ($r(27) = .15, p = .45$). Thus, all models included GORT Fluency scores separate from Comprehension scores.

Growth analyses may center the outcome score variables at the grand mean or the group mean (Hawkins, Guo, Hill, Battin-Pearson, & Abbott, 2001). GORT Fluency and Comprehension scores were grand mean centered for all models. To do so, the sample mean was subtracted from each score. The intercept thus became the expected score for the outcome variable when the value of the predictor is the grand mean.

The number of sessions in which students participated was also centered. The y-intercept was centered at the average number of sessions students received and the slope was estimated backwards to the initial session. Doing so made it possible to use the average number of sessions students received to predict the change in performance on the GORT

Fluency and Comprehension subtests at the end of the school year.

For all models, preliminary analyses, known as the fully unconditional model, were conducted to determine whether there was sufficient variability at Levels 1 and 2 to warrant continuation of analyses (Raudenbush & Bryk, 2002). The slopes were constrained for all subsequent models to utilize the best fitting model for each variable.

Each of the three types of gains as they were related to the difference in students' pre- to post-test GORT scores were analyzed using the same series of models. For example, with students' generalization of passage gains (GPGs), the following linear model was conducted (Model 1):

$$\text{GPG}_{it} = \beta_{0it} + \beta_{1it} (\text{SESSION-50}) + r_{it} \quad (1.1)$$

$$\beta_{0it} = \gamma_{00} + \gamma_{01} (\text{GORT FLUENCY CHANGE}) + u_{0i} \quad (1.2)$$

$$\beta_{1it} = \gamma_{10} + \gamma_{11} (\text{GORT FLUENCY CHANGE})$$

In Equation 1.1, GPG_{it} represents the level of GPGs for student i at time t . The variable (SESSION-50) indicates that the analysis is centered at session number 50. Thus, (SESSION-50) equals 0, -45, -40 and so on, corresponding to $t = 0, 1, 2$, etc. The expected level of growth for student i at session 50 is represented by β_{0it} , and β_{1it} is the expected rate of change over time as students are exposed to increasing intervention sessions. In addition, r_{it} is the random within-subjects error of prediction for student i at time t , based on that student's change parameters β_{0it} and β_{1it} . Within-subjects errors are assumed to be mutually independent and normally distributed with a mean of zero (Hawkins et al., 2001).

In the Level 2 model, the individual intercepts (β_{0it}) and slopes (β_{1it}) become the outcome variables. The average number of WCPM for the sample at baseline is represented

by γ_{00} and the average change over time for the sample is represented by γ_{10} . The relationship between students' GPGs (or IPGs or RPGs) and the difference in students' pre- to post-test scores (change scores) is represented by γ_{01} , and the interaction between time (SESSION) and students' change scores is represented by γ_{11} . The extent to which students' vary from the sample average of GPGs is represented by u_{0i} .

Model 2 is identical to Model 1 with the exception of the use of the GORT

Comprehension change score as the Level 2 variable:

$$\text{GPG}_{it} = \beta_{0it} + \beta_{1it} (\text{SESSION}) + r_{it} \quad (2.1)$$

$$\beta_{0it} = \gamma_{00} + \gamma_{01} (\text{GORT COMPREHENSION CHANGE}) + u_{0i} \quad (2.2)$$

$$\beta_{1it} = \gamma_{10} + \gamma_{11} (\text{GORT COMPREHENSION CHANGE})$$

Models 3 and 4 analyzed students' Immediate Passage Gains (IPGs), and Models 5 and 6 evaluated students' Retention Passage Gains (RPGs) as the dependent variables.

In keeping with Raudenbush and Bryk's (2002) recommendations, preliminary analyses (fully unconditional model) were conducted for each model to determine whether there was significant variability at Level 1 and Level 2 to warrant further analyses. An example of the fully unconditional model for GPGs is below:

$$\text{GPG}_{it} = \beta_{0it} + r_{it}$$

$$\beta_{0it} = \gamma_{00} + u_{0i}$$

Chapter 4:

Results

Generalization of Passage Gains

Results from the fully unconditional model for students' GPGs indicated that 53% of the variability in students' WCPM scores for the GPGs for each new passage was between students, and 47% of the variability was within students.

The first analysis examined whether there were statistically significant changes over time in students' generalization of passage gains (GPGs), as measured by students' slope of GPGs. This model also evaluated whether the changes in students' GPGs over time were associated with the differences in students' pre- to post-test scores for measures of reading fluency (Model 1) and comprehension (Model 2) as assessed by the GORT.

The intercept (γ_{10}) for the slope of the number of sessions in which students participated indicates students' WCPM scores (see Figure 1) and, therefore, their GPGs increased over time as they read new passages (see Table 1). In addition, students' WCPM score from the last intervention session was significantly related to, and was a statistically significant predictor of, the change in their GORT Fluency score (γ_{01} , Table 1). However, the interaction between time (the number of sessions in which students' participated) and students' pre- to post-test differences was not statistically significant (γ_{11} , Table 1). This result suggests that, although students experienced gains in GPGs over time, the number of intervention sessions in which students participated did not significantly contribute to the difference in students' Fluency scores.

As with the previous model, Model 2 indicates there was a relationship between GPGs and time, such that students' GPGs significantly increased over time (γ_{10} , Table 1). However, students' GPGs and their Comprehension scores were not significantly related, indicating that the gains students may have experienced did not have a statistically significant effect on their GORT Comprehension scores (γ_{01} , Table 1). The interaction between time and test scores was also non-significant (γ_{11} , Table 1).

Immediate Passage Gains

Results from the fully unconditional model for students' IPGs indicated that 16% of the variability in students' WCPM scores for the IPGs for each new passage was between students, and 84% of the variability was within students.

The models for students' IPGs examined whether there were statistically significant changes over time in students' IPGs, as measured by the slope of IPGs. An additional question was whether the changes over time in students' IPGs were associated with the differences in students' pre- to post-test scores for GORT measures of reading fluency and comprehension. The equations for these two models (Model 3 and 4) are below:

Model 3:

$$\begin{aligned} \text{IPG}_{it} &= \beta_{0it} + \beta_{1it} (\text{SESSION}) + r_{it} \\ \beta_{0it} &= \gamma_{00} + \gamma_{01} (\text{GORT FLUENCY CHANGE}) + u_{0i} \\ \beta_{1it} &= \gamma_{10} + \gamma_{11} (\text{GORT FLUENCY CHANGE}) \end{aligned}$$

Model 4:

$$\begin{aligned} \text{IPG}_{it} &= \beta_{0it} + \beta_{1it} (\text{SESSION}) + r_{it} \\ \beta_{0it} &= \gamma_{00} + \gamma_{01} (\text{GORT COMPREHENSION CHANGE}) + u_{0i} \\ \beta_{1it} &= \gamma_{10} + \gamma_{11} (\text{GORT COMPREHENSION CHANGE}) \end{aligned}$$

The first model in Table 2 shows that there was a statistically significant change in students' IPGs, such that IPGs increased over time (γ_{10} , Table 2). However, students' Fluency scores at post-test were not affected by their gains over time (γ_{01} , Table 2). Although students experienced gains, there was no statistically significant interaction between the number of sessions in which students' participated and their GORT Fluency score (γ_{11} , Table 2).

As with the previous, related model, Model 4 indicates that students' IPGs significantly increased over time (γ_{10} , Table 2). Likewise, students' IPGs were not significantly related to their Comprehension scores at the end of the year (γ_{01} , Table 2). The interaction between time and test score suggests that students' gains declined over time, though this decline was not statistically significant (γ_{11} , Table 2).

Retention of Passage Gains

Results from the fully unconditional model for students' RPGs indicated that 3% of the variability in students' WCPM scores for the RPGs for each new passage was between students, and 97% of the variability was within students.

The models for students' RPGs examined whether there were statistically significant changes over time in students' RPGs, as measured by the slope of RPGs. Additionally, this model evaluated whether the changes over time in students' RPGs were associated with the differences in students' pre- to post-test scores for measures of reading fluency and comprehension as assessed by the GORT. The equations for these two models (Model 5 and 6) are below:

Model 5:

$$\begin{aligned} \text{RPG}_{it} &= \beta_{0it} + \beta_{1it} (\text{SESSION}) + r_{it} \\ \beta_{0it} &= \gamma_{00} + \gamma_{01} (\text{GORT FLEUNCY CHANGE}) + u_{0i} \\ \beta_{1it} &= \gamma_{10} + \gamma_{11} (\text{GORT FLUENCY CHANGE}) \end{aligned}$$

Model 6:

$$\begin{aligned} \text{RPG}_{it} &= \beta_{0it} + \beta_{1it} (\text{SESSION}) + r_{it} \\ \beta_{0it} &= \gamma_{00} + \gamma_{01} (\text{GORT COMPREHENSION CHANGE}) + u_{0i} \\ \beta_{1it} &= \gamma_{10} + \gamma_{11} (\text{GORT COMPREHENSION CHANGE}) \end{aligned}$$

Model 5, in Table 3, indicates that, though there was a decline in RPGs over time, the relationship between RPGs and time was not statistically significant (γ_{10} , Table 3). Likewise, the relationship between students' RPGs and their GORT Fluency scores was not statistically significant (γ_{01} , Table 3), indicating that the change in test scores over time was not related to students' RPGs. Finally, based on this model, the interaction between the number of sessions in which students participated and their Fluency score was not statistically significant (γ_{11} , Table 3).

The second model in Table 3 is similar to the first. Students' experienced a decline in RPGs over time, though the decline was not statistically significant (γ_{10} , Table 3). Students' RPGs and Comprehension scores were not significantly related (γ_{01} , Table 3), indicating that students' retention gains, or lack of gains, over time had no bearing on their post-test Comprehension performance. Finally, the interaction between time and changes in test scores was not statistically significant (γ_{11} , Table 3), indicating that students' RPG performance over time was not related to their overall performance on the GORT.

Further Analyses

Although an increase in growth is characteristic of an entire trajectory, that status at baseline and the instantaneous growth rate parameters depend on the choice of value for time (Raudenbush & Bryk, 2002). The linear models for GPGs and IPGs displayed trajectories of upward growth, indicating an increase in gains over time. These observations suggested the use of a quadratic growth model to determine whether the trend would continue (Raudenbush & Bryk, 2002), as well as the acceleration or deceleration of fluency growth (Hawkins et al., 2001). Finally, although students' experienced some declines with retention gains, the quadratic model was used to determine whether students would eventually experience an increase in gains over time.

The quadratic models are similar to the linear models with an added element. To illustrate, the quadratic model that includes GPG and changes in GORT Fluency is as follows:

$$\begin{aligned} \text{GPG}_{it} &= \beta_{0it} + \beta_{1it} (\text{SESSION}) + \beta_{2it} (\text{SESSION}^2) + r_{it} \\ \beta_{0it} &= \gamma_{00} + \gamma_{01} (\text{GORT FLUENCY CHANGE}) + u_{0i} \\ \beta_{1it} &= \gamma_{10} + \gamma_{11} (\text{GORT FLUENCY CHANGE}) \\ \beta_{2it} &= \gamma_{20} + \gamma_{21} (\text{GORT FLUENCY CHANGE}) \end{aligned}$$

In Level 1, SESSION^2 is added to represent the acceleration or deceleration of growth in students' WCPM scores over time. The added component at Level 1 adds an additional outcome variable at Level 2. Thus, β_0 is the status at baseline, β_1 is the instant growth rate for student at time t , and β_2 is the "curvature or acceleration in each growth trajectory" (Raudenbush & Bryk, 2002, p. 169).

Generalization of Passage Gains

As with the linear model, there was a significant relationship between GPGs and time, such that students' gains significantly increased over time (γ_{10} , Table 4). However, the quadratic analysis of time was not significant (γ_{20} , Table 4). In other words, quadratic analyses suggest that students will eventually reach a ceiling in the effects of the intervention. There was also a statistically significant relationship between students' GPGs and their GORT Fluency score (γ_{01} , Table 4), such that students with greater gains also experienced a greater pre- to post-test difference. The interactions between the number of sessions in which students participated and the changes in their GORT scores from pre- to post-test were non-significant for both the linear and the quadratic parameters (γ_{11} , γ_{21} , Table 4). These results suggest that participation in a greater number of sessions had little to no impact on students' GORT Fluency scores from pre- to post-test.

Table 4 also includes the results for GPGs related to GORT Comprehension scores. There was a statistically significant relationship between students' GPGs and time, such that their gains increased over time (γ_{10} , Table 4). However, these gains tended to level off within the quadratic parameter (γ_{20} , Table 4). GPGs and Comprehension scores were not significantly related. In other words, though students experienced an increase in fluency gains over time, their Comprehension scores did not significantly increase from pre- to post-test as a result (γ_{01} , Table 4). Finally, as with the GPG Fluency model, neither the linear nor the quadratic interactions were statistically significant (γ_{11} , γ_{21} , Table 4).

Immediate Passage Gains

Results from this quadratic model, as summarized in Table 5, indicate that students' gains increased over time, though this increase was not statistically significant (γ_{10}). There was a negative relationship between students' IPGs and gains in their Fluency subtest scores (γ_{01} , Table 5). In other words, students with a greater change in their Fluency score from pre- to post-test experienced a decrease in IPGs over time, but this was not significant. The quadratic interaction between time and students' Fluency scores was statistically significant (γ_{20} , Table 5). Students who had a greater pre- to post-test difference in score experienced fewer IPGs over time (γ_{21} , Table 5) when compared to students whose scores did not increase as much (see Figure 2). However, all students experienced IPGs over time.

The model containing students' IPGs as related to their Comprehension score is also summarized in Table 5. Results from this model indicate that, though not statistically significant, there was an increase in students' IPGs over time (γ_{10} , Table 5). As with the previously described model, there was a negative relationship between students' IPGs and Comprehension subtest scores, such that students who experienced greater pre- to post-test increase in scores, experienced a decrease in gains over time. However, these decreases were not statistically significant (γ_{01} , Table 5). The interaction between time and students' Comprehension scores was statistically significant (γ_{21} , Table 5). Students who had a greater pre- to post-test difference in score experienced a decline in IPGs over time, whereas students whose scores did not increase as much continued to experience gains over time (see Figure 2).

Retention of Passage Gains

Results from the linear parameter for students' gains and time indicated a loss in gains over time, though this loss was not statistically significant (γ_{10} , Table 6). The quadratic parameter, however, showed that with an increased number of sessions, students' gains eventually increased, though the increase was not statistically significant (γ_{20} , Table 6). Likewise, the relationship between RPGs and Fluency scores was not statistically significant (γ_{01} , Table 6). However, the interaction between quadratic time and students' GORT Fluency scores was statistically significant (γ_{21} , Table 6). Students with the greatest difference between their pre- to post-test Fluency score also experienced an increase in RPGs over time. Students whose post-test scores did not increase as much as their peers experienced a decrease in RPGs over time (see Figure 4).

Table 6 also shows the results for the model related to students' Comprehension scores. Although neither were statistically significant, results indicated that students' RPGs declined somewhat over time (γ_{10}) and then eventually increased (γ_{20}). Likewise, there was no statistically significant relationship between students' RPGs and GORT Comprehension scores (γ_{01} , Table 6). The interaction between time and students' Comprehension scores (γ_{21} , Table 6) indicated that, as students participated in more intervention sessions, those with greater RPGs also scored higher on the Comprehension subtest at post-test than those who did not experience similar gains over time (see Figure 5).

Summary of Results

Table 7 provides a summary of the levels of significance of the relationships between

each type of gain and growth over time as well as students' GORT Fluency and Comprehension scores. Results from students' GPGs were significant for both the linear and quadratic models, indicating that students' gains over time were significant and that these gains were related to the difference in their pre- to post-test GORT scores for both the Fluency and, within the quadratic parameter, Comprehension subtests.

Results from students' IPGs indicated that students experienced statistically significant gains over time, but that these gains were not significantly related to their performance on the GORT at post-test. Interestingly enough, results from students' RPGs were not statistically significant until the quadratic parameters for RPGs and GORT Fluency and Comprehension scores. Thus, it is possible that, over time, students' retention gains will have an impact on their outcome scores.

Chapter 5

Discussion

The purpose of this study was to explore the psychometric characteristics and predictive validity of the three forms of CBM-R progress monitoring that are built within the HELPS Program. Specifically, this study sought to answer the following questions related to each form of progress monitoring: (1) are there statistically significant changes over time in students' gains as measured by students' slope of gains, and (2) are the changes in gains experienced over time associated with the differences in students' pre- to post-test scores for measures of reading fluency and comprehension as assessed by the GORT? It was hypothesized that, as a result of participating in HELPS, students' slopes for Generalization of Passage Gains (GPGs), which closely mirror CBM-R data, will be positive over time. It was hypothesized that there would be a relationship between students' CBM-R (i.e., GPG) scores and the differences in students' pre- to post-test scores for the GORT Fluency and Comprehension subtests.

The research questions related to students' Immediate Passage Gains (IPGs) and Retention Passage Gains (RPGs) were more exploratory in nature. There is limited research to evaluate whether IPGs and RPGs are associated with overall increases in reading fluency over time. It is possible, though, that as students' level of fluency increases, they ultimately reach a maximum threshold for growth, and their gain scores begin to decrease over time. Thus, it was hypothesized that, for both IPGs and RPGs, students' scores would decrease over time, resulting in a negative slope. Likewise, IPG and RPG data have not been used to

predict outcome scores on measures of reading fluency and comprehension. This question, too, was exploratory in nature. However, if students' IPG and RPG slopes decrease over time, ultimately indicating gains in fluency, it was reasonable to predict that these negative slopes would be associated with changes on outcome measures such as the GORT.

Based on the results from students' GPGs, students experienced statistically significant gains over time. This score was also significantly related to GORT Fluency outcomes. There was not a significant relationship with GORT Comprehension outcomes. Although the latter is true, it is possible that GPGs could be used as a marker for predicting outcomes on standardized measures of reading achievement. Put simply, as a teacher implements HELPS with a student, he/she can observe the student's GPG scores over time, and if the student shows an increasing trend with GPG scores, the teacher can feel confident that the improved GPG scores would translate to improvements on standardized measures of reading fluency, such as the GORT.

It should be noted that, in this study, the number of sessions in which students participated did not significantly contribute to students' overall gains in outcome scores. This finding may seem contrary to the results from Begeny (2009). In that study, students who received HELPS three times per week significantly outperformed the control group on a measure of fluency and comprehension, whereas the group of students who received HELPS once or twice per week only outperformed the control group on the measure of fluency. The probable explanation for the finding in the present study is that students received approximately the same number of sessions (mean = 50.47; range = 44-58). In the Begeny

(2009) study, students in one group received twice the number of sessions than the students in the other group. Collectively, it seems there are advantages of receiving HELPS three times per week, but on occasion, if a session is missed, this should not have a meaningful impact on the overall effectiveness of the program with respect to improvements of fluency and comprehension. Ultimately, however, this study suggests that as students' reading abilities increase, their ability to read new passages also improves with time, and, in turn, their test scores tend to increase from pre- to post-test.

The Immediate Passage Gains (IPGs) variable also indicated that students experienced statistically significant gains in reading fluency over time. These gains were not related to students' test scores in the linear model. However, the quadratic model indicated that these gains were significantly related to test scores in two different ways. For students with less of an increase in their pre- to post-test scores (i.e., a change score one standard deviation below the centered mean), the number of sessions in which these students received seemed to have an impact, as these students continued to experience IPGs over time. Put another way, these students consistently experienced greater IPG scores over time when compared to students with a greater pre- to post-test increase in outcome scores (i.e., a change score one standard deviation above the centered mean). Students who experienced greater pre- to post-test increases on the GORT Fluency test continued to experience an increase in IPGs over time. However, this increase was lower compared to students' whose test scores did not increase as much (see Figure 2).

Interestingly, on the GORT Comprehension measure, the students' trends differed.

Students with a smaller increase in scores from pre- to post-test continued to experience an increase in IPGs as they did in the model with the Fluency measure. In contrast, students who experienced a greater increase in post-test scores also experienced a decline in IPGs over time. In other words, students' IPGs decreased over time as their test scores increased (see Figure 3).

According to the quadratic parameter, students whose Fluency and Comprehension scores increased over time also experienced an increase in RPGs (see Figure 4). Students who did not have a large increase in pre- to post-test scores experienced a decline in RPGs over time (see Figure 5). However, these results are of limited interpretive significance due to the fact that there was considerable within-student variability in RPGs over time and therefore the models for this variable were not a good fit. Consequently, students' RPGs, within the scope of this study, were not a reliable source of data to determine reading growth over time. RPGs may still be important as a means of ensuring that students retain practiced material, which is important for the overall structure of the HELPS instructional strategies. This source of data, however, may not be the measure of choice for reliably and validly assisting teachers with monitoring student progress over time.

Practical Implications

Much of the CBM-R research assesses student growth in reading fluency in linear terms (e.g., Ardoin & Christ, 2008; Francis et al., 2007; Graney & Shinn, 2005; Griffiths, VanDerHeyden, Skokut, & Lilles, 2009). In addition, research evaluating the use of CBM-R for the purpose of predicting reading outcomes (e.g., End of Grade test scores) also utilizes a

linear approach (Good et al., 2001; Hintze & Silberglitt, 2005; McGlinchey & Hixson, 2004; Stage & Jacobsen, 2001). For the purposes of this study, a quadratic component was added to determine whether it contributed to the modeling of student growth over time. Although results from several of the quadratic models indicated statistically significant relationships between different types of gain scores and growth over time and testing outcomes, the linear model tends to be a better fit when evaluating growth data such as those evaluated here.

This conclusion is similar to that found by Fuchs and colleagues (1993) when they evaluated CBM-R growth data with both a linear and quadratic model. Their findings indicated that only a small percentage of CBM-R scores had significant relationships with student outcomes within the quadratic model. In contrast, the linear relationship significantly contributed to the description of student increases over the course of the academic year. In other words, the linear model was a better fit for evaluating student growth using CBM-R.

The linear model for GPGs was the best fit overall for demonstrating statistically significant growth over time as well as being a significant predictor of outcome scores. Students' GPGs also mirror CBM-R. In other words, GPGs within the context of the HELPS Program are scores that result from the same type of process used in collecting CBM-R scores. A student reads an unfamiliar passage at designated points in time, the WCPM score is recorded, and the score can then be used to track student growth over time, as well as predict student outcomes as previously mentioned. Thus, it appears as though the data related to students' GPGs are the most informative for evaluating student growth and predicting outcomes when compared to students' IPGs and RPGs.

As part of the HELPS Program, GPG data are also more user-friendly relative to IPG and RPG data because, during every HELPS session, the teacher records a student's GPG scores in ways that make it easy to visually observe student progress over time. This is accomplished with the student's Graph (Appendix D) and Progress Tracking Form. With the Graph, GPG scores are circled so that teachers can easily observe whether circled data points are increasing over time. With the Progress Tracking Form, the column designated for Passage B, Reading #1 indicates the student's GPG scores. Thus, this makes it easy for teachers to visually scroll the WCPM scores in that column to look for increases over time. Although teachers also record information related to IPGs and RPGs on the Graph and Progress Tracking Form, calculations are needed to obtain the true RPG and IPG scores, making it slightly more complicated to observe student increases over time.

Furthermore, the use of GPG data as a basis for determining student progress also makes more intuitive sense. In other words, if students are experiencing overall gains in reading fluency, it stands to reason that they are better able to generalize those gains to unfamiliar text. As GPGs continue to increase, one can reasonably assume that the student continues to experience gains in reading fluency. To better determine whether a student continues to make gains, the teacher should also administer the appropriate seasonal ORF benchmark assessment. If the student's benchmark scores show an increasing trend toward the expected grade-level WCPM scores, , the teacher would have further evidence that the student is making meaningful progress in his/her reading development. In such a situation, the teacher would likely want the student to continue receiving the HELPS Program with the

idea that the student would continue to benefit from HELPS and continue on the upward trajectory toward grade-level expectations. If at the next seasonal benchmark assessment the teacher finds the student reading at grade-level expectations, the teacher could then make an informed decision about whether to continue HELPS with the student.

Within the context of this study, data derived from students' IPGs were used in exploratory analyses to determine whether they were useful in tracking student growth or predicting outcomes. Although results from the linear model indicated IPGs were not useful for predicting students' outcomes on the GORT, results did reveal that IPGs were statistically significant for showing growth over time. Even then, however, students' IPGs eventually plateaued. This plateau could be misinterpreted as the student no longer benefitting from the HELPS Program, but it seems more likely an indicator that the student is continuing to make overall reading improvements and is reaching their maximum threshold for ORF (Martens et al., 2007). In other words, as with nearly every skill, there is a maximum threshold for fluency. Thus, as GPG scores increase (i.e., the first reading of a new passage progressively increases), and a student continues to meet that maximum threshold for fluency, IPG scores are likely to plateau or even decrease. The scores for IPGs are obtained by subtracting the WCPM scores for the student's first reading of a passage from the student's last reading of that passage in that session. As students' proficiency with reading improves, the differential between these two scores will ultimately decrease.

There was no significant relationship between RPGs and outcome scores within this study. It should be noted that, due to the lack of fit with the method of analysis used, these

data are not as useful for monitoring student growth and potential outcomes on more broad, standardized measures of reading. These data do show that students' RPGs will likely fluctuate session to session, and these fluctuations should be anticipated.

Limitations

The results from this study indicate that GPG scores, obtained as a result of implementing the HELPS Program, successfully predict students' testing outcomes on a reliable and valid measure of reading fluency and comprehension (i.e., the GORT). Although this represents a promising use of HELPS as an assessment tool in addition to an instructional tool, this finding is somewhat limited because the GORT does not necessarily evaluate reading in all the ways desired by every school district. For example, many school districts have strong interest in students' state-mandated test scores, such as the end-of-year standardized achievement test (Wang, Beckett, & Brown, 2006). It is reasonable to assume that students' scores on state-mandated measures will improve if they also show improvement on other valid measures of reading ability, but future research should specifically examine to what degree students receiving the HELPS program demonstrate improvement on statewide, end-of-grade tests.

Multilevel modeling (MLM) was an appropriate method of analysis for this particular study in determining whether the various types of gains were related to gains in fluency over time and to the change in students' pre- to post-test scores. However, due to the specialized nature of these analyses, MLM is not a method that is accessible to the general practitioner. In other words, educational professionals are likely not trained in this area of statistics and

will have difficulty employing MLM in the data-based decision making process. What might be useful for practitioners is an equation in which educators can insert the appropriate data in order to determine whether students are continuing to make gains. Unfortunately, the current study cannot provide such information. Ideally, future research will incorporate a method of determining student growth using analyses that are easily applied by educational professionals.

Future Research Directions

Results from this study indicate that students' GPGs are useful for tracking student growth, as well as predicting student outcomes on standardized measures of reading fluency (i.e., the GORT). Results also indicate that students' IPGs model student growth, but it is unclear as to how IPG data can be practically useful for educators. Further research is needed to determine the clinical utility of IPG data. More specifically, further research might evaluate the way in which IPG data demonstrate growth, as well as what a plateau in growth may mean. Furthermore, future research might examine whether benchmark assessments, GPG data obtained from HELPS sessions, and IPG data from HELPS sessions can be combined to better monitor and predict students' reading development.

The GORT is a sound measure of reading fluency and comprehension, and it can provide specific information regarding students' level of ability in these two domains. State-mandated assessments are not a specific measure of reading fluency, though they may include subtests measuring comprehension. The latter form of assessment does, however, require the student to utilize his or her fluency and comprehension abilities in order to

perform well and may therefore be considered a strong measure of generalized reading ability. In other words, students are required to generalize their reading skills, not only for the purposes of completing what has become an increasingly important achievement test, but also for experiencing success academically. Thus, it would be useful to evaluate GPGs and IPGs with respect to their ability to predict student outcomes on broader measures of reading, such as state-mandated assessments.

It should be noted that state-mandated assessments typically do not begin until the end of Grade 3. The sample for this study solely included second graders. Thus, future research related to the utility of HELPS in predicting outcomes on state-mandated achievement tests will also need to take into account the timing of assessment as it relates to student performance in the HELPS Program. In other words, participation in HELPS may need to extend into third grade in order to determine whether HELPS can aid in predicting student outcomes on state-mandated tests.

In conclusion, the results from this study further the research on the HELPS Program, confirming again that students do experience gains as a result of participating in this program. In addition, the GPG and IPG assessment components of HELPS can potentially be used to track students' overall fluency gains, and GPGs can be used to predict outcomes on standardized measures of reading. Further research to determine the broader clinical utility of the HELPS assessment components will likely strengthen its overall effectiveness for students and usefulness for educational professionals.

References

- AIMSWeb, Inc. (2008). Reading Curriculum Based Measurement. Retrieved April 1, 2009, from <http://www.aimsweb.com/measures-2/reading-cbm/>
- Ardoin, S. P. (2006). The response in Response to intervention: Evaluating the utility of assessing maintenance of intervention effects. *Psychology in the Schools, 43*, 713-725.
- Ardoin, S. P., & Christ, T. J. (2008). Evaluating curriculum-based measurement slope estimates using data from triannual universal screenings. *School Psychology Review, 37*, 109-125.
- Ardoin, S. P., Witt, J. C., Connell, J. E., & Koenig, J. L. (2005). Application of a three-tiered response to intervention model for instructional planning, decision making, and the identification of children in need of services. *Journal of Psychoeducational Assessment, 23*, 362-380.
- Begeny, J. C. (2009). *Effects of the Helping Early Literacy with Practice Strategies (HELPS) reading fluency program when implemented at different frequencies*. Manuscript in preparation.
- Begeny, J. C., Groce, M. K., Krouse, H. E., & Mann, C. M. (2008a). *Teacher judgments of students' reading abilities across a continuum of rating methods and achievement measures*. Manuscript submitted for publication.

- Begeny, J. C., Laugle, K. M., Krouse, H. E., Lynn, A. E., Parker, M., & Stage, S. A. (2008b). *A control-group comparison of two reading fluency programs: The Helping Early Literacy with Practice Strategies (HELPS) program and the Great Leaps K-2 reading program*. Manuscript submitted for publication.
- Begeny, J. C., & Martens, B. K. (2006). Assisting low-performing readers with a group-based reading fluency intervention. *School Psychology Review, 35*, 91-107.
- Begeny, J. C., Mitchell, R. C., Whitehouse, M. H., & Harris, C. F. (2009). *Effects of the Helping Early Literacy with Practice Strategies (HELPS) reading fluency program when implemented by classroom teachers with low-performing second-grade students*. Manuscript in preparation.
- Begeny, J. C., & Silber, J. M. (2006). An examination of group-based treatment packages for increasing elementary-aged students' reading fluency. *Psychology in the Schools, 43*, 183-195.
- Big Ideas in Beginning Reading (2005). Analysis of Reading Assessment Instruments for K-3. Retrieved January 15, 2009 from <http://reading.uoregon.edu/assessment/analysis.php>
- Busch, T. W., & Reschly, A. L. (2007). Progress monitoring in reading: Using curriculum-based measurement in a response-to-intervention model. *Assessment for Effective Intervention, 32*, 223-230.
- Caplan, G. (1964). *Principles of Preventive Psychiatry*. New York: Basic Books.

- Chard, D. J., Vaughn, S., & Tyler, B. J. (2002). A synthesis of research on effective interventions for building reading fluency with elementary students with learning disabilities. *Journal of Learning Disabilities, 35*, 386-406.
- Christ, T. J., & Hintze, J. M. (2007). Psychometric considerations when evaluating response to intervention. In Jimerson, S. R., Burns, M. K., & VanDerHeyden, A. M. (Eds.), *The handbook of response to intervention: The science and practice of assessment and intervention*. New York: Springer.
- Christ, T. J., & Silberglitt, B. (2007). Estimates of the standard error of measurement for curriculum-based measures of oral reading fluency. *School Psychology Review, 36*, 130-146.
- Colón, E. P., & Kranzler, J. H. (2006). Effect of instructions on curriculum-based measurement of reading. *Journal of Psychoeducational Assessment, 24*, 318-328.
- Crawford, L., Tindal, G., & Steiber, S. (2001). Using oral reading rate to predict student performance on statewide achievement tests. *Educational Assessment, 7*, 303-323.
- Daly, E. J., Martens, B. K., Barnett, D., Witt, J. C., & Olson, S. C. (2007). Varying intervention delivery in response to intervention: Confronting and resolving challenges with measurement, instruction, and intensity. *School Psychology Review, 36*, 562-581.
- Daly, E. J., Martens, B. K., Hamler, K. R., Dool, E. J., & Eckert, T. L. (1999). A brief experimental analysis for identifying instructional components needed to improve oral reading fluency. *Journal of Applied Behavior Analysis, 32*, 83-94.

- Deno, S. (1985). Curriculum-based measurement: The emerging alternative. *Exceptional Children, 52*, 219-232.
- Deno, S. L., Marston, D., & Tindal, G. (1986). Direct and frequent curriculum-based measurement: An alternative for educational decision making. *Assessment of Exceptional Children, 2*, 5-27.
- Fletcher, J. M., Coulter, W. A., Reschly, D. J., & Vaughn, S. (2004). Alternative approaches to the definition and identification of learning disabilities: Some questions and answers. *Annals of Dyslexia, 54*, 304-331.
- Fletcher, J. M., Denton, C., & Francis, D. J. (2005). Validity of alternative approaches for the identification of learning disabilities: Operationalizing unexpected underachievement. *Journal of Learning Disabilities, 38*, 545-552.
- Florida Department of Education (2005). *Primary and secondary diagnostic instruments*. Tallahassee, FL: Author. Available:
<http://www.justreadflorida.com/educators/PrimSecDiagChart.asp>
- Francis, D. J., Santi, K. L., Barr, C., Fletcher, J. M., Varisco, A., & Foorman, B. R. (2008). Form effects on the estimation of students' oral reading fluency using DIBELS. *Journal of School Psychology, 46*, 315-342.
- Fuchs, D., & Fuchs, L. S. (2006). Introduction to response to intervention: What, why, and how valid is it? *Reading Research Quarterly, 41*, 93-99.
- Fuchs, L. S. (2003). Assessing intervention responsiveness: Conceptual and technical issues. *Learning Disabilities Research and Practice, 18*, 172-186.

- Fuchs, L. S., Deno, S. L., & Mirkin, P. K. (1984). The effects of frequent curriculum-based measurement and evaluation on pedagogy, student achievement, and student awareness of learning. *American Educational Research Journal, 21*, 449-460.
- Fuchs, L. S., & Fuchs, D. (1998). Treatment validity: A unifying concept for reconceptualizing the identification of learning disabilities. *Learning Disabilities Research and Practice, 13*, 204-219.
- Fuchs, L. S., Fuchs, D., & Compton, D. L. (2004). Monitoring early reading development in first grade: Word identification fluency versus nonsense word fluency. *Exceptional Children, 71*, 7-21.
- Fuchs, L. S., Fuchs, D., Hamlett, C. L., Walz, L., & Germann, G. (1993). Formative evaluation of academic progress: How much growth can we expect? *School Psychology Review, 22*, 27-28.
- 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*, 239-256.
- Glover, T. A., & DiPerna, J. C. (2007). Service delivery for response to intervention: Core components and directions for future research. *School Psychology Review, 36*, 526-540.
- Good, R. H., & Kaminski, R. A. (Eds.). (2002). *Dynamic Indicators of Basic Early Literacy Skills* (6th ed.). Eugene, OR: Institute for the Development of Educational Achievement. Available: <http://dibels.uoregon.edu/>

- Good, R. H., Simmons, D. C., & Kame'enui, E. J. (2001). The importance of 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*, 257-288.
- Graney, S. B., & Shinn, M. R. (2005). Effects of reading curriculum based measurement (R-CBM) teacher feedback on general education classrooms. *School Psychology Review, 34*, 184-201.
- Gresham, F. M. (2007). Evolution of the RTI concept: Empirical foundations and recent developments. In Jimerson, S. R., Burns, M. K., & VanDerHeyden, A. M. (Eds.), *The handbook of response to intervention: The science and practice of assessment and intervention*. New York: Springer.
- Griffiths, A., VanDerHeyden, A. M., Skokut, M., & Lilles, E. (2009). Progress monitoring in oral reading fluency within the context of RTI. *School Psychology Quarterly, 24*, 13-23.
- Hasbrouck, J. E., & Tindal, G. (1992). Curriculum-based oral reading fluency norms for students in grades 2 through 5. *Teaching Exceptional Children, 24*, 41-44.
- Hasbrouck, J. E., & Tindal, G. (2006). Oral reading fluency norms: A valuable assessment tool for reading teachers. *The Reading Teacher, 59*, 636-644.
- Hawkins, J. D., Guo, J., Hill, K. G., Battin-Pearson, S., & Abbott, R. D. (2001). Long-term effects of the Seattle social development intervention on school bonding trajectories. *Applied Developmental Science, 5*, 225-236.

- Hintze, J. M. ,& Christ, T. J. (2004). An examination of variability as a function of passage variance in CBM progress monitoring. *School Psychology Review, 33*, 204-217.
- Hintze, J. M., Ryan, A. L., & Stoner, G. (2003). Concurrent validity and diagnostic accuracy of the Dynamic Indicators of Basic Early Literacy Skills and the Comprehensive Test of Phonological Processing. *School Psychology Review, 32*, 541-556.
- Hintze, J. M., & Shapiro, E. S. (1997). Curriculum-based measurement and literature-based reading: Is curriculum-based measurement meeting the needs of changing reading curricula? *Journal of School Psychology, 35*, 351-375.
- Individuals With Disabilities Education Improvement Act (2004). Public Law 108-446 (20 U.S.C. 1400 *et seq.*).
- Jacob, S., & Hartshorne, T. S. (2007). *Ethics and law for school psychologists* (5th ed.). Hoboken: John Wiley and Sons.
- Jenkins, J. R., Hudson, R. F., & Johnson, E. S. (2007). Screening for at-risk readers in a response to intervention framework. *School Psychology Review, 36*, 582-600.
- Jenkins, J. R., Zumeta, R., Dupree, O., & Johnson, K. (2005). Measuring gains in reading ability with passage reading fluency. *Learning Disabilities Research and Practice, 20*, 245-253.

- Jimerson, S. R., Burns, M. K., & VanDerHeyden, A. M. (2007). Response to intervention at school: The science and practice of assessment and intervention. In Jimerson, S. R., Burns, M. K., & VanDerHeyden, A. M. (Eds.), *The handbook of response to intervention: The science and practice of assessment and intervention*. New York: Springer.
- Kuhn, M. R., & Stahl, S. A. (2003). Fluency: A review of developmental and remedial practices. *Journal of Educational Psychology, 95*, 3-21.
- LaBerge, D., & Samuels, S. J. (1974). Toward a theory of automatic information processing in reading. *Cognitive Psychology, 6*, 293-323.
- Lee, J., Grigg, W. S., & Donahue, P. L. (2007). *The nation's report card: Reading 2007*. U. S. Department of Education, National Center for Education Statistics. Washington, DC: Government Printing Office.
- Lyon, G. R. (1995). Research initiatives in learning disabilities: Contributions from scientists supported by the National Institute of Child Health and Human Development. *Journal of Child Neurology, 10*, S120-S126.
- Marston, D. B. (1989). A curriculum-based measurement approach to assessing academic performance: What it is and why do it. In Shinn, M. R. (Ed.), *Curriculum-Based Measurement: Assessing Special Children* (pp. 18-78). New York: Guilford.

- Martens, B. K., Eckert, T. L., Begeny, J. C., Lewandowski, L. J., DiGennaro, F. D., Montarello, S. A. et al. (2007). Effects of a fluency-building program on the reading performance of low-achieving second and third grade students. *Journal of Behavioral Education, 16*, 39-54.
- McGlinchey, M. T., & Hixson, M. D. (2004). Using curriculum-based measurement to predict performance on state assessment in reading. *School Psychology Review, 33*, 193-203.
- Merrell, K. W., Ervin, R. A., & Gimpel, G. A. (2006). *School psychology for the 21st century: Foundations and practices*. New York: Guilford.
- National Institute of Child Health and Human Development (2000). *Report of the National Reading Panel: Reaching children to read: An evidence based assessment of the scientific research literature on reading and its implications for reading instruction: Reports of the subgroups* (NIH Publication No. 00-4754). Washington, DC: U. S. Government Printing Office.
- Neupert, S. D., Miller, L. M., & Lachman, M. E. (2006). Physiological reactivity to cognitive stressors: Variations by age and socioeconomic status. *International Journal of Aging and Human Development, 62*, 221-235.
- No Child Left Behind Act (2001). U. S. Congress. Public Law 107-110. January 8, 2002.
- O'Connor, R. E., White, A., & Swanson, H. L. (2007). Repeated reading versus continuous reading: Influences on reading fluency and comprehension. *Exceptional Children, 74*, 31-46.

- Pearson, P. D., Johnson, D. D., Clymer, T., Indrisano, R., Venezky, R. L., Baumann, J. F., Hiebert, E., & Toth, M. (1989). *Silver, Burdett, and Ginn*. Needham, MA: Silver, Burdett, and Ginn, Inc.
- Poncy, B. C., Skinner, C. H., & Axtell, P. K. (2005). An investigation of the reliability and standard error of measurement of words read correctly per minute using curriculum-based measurement. *Journal of Psychoeducational Assessment*, Response to Intervention, 23, 326-338.
- Powell-Smith, K. A., & Bradley-Klug, K. L. (2001). Another look at the 'C' in CBM: Does it really matter if curriculum-based measurement reading probes are curriculum-based? *Psychology in the Schools*, 38, 299-312.
- Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical linear models: Applications and data analysis methods* (2nd ed.). Thousand Oaks, CA: Sage.
- Riley-Heller, N., Kelly-Vance, L., & Shriver, M. (2005). Curriculum-based measurement: Generic versus curriculum-dependent probes. *Journal of Applied School Psychology*, 21, 141-162.
- Shinn, M. R. (1989). Identifying and defining academic problems: CBM screening and eligibility procedures. In Shinn, M. R. (Ed.), *Curriculum-Based Measurement: Assessing Special Children* (pp. 18-78). New York: Guilford.

- Shinn, M. R. (2002). Best practices in using curriculum-based measurement in a problem solving model. In A. Thomas & J. Grimes (Eds.), *Best practices in school psychology IV* (Vol. 1, pp. 671-697), Bethesda, MD: NASP Publications.
- Shinn, M. R. (2007). Identifying students at risk, monitoring performance, and determining eligibility within response to intervention: Research on educational need and benefit from academic intervention. *School Psychology Review, 36*, 601-617.
- Silberglitt, B., & Hintze, J. (2005). Formative assessment using CBM-R cut scores to track progress toward success on state-mandated achievement tests: A comparison of methods. *Journal of Psychoeducational Assessment, 23*, 304-325.
- Silberglitt, B., & Hintze, J. M. (2007). How much growth can we expect? A conditional analysis of R-CBM growth rates by level of importance. *Exceptional Children, 74*(1), 71-84.
- Snow, C. E., Burns, S. M., & Griffin, P. (Eds.). (1998). *Preventing reading difficulties in young children*. Washington, DC: National Academy Press.
- Spache, G. (1953). A new readability formula for primary-grade reading materials. *The Elementary School Journal, 53*, 410-413.
- Speece, D. L., Case, L. P., & Molloy, D. E. (2003). Responsiveness to general education instruction as the first gate to learning disabilities identification. *Learning Disabilities Research and Practice, 18*, 147-156.
- Stage, S. A. (2001). Program evaluation using hierarchical linear modeling with curriculum-based measurement reading probes. *School Psychology Quarterly, 16*, 91-112.

- Stage, S. A., & Jacobsen, D. M. (2001). Predicting student success on a state-mandated performance-based assessment using oral reading fluency. *School Psychology Review, 30*, 407-419.
- Therrien, W. J. (2004). Fluency and comprehension gains as a result of repeated reading: A meta-analysis. *Remedial and Special Education, 25*, 252-261.
- U.S. Department of Education (2002). *Guidance for the reading first program*. Washington, DC: Author. Available: <http://www.ed.gov/programs/readingfirst/guidance.pdf>
- Wang, L., Beckett, G. H., & Brown, L. (2006). Controversies of standardized assessment in school accountability reform: A critical synthesis of multidisciplinary research evidence. *Applied Measurement in Education, 19*, 305-328.
- Weiderholt, J. L., & Bryant, B. R. (2001). *Gray Oral Reading Test*, (4th ed.). Austin, TX: Pro-Ed.

APPENDICES

Appendix A

HELPS Program: Implementation Protocol

- Establish rapport and describe/review the goals of program (*0.25 minutes*)
 - Goals of program are to: remember what happens in the in the story; remember the difficult words that you will practice; and remember to read with speed, accuracy, and good expression (like your teacher when s/he reads in class).
- Student reads indicated story after teacher provides directions (*1 to 1.5 minutes*)
 - NOTES: for ALL student readings of the story, the “one-minute bracket” should be indicated. Students that meet the WCPM criterion can be stopped at 1 min, and those who do not should read 1.5 minutes to encourage additional practice.*
- Teacher asks student to say what he/she remembers about the story, asking the student to recall what happened in appropriate sequential order (*0.5 to 0.75 minutes*)—see protocol for directions

IF STUDENT READS 100 WCPM OR MORE (SEE BACK PAGE IF STUDENT DOES NOT MEET GOAL)

- If student (a) reads 100 WCPM or more, (b) reads with 3 or fewer errors, **and** (c) can recall parts of the story for at least 30 seconds (with reasonable sequential order), teacher provides praise for meeting the goal and the teacher immediately charts the student’s performance (*0.25 minutes*)
 - Teacher tells student he/she will earn at least one star at the end of the session for meeting the reading goal of 100 WCPM
 - Teacher obtains the next story in the sequence
- Student reads next story in the curriculum sequence after teacher provides directions (*1.5 to 2 minutes*)
- Teacher implements phrase-drill error correction on all incorrectly read words (*0.25 to 0.75 minutes*)
 - Teacher follows protocol for Phrase-drill
- Student reads the same story a second time for no more than one minute (*1 minute*)
- Teacher models fluent oral reading of story while student follows along with his/her finger (*1 to 1.5 minutes*)
 - Teacher follows protocol for modeling fluent reading
- Student reads the same story a third time for no more than one minute (*1 minute*)

- Teacher provides ENTHUSIASTIC praise and feedback on student's reading and praises effort as applicable (*0.5 to 1 minute*)
 - Teacher charts the number of WCPM and WIPM for the student's first and third reading of the NEW story, and praises student for reading improvements if applicable
 - Teacher gives one star if student met goal on first story
 - Teacher gives two stars if student meets a first reading goal and reads 100 WCPM or more on a reading of the new story (*note: 100 WCPM or more on the new story does not mean the student graduates to a new story*)
- Teacher records all information accurately on student's progress tracking form and indicates on the tracking form which story probes will be read in the next session

Total time of implementation if student does meet goal: 7.25 to 10 minutes

IF STUDENT DOES NOT READ 100 WCPM OR MORE

- Teacher models fluent oral reading of story while student follows along with his/her finger (*1 to 1.5 minutes*)
 - Teacher follows protocol for modeling fluent reading
- Student reads Story a second time for no more than one minute (*1 minute*)
- Teacher implements phrase-drill error correction on all incorrectly read words (*0.25 to .75 minutes*)
 - Teacher follows protocol for Phrase-drill
- Student reads Story a third time for no more than one minute (*1 minute*)
- Teacher implements phrase-drill error correction on all incorrectly read words (*0.25 to .75 minutes*)
 - Teacher follows protocol for Phrase-drill
- Teacher provides ENTHUSIASTIC praise and feedback on student's reading and praises effort as applicable (*0.5 to 1 minute*)
 - Teacher charts the number of WCPM and WIPM for the student's FIRST and THIRD reading of the story, and praises student for reading improvements if applicable
 - Teacher gives one star if student reads at least 100 WCPM during one of the readings (*note: 100 WCPM or more on the 2nd or 3rd reading does not mean the student graduates to a new story*). The teacher may also give one star if the student does not meet the above criterion but clearly puts forth effort during the session.
- Teacher records all information accurately on student's progress tracking forms and indicates on the tracking form which story probes will be read in the next session

Total time of implementation if student does not meet goal: 5.75 to 8.5 minutes

Appendix B

Tips and Reminders for Implementing the HELPS Program

- *Preparation*
 - *Make sure you have all necessary materials ready prior to starting the session with your student.*
 - *Stop watch, examiner copy, student copy, dry erase marker, graphing sheet, progress tracking form, star chart, bonus bag, and prize box.*
- *Praise*
 - *With enthusiasm, praise specific reading behaviors (e.g., nice job reading accurately and with good expression; I like how you corrected words you missed previously), and always praise the student for something related to her reading performance at the end of the session.*
 - *Frequently use the student's name when using praise.*
 - *Using praise and creating a positive experience during each HELPS session is a primary component of this Program.*
- *Scoring Passages*
 - *Indicate the student's time of completion IF the student reads the entire story in less than one minute. Do not record WCPM if the student finishes reading a story in less than one minute; simply record the time it took for the student to finish the passage.*
 - *If the student reads beyond the line denoting where the tutor stops in the modeling phase, allow the student to keep reading until the minute is up.*
 - *When using the same scoring area on a passage, remember to put the appropriate number (i.e., 1, 2, or 3) next to the bracket representing where the student stops after his first, second, and 3rd reading (as applicable).*
 - *Be sure to follow rules for how to score errors depending on whether it's the first, second, or third reading. Mark the errors differently during each reading (first-slash, second-underline, third-circle)*
- *Determining if the Student Meets His/Her Goal*
 - *In order to meet the reading goal (and remember, there's only one "goal" that can be met per day), the student must 1) pass the comprehension/retell, 2) read at least 100 WCPM, and 3) make 3 or fewer errors. (Note, reading more than 100 WCPM after the initial reading of a session is not considered the reading "goal." However, doing this does earn the student a star.*
 - *Remember to do the comprehension component immediately after the student reads the 1st story of the session. This is probably the most frequently missed step.*
 - *During comprehension, use follow-up questions if appropriate. In other words, it is okay to have some discourse about the story, but keep all of this within the 30-45 second timing. Ideally, the student will be able to recall 30-45 seconds of information, and if so, no prompting or discourse is necessary. You should also provide feedback regarding the student's ability to recall the story in correct sequential order.*
 - *Also during the comprehension procedure, do not allow the student to see or review the story. Instead, turn the student reading booklet over or otherwise move it out of sight.*
 - *If a student does not go to a new story because of the comprehension criterion, please make a specific note of this on the tracking form.*
- *Graphing Performance*
 - *Circle the data point and session number when a new story/probe is started.*
 - *Do not connect lines between different stories.*

- Connect lines for the student during the session (when applicable). This usually helps students visualize the growth they have made.
- You will chart 2-3 readings each day (3 if the goal is met, 2 if the goal is not met)
 - Always chart the first and last reading of *any* story read during a given session.
- *Using the Reward System*
 - When talking about the program and the “benefits” of it to your students, do not focus on earning stars and prizes. Rather, focus on students becoming better readers, and use these incentives as tangible ways to praise students’ reading effort and improvement.
 - Students may record the stars on their charts if they like to do so, they are fast, and if time permits.
 - When the student lands on OR passes a shaded square on the chart, she is allowed to pick a ticket from the bonus bag. Based on what the ticket says, add those stars (or give a reward) immediately. If a student happens to select a “bonus PRIZE” ticket AND gets to the end of a row on the same day, the student should be allowed to choose two prizes.
 - When giving students stars on their charts, remember to ACCURATELY tell the student why he/she earned EACH star. As noted above, we want to use specific praise and feedback.
- *Phrase-Drill (PD) Procedural Reminders*
 - During the phrase-drill phase, you must “drill” on the missed words. However, if the student misses 2 or fewer, but reads some section(s) of the story non-fluently, do phrase drill on that section and simply tell the student: “You read this part correctly, but let’s practice reading it with better speed and expression.”
 - Use “logical” phrases during PD.
 - Also during phrase-drill, (a) tell the student you want her to “READ” the difficult words (in contrast to “SAYING” or “REPEATING” the words that were missed; and (b) remember to point (or have the student point) to each word being read as the student reads the phrase.
- *Listening Passage Preview Procedural Reminders*
 - When reading the story aloud to the student: (a) read at a pace that is just a little faster than the student’s reading ability, (b) remember to read with good expression, and (c) read at a volume that the student can clearly hear you.
- *Additional Procedural Reminders*
 - After 3 or 4 sessions with a student, they will probably learn the procedures quite well. When you are *sure* the student you are working with knows the procedures, you do not need to read the directions from the scripted protocol verbatim. Rather, just give brief directions, and perhaps just paraphrase the script as needed. For example, when telling the student to read the passage, you could say, “Now you are going to read the story, just like you’ve done before.” Or when telling the student to follow along as you read, you could say, “Now I’m going to read the story to you. As always, read silently and follow along with your finger.”
 - On occasions when a student starts reading so fast that he uses no expression or makes frequent errors, it is okay to restart the timing (*only if caught **within 1-5 seconds***) and then say to the student something like, “Slow down; you’ll see that by going slower you’ll actually read more words correctly and faster.”

➤ *Tracking Sheet Reminders*

- After finishing up a session with your student, you **MUST** complete the progress tracking sheet. Make sure to fill out all required information in the chart **BEFORE** erasing the data from your examiner copy.
- Each session you will record 2 or 3 WCPM and WIPM scores on the tracking sheet. If a student meets her goal on Passage A, you should record WCPM/WIPM in the first Passage A column and then record the student's first and third reading of Passage B in the appropriate Passage B columns (thus, you make 3 recordings if the student meets his goal). If the student does not meet her goal, you would only record WCPM/WIPM twice by recording scores in the two columns represented by Passage A, and nothing would be recorded in the Passage B columns.
- On the tracking form, DO NOT record absences as a session; the session numbers on the tracking form should only reflect the number of times you actually work with the student.

Scripted Protocol—HELPS

Describing the reading goals:

<Student Name>, you're going to be doing some reading with me today. As you read, I want you to try to remember what happens in the story and try to remember the difficult words that we practice. Also, I want you to read as quickly as you can without making errors, and try to read with good expression (like your teacher reads during story time).

Directions to administer before student reads story:

1. Place the examiner copy of the reading passage in front of you but shielded so the student cannot see what you record.
2. Present the student copy of the reading passage to the student.
3. Say to the student, "Here is a story that I would like you to read. When I say 'Begin', start reading aloud at the top of the page and read across the page. Try to read each word. If you come to a word you don't know, I'll tell it to you. Do you have any questions? Be sure to do your BEST reading."
4. Say, "Begin!" and start the stopwatch when the student says the first word.
5. If the student hesitates on a word for more than 3 seconds, say the word.
6. At the end of one minute, place a closed bracket after the last word.
7. *If the student reads so fast that no expression is given, remind the student that when he/she reads the next story, you want him/her to read at a comfortable rate (i.e., with good expression, like when talking).*
8. Remove both copies of the reading passage.

Directions for administering retell/comprehension:

1. Say to the student, "Now I want you to tell me everything you remember about the story you just read. Try to tell me what happened in the correct order."
2. Start your stopwatch and stop the retell activity in 30-45 seconds. Use prompts or follow-up questions as appropriate.
3. If student clearly struggles to remember parts of the story during his/her retell, note this on the student's tracking sheet and use this information when determining whether the student met his/her reading goal.

Directions for administering Phrase-drill:

1. Say to the student, "Now we are going to practice some of the words you missed."
2. Point to the first error word, say the word, and then say, "Read this after I do, <read the 2-5 word phrase containing the error word>. Again, Again." In essence, allow the student to read the phrase three times. Make sure the student points to the words being read; students will sometimes just "memorize" the phrase and repeat it. (We want students to read rather than recite).
3. Repeat the above procedure for all *unique* error words in the passage.
4. Praise the student for every two to three sets of phrase drills.

Directions for teacher to model fluent reading:

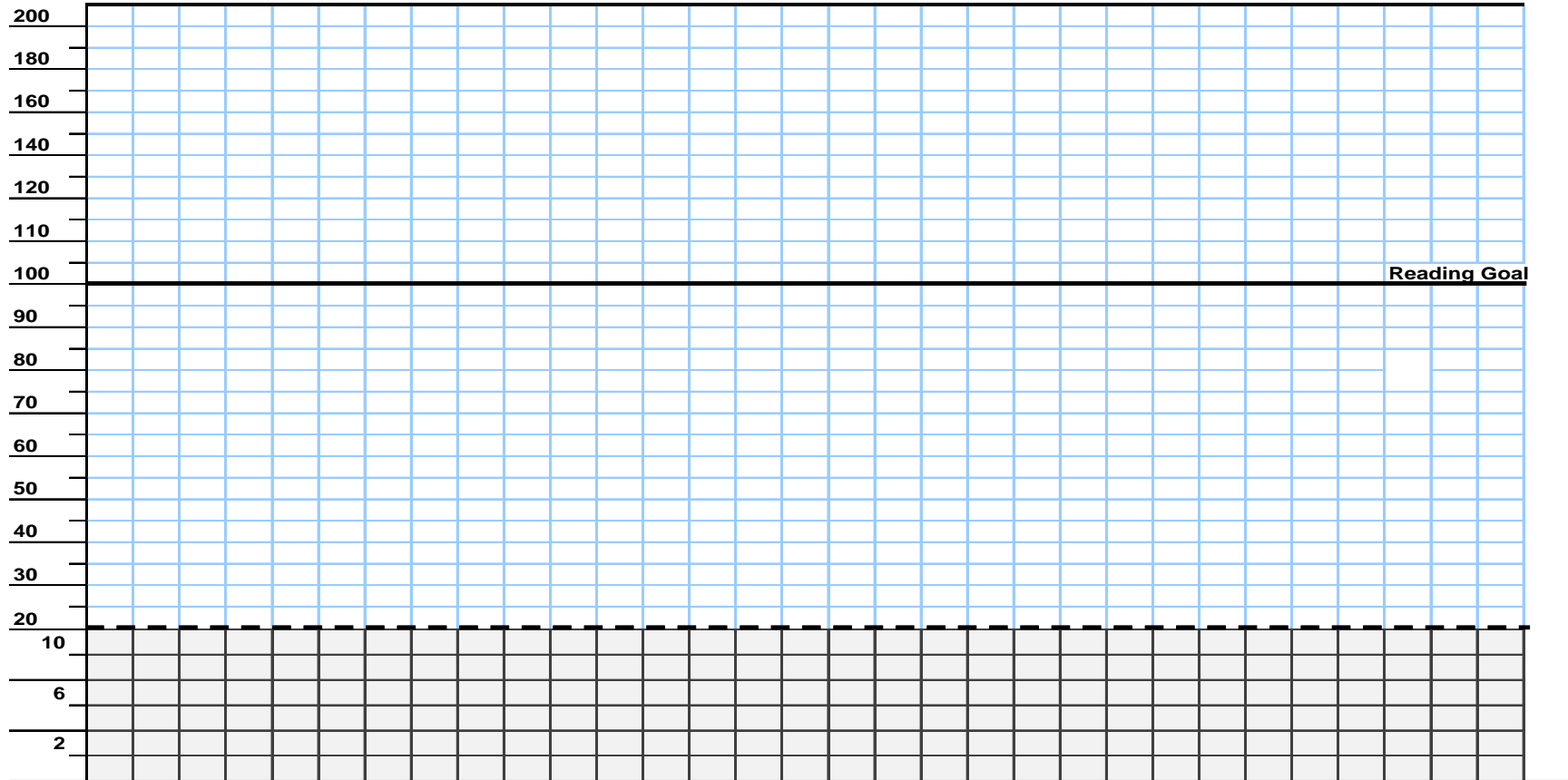
1. Say to the student, "Now I am going to read today's story to you. Please follow along with your finger, reading the words to yourself as I say them."
2. Read the passage at a comfortable reading rate and with good expression until the stopping point indicated on the examiner copy of the passage. Make sure the student is following along with his/her finger and prompt the student to do this (as necessary).

Appendix D

NAME: _____

Standard HELPS Graph for Grade 2

*Notes: Circle session # when you begin a new story. Indicate date with a line connecting to the respective session number.
(Mark WCPM with dot; Mark WIPM with X) Only connect dots & Xs for readings of the same story.*



HELPS Program Prize Chart

(15 Stars = 1 prize from the special prize box)

Student Name: _____

			■						■				
		■				■							
				■								■	
							■						
		■							■				
		■				■							
				■								■	
							■						
			■						■				
		■				■							

Appendix F

Description of the HELPS Session

At the beginning of the intervention session, the tutor establishes rapport with the student and outlines the objectives for the session per the scripted protocol. The tutor then obtains the first story (passage A) for the day and, according to the protocol, instructs the student to read the story aloud. The reading is timed for 60 seconds and the tutor marks errors per the CBM protocol. At the end of the minute, the tutor removes the story from the child's view and asks him or her to recall as much of the story as possible in sequential order. The student should be able to sufficiently recall details from the story for 20-45 seconds, depending on the length of the reading. The tutor then determines whether the student met his or her goal for the day's first reading. Goal criteria are reading 100 words or more correctly, with 3 or fewer errors, and an adequate retelling of the story.

If the student meets goal on passage A, the tutor immediately graphs the number of words read correctly, as well as the number of errors. The tutor informs the student that they will receive a star on their star chart and also provides enthusiastic praise of the student's performance on the first story. The tutor then obtains the next story in the sequence (passage B) and the student reads it for 1 minute. Following the first reading of passage B, the tutor administers the phrase drill, following the scripted protocol. The phrase drill focuses on any words the student reads incorrectly or phrases the student does not read fluently. The student then reads passage B a second time for no more than 1 minute. Subsequently, the tutor instructs the student to follow along with his or her finger as the tutor reads the story out loud

to the student, at a slightly faster pace. After the model reading, the student reads passage B for the third and final time for that session. The tutor then provides praise, and graphs the student's reading of the first and third readings of passage B, while the student watches. The tutor then draws (or allows the student to draw) two stars on the student's star chart: one for meeting goal on passage A and one for their hard work on passage B. The tutor fills out the tracking form and the session is complete.

If the student does not meet his or her goal, the treatment sequence is different. The tutor determines that the student does not meet goal, so the student reads passage A two more times during the session instead of moving on to a new story. The tutor performs the model reading of passage A and then instructs the child to read the story a second time. Following that reading, the tutor administers the phrase drill, the story is read a third time and a second phrase drill is completed. At the end of that phrase drill, the tutor charts the student's words correct and incorrect for the first and third readings of passage A. The student receives one star on the star chart for their effort during the session. The tutor fills out the tracking form and the session is complete.

The student's next session with the tutor begins with the last story read in the previous session. The intervention process then repeats itself as described. Each experimental session lasts approximately 8 to 12 minutes.

During all passage readings, CBM-R procedures are used in marking errors and computing the number of words read correctly per minute (WCPM) and the number of words read incorrectly per minute (WIPM). At the end of the session, the student's performance is

recorded on a tracking sheet, and the WCPM and WIPM are plotted on a graph while the student watched. Data pertaining to the comprehension piece are collected insofar as the student met the retell criteria as part of meeting his or her goal for passage A.

Table 1

Unstandardized Coefficients (and Standard Errors) for Linear Multi-level Models of Generalization of Passage Gains

Fixed Effects	Model 1	Model 2
GPG, β_0		
Intercept, γ_{00}	94.28*** (1.83)	94.34*** (2.83)
Fluency Change Score, γ_{01}	5.14***(.72)	
Comprehension Change Score, γ_{01}		1.42 (.93)
Slope of Session Number, β_1		
Intercept, γ_{10}	.31***(.03)	.32***(.03)
Session*Fluency Score, γ_{11}	.02 (.01)	
Session*Comprehension Score, γ_{11}		.01 (.01)
Random Effects		
GPG (τ_{00})	77.74** (22.70)	213.08** (59.64)
Within person fluctuation (σ^2)	179.62*** (7.73)	179.99*** (7.75)
Note: ** $p < .001$, *** $p < .0001$		

Table 2

Unstandardized Coefficients (and Standard Errors) for Linear Multi-level Models of Immediate Passage Gains

Fixed Effects	Model 3	Model 4
IPG, β_0		
Intercept, γ_{00}	27.20*** (1.45)	27.22***(1.43)
Fluency Change Score, γ_{01}	.71 (.57)	
Comprehension Change Score, γ_{01}		.35 (.47)
Slope of Session Number, β_1		
Intercept, γ_{10}	.11*** (.03)	.11***(.03)
Session*Fluency Score, γ_{11}	.01 (.01)	
Session*Comprehension Score, γ_{11}		-.01 (.01)
Random Effects		
IPG (τ_{00})	42.20** (12.75)	40.89**(12.40)
Within person fluctuation (σ^2)	232.32** *(8.80)	232.44*** (8.81)
Note: ** $p < .001$, *** $p < .0001$		

Table 3

Unstandardized Coefficients (and Standard Errors) for Linear Multi-level Models of Retention Passage Gains

Fixed Effects	Model 5	Model 6
RPG, β_0		
Intercept, γ_{00}	-1.81 (.94)	-1.84 (.93)
Fluency Change Score, γ_{01}	-.34 (.37)	
Comprehension Change Score, γ_{01}		.17 (.31)
Slope of Session Number, β_1		
Intercept, γ_{10}	-.04 (.03)	-.04 (.03)
Session*Fluency Score, γ_{11}	-.01 (.01)	
Session*Comprehension Score, γ_{11}		.01 (.01)
Random Effects		
RPG (τ_{00})	7.45*(3.18)	7.24*(3.13)
Within person fluctuation (σ^2)	204.27*** (7.75)	204.09*** (7.74)
Note: * $p < .05$, ** $p < .001$, *** $p < .0001$		

Table 4

Unstandardized Coefficients (and Standard Errors) for Quadratic Multi-level Models of Generalization of Passage Gains

Fixed Effects	Model 7	Model 8
GPG, β_0		
Intercept, γ_{00}	94.68*** (2.02)	94.38*** (2.95)
Fluency Change Score, γ_{01}	4.69*** (.80)	
Comprehension Change Score, γ_{01}		1.93 (.97)
Slope of Session Number, β_1		
Intercept, γ_{10}	.36*(.11)	.33*(.11)
Session*Fluency Score, γ_{11}	-.03 (.04)	
Session*Comprehension Score, γ_{11}		.07*(.04)
Slope of Session Number ² , β_2		
Intercept, γ_{20}	.001 (.002)	.0002 (.002)
Session ² *Fluency Score, γ_{21}	-.001 (.001)	
Session ² *Comprehension Score, γ_{21}		.001 (.001)
Random Effects		
GPG (τ_{00})	77.85*** (22.73)	213.15*** (59.66)
Within person fluctuation (σ^2)	179.68*** (7.74)	179.81*** (7.75)

Note: * $p < .05$, ** $p < .001$, *** $p < .0001$

Table 5

Unstandardized Coefficients (and Standard Errors) for Quadratic Multi-level Models of Immediate Passage Gains

Fixed Effects	Model 9	Model 10
IPG, β_0		
Intercept, γ_{00}	26.49*** (1.67)	26.42***(1.66)
Fluency Change Score, γ_{01}	-0.31 (.65)	
Comprehension Change Score, γ_{01}		-0.43 (.56)
Slope of Session Number, β_1		
Intercept, γ_{10}	.02 (.11)	.01 (.11)
Session*Fluency Score, γ_{11}	-.11*(.04)	
Session*Comprehension Score, γ_{11}		-.10*(.04)
Slope of Session Number ² , β_2		
Intercept, γ_{20}	-.002 (.002)	-.002 (.002)
Session ² *Fluency Score, γ_{21}	-.003*(.001)	
Session ² *Comprehension Score, γ_{21}		-.001*(.001)
Random Effects		
IPG (τ_{00})	41.63*** (12.59)	40.63*** (12.33)
Within person fluctuation (σ^2)	230.88*** (8.75)	231.50*** (8.78)
Note: * $p < .05$, ** $p < .001$, *** $p < .0001$		

Table 6

Unstandardized Coefficients (and Standard Errors) for Quadratic Multi-level Models of Retention Passage Gains

Fixed Effects	Model 11	Model 12
RPG, β_0		
Intercept, γ_{00}	-1.77 (1.30)	-1.68 (1.30)
Fluency Change Score, γ_{01}	.93 (.53)	
Comprehension Change Score, γ_{01}		.90 (.47)
Slope of Session Number, β_1		
Intercept, γ_{10}	-.02 (.11)	-.02 (.11)
Session*Fluency Score, γ_{11}	.13*(.04)	
Session*Comprehension Score, γ_{11}		.09*(.04)
Slope of Session Number ² , β_2		
Intercept, γ_{20}	.0004 (.001)	.0005 (.002)
Session ² *Fluency Score, γ_{21}	.003**(.001)	
Session ² *Comprehension Score, γ_{21}		.001*(.001)
Random Effects		
RPG (τ_{00})	7.31**(3.13)	7.28*(3.13)
Within person fluctuation (σ^2)	202.93*** (7.70)	203.69*** (7.73)
Note: * $p < .05$, ** $p < .001$, *** $p < .0001$		

Table 7

Summary of the Significance of the Relationship Between Types of Gains and Gains Over Time and Outcome Scores

Type of Gain	Significance of Gains Over Time	Relationship to Fluency Outcomes	Relationship to Comprehension Outcomes
GPG (linear)	< .001	< .001	.14
GPG (quadratic)	.66	< .001	.05
IPG (linear)	< .001	.22	.47
IPG (quadratic)	.33	.64	.45
RPG (linear)	.12	.37	.58
RPG (quadratic)	.84	.09	.06

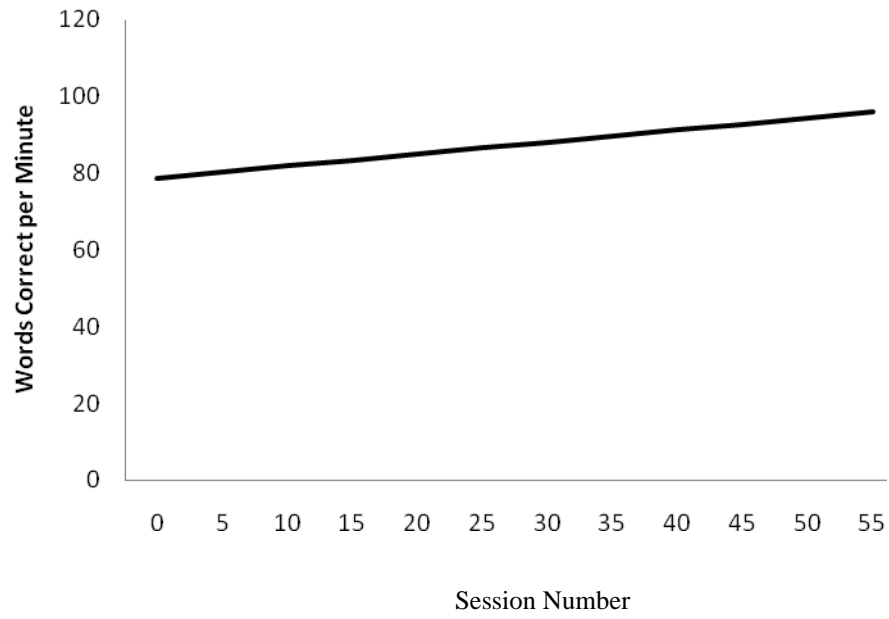


Figure 1 Growth for Generalization of Passage Gains

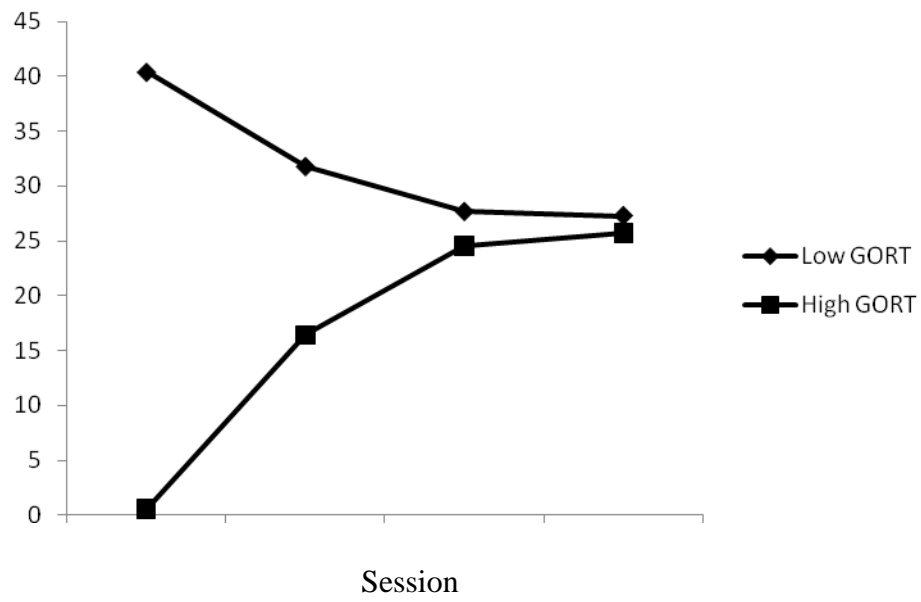


Figure 2 Quadratic Interactions for Immediate Passage Gains—Fluency

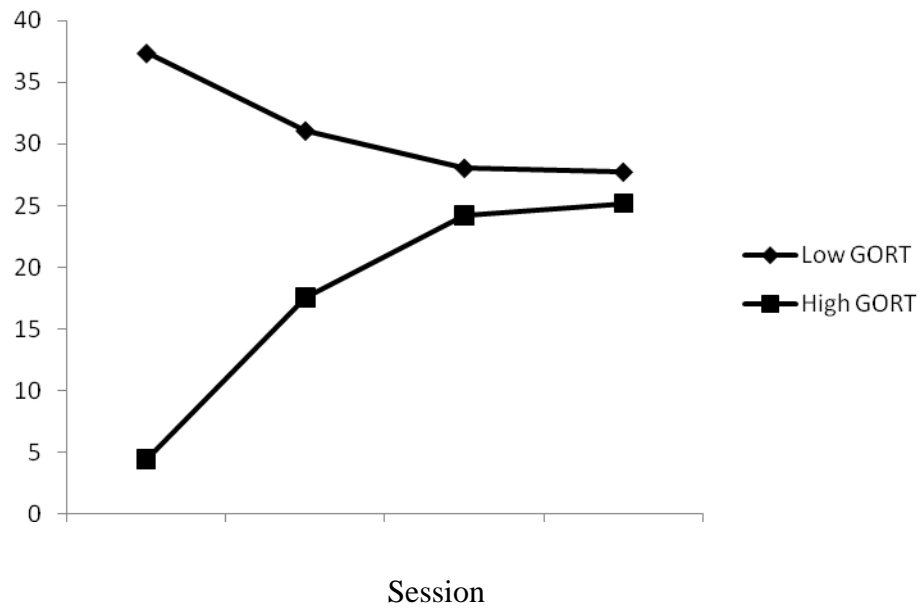


Figure 3 Quadratic Interactions for Immediate Passage Gains –Comprehension

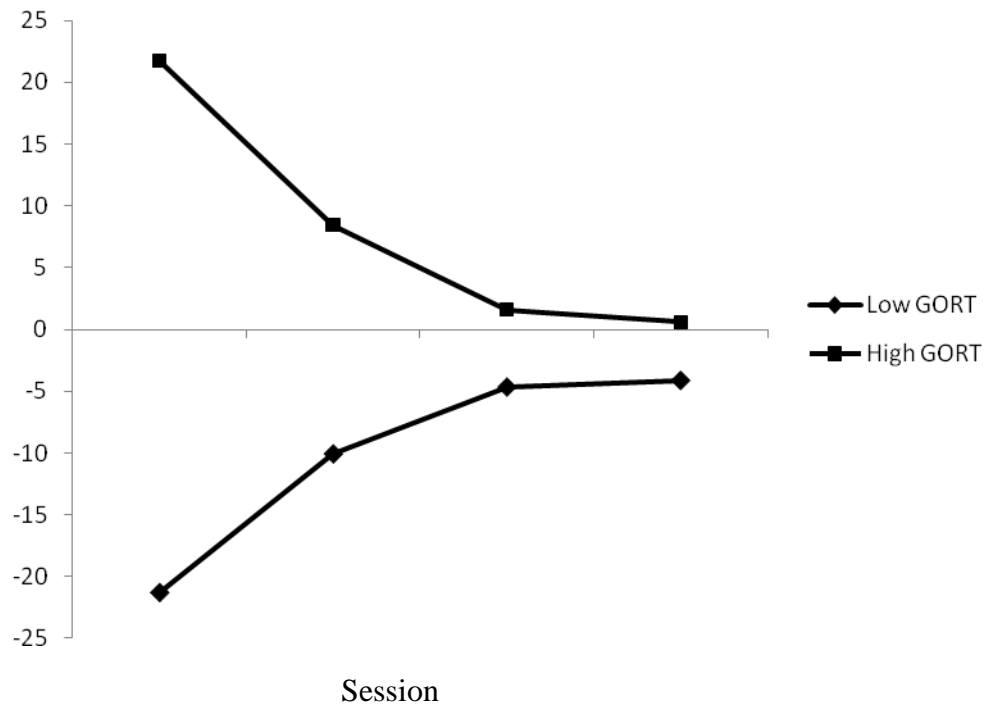


Figure 4 Quadratic Interactions for Retention Passage Gains—Fluency

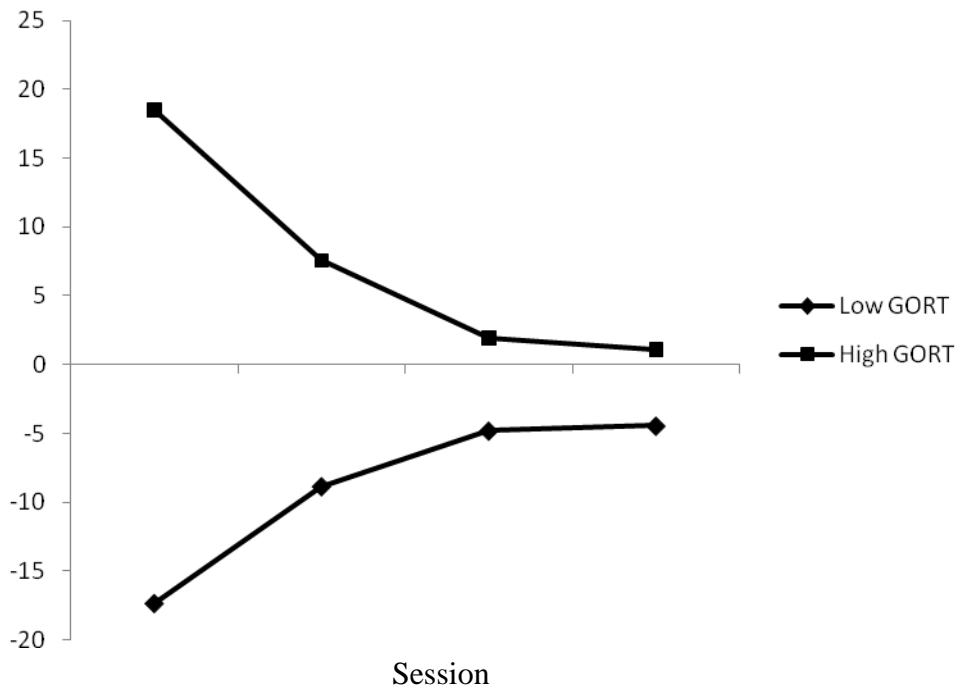


Figure 5 Quadratic Interactions for Retention Passage Gains—Comprehension