

IDENTIFYING STUDENTS AT RISK FOR ACADEMIC FAILURE:
THE APPLICATION OF A PREREFERRAL SCREENING MODEL
INCLUDING RESPONSIVENESS TO INTERVENTION

A Dissertation

Submitted to the Graduate Faculty of the
Louisiana State University and
Agricultural and Mechanical College
in partial fulfillment of the
Requirements for the degree of
Doctor of Philosophy

In

The Department of Psychology

by

Susan Lockhart Gatti
B.A., Louisiana State University, 1997
M.A. Louisiana State University, 2000
May 2004

Table of Contents

Abstract	iv
Introduction	1
Areas of Investigation	1
Rationale and Purpose of the Current Study	2
Review of the Literature	6
Learning Disabilities	6
Screening with Curriculum-Based Measurement	14
Responsiveness to Intervention	23
Purpose of the Current Study	29
Method	30
Setting	31
Participants in the School’s STEEP Process	32
Assessor, Intervention Provider, and Teacher Training	32
Measures	34
Procedures	39
Data Collection for this Study: Retrieval of Archival Records	41
Data Analyses	43
Reliability of Measures and Procedural Integrity	48
Results	50
Outcomes of Data Analyses	50
Reliability of Measures and Procedural Integrity	61
Discussion	62
Limitations of the Study	69
Future Directions	71
Summary	73
References	75
Appendix	
Appendix A: University Course Readings	82
Appendix B: STEEP Instructions for Math	84
Appendix C: Sample Math Probe	85
Appendix D: STEEP Instructions for Reading	86
Appendix E: Sample Reading Probe, Examiner Copy	87

Appendix F: Instructions for Can't Do/Won't Do Assessment in Math	88
Appendix G: Instructions for Can't Do/Won't Do Assessment in Reading	89
Appendix H: Integrity Checklist for Math Probe Administration	90
Appendix I: Integrity Checklist for Reading Probe Administration	91
Appendix J: PPRI Implementation Procedures	92
Appendix K: Treatment Integrity Checks for PPRI	93
Vita	94

Abstract

Increasing demands are continually placed on our education system to improve the educational outcomes of **all** children. To provide the best services to our children, appropriate screening, assessments, and intervention services need to be implemented within the school setting. The current study evaluated the relationship of a screening procedure (Screening to Enhance Educational Performance, STEEP; Witt, 1996) using curriculum-based measurement (CBM) and compared the results of the screening procedure to other commonly used problem nomination procedures for second, third, and fourth graders. The data were obtained from archival sources that resulted from a service and training project conducted jointly by a university and its associated elementary level professional development school. The students who demonstrated the greatest need for remediation of their reading skills were provided an intensive reading intervention and the improvement in their reading skills relative to their peers was evaluated to determine each individual's response to the intervention, as was the overall effectiveness of the intervention for improving reading fluency for all students who participated in the intervention. Results generally indicated that STEEP was concordant with other problem identification methods such as teacher referrals and criterion-referenced tests. Furthermore, students' response to intervention services was evaluated based on their performance when compared to peers and on the demonstrated efficacy of the intervention that was implemented. The data were also analyzed to determine the number of intervention sessions necessary to predict long-term outcomes based on students' slope of their regression estimates. The study suggests that the process implemented by the school that included broad general screening, more detailed assessment of at-risk

students, and early intervention for low performing students may be a viable approach for secondary prevention that is generally, but not precisely, concordant with other methods of defining educational problems. Models similar to the one implemented in this case may contribute to the professional evaluation of the determination of students' need for special education services.

Introduction

There is a significant increase in the number of students identified as learning disabled in school systems, which is a source of concern given that the educational outcomes of learning disabled students are poor (U.S. Department of Education, 1998). Recent legislation mandates that our educational systems be responsible for educating *all* of our students, including those in general education and special education. This mandate is most clearly evident in the recent legislation that has been described as the *No Child Left Behind Act* (PL 107-110, 2001). The apparently dismal results of the students who are educated in special education curricula indicate that we need to evaluate not only the special education programs in our country, but also the classification procedures by which students are initially identified for these services. Society has increasingly emphasized the importance of identifying students in need of extra assistance early. With intense intervention, the growth rates of at-risk students can approximate that of their normal peers such that their reading performance is on an average level (Velluntino, Scanlon, & Tanzman, 1998). Early intervention services for students can lead to improved long-term performance and potentially prevent their placement in special education. This is critical since the law mandates that we educate our students in the least restrictive environment possible (Individuals with Disabilities in Education Act, 1990). The combination of these two factors points to the need for a reliable early screening procedure.

Areas of Investigation

Classification procedures can be viewed as a combination of several activities that are conducted when determining a student's placement, including prereferral screening

procedures, standardized assessment procedures, and the determination of a student's response to remediation services. Several research questions are posed in the current study to examine specific activities conducted during the classification procedure. Specific research questions will address the concordance between different problem identification methods, the effectiveness of a reading intervention, the response of students to the intervention, and the predictability of student performance throughout the intervention to long-term outcomes. The goal for the current study is to examine the effectiveness of methods that can be used during disability classification procedures and special education eligibility determinations that would rule out possible explanations for academic deficits. By ruling out alternative reasons for academic failure (such as motivation and material that is too difficult) through objective data to reduce bias, students who proceed through the activities and continue to experience significant difficulty will be more likely to demonstrate the characteristics of a student with a disability. Background information is provided in the areas of learning disabilities (LD), screening tools, and responsiveness to intervention to address the relevant issues to the current study.

Rationale and Purpose of the Current Study

The educational system must have an efficient and effective method to identify students in need of additional assistance in the classroom and to identify students with high-incidence disabilities. Currently there is significant dependence on teacher referral to identify students in need of additional assistance. However, teachers are susceptible to bias and have different expectations and tolerance levels for their students' problems. The Screening to Enhance Educational Performance (STEEP) model uses curriculum-based

measurement (CBM) to screen all students, which allows comparisons of student performance on many levels (individual, class, and national standards; Witt, 1996). The STEEP model has been implemented in many areas of the state of Louisiana and has demonstrated positive results within the Pupil Appraisal Model and the Problem Validation Screening (PVS) (VanDerHeyden, Witt, & Naquin, 2003; Witt, Adler, Thomas, & Naquin, 1999).

The purpose of this study is to supplement the literature and provide empirical evidence of several classification activities. Additionally, this study is unique in that it is based on data that occurred as the result of a complete implementation of STEEP with a public school. The accuracy of STEEP as a screening instrument is assessed by comparing the results of STEEP to the results of criterion-referenced tests that are administered to students annually to determine the concordance between the measures. Also, the students nominated for additional assistance by STEEP were compared to students nominated for additional assistance by teachers. Test performance and teacher referral are typically used when deciding to refer a student for additional assistance at the school level. Therefore, the extent to which STEEP scores are related to these problem nomination procedures was examined.

Once students are identified as demonstrating reading difficulties, there is a growing trend of research to indicate that their response to an appropriate and strong intervention is important to determining the extent that lack of instruction has contributed to their reading problem (Gresham, 2001; Speece, Case, & Molloy, 2003; Vaughn, Linan-Thompson, & Hickman, 2003). A reading intervention that included the empirically-validated components modeling, drills, error correction, and reinforcement

was provided to low performing students by their school district. Based upon the progress monitoring data collected, this provided the opportunity to determine the extent to which the low reading performers responded to the intervention by increasing their reading fluency. The overall effectiveness of the intervention to improve reading performance of the participants and how soon the effectiveness of the intervention can be reliably predicted is determined. Whereas the Fuchs and Fuchs (1997) treatment utility model emphasizes adaptations that can be implemented in the general education environment, the current study evaluates students' responses to a very strong, individualized intervention. To determine who responded to the intervention successfully and who did not respond to the intervention successfully, an evaluation was conducted based primarily on students' rates of improvement. The extent to which STEEP identifies students who are resistant to the intervention according to the proposed model's definition is also addressed.

The current study addresses the above topics through the following questions. (1) What is the stability of class rankings based upon CBM over several months? This question addresses the extent to which low class rank is a stable assessment result. In order for an assessment to **suggest** disability it must be reasonably stable over time. (2) What is the concordance between differing methods of defining academic deficits? Although precise concordance is not expected between different ways of defining problems, if a new approach is adopted, policy makers and clinicians need to be aware of the degree of concordance between historic and new systems. It is particularly important that the nature of discordance be clarified. (3) Which students were responsive to intervention? This is a central element of recently proposed methods for identifying

students who are disabled. (4) Does the Progressive Practice Reading Intervention (PPRI) remain effective in this clinical replication? (5) How many intervention sessions are necessary to reliably predict the treatment outcome? The answer to this question provides critically needed guidance for applied decision making that has not received empirical scrutiny.

Review of the Literature

Learning Disabilities

The initial framework describing students with LD was based on the view that the problem lies within the student (MacMillan & Speece, 1999). The discrepancy approach used to classify students with a LD is based on the descriptions of students by Kirk (1962) and Bateman (1965). Bateman (1965) particularly emphasized the significant difference between potential and actual academic performance. Rutter and Yule's (1975) research hypothesized two types of impaired readers based on relationships that were demonstrated between IQ (potential performance) and achievement scores (actual academic performance). The "specific reading retardation" type demonstrated significant discrepancies between expected and observed reading scores without any general learning problems (which is now interpreted as cognitive skills deficits). The second type of reading impaired students was defined as having "general reading backwardness," which is a lack of discrepancy between expected and observed reading skills with general learning problems. The research of Rutter and Yule (1975), as well as Kirk (1962) and Bateman's (1965) conceptualization of learning disabilities, primarily established the rationale and the basis for the IQ-achievement discrepancy classification method that is still widely used today.

However, the essential assumptions in the discrepancy approach (i.e. students with significant discrepancies between their achievement and aptitude scores perform differently than students with low achievement and aptitude scores) have not been demonstrated to be valid, and Rutter and Yule's (1975) work has not been successfully replicated (Velluntino, Scanlon, & Lyon, 2000). In order for the discrepancy approach to

have merit, general intellectual scores must be predictive (1) of performance on reading measures and/or (2) of a child's responsiveness to remediation. More recent research has demonstrated that low performing readers who were shown to have significant aptitude-achievement differences and low performing readers who did not have a significant aptitude-achievement differences performed no differently on independent measures of reading achievement (Stanovich & Siegel, 1994). In fact, language and language-based skills are better predictors of reading performance than IQ scores (Siegel, 1989).

Velluntino, Scanlon, and Lyon (2000) further replicated that IQ scores are not highly predictive of reading ability at the beginning stages in a longitudinal study, and IQ scores also do not differentiate between impaired readers whose poor skills are easily remediated versus those impaired readers whose skills are more resistant to remediation.

Given the lack of support for the discrepancy approach to define learning disabilities, it is not surprising that the referral and classification procedures in the school setting are highly variable between settings and often confusing to consumers. The lack of a consistent classification procedure and LD definition has also impeded effective research and practice (Bocian, Beebe, MacMillan, & Gresham, 1999; Lyon, 1996).

However, according to Bocian, Beebe, MacMillan, and Gresham (1999) there are three essential and rational steps in identifying a disabled student that are consistent in the school setting: a referral by the general education teacher, a psychological evaluation of the student providing psychometric scores that are compared to state criteria for placement, and a team recommendation for placement. The above steps occur in a sequence, and it is necessary to pass through one gate to enter the next gate. Furthermore, various factors affect decisions at each step of the process, such as the question being

addressed in the decision-making process, the role of professional judgment, the use of local versus national norms, and the influence of social and cultural factors. The above factors result in the three steps being “competing paradigms” during the classification of LD students and are likely relevant in the differentiation between the students who qualify for LD classification according to school and research criteria. For example, teachers often use local comparisons for students’ performance and often will consider the sociocultural and contextual factors when using professional judgment in deciding whether or not to refer students. The psychological evaluation inevitably employs a comparison of the student to national norms and then to the local eligibility criteria. The multidisciplinary team also exercises the judgment of many professionals while examining the data, which should include a comparison of the student’s performance to local norms and national norms, as well as consideration of the social circumstances of the student. A teacher nominates a student as potentially having a disability according to this classification procedure, and the psychological testing and the multidisciplinary team decision follow to verify the teacher’s judgment. The current classification procedure requires extensive assessments at each of the three steps while the competing paradigms could be restructured in a procedure that is more considerate of the time of professionals and the needs of students (Bocian et al., 1999).

Arguments for a more efficient process to classify LD students vary from trusting teacher judgments about a student’s abilities to staying with the traditional approach of demonstrating a significant discrepancy between aptitude and academic achievement. The discrepancy approach attempted to differentiate a student’s potential performance from actual performance, and four major methods have been used to measure this

discrepancy (Berninger & Abbott, 1994). The methods include (1) deviation from grade level, (2) expectancy formulas using grade-equivalent scores while controlling for intelligence level, (3) simple standard score difference between achievement and intelligence measures, and (4) standard regression analysis, which accounts for measurement errors in the simple standard score difference method. The majority of the above methods assume that the intelligence scores accurately measure a student's potential achievement. It has been argued that intelligence tests are not the best measure of student potential and are not useful for instructional or intervention planning (Gresham & Witt, 1997; Siegel, 1989). The authors suggest replacing intelligence tests with assessment tools with demonstrated treatment utility since the information from intelligence tests does not have utility for planning, implementing, or evaluating instructional interventions and does not demonstrate utility for making differential diagnoses for students with mild problems. The measurement problems that are involved with the classification of LD and the time and expense of classification procedures call for a more efficient, data-based, and streamlined referral and classification procedure that pivots on a teacher's judgment. Gerber and Semmel (1984) recommend that psychoeducational assessments of disabilities be dismissed and that we trust teachers' judgments of student problems while providing them with additional support in the classroom. The idea is that teachers are "imperfect tests" for identifying students in need of special services (Gerber & Semmel, 1984).

Referral and Classification Procedures in the Schools

Traditional screening methods used by the public school system to identify students with disabilities appear to be confusing, inconsistent, and highly reliant on

information provided by teachers. The procedures used by teachers for referring students vary from school to school and do not follow standardized procedures or tests. Once a student is referred for a special education evaluation, his placement is probable. Results from a study by Algozzine, Ysseldyke, and Christenson (1983) indicated that 92% of students who were referred for a psychoeducational evaluation were evaluated, and 73% of students who were referred for a psychoeducational evaluation were identified as eligible to receive special education services. Furthermore, the number of students receiving special services under the Individuals with Disabilities Act that are served as LD has increased dramatically (U.S. Department of Education, 1998). According to the 1998 United States Department of Education report, the number of students served as LD increased from 797,213 to 2,259,000 between 1976-77 and 1996-97 representing an increase to a level that was 283% of the estimates from 1976-77, and the students classified as LD represent 52% of all the students served.

There is much debate in the public schools, as well as in the scientific literature, about classification and assessment procedures for students with a learning disability (Bocian et al., 1999). A substantial proportion of students (52% to 70%) who are classified as having a learning disability (LD) by the public school system fail to meet criteria for eligibility according to state or federal guidelines (Lyon, 1996; MacMillan, Gresham, & Bocian, 1998; MacMillan & Speece, 1999; Shaywitz, Shaywitz, Fletcher, & Escobar, 1990; and Shepard, Smith, & Vojir, 1983). Keogh (1994) noted the difference could be due to comparing students who were classified for various purposes: advocacy, services, or scientific study. Nonetheless, the significant difference is indicative of the

need for a consistent and streamlined prereferral screening instrument and classification system that is applicable in the school setting.

Accuracy of teacher referrals. The significant increase in the rates of LD classification makes a reliable and valid referral system a necessity in the school system. A teacher's decision as to whether or not to refer a student to the assistance team at their school is a necessary, but not sufficient, first step to classifying a student with a disability. There is evidence to support the teachers' decisions and judgments, as well as evidence to question their accuracy (Shinn, Tindal, & Spira, 1987). Research has indicated that teacher ratings are related to psychoeducational assessments (Gresham & Witt, 1997) and that teachers are as accurate at differentiating between non-handicapped students and LD students using a rating scale as a standardized test (Gresham, Reschly, & Carey, 1987). However, the accurate differentiation in the above study may have been biased by prior knowledge of the placement of LD reevaluation students that participated in the study. In a study using students that had been referred and not yet identified, it was concluded that teachers "are accurate and defensible 'tests' for identifying and classifying children into psychometrically defined at-risk groups of LD, Low IQ, and LA"(Low Achiever) (pp. 56-57, Gresham, MacMillan, & Bocian, 1997).

There is also evidence that teachers are biased in their referral. Shaywitz, Shaywitz, Fletcher, and Escobar (1990) demonstrated that males and students with behavior problems were overrepresented, while females were underrepresented in teachers' referrals of reading difficulties. Teachers have also nominated black students more often for referral than white students (Bahr, Fuchs, Stecker, & Fuchs, 1991). Although this may be a result of poorer academic performance from black students (Bahr

et al., 1991), it could also be a factor in the overrepresentation of minority students in special education. The problem of minority overrepresentation is a concern of educators and administrators at local and national levels, and it has been proposed that the overrepresentation might be a result of bias in tests and their procedures and/or teacher selection of referred students (Bahr et al., 1991). Given the identification rates of black males, especially in the areas of emotional disturbance and mental disabilities (U.S. Department of Education, 2002), it is critical that teachers have reliable, valid, and unbiased data to assist in their decision-making when determining whether or not a student will be referred for special education services. Data-based decision making procedures are critical in the referral process, especially since teachers concede that they are influenced by factors both external and internal to their school environment (Christenson, Ysseldyke, & Algozzine, 1982).

Classification differentiation. The literature referring to the accuracy of teachers' judgments and the accuracy of test-based discrepancy models to reliably identify students with disabilities is variable. One reason for the variability in the classification of school-identified LD students is inevitably a result of the "competing paradigms" in the school setting that result in inconsistent classification procedures (Bocian et al. 1999).

MacMillan, Gresham, and Bocian (1998) demonstrated that fewer than half of students referred for a learning disability that were classified at the school level met the aptitude-achievement discrepancy required by the state. The consistencies in the application or misapplication of the law were to exclude students with higher scores in spite of a significant aptitude-achievement discrepancy from LD classification and to classify students without a significant aptitude-achievement discrepancy as LD who were lower

performing or had a sufficiently low IQ and profile to qualify for a mental disability. This practice of classifying mentally disabled students with a learning disability is one hypothesis to explain why the numbers of students classified as LD is increasing while the number of students classified with a mild mental disability is decreasing (U.S. Department of Education, 1998). This phenomenon likely represents the relative social undesirability of the two labels. It may be less distressing for parents to describe their children as LD than mentally retarded.

Many studies have compared school-identified LD students and low achieving students to determine if there are any significant differences between the groups. A pivotal study in this area reported that LD students could not be reliably differentiated from low achieving students on a battery of psychoeducational measures (Ysseldyke, Algozzine, Shinn, & McGue, 1982). An average of 96% of the scores (range from 82% to 100%) fell within a common range on the measures. The primary concerns that resulted from this study were either schools are not classifying students who are LD or schools are classifying students as LD who do not have the disability. Kavale, Fuchs, and Scruggs (1994) questioned the above results primarily because the methodology did not take into account each group's variability. When using a methodology that accounted for the variability of each group, the study resulted in an average of 37% overlap of scores and discriminated 63% of the LD group from the LA group. Kavale et al. (1994) concluded that students with learning disabilities could be clearly differentiated from low achieving students even when using data from research that has argued against the differentiation.

Conflicting results continue to fuel the differentiation debate. Shaywitz, Fletcher, Holahan, and Shaywitz (1992) conducted a longitudinal study that compared the two

definitions of LD that were originally outlined by Rutter and Yule (1975). The study identified second-grade students according to the traditional discrepancy-based model and the low reading achievement model and compared the groups on parent, teacher, and student measures (completed retrospectively to kindergarten and continuing into the fifth grade). The findings suggest that although students who qualify according to the discrepancy-based model have higher verbal, performance, and full scale IQ scores, there are more similarities than differences between the two groups. The authors advocate that both groups of students demonstrate reading disabilities and should be classified as learning disabled.

Inconsistencies within the LD field in the areas of the underlying assumptions about the disability, the definition of the disability, the classification procedures for identification, and the differentiation of characteristics of students with LD versus low achievers lead to conflicting views from experts in the LD field, general confusion about the disability, and the potential misapplication of the disability label in school settings. The traditional problems in special education and in the area of LD have led educators to emphasize the need for early and accurate identification of students with disabilities and to emphasize the extent of services needed to help the students thrive in their environment.

Screening with Curriculum-Based Measurement

The traditional model of learning disabilities that is based on a psychometric-exclusionary approach, which relies on standardized scores from psychoeducational assessments to diagnose reading disabilities is problematic for several reasons. Velluntino, Scanlon, and Tanzman (1998) note that the psychometric-exclusionary

approach (1) does not differentiate between impaired readers who have difficulty due to cognitive deficits or to inadequate instruction and/or preliteracy experiences; (2) does not outline specific criteria for the disability that are empirically-validated; (3) places too much emphasis on the aptitude-achievement discrepancy that has not been consistently demonstrated in the literature; and (4) does not assist in the development of specific interventions to meet the child's needs. The authors note the importance of providing early and intensive intervention for students who are at-risk for reading deficits. It naturally follows that schools are in need of screening instruments that can accurately and effectively identify students who are in need on early intervention services.

Traditionally, schools have employed a wait-and-see strategy that allows students to fail prior to referring them for additional assistance in spite of the administration of assessment instruments throughout their schooling history. Many preschoolers and kindergartners are given a developmental screening instrument upon entering school; however, the data from the screening instrument are seldom integrated into instructional plans for the class or the individual. Also, school districts require the administration of a norm-based achievement test at the conclusion of the school year. The late administration, delay in receiving test results, and global nature of the achievement test also make it more difficult for teachers to use for instructional planning for whole classes and for individual students. Furthermore, the assessment instruments administered in the school often provide comparisons of a student's performance to that of a national sample, and do not often provide a readily available comparison to local peers. Teacher judgment is a primary factor when deciding to refer a student, and teacher referral is the first step in identifying a student with a disability within the school system. Data from all of the tests

administered can be used to validate teacher concerns that originated in classroom performance. Data from an efficient screening procedure for all children that allows comparison to classroom peers and is relevant to specific instructional planning is a more efficient process of determining who is in need of remedial instruction in the classroom.

Curriculum-based Measurement (CBM)

An assessment tool with standardized procedures that allows for comparisons on individual and group levels is important for the effective screening of students. Curriculum-based measurement (CBM) is an assessment tool that meets the above criteria, and it is becoming an increasingly common practice to conduct CBM in schools while screening for students who are at risk for academic failure (Deno, 2003). CBM is a methodology for measuring academic performance and progress. The purpose of CBM was to create a measurement system that could (1) be used efficiently by teachers, (2) produce accurate information that records academic performance and progress in a meaningful way, (3) provide feedback about the impact of instruction and intervention on student progress, and (4) produce information that could improve instructional planning (Deno, 1985; Deno, Fuchs, Marston, & Shinn, 2001). CBM is an efficient and inexpensive assessment procedure that can be administered multiple times in parallel forms throughout the school year to all students (Shinn, 1989). This measurement allows for assessment of basic academic skills, which is directly linked to instructional planning for educators. CBM methods produce reliable and valid information about students' academic performance and their skill acquisition or mastery level at a given point in time (Marston, 1989).

Reading fluency measures have been demonstrated to be predictive of students' reading comprehension performance (Fuchs, Fuchs, & Maxwell, 1988). Students' performance on statewide achievement tests has also been predicted through moderate correlations with CBM scores from the same year and the previous year (Crawford, Tindal, & Stieber, 2001). CBM has also successfully discriminated between students with typical achievement and students with specialized instruction needs (Marston & Magnusson, 1988). The results support the use of benchmarks to assist in identifying students at risk for failure on high-stakes tests and support the stability of CBM measures over time.

In addition to criterion-related comparisons and comparisons to national standards of appropriate fluency levels and growth rates (Deno et al., 2001), CBM allows comparison of academic performance and rates of improvement in performance to classroom peers and the student's own progress. For example, CBM has been used to individually set goals for student performance based on their growth rates. As a result, teachers set higher standards for students, and student performance improves over goals that are not modified based on student performance (Fuchs, Fuchs, & Hamlett, 1989). CBM data have been used effectively to establish norms for the local area and literature-based curricula, allowing comparison to peers on multiple levels (Hartman & Fuller, 1997; Shinn, 1989).

CBM is an effective tool for assessing student achievement and progress that can be readily employed during screening assessments because it is efficient and targets specific skills that are predictive of academic achievement. The advantages of using CBM address many of the traditional problems outlined above. CBM uses measures that

can address preliteracy skills that have been demonstrated to predict reading deficits such as basic phonics skills (Kaminski & Good, 1996; Velluntino et al. 1998). Furthermore, CBM can also be used to design effective interventions for a student and to evaluate a student's responsiveness to remedial instruction because it addresses specific skills and is sensitive to change in student performance (Marston, 1989). The legal requirement and growing emphasis on prereferral interventions for students, as well as the use of CBM in evaluations of eligibility determinations, makes the practice of CBM a critical component of school activities during the referral process (Individuals with Disabilities Education Act, 1990; Pupil Appraisal Handbook, 2000).

Screening and Prereferral Assessment Models

Elliott and Fuchs (1997) suggest that CBM be considered as an alternative to traditional achievement and aptitude tests in an eligibility framework consisting of a series of questions based on Messick's (1984, 1995) framework for treatment validity. Messick (1984, 1995) defined the validity of assessment instruments as not only the meaning and interpretation of the test scores, but also the associated social consequences (both intended and unintended) of the assessment for the individual. Three questions are critical when considering a student's performance and suggest the utility of CBM to answer the questions (Elliott & Fuchs, 1997). (1) Is the level of learning acceptable in the current instructional environment? (2) If not, can the instructional environment be modified to promote acceptable learning? (3) If not, does the student demonstrate a need for special education support to produce acceptable learning? This model is further elaborated on an individualized level by Fuchs and Fuchs (1998) who suggest that school-level teams use a dual-discrepancy model when determining eligibility. The

model consists of (1) comparing an individual student's achievement level and growth during an intervention to that of his or her classroom peers, (2) determining if the individual student's rate of learning given an **appropriate** intervention is unacceptable relative to his classroom peers, and (3) ensuring that special education placement will result in improved academic achievement for the individual student. The dual discrepancy model involves assessing both the level and the trend of student performance compared to classroom peers (1 standard deviation for each; Fuchs & Fuchs, 1997, 1998).

The Minneapolis Public Schools began to incorporate CBM into the prereferral assessment procedures for special education eligibility. Subsequently, the number of students who were referred for special education that were classified dropped significantly from previously reported rates in the literature (Algozzine et al., 1983; Marston & Magnusson, 1988). The model called for ongoing CBM data collection while adaptations were made to instructional environments. The CBM data were used to determine the extent to which the students' learning could improve and their needs could be met given the implemented adaptations. Only if a student did not demonstrate improvement following the adaptations was he or she eligible for special education services. From the students who were initially referred for special education services, only 25 to 45% were considered eligible for special education services.

The Problem Solving Model (PSM) is an updated program that has been officially implemented in the Minneapolis Public Schools since September 1992 (Marston, Muyskens, Lau, & Canter, 2003). CBM was an integral part of the PSM that was used to develop interventions in the general education setting, to determine who to refer to special education, and to evaluate students suspected of having a high-incidence

disability. The following sequence of problem solving steps is implemented in the model following a school-wide screening. First, the student's problem is specifically defined, including an evaluation of strengths and weaknesses. Second, the staff generates instructional interventions to match the student's need and implements the intervention with integrity. Third, the student's progress is monitored (typically through CBM) and evaluated every six to eight weeks. Finally, the above cycle is continued since the model is based on a "teach-test-teach-test" model. Most of the emphasis of the model is in the above problem-solving first step within the general education setting and with the general education classroom teacher. A multidisciplinary team then meets about the student's specific needs and develops possible solutions in the second step. The final step is a special education evaluation for those students who do not demonstrate sufficient progress in response to the implemented interventions.

Results of the PSM are positive. The implementation of the PSM resulted in significantly higher numbers of students being referred to the multidisciplinary committee (step 2); however, the number of students identified with high-incidence disabilities did not significantly increase. Meanwhile, the student achievement performance level and growth of the students identified through PSM and LD students were similar and indicated that the students were likely to pass the local standards test. The implementation of PSM resulted in the decreased likelihood of placing an African American student in special education versus a white student, which positively impacts the issue of disproportionate placement of African-American students in special education. Results indicated that 68.9% (n=184) of students placed in special education were African-American in 1997-1998 versus 55.4% (n=124) of students placed in special

education were African-American in 2000-2001 (Marston et al., 2003). An independent evaluation comparing PSM to traditional approaches concluded that superior prereferral interventions were implemented; students received special education services earlier; non-discriminatory practices were implemented; and an overlap of 75% of students identified through PSM and likely to be identified by traditional approaches occurred. The systematic, data-driven process of PSM, which is based in functional assessment and collaborative teaming, resulted in improved decision-making and assessment for special education while focusing on improving interventions implemented with better identified students in general education in need of assistance (Marston et al., 2003).

Both the PSM (Marston et al. 2003) and the PVS (VanDerHeyden et al., 2003) focus on reducing bias in the referral system. The PVS is an additional screening procedure whose goal is to screen all students in a classroom or school using CBM assessments prior to considering them for a full evaluation and eligibility determination (VanDerHeyden et al., 2003). The PVS is based on a problem-solving approach (Fuchs & Fuchs, 1998, Good & Kaminski, 1996, Shinn, 1995, & Shinn et al., 1998), including problem identification and problem certification, to identifying students for special education services (Shinn, 1995). The model calls for the use of CBM while conducting local comparisons for students' performance, brief implementation of an appropriate intervention, and dual discrepancy comparisons on level and trend of student performance. Procedures for validating student problems proposed by Witt, Daly, and Noell (2000), which include collecting classwide CBM data, directly observing the target student and his peers, assessing the effect of offering an incentive on academic

performance, and intervening in the natural setting while monitoring treatment integrity, are the basis for PVS.

PVS is a gated procedure, where a student proceeds through sequential steps in the process based on their performance (VanDerHeyden et al., 2003; Witt et al., 2000). First, classwide academic assessments using standardized CBM procedures in the areas of reading and math were conducted. Second, a performance/skill deficit assessment was conducted for lowest performing students. The assessment is supported by research that has demonstrated some students need only incentives to improve performance and not additional instruction (Noell et al., 1998). If a student was determined to have a primarily motivational deficit, then the student was not referred on to the intervention phase and not considered for a full evaluation. The final phase of PVS, as described by VanDerHeyden et al. (2003), consisted of a brief instructional session. The instructional session was to clarify whether a student simply did not understand the task or the direction or whether a student might respond quickly to intervention. The instructional session consisted of a teacher re-administering a CBM probe, reviewing task directions, modeling correct responses, allowing the student to correct their errors, and re-administering the CBM probe. The students' performances during these procedures were compared to multiple standards (including standardized test scores, school-based assessments, teacher referral, and the criterion standard of a student's responsiveness to an extended intervention) used when deciding whether or not to refer a student for a full evaluation. When comparing students' performance on the PVS to their performance on the Iowa Test of Basic Skills (ITBS; Hoover, Hieronymus, Frisbie, & Dunbar, 1993) and on the extended intervention, PVS demonstrated the best predictive power of student

performance. Only 19% of the students that teachers referred for the school-level committee for help demonstrated a validated problem during the criterion assessment, and teachers failed to identify 11% of the students who did demonstrate a validated problem according to the criterion assessment. Overall, PVS correctly identified 87% of the students with validated problems. Additionally, CBM data obtained strongly correlated with ITBS scores. It is noted that not all of the students in the school participated in the criterion assessment that is used as the criterion for comparison. The results indicate that the overreliance on teachers to initiate the referral process needs to be reevaluated. Additionally, the authors conclude that the results support the need for universal screening of all students, the importance of the performance/skill deficit assessment, the use of a class-level of comparison for student performance, and the use of a problem-solving model when evaluating students.

Responsiveness to Intervention

Responsiveness to intervention (RTI) is the extent to which a change in behavior or performance occurs as a result of an intervention (Gresham, 1991). Conceptualizing the identification and severity of disorders in terms of a student's responsiveness to intervention takes the focus away from the within-child view of a problem (MacMillan & Speece, 1999) and focuses on the contribution of the environment and instruction to which a student is exposed (Gresham, 1991). When behavior or performance is resistant to improvement following the implementation of an empirically-validated and appropriate intervention, there is evidence that more intensive services are needed to improve the problem.

Rationale of Responsiveness to Intervention Model

Resistance to intervention is based on the behavioral momentum literature, which examines a behavior's resistance to change using the language and terminology of physics (Nevin, 1988). Behavior is compared to a moving body that possesses both mass and velocity. The body will maintain constant velocity given constant conditions and will change proportionate to an external force and inversely proportionate to its own mass. Once the body is in motion, its tendency is to stay in motion. Gresham (1991) compares the strength of a response to the mass of a moving body and the implementation of an intervention to the application of an external force. The more problematic a student's behavior or academic performance is initially, the more intense the intervention that is applied needs to be. The responsiveness to the intervention is theoretically the strength of a response as a function of an intervention that was implemented to improve that response. The strength of the response is the difference between the behavior or performance at baseline levels and following the intervention. Therefore, responsiveness to intervention also takes a discrepancy approach to identifying students with learning disabilities by comparing their performance before and after the implementation of an appropriate intervention.

Treatment Utility

An essential component of the RTI model is the implementation of an appropriate and empirically-validated intervention. The assessment of learning disabilities then should lead to a more effective intervention. The treatment utility of assessment is "the degree to which assessment is shown to contribute to beneficial treatment outcome" (Hayes, Nelson, & Jarrett, 1987, pg. 963). The traditional aptitude-achievement

discrepancy approach does not assist in identifying students whose reading difficulties primarily are due to inadequate or absent instruction, does not predict response to remediation, and does not discriminate between disabled and non-disabled readers according to Velluntino et al. (1996, 2000). Fuchs and Fuchs (1997) propose that in order for the LD assessment process to demonstrate treatment utility, it must be able to model academic growth, to discriminate between inadequate instruction and unsatisfactory levels and rates of individual learning, to contribute to improved instructional decisions, and to be sensitive to growth for monitoring treatment effects.

The discrimination between a skill deficit and performance deficit is an important distinction when identifying students with disabilities and when developing instructional plans and interventions. A skill deficit indicates that a student is lacking the necessary knowledge or skills to perform a given academic task, and a performance deficit indicates that a student is lacking the necessary motivation to perform an academic task. One example of an assessment tool that has been proposed to have treatment utility is the use of CBM to determine if a child exhibits a skill deficit or a performance deficit (Noell & Witt, 1998). Noell and Witt (1998) recommend comparing a student's reading fluency first using standard CBM conditions and then using a reward for improved performance over the score that is specified from the standard conditions. A significant improvement would indicate that the student has the skills and is not motivated to perform, and an insignificant level of improvement would suggest that the student does not have the prerequisite skills to perform the task. Whereas additional and appropriate instruction is necessary to remediate skill deficits, motivating contingencies for performance are necessary to remediate performance deficits (Lentz, 1988; Noell & Witt, 1998). The

above example demonstrates the importance of using assessment tools when evaluating students for LD that directly lead to effective interventions, particularly within the RTI model.

RTI into Practice

In recent literature there is an increased focus on RTI as a model to evaluate eligibility for special education services. The model presented by Fuchs and Fuchs (1997) has been specifically reviewed and analyzed (Gresham, 2001; Speece, Case, & Molloy, 2003; Vaughn & Fuchs, 2003). Fuchs and Fuchs presented a dual discrepancy model that emphasizes the importance of treatment utility and CBM. The model also emphasizes the screening of all students (as mentioned in a prior section) and the comparison of a student's performance level and rate of growth to that of their classroom peers through three phases. The purpose of the first phase is to determine if the overall instruction in the classroom is sufficient to foster academic progress for the students by assessing the performance and rate of growth of all students in a classroom. After the classroom environment demonstrates its efficacy, the students who are performing at a significantly lower level and progressing at a slower rate than their classroom peers (dual discrepancy of level and rate of growth) are identified in the second phase. The identified students are considered to be resistant to improvement in an effective general education setting. The third phase consists of systematically enhancing instruction and implementing classroom adaptations in the general education setting that are designed to improve individual student performance and to be applicable for long-term implementation by their teacher. For those students who do not demonstrate adequate growth, special education may then be considered. Most recently, the model has revisited its initial fourth phase, which

verifies the effectiveness of the special education program prior to determining a student is eligible for placement (Fuchs, Fuchs, & Speece, 2002). The Fuchs and Fuchs (1997) model begins with screening all children and progresses to assessing students at risk for academic progress for their responsiveness to adaptations of the general education setting and environment.

RTI in practice has yielded positive results. Speece, Case, and Molloy (2003) reviewed three studies that evaluated students' responses to general education in the area of reading to determine if more intensive instruction and remediation were necessary. The authors found that a valid group of poor readers could be identified using CBM reading fluency while implementing the dual-discrepancy model. More severe academic and behavioral deficits were evidenced in students that demonstrated a significant difference in their performance level and rate of growth than students who demonstrated an aptitude-achievement discrepancy. Also, students who were identified at risk for learning difficulties and participated in specially designed general education instruction resulted in better academic performance and required fewer additional services than did their at-risk peers who did not participate in the specialized instruction. The authors suggest that RTI has a promising beginning and warrants further investigation and study as a method for identifying students at-risk for learning difficulties and students with disabilities early in their school career.

As the recognition and research of the LD population continues to grow, there is a growing demand to redefine the disability and to reexamine the assessment process for eligibility. Vaughn and Fuchs (2003) evaluate the "promise" and the "potential pitfalls" of implementing responsiveness to intervention into common practice as an alternative

assessment process for LD. A primary benefit to a RTI model as outlined by Fuchs and Fuchs (1997) is the identification of students with learning difficulties early in their academic careers and the remediation of academic deficits early in their careers for both students with disabilities and for students without disabilities who are at risk for failure in school. The screening and assessment process also assists in reducing bias in referrals for special education evaluation, and has direct implications for improved, individualized instructional planning. The authors outline many potential complications with the implementation of RTI. First, the emphasis needs to be placed on the purpose of RTI assessments to rule out contextual variables as reasons for poor academic performance since disabilities are traditionally viewed as a deficit within the individual and no cognitive assessments are necessary for eligibility. Second, different levels of validated interventions are available for different content areas at different ages, and validated assessments and interventions are a necessity in RTI. Guidelines about the intensity level of the interventions and the sufficient number and length of interventions are also not established in the literature at this time. Third, the authors argue that the assessment needs to go beyond being nonresponsive to intervention and into the value and efficacy of the special education program for the individual. Final “potential pitfalls” include (a) the large number of personnel that would need the knowledge and skills in numerous areas (CBM, intervention implementation, knowledge/implementation of problem solving models, eligibility decisions, etc.) and (b) the ambiguous point of entry for parental involvement and due process. Vaughn and Fuchs (2003) presented several obstacles to the implementation of RTI and several benefits that make empirical investigations of the RTI model for assessing LD worthwhile.

Purpose of the Current Study

The current study was designed to examine a number of variables and outcomes of a naturally occurring implementation of a RTI process within a public elementary school. The school had adopted the STEEP model to try to better serve at-risk students in collaboration with a local university. In the context of this implementation, data were available that potentially could shed light on the extent to which varying methods of defining problems suggest the same result, as well as the extent to which the results are stable over time, suggesting enduring problems. Additionally, the data permit examination of the efficacy of the intervention the school adopted in a clinical replication, the prevalence of resistance to intervention in a natural implementation of RTI, and the number of progress monitoring sessions needed to estimate student response to intervention.

Method

The data were collected at an elementary school as a part of the school's educational practices that occur as a result of its direct link with a local university and its efforts to better serve at-risk students. The school is a university affiliated professional development school (PDS), which involves a direct relationship with a university to maximize student achievement and performance (American Association of Colleges for Teacher Education, 2004). This specific school was formed uniquely for the purpose of becoming a PDS with new administrators, staff, and students. As a PDS, there is ongoing training for university students and the opportunity for faculty, students, and parents to work together on various projects at group and individual levels. The PDS distributes information to parents about their link with the university and ongoing activities through newsletters. The university provides the PDS with professional development activities and follow-up consultation activities to promote best practices in the PDS setting. The university is collaboratively involved with the PDS in several aspects, including designing curriculum and serving as on-going sources of support. The active participation of the university faculty and students in programs for students is a routine educational practice at the PDS.

The current study gathered, systematized, and analyzed the data that were collected by the PDS and the university personnel who implemented the project as a field experience course for university students, and the PDS's efforts to better serve at-risk students. The significant amount of information that had been collected as part of the staff training and service effort was not being analyzed or recorded formally and no plans were in place to examine the data. Consent was gained from the university instructor,

PDS associated university faculty, and the PDS school administrator to use the data in a formal evaluation of its effectiveness. The analyses would determine the effectiveness of the activities for identifying students who are at risk for academic failure, and specifically to validate the reading intervention that was implemented.

The school implemented STEEP in the 2002-2003 academic school year. The data were gathered and analyzed in the 2003-2004 academic year. The data were gathered from multiple electronic and paper sources from both the university and PDS. For all data collection, identification codes were used for students that have no meaning outside the study. No record linking student names to these codes was retained once data collection was completed. The data collection and analysis procedures are described below following description of the program implemented by the PDS. The data collection procedures can be more meaningfully understood following description of the program that produced the data.

The study design was screened by the Institutional Review Board (IRB) for Human Research Protection and found to fall within a class of studies that are exempt from IRB oversight.

Setting

The assessment and intervention activities were conducted at the formerly described urban PDS. University personnel collaborate with school initiatives while the supervised university students interact with the faculty and students of the PDS for field experience courses. The PDS was chosen since it implemented STEEP CBM procedures and individualized PPRI prereferral services and retained records of student performance as a part of their routine educational practices and their association with a university. The

selected PDS and its associated university were receptive to opportunities for analyzing the wealth of information they collected from their school initiatives to guide their school improvement process. Also, the association of the school with a local university promoted an atmosphere that was conducive to data collection and data analysis. The school consisted of 100% African-American students, and 85.86% of students who attended this school received a free or reduced price lunch.

Participants in the School's STEEP Process

All students in the second, third, and fourth grade classes at the school participated in intervention and assessment activities. Eleven classrooms were included in the activities. The participating grades were selected by the school faculty based on available resources and prioritized grade levels. One-hundred seventy-seven (177) students participated in the assessment component, and 22 students participated in the intervention. Two (2) students participating in intervention services transferred schools during the intervention and did not participate in the full length of PPRI or final spring STEEP assessment. The number of students included in the analyses of each research question varies because of the requirements for the various analyses. The criteria for inclusion in each data analysis are provided in the description of the statistics used for each research question.

Assessor, Intervention Provider, and Teacher Training

Graduate and undergraduate students enrolled in coursework at the affiliated local university acted as assessors and intervention providers for the implementation of STEEP at the PDS. The university students were selected by university personnel based on their grade point average and experience to ensure the quality of the course participants. Also,

several employees of the school district who expressed interest in learning about STEEP and PPRI activities enrolled in the course. These university students were enrolled in a course that included training in the areas of assessment, intervention, and field experiences in the implemented STEEP and PPRI activities. Eleven university students were enrolled in the course, seven (7) of which were also employed by Pupil Appraisal Services in the district.

The university students were provided classroom training in CBM and academic interventions, as well as STEEP and PPRI procedures. Background information was provided about CBM and academic interventions, and university students were required to read information in the areas. A list of the required readings is provided in Appendix A.

STEOP training occurred first for the university students who provided the assessment at the PDS. A “Tell, Show, Do” model of training was used to instruct the university students how to administer CBM to students according to STEOP procedures using student materials. That is, the instructor first described the STEOP procedures to the university students and provided them with the STEOP manual. Samples of information and training materials from the STEOP manual and student samples are provided in the Appendixes B-G. The instructor then modeled the implementation of STEOP for the university students. Finally, the university students practiced administering STEOP to each other. The class training time for CBM and STEOP training totaled 12 hours. The university students were then responsible for the administration of STEOP for the winter assessment. The instructor of the university class and additional trained persons monitored the university students as they administered the assessment

using the steps provided in the integrity checklists provided in Appendixes H and I. University students also scored and recorded the data.

The classroom teachers were trained to administer the spring assessment through in-service trainings. The teachers were trained in approximately 4 hours using the above “Tell, Show, Do” model of training, and the classroom teachers administered the second assessment with the instructor and experimenters monitoring and assisting as needed using the steps in the integrity checklists as a guide.

University students were also trained to implement the Progressive Practice Reading Intervention (PPRI) using the same method of instruction. The intervention procedures were administered to the university students and were described by the instructor. The PPRI procedures that were administered are provided in Appendix J. The instructor then modeled the PPRI procedures for the university students. The university students then practiced administering the PPRI to each other. The training sessions for PPRI totaled 6 hours. The university students subsequently were responsible for administering the intervention to the identified elementary school students at the PDS. The university students were closely monitored during their interactions with the students and required to record their activities as well as the student’s performance and progress. They were also required to record personal notes about their interactions with the PDS students.

Measures

The school used CBM to assess every student in second, third, and fourth grades in the areas of reading and mathematics. The CBM procedures were used to identify students who were performing poorly in core academic subjects and to develop

hypotheses as to whether students exhibited a performance or a skill deficit. These measures were developed by Witt (1996) and have been referred to as STEEP. Following identification of students with hypothesized skill deficits in the area of reading, an intensive instructional intervention was implemented. The university students and teachers at the school were trained to administer the assessment procedures, and the university students were trained to administer the intervention. Students' performance on the assessment and intervention components of the study are compared to student performance on the LEAP test and to nominations by teachers for referral and retention. Following is a detailed description of each of the above measures.

Teacher Referral

Teacher referral is a primary factor and a screening tool used in the school systems to identify children who need special education (Algozzine et al., 1983). Teachers are asked to refer the students who are perceived as the lowest performing to a multidisciplinary team for assistance in implementing appropriate accommodations and interventions for the student to improve their educational performance. Teacher ratings of the student performance have been found to correlate with more direct psycho-educational assessments (Gresham & Witt, 1997). Teachers were asked to nominate students who they were likely to refer to a School Building Level Committee (SBLC) within the next month based on their perceptions of the students' academic performance prior to the CBM assessments. The SBLC is a multidisciplinary team of professionals who meet to determine recommendations for students exhibiting difficulty in school and is the first step in the process of referring a student for special education services.

Screening to Enhance Educational Performance

STEEP is a pre-referral screening model that includes classroom-wide academic assessments and assessments of performance and skill deficits using CBM (Witt, 1996).

Classroom-wide Academic Assessment. The first step in the screening process is to administer curriculum-based measures to each student in the areas of reading and mathematics. The procedures for math and reading CBM administration and scoring are included in Appendices B and D. The reading assessment consists of an oral reading fluency measure and is an individually-administered assessment. The administrator allows each student to read a passage, which is controlled for readability level, for one minute. The number of words read correctly in one minute is the child's score in the area of reading.

CBM reading in the form of oral reading fluency has documented reliability and validity that supports its use (Marston, 1989). Oral reading fluency measures have been found to be highly correlated with teachers' judgments, generally accepted norm-referenced tests of reading ability, and basal reading series (in general, correlations above .80). Oral reading fluency has also been demonstrated to reliably differentiate between distinguished intact groups of readers and to validly measure sensitivity to growth. Reliability estimates for test-retest (most correlations above .90, range .82 to .97), parallel forms (most correlations above .90, range .84 to .96), and interrater reliability (.99) are also all excellent.

The mathematics assessment is group-administered. A predetermined mathematics skill (i.e. subtraction from 18, addition with regrouping) that was identified by the teachers at each grade level as appropriate was administered at each grade level.

The mathematics worksheet consisted of problems using the predetermined skill presented in a randomized order. The classrooms were given two minutes to complete as much of the worksheet as possible, and their score was the number of digits the children correctly produced. CBM in the area of mathematics (fluency measure of digits correct) has demonstrated content validity (Shinn & Marston, 1985). Test-retest (.78 to .93), parallel form (.48 to .72), and interrater (.90 to .99) reliability for single administration are also acceptable (Marston, 1989).

STEEP outcome measures for each student included fluency scores and class ranks. Students are ranked according to their score for each skill assessed. The outcomes of the STEEP assessment were organized by class to rank each student according to their score in each area assessed.

Motivational Assessment. The children who scored in the bottom 16% of their class for each skill were then administered a motivational assessment, otherwise referred to as a “Can’t Do/Won’t Do” assessment (Witt et al., 2000). Children were offered an incentive to improve their score from the class-wide assessment. The incentive was their choice of a prize from a treasure chest that included items such as pencils, erasers, stickers, and small toys. Scripted instructions for administration of the motivational assessment in math and reading are located in Appendices F and G. The children who improved their score to reach at least the classroom median were considered to have a motivational problem (“won’t do problem”). The children who did not improve their score or improved their score only marginally were considered to have a skill deficit (“can’t do problem”). The students with the two lowest scores (either on the class-wide assessment or the motivational assessment) in each class were nominated to receive

individualized intervention. Previous research based on this general model suggests its utility in identifying effective intervention components for students referred for reading difficulties (Gansle, Noell, & Freeland, 2002; Noell, Freeland, Witt, & Gansle, 2001)

Progressive Practice Reading Intervention

The Progressive Practice Reading Intervention (PPRI) is a structured intervention in which children progress through a series of increasingly difficult passages based on their fluency rate on those passages while generalization probes are administered to monitor their overall progress. There are 12 lessons for each grade level of reading. Each lesson consists of three passages with overlapping content. The intervention includes listening passage preview (the passage is read to the child) and oral passage preview (the student reads the passage while receiving corrective feedback) for all of the reading passages in each successive lesson. Following each of the 3 passages in the lesson, multiple choice word identification problems and comprehension questions are completed. The final reading passage is also used to measure the student's fluency. The fluency score determines whether the student repeats the lesson or progresses to the next lesson during the next session. Generalization probes are also administered to monitor the child's progress on non-rehearsed passages. PPRI instructions are located in Appendix J. Previous research using single subject methodology has supported its effectiveness (Noell et al, 1998; Noell et al., 2001), and its components have also been empirically supported (Daly, Lentz, & Boyer, 1996).

The Louisiana Educational Assessment Program for the 21st Century (LEAP)

The Louisiana Educational Assessment Program for the 21st Century (LEAP) is a criterion-referenced test that assesses English Language Arts (ELA), Mathematics,

Science, and Social Studies. LEAP is administered to fourth grade students. The test was developed and aligned with the state's benchmarks and content standards for the fourth grade. Students receive one of five achievement ratings (Advanced, Proficient, Basic, Approaching Basic, and Unsatisfactory). Students' performance on the test is socially significant since fourth graders must score on the Approaching Basic Level or better in the areas of English Language Arts (which includes writing, reading and related skills) and Mathematics to be promoted to the fifth grade. The English Language Arts section has a mean Cronbach's alpha of .85, and the Mathematics section has a mean Cronbach's alpha of .92. Both of these scores fall within the excellent range (Data Recognition Corporation, 2003).

Teacher Nominations for Retention

At the completion of the school year, teachers nominate students for retention through the School Building Level Committee. A list of students nominated for retention in grades three and four was obtained for comparison to results of the other measures obtained.

Procedures

Assessment Procedures

Teachers were asked to nominate those students that they believed they would refer to their school-building level committee within the next month. The first STEEP assessment was then administered and scored by the university students under the supervision of the faculty member of the course they were enrolled in. The math assessment was administered to the class as a group first, and then each student read individually with a university student for one minute. The students were ranked in their

class on both their math and reading fluency scores. Once the data were aggregated, the bottom 16% in each class for each skill area was identified. The bottom 16% represents the percentage of the normal population who would score below one standard deviation below the mean. On the same day as the initial assessment, the low performing students were given parallel forms of the probes administered to the class for the motivational assessment. In the spring, the PDS students were administered different forms of assessment tools that were used during the initial assessment appropriate for their current grade level. The same procedures were followed (group administration of math, individual administration of reading, and class rankings), and the students' teacher administered the assessment. The teacher also scored the materials and followed up with a "Can't Do/Won't Do" assessment for the bottom 16% of the students in each skill area for their class.

Intervention Procedures

The students with the lowest fluency scores (either on the classwide assessment or on the motivational assessment) were then targeted for intervention. The students participated in eight (8) weeks of intervention. Each individualized intervention session lasted approximately 20-25 minutes, and the sessions were scheduled to occur three times a week. The passages presented during the instructional sessions were on the students' instructional level based upon the instructional placement recommendations provided by Deno and Mirkin (1977). Instructional level was determined prior to intervention by sequentially administering lower level passages until the student's fluency score fell within an instructional range at that grade level (Deno & Mirkin, 1977). The students' progress was monitored using fluency scores obtained during intervention sessions and

during progress monitoring on generalization passages. The generalization passages are unrehearsed passages presented on the same readability level. If a student proceeded through the 12 lessons on their instructional level with appropriate fluency scores, the materials in which they were being instructed were changed to the next higher grade level.

Data Collection for this Study: Retrieval of Archival Records

Teacher Referrals

Each teacher submitted the list of students they were considering for referral to the multidisciplinary committee to his or her administrator. The administrator retained the original lists and provided a copy of the information to the university instructor for purposes of coordinating services. The information was collected for this study from the university instructor.

STEEP

The student probes were generated from the winter and spring assessments. Twelve weeks occurred between the initial and final assessment. Following both winter and spring STEEP assessments, the teachers provided the PDS students' math and reading probes to the university instructor. The university students were responsible for scoring the reading and math probes. Once the probes were scored, each student's classroom rank was determined for each skill. The information was used to identify the students who participated in the motivational assessment, and the fluency scores for the motivational assessment were kept by the university instructor. The data were returned to the teachers in a graphical format that indicated the students' scores on the classwide assessment and the motivational assessment for their class. The university retained the

original data in a collection of Microsoft Excel ® databases. These were gathered and compiled with the other data to create the multivariate database analyzed in this study. The original math probes were collected from the university to conduct reliability checks.

PPRI

Each student who participated in PPRI had a folder containing all of his or her intervention materials. The folders were located centrally at the school where the university instructor, the university students, and the teachers had access to their students' information. Following each intervention session, the PDS students' updated information was recorded in their folder on a summary sheet. This allowed for each student's progress to be readily monitored by his or her teacher. The data used in this study were gathered from this repository.

LEAP

Each student in the fourth grade participated in the LEAP test, and their results were provided to their school administrator and teacher. A university instructor had been provided the LEAP information and the LEAP scores were obtained from the instructor for the purposes of this study.

Teacher Nominations for Retention

The administration at the school maintains records of students referred to the multidisciplinary committee at the end of the school year for consideration for retention in their current grade. With the cooperation of the university instructor and the administration, the nominations of students for retention were added to the data set.

Data Analyses

The following research questions were examined in the current study using the statistical methodologies described below. The data were analyzed using the SPSS statistical program.

Research question 1. What is the stability of class rankings based upon CBM over several months?: Correlation of Class Rankings

Each second, third, and fourth grade student that was present in school participated in the winter and spring STEEP assessments. Only the students who participated in both the winter and spring assessments were selected for each area (with the exception of PPRI participants). One-hundred forty-one (141) students were included in the reading correlations, and 155 students participated in the math assessment. This is the number of students for whom complete data were available permitting contribution to the correlation analysis. The students were ranked in their class according to their score on each skill assessed. The ranking of each student on the first assessment was compared to the student's ranking on the second assessment for each subject area using a Kendall's *tau b* correlation. Kendall's *tau b* was selected because of its effectiveness for correlating ordinal data and to attempt to replicate previous findings (VanDerHeyden et al., 2003). Additionally, Pearson *r* correlations were conducted for the fluency scores from the winter to the spring in each area, and the means and standard deviations of the fluency scores will also be reported. Pearson *r* correlations are the standard for use with interval scale data (Gravetter & Wallnau, 1996). For the reading assessments, those students who participated in the PPRI were excluded from this analysis.

Research question 2. What is the concordance between differing methods of defining academic deficits?: Chi-square Analyses of Assessment Outcomes

Several screening tools were assessed during this project. The nomination of a student to have a significant problem according to the STEEP and motivational assessments was compared to the nomination of a student to have a significant problem based upon teacher referral, teacher nomination for retention, and LEAP scores in the areas of English Language Arts (ELA) and mathematics. Chi-Square analyses were run to determine the concordance of STEEP and motivational assessments with referral for special education evaluation and retention as well as performance on a criterion-referenced test (LEAP). Chi-square analyses were chosen since they allow a comparison of the similarities between the given problem nomination techniques while analyzing nominal data (Gravetter & Wallnau, 1996). A 2 x 2 classification table was constructed for STEEP nominated problem or no problem in comparison to the nomination of a problem according to the other tools (teacher nomination for SBLC, retention nomination, and LEAP performance in ELA and mathematics). The number of students included in each chi-square analysis varies due to the criteria for inclusion based on available data with each of the problem nomination techniques as explained below. Students scoring below the Approaching Basic Level on the LEAP were defined as having a significant problem. Only fourth graders who participated in LEAP testing were included in this analysis. STEEP performance in the bottom 16 percent of the class (without improvement to the class median score and in the standard instructional range following the motivational assessment) was defined as having a significant problem. The students below the 16th percentile for their class were chosen as this is the percentage of the normally distributed population who score greater than one standard deviation below

the mean. Students who participated in the winter STEEP assessment in the respective skill areas were included. If a student scored below the 16th percentile for a skill and there were no data from a motivational assessment, they were excluded from the analysis. Also, students nominated for referral or retention were defined as having a significant problem.

Research question 3. Which Students were Responsive to Intervention? A comparison of PPRI participants and their grade-level peers.

School-wide STEEP assessments occurred in the winter and the spring for students in grades 2, 3, and 4. The students who participated in both the winter and spring assessments for each grade level were included in this analysis. For those students who did not participate in the intervention, the winter and spring fluency scores were graphed on a scatterplot and the regression line was determined and graphed (n=46 for 2nd grade; n=30 for 3rd grade; & n=65 for 4th grade). The standard error of the regression line was calculated. Lines were then graphed that accounted for the area in which 68% of scores were likely to fall based on the standard error of the regression line. The regression line was used since it serves to provide the best prediction of a posttest score given the pretest score (Gravetter & Wallnau, 1996). The standard error lines which account for 68% of the distribution of scores were chosen to be representative of students who fell within one standard deviation of their peers.

The data points of the PPRI participants were graphed and then analyzed to determine who fell above the standard error of the regression line. Five (5) PPRI participants in the second and third grades and 8 PPRI participants in the fourth grade were included in the comparison. The students whose scores fell above the standard error line were determined to be responders to the intervention. The students whose scores fell

within the standard error of the regression line and below were determined to be nonresponders to the intervention. If a child is improving his performance at a greater rate than his average student, then he could theoretically “catch up” with his peers; therefore, he is not resistant to intervention according to this analysis. However, if a child is receiving additional individualized instructional time and continues to improve his performance at the same rate as his peers or at a lesser rate, then that student is demonstrating resistance to the given intervention according to this analysis.

Measurement of the trend of a student’s performance compared to his or her local peers is an important component in the dual-discrepancy model that is used to determine whether or not a student has a disability (Fuchs & Fuchs, 1998).

Research Question 4. Did the PPRI remain effective in this clinical replication?: Analysis of Regression Discontinuity

Students were selected for intervention based on their performance on the initial STEEP assessment. Since the assignment to the treatment group was systematic and nonrandom, the Regression Discontinuity design will be used to evaluate the treatment effects (Shadish, Cook, & Campbell, 2002). The Regression Discontinuity design allows for the evaluation of treatment effects for participants selected based upon need while providing unbiased causal estimates (Shadish et al., 2002). The design allows for assignment to a treatment group based upon a cutoff score and analyzes the difference of the regression of pre-post scores between the treatment and control groups. A cutoff score was selected that accounts for the majority of the bottom 16% in each class across the grade levels. Students who participated in the intervention and had a pretest (winter reading assessment) score of 41 words read correctly per minute or less were included in the PPRI participant group. Students who did not participate in the PPRI and had a pretest

score greater than 41 were also included in the evaluation. The PPRI participant group consisted of 17 students. The comparison group of students who did not participate in the intervention consisted of 121 students. The scores for the students receiving intervention were compared with those students who did not receive intervention services on their initial and final STEEP assessment scores following the regression discontinuity model.

Research Question 5. How many intervention sessions are necessary to reliably predict the treatment outcome?: Correlation of regression estimates

During each instructional session, the university students concluded the session with a timed reading sample that yielded a fluency rate on the reviewed material (treatment fluency score). Approximately two treatment sessions occurred each week.

The regression line for the students' progress was based on the treatment fluency scores from each of the sessions to provide the best estimate of performance for any given session (Gravetter & Wallnau, 1996). This regression line following 11 sessions was used as the standard for each student's treatment outcome. The number of sessions was selected to maximize the number of participants that could be used in the analysis. Students often received varying numbers of treatment sessions due to missing intervention sessions because of student absences, school holidays, and similar factors. Seventeen (17) students had data for a minimum of 11 sessions and were included in the analysis. Regression estimates and the slope of these lines were then calculated beginning with three treatment fluency scores and recalculated for each additional treatment fluency score from additional sessions. Each of the regression estimates from three sessions and more for treatment fluency scores were then correlated to the treatment outcome line of regression to determine the strength and significance of the linear relationship following each session (Gravetter & Wallnau, 1996). The data were then analyzed to determine at

which session in the treatment the students' long-term outcome could be reliably predicted.

Reliability of Measures and Procedural Integrity

Reliability of Measures

Two independent scorers scored the math probes. Thirty-six percent of the probes were randomly selected and scored. The interobserver agreement (IOA) was calculated as the number of agreements divided by the number of agreements plus the number of disagreements for each case. The total estimated inter-observer reliability is the average IOA score across all cases.

Procedural Integrity of School Based Procedures

Scripted instructions were provided for the administration of the in-class assessments and the "Can't Do/Won't Do" assessments. Each university student was required to demonstrate the process with 100% accuracy prior to administering the STEEP assessment and the PPRI. University instructors monitored the administration and corrected any deviations from the script during the assessment and the intervention procedures. Furthermore, the university students were required to note each step of the intervention on a tracking chart. Frequent checks of the integrity of implementation were integral to the university's and school's efforts to assure that university students had mastered the targeted skills and to assure that the services were being provided to students as designed.

For purposes of the current study, each step of the intervention sessions was scored as administered correctly or incorrectly. The procedural integrity for each intervention session was formally calculated by dividing the number of correct steps by

the number of correct steps plus the number of incorrect steps. The mean procedural integrity was estimated by calculating the average procedural integrity score of each of the intervention sessions. Procedural integrity checks were conducted for 37.4% of randomly selected intervention sessions based on a review of the intervention logs. The procedural integrity checklist that was used is provided in Appendix K.

Results

The collected data included information about student performance based upon teacher judgment, STEEP assessment, responsiveness to intervention, and criterion-reference testing. While all students were assessed using CBM in the areas of reading and math and were able to be considered for teacher referrals and nominations, only fourth graders participated in the criterion referenced assessment and only the lowest performing students participated in motivational assessments and intervention services. The students with the lowest reading fluency scores also received PPRI to help remediate their reading deficits and to assess their responsiveness to PPRI intervention services when determining the severity of their learning problem.

Outcomes of Data Analyses

Research question 1. What is the stability of class rankings based upon CBM over several months?: Correlation of Class Rankings

Correlations of student class rankings from the winter to spring assessment were used to establish the stability of CBM within the STEEP procedures. The students who received intervention services were excluded from the analyses of the reading correlations since they received significantly different instruction between the initial and final assessments. The class rankings from the winter and spring assessments were significant for both reading and math. One-hundred forty-one (141) students participated in the winter and spring reading assessments (and did not participate in PPRI), and the Kendall's tau b correlation for their class ranking was 0.63. The Kendall's tau b correlation for the class rankings on CBM math scores was 0.46 (n=155). The correlation for the class rankings in reading and math are both significant at a 0.01 level ($p < 0.01$ for both reading and math). Descriptively, the scores improved over time. The average

winter reading score for students not receiving intervention services was 73.97 (sd=26.61) and the average reading score for the nonparticipants in the spring was 82.16 (sd=27.59). The average math fluency score for a two-minute assessment was 12.05 (sd=7.96) in the winter assessment and was 19.23 (sd=10.12) in the spring assessment.

Additionally, a Pearson's r correlation was conducted to compare the winter reading fluency scores of students who did not receive intervention services to their spring assessments, and to compare the students' winter and spring math fluency scores. The correlation was 0.78 (n=141) for reading fluency scores and 0.60 (n=155) for math fluency scores, which are both significant at a 0.01 level.

Research question 2. What is the concordance between differing methods of defining academic deficits?: Chi-square Analyses of Assessment Outcomes

Chi-square analyses were conducted to determine if there is a relationship between the problem identification method of STEEP and other problem identification methods. The students who scored in the bottom 16% of their class and did not improve significantly with intervention were nominated as having a problem based upon STEEP. A significant chi-square value allows a rejection of the null hypothesis that there is no relationship and indicates a concordance between the problem identification methods. Each of the Chi-square analyses indicated a concordance between STEEP nominated problems and other identification methods (LEAP scores, teacher referrals, teacher retention nominations). The number of students included in each analysis varied according to the criteria previously described for each problem nomination area included.

STEER Reading Fluency versus LEAP 21 ELA scores. The students nominated as having a significant reading problem based upon the STEEP results were compared with students who performed unsatisfactorily on the ELA portion of the LEAP exam. This

analysis included fourth graders only because they are the only elementary students who participate in LEAP testing. The relationship between the two problem nomination methods is statistically significant when comparing these nominations in a chi-square analysis, and the Pearson chi-square value is 17.45 ($n=77$; $df=1$, $p<0.01$). Both the LEAP ELA and the STEEP reading fluency assessments agreed that 56 students did not have significant problems and that 8 students did have significant problems. However, ten students did not pass the ELA portion of the LEAP that were not indicated as having a reading problem according to STEEP, and 3 students did pass the ELA portion of the LEAP that did demonstrate a reading problem according to the STEEP assessment.

Table 1
Chi-Square Analyses Comparing STEEP Reading Performance
to LEAP ELA Performance

		LEAP ELA Problem	
		No	Yes
STEPP Reading Problem	No	56	10
	Yes	3	8
		Value	Significance
Pearson Chi-Square		17.45	<0.01
Phi		0.48	

STEPP Math Fluency versus LEAP 21 Math scores. A significant relationship occurred between the nomination of a math problem according to STEEP criteria and LEAP Math criteria. Only fourth graders were administered the LEAP exam and were included in this analysis. The Pearson chi-square value was 4.78 ($df=1$, $p=0.03$), which is significant at a 0.05 level. Of the 75 students included in the analysis, the problem nomination methods agreed on the occurrence of a problem in 7 students and the absence of a problem in 43 students. The problem nomination procedures were not in agreement

for 25 of the students. Three students that were identified with the STEEP procedures scored within an acceptable range on the LEAP Math portion. A greater number of students (22) performed unsatisfactorily on the LEAP Math test and were not nominated as having a difficulty according to the STEEP assessment.

Table 2
Chi-Square Analyses Comparing STEEP Math Performance
to LEAP Math Performance

		LEAP Math Problem	
		No	Yes
STEPP Math Problem	No	43	22
	Yes	3	7
		Value	Significance
Pearson Chi-Square		4.78	0.03
Phi		0.25	

STEPP Problem Identification versus Teacher Referral of Problems. The difference between the students nominated as having either a math or reading problem was compared to the students who were nominated by teachers to be referred to a multidisciplinary team (School Building Level Committee) for consideration of special education placement. A significant relationship was demonstrated between these two problem nomination procedures according to a chi-square analysis. The concordance of the procedures is significant at a 0.01 level with a chi-square value of 26.38 (df=1, $p < 0.01$). Of the 158 students included in the analysis, both methods indicated that 89 students did not have a significant academic problem while they indicated that 29 students did have a significant academic problem. Teachers referred 21 students who did not have a significant academic problem according to STEEP and did not refer 19 students who did have a significant academic problem according to STEEP.

Table 3
Chi-Square Analyses Comparing STEEP Performance to Teacher Referrals

		Teacher Referral to Multidisciplinary Team	
		No	Yes
STEEP Reading or Math Problem	No	89	21
	Yes	19	29
		Value	Significance
Pearson Chi-Square		26.38	<0.01
Phi		0.41	

STEEP Problem Identification versus Teacher Nominations for Retention. At the end of the school year, the teachers nominated students to the administration that they believed needed to be retained in their current grade for academic concerns. The list of nominated students was compared to the students identified as having academic deficits according to the STEEP Model.

Table 4
Chi-Square Analyses Comparing STEEP Performance to Teacher Retention Nominations

		Teacher Nomination for Retention	
		No	Yes
STEEP Reading or Math Problem	No	65	45
	Yes	11	37
		Value	Significance
Pearson Chi-Square		17.52	<0.01
Phi		0.33	

Of the 158 students evaluated, teachers nominated 82 (51.9%) students to be retained. STEEP indicated that 37 of the teacher nominated students had academic deficits and did not indicate that 45 of the teacher nominated students had academic deficits. Both STEEP and teacher retention nominations indicated that 65 students did not have academic

problems, and STEEP indicated that 11 students had academic problems that were not nominated for retention by the teachers. A significant relationship between STEEP nominated problems and teacher nominations for academic problems was confirmed by a significant chi-square value of 17.52 ($df=1, p<0.01$).

Research question 3. Which Students were Responsive to Intervention? A comparison of PPRI participants and their grade-level peers.

Scatterplot graphs were created for fluency scores from the winter and spring assessments (conducted 12 weeks apart from each other) for students who did not participate in PPRI at each grade level. A linear regression line was then determined that best accounted for the variability in the scores. The regression line for the second grade had a slope of 0.75 ($n=46$). The regression line for the third grade had a slope of 1.24 ($n=30$). For fourth graders, the regression line slope was 0.69 ($n=65$). The standard error lines composed around each regression line accounted for 68% of the distribution for each grade level. The graphs of the regression lines are demonstrated below. The data points for the individuals who participated in the PPRI were then graphed on the scatter plots and analyzed visually.

The PPRI participants in the second grade performed well relative to their grade-level peers. Four of the five participants performed better than the estimate of their peers. Three of the students performed at a level greater than one standard deviation above the regression line of their peers. The three students who had the greatest rate of improvement were defined as responders to the intervention. Figure 1 depicts the placement of the PPRI participants in the second grade and their peers, including their regression lines.



Figure 1
 Winter and Spring Reading Fluency Scores for Second Grade Including
 Regression Line with 68% Standard Error Lines for Nonparticipants

The PPRI participants in the third grade all performed the best relative to their grade-level peers. Each of the five participants performed better than the estimate of their peers. Three of the students performed at a level greater than one standard deviation above the regression line of their peers. The three students who had the greatest rate of improvement were defined as responders to the intervention. The placement of the responders to intervention relative to their peers according to their reading fluency scores and respective regression and standard error lines are depicted for third graders in Figure 2 below.

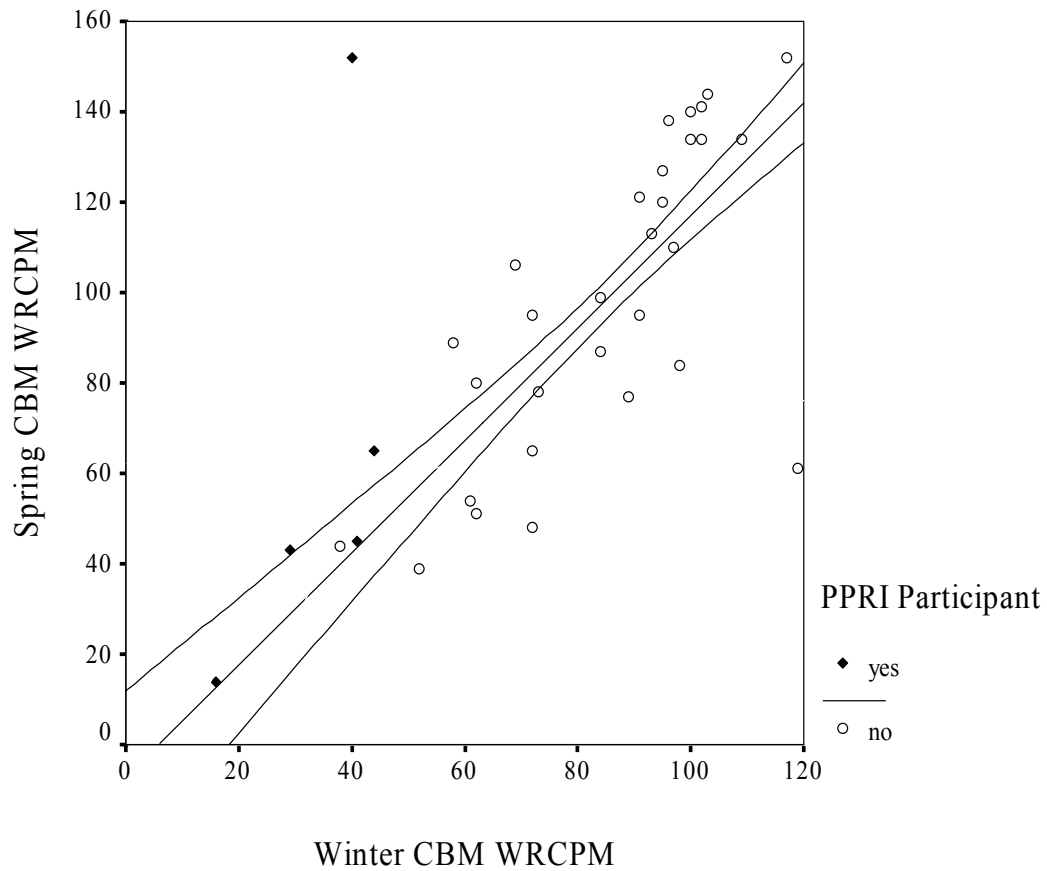


Figure 2
 Winter and Spring Reading Fluency Scores for Third Grade Including
 Regression Line with 68% Standard Error Lines for Nonparticipants

The PPRI participants in the fourth grade had the least amount of improvement when compared to their grade-level peers. Two of the eight participants performed better than the estimate of their peers. One of the students performed at a level greater than one standard deviation above the regression line of their peers. This student was defined as a responder to the intervention. Six students did not demonstrate sufficient progress to meet the regression line of their peers. The graph of fourth graders reading fluency scores, and their regression lines of the students who did not participate in the PPRI is depicted in Figure 3.

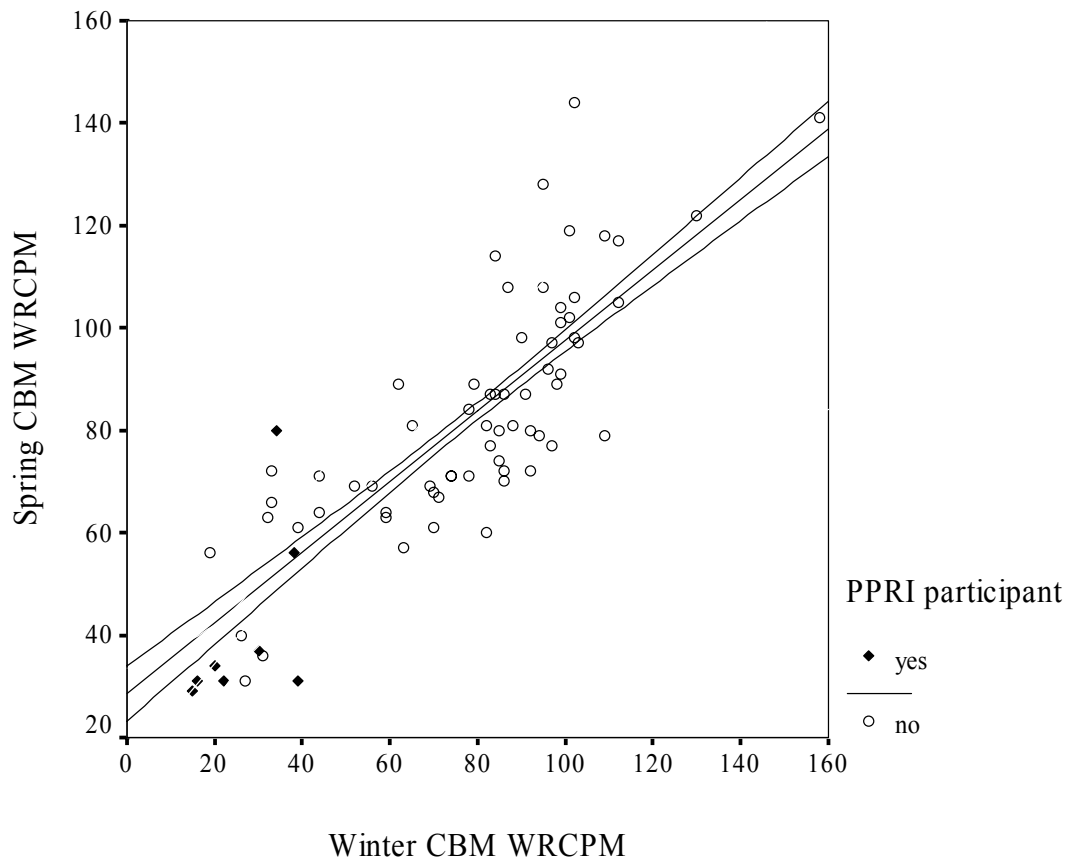


Figure 3
 Winter and Spring Reading Fluency Scores for Fourth Grade Including
 Regression Line with 68% Standard Error Lines for Nonparticipants

Research Question 4. Did the PPRI remain effective in this clinical replication?:
 Analysis of Regression Discontinuity.

A Regression Discontinuity Analysis was conducted to determine if the regression estimate of the winter and spring reading assessments (conducted 12 weeks apart from each other) for the students participating in PPRI differed from the regression estimate of the reading assessment for students who did not participate in the intervention if the regression estimates differ at the point of the cutoff score. The slope of the regression line for the PPRI participants was 2.07 (n=17). For the students who did not participate in PPRI, the slope of the regression line of their scores was 0.87 (n=121).

A significant difference was demonstrated between the two groups between the slope of their regression lines and at the point of the cutoff score. The interaction F value was equal to 4.53 ($df=1, p=0.03$), which is significant at the 0.05 level. The significance between the two groups is determined by an interaction variable that is derived from the intervention participation variable multiplied by the posttest scores and the pretest scores (with the cutoff score subtracted from the pretest score). The significant interaction is indicative of significantly different slope for the two groups. The F value for the difference between the groups was 43.36 ($df=1, p<0.01$). This score is indicative of the significance of the difference of the regression lines of the two groups at the point of the cutoff score where the two groups are theoretically the most similar. Depicted in Figure 4 is a visual representation of the difference between the regression lines of the PPRI participants and nonparticipants.

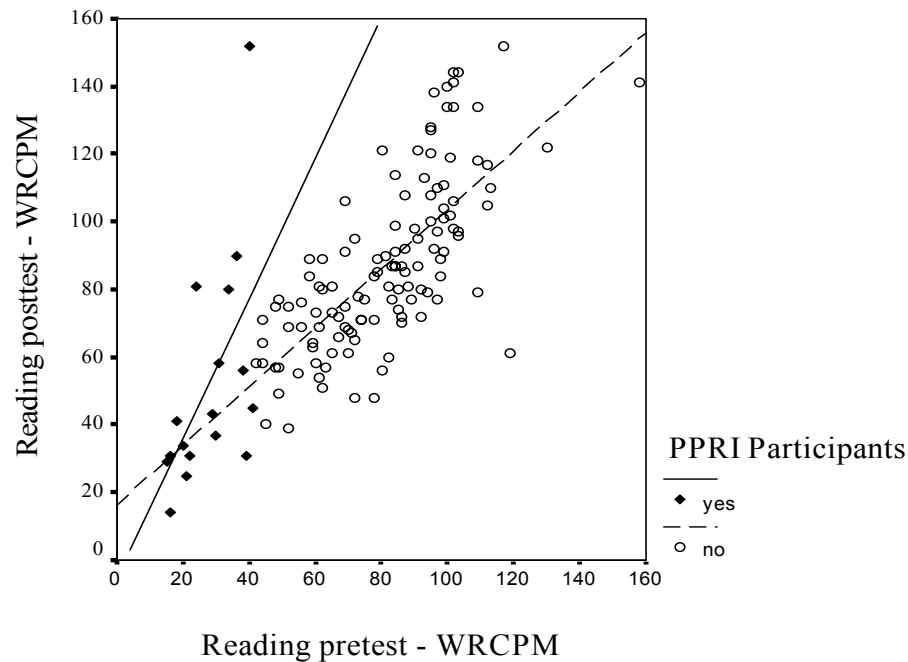


Figure 4
Regression Estimates for PPRI Non-participants and
PPRI Participants for Regression Discontinuity Analysis

The difference between the two groups is significant because of the difference of 28.3 in the intercept values (participants: $B=79.24$, standard error= 12.35 ; nonparticipants: $B=50.93$, standard error= 3.77) at the point of the cutoff score. The intercept values are representative of the point at which the regression line crosses the cutoff score and were obtained by subtracting the cutoff score from the winter assessment score to make zero the cutoff point between the two groups.

Research Question 5. How many intervention sessions are necessary to reliably predict the treatment outcome?: Correlation of regression estimates

The slope of the regression line of treatment fluency scores for 11 sessions was the standard to which other regression lines were compared. The treatment fluency scores of 17 intervention participants were used in the analysis. The Pearson r correlation comparing the slope of the regression lines for the treatment fluency scores to the slope of the standard regression line was not significant for the first 6 sessions. The correlations ranged from 0.27 to 0.475 ($p=0.05$ to $p=0.30$). Refer to Table 9 for the specific Pearson r correlations and their significance for the slope of each successive session.

Table 5
Correlation of Regression Slopes to Slope Standard

	Intervention Session Number							
	3	4	5	6	7	8	9	10
Pearson Correlation	0.27	0.47	0.45	0.44	0.61	0.84	0.82	0.86
Significance Level	0.30	0.05	0.07	0.08	0.01	<0.01	<0.01	<0.01

Beginning with session 7, the slopes of the regression lines were significantly correlated with the standard. The Pearson r correlation value 0.61 was significant for the slope of the regression line following 7 intervention sessions ($p<0.01$). The correlation of

the slope values continued to be significant ($p < 0.01$) for each successive session and reached a maximum Pearson r value of 0.86.

Reliability of Measures and Procedural Integrity

Reliability of Measures

For the math winter and spring assessments, 36% (118 of 328) of the math probes were scored for interrater reliability. The math probes were selected by randomly choosing 6 numbers between 1 and 20. Each class folder for the winter and spring assessments was then collected and the math probes were selected that fell in the sequential order of the selected numbers was scored by an independent examiner. This allowed a sampling from each of the classes and examiners that administered the CBM probes. The average interrater reliability was 99.7% (81.8% to 100%). Only four of the selected probes had an interrater reliability of less than 100%.

Procedural Integrity

A total of 187 PPRI sessions were included in analyses during the current study, and 37.4% (70) were randomly selected to conduct procedural integrity checks. The random selection was similar to the reliability measures. A random set of numbers was selected, and the sessions that fell in the sequential order of the selected numbers for each PPRI participant were selected. This process allowed sessions for each student to be examined. The mean procedural integrity was 99% (77.8% to 100%). Only five of the selected sessions were conducted with a procedural integrity score of less than 100%.

Discussion

The current study has demonstrated support through innovative analyses for many activities conducted during classification procedures from prereferral assessment to intervention services. The progress of students through the activities, which increase in intensity, follows a logical and data-based sequence to determine if a student exhibits the characteristics of a disability. The activities allow for comparison of student performance to their school-wide, grade-level, and classroom peers, as well as noting their progress compared to their own previous performance. At the end of the process, multiple assessment outcomes contribute to available information that can be considered when determining the eligibility of a student to receive special education services. Support was demonstrated for several activities that can be used by the school systems when evaluating students for a disability to identify or reject alternative explanations for academic failure.

Multiple methods for the identification of students at risk for academic failure are implemented within education; however, how they are implemented varies (VanDerHeyden et al., 2003). Teacher referral of students perceived as having significant difficulties learning in the classroom to multidisciplinary teams is the method that is most broadly employed in educational settings. The process of the teacher being responsible for identifying problems has been referred to as the “teacher as the imperfect test” (Gerber & Semmel, 1984). While some research supports the use of teacher referral as an efficient method to identify students at risk for academic failure and learning problems (Gresham et al., 1997; Gresham et al., 1987), the need for objective data to guide and support teacher decisions is critical for the process to be unbiased. The STEEP process

was demonstrated to have a substantial and statistically significant overlap with teacher referral, but the overlap was incomplete. Of particular note for future research is the finding that teachers were generally nominating a higher proportion of students as exhibiting problems than the upward bound of 16% identified students employed within STEEP. This remains a fertile area for future research examining the prevalence of academic concerns and how this prevalence may be moderated by the identification process and contextual factors such as poverty.

The STEEP process remains promising based on the incorporation of objective data that are proximal to the concern in identifying students in need of assistance. The STEEP problem identification methods in reading and math assessment were also demonstrated to be in concordance with LEAP scores in the areas of ELA and math to a statistically significant degree. The relationship of STEEP to the socially significant, high-stakes test of LEAP lends support for its usefulness to guiding instructional planning and teacher referrals. However, a substantial degree of disagreement occurred. This appears primarily to result from differing base rates between the methods. The STEEP was limited to identifying 12 students at most (i.e., 16% of 75) at the fourth grade level. In contrast, the LEAP identified 18 students in language arts and 29 in mathematics. Again, this suggests lines of additional research and consideration of the possibility that the percentile used within STEEP may need to be adjusted in high-poverty schools. With a higher proportion of students being permitted to be identified, STEEP may be a practical means for teachers to identify students who are at high risk for failing this high stakes tests such as LEAP. Furthermore, the STEEP and CBM data can then be used by teacher, administrators, and multidisciplinary teams to assist in defining the problem and

identifying appropriate accommodations, interventions, and instructional planning methods to insure the educational experience is maximized for the students.

The significant relationship that was demonstrated between STEEP and teacher nominations for retention suggests that this process may be useful in identifying students at high risk for retention. However, the same issues of base rate that are discussed above arise in a slightly different fashion. If there was no overlap between reading and mathematics problems and no motivational deficits the maximum at-risk rate for STEEP in this sample would be 50 students ($2 \times 16\% \times 158$ students, see Table 4). However, teachers recommended 82 students for retention, precluding precise concordance.

Summarizing across the concordance assessments, it is noteworthy that STEEP problem identification procedures demonstrated a significant relationship with each of the problem nomination techniques to which they were compared since there is a significant difference in the percentage or number of students allowed to be nominated by each technique. At most, STEEP problem identification procedures would allow 16% of the student population to be identified as having a significant concern in each area assessed. However, there is no theoretical limit on the number of students that can be referred, nominated for retention, or fail the LEAP exam. For example, over approximately 50% of the population was nominated for grade-level retention by teachers. The basic difference in the numbers of students available for problem nomination, in part, provides an explanation for the numbers of students that were identified by other problem nomination techniques and not STEEP. The most critical is the number of students who were not identified by STEEP and were identified by other methods. STEEP did not identify 10 students who failed the LEAP ELA, 22 students who failed the LEAP Math, 21 students

referred by teachers, and 45 students nominated for retention. Even with the limit on the number of students who can be identified with STEEP, other problem nomination techniques did not indicate problems with some of the STEEP identified students. For example, teachers did not indicate a significant concern through a referral for 19 students who performed poorly according to STEEP.

Each of the problem nomination techniques has social significance for students who are nominated and as a result these disagreements can be critical. Despite the absence of a standard problem nomination technique, educational systems cannot afford to overlook students with significant academic deficits or to invest significant extra resources for students who would be performing well without them. Therefore, it is critical to consider a plethora of information when determining the instructional supports needed to support student achievement. The current study lends some support for the use of STEEP as objective data to supplement and guide educational decisions since a significant relationship with each of the traditionally employed techniques was demonstrated. However, additional study examining the nature and meaning of disagreements between methods is clearly needed.

The STEEP prereferral screening procedure conducted in the current study is based on CBM principles and procedures that have been well established in the literature (Deno, 2003; Marston, 1989). The stability of the measures over time demonstrated in the current study supplements previous literature documenting the validity of CBM (Marston, 1989; VanDerHeyden et al., 2003). The efficiency of CBM and ease of implementation procedures recommend it as an easy tool for educators. There is strength in using the same unit of measurement for prereferral screening, motivational

assessments, and intervention progress monitoring. The frequent use of CBM allows for multiple comparisons of at-risk students to their peers and their own changes in performance throughout the school year.

Arguments for moving beyond traditional assessment and toward evaluating students' responsiveness to intervention are developing in the literature; however, there are few empirical investigations supporting the use of RTI (Gresham, 2001; Speece et al., 2003; Vaughn & Fuchs, 2003). For those students who respond to intervention, the resistance theory put forth by Gresham (1991) indicates that the problem within the responding students is less severe than those students who do not respond to intervention. The evaluation of student performance relative to their grade-level peers was analyzed in an effective and unique way to rule out potential learning problems or disabilities with students in the current study. While students were placed in PPRI based on the level of their winter reading fluency score relative to their peers, the response of the students was evaluated based on relative improvement in their fluency scores. By comparing PPRI participating students' individual performance to that of their peers based on a regression line from pretest and posttest scores, it minimized the number of assessments necessary to make this comparison and allowed for comparison of performances in both level and trend of student performance. Through this analysis, many students responded positively to the intervention and according to their learning trends and levels. The data indicated that several students could "catch up" with their peers given continued intervention services. As a result of moving beyond assessment to analyzing students' responsiveness to intervention, significant learning problems and disability eligibility were ruled out for several participants. By identifying students who do not have disabilities earlier in the

classification process, fewer full evaluations are required by the school system. The study provided a unique analysis for use while evaluating RTI when ruling out potential explanations for academic deficits and determining who is appropriate for a special education referral. For several students who were highly responsive, they moved from the low end of the distribution the high end with intervention, the PPRI in the context of STEEP provides strong evidence that these students are **not** disabled in the sense that they do profit from systematic instruction.

Evidence-based and empirically validated interventions are critical when evaluating RTI (Vaughn & Fuchs, 2003). If a student does not respond to an inappropriate and ineffective intervention, then the evaluation is meaningless. Therefore, the overall effectiveness of the reading intervention that was implemented is an important factor when evaluating RTI for students who received intervention services. The current study placed the lowest performing students (based on a cutoff reading fluency score) in an intensive reading intervention. The Regression Discontinuity Design (Shadish et al., 2002) used in this study evaluates students' performance, employing proven statistical analytic techniques, while providing services to students who demonstrated the greatest need for remediation. The Regression Discontinuity Design is an effective, data-based statistical analysis that is underutilized in the educational environment. Since it is important to remediate our lowest performing students' skills and provide empirical evidence of the impact of supports provided, the Regression Discontinuity Design would meet the needs of increasing political and legal requirements of our education system since it allows for nonrandom assignment based on data to students with the greatest need. In the current study, evidence of the effectiveness of PPRI is demonstrated through

Regression Discontinuity Analysis. The outcomes were significant in two ways. The regression estimates of student improvement were significantly different for participants versus nonparticipants. The slopes of the regression line supported that the students receiving PPRI services improved at a significantly greater rate than their peers. The effectiveness of PPRI is further supported by the significant difference in the regression lines for participants and nonparticipants at the point of the cutoff score where the two groups are theoretically the most similar. The demonstrated effectiveness of PPRI as an intervention package is of critical importance. While several components of PPRI have been validated (e.g., Noell et al., 2001; Noell et al., 1998), the results provide evidence of the overall effectiveness of the combination of the components of PPRI and its procedures in a clinical replication. This issue has not been studied previously. The current study provides support for the use of PPRI to remediate reading skills for students with reading fluency deficits.

Along with implementing an effective intervention, the appropriate length of an intervention before determining a student's response is an important factor when evaluating RTI (Gresham, 2001; Vaughn & Fuchs, 2003). However, there is no empirical evidence indicating what the appropriate length for an intervention is, nor is there a specific methodology recommended for making this determination. The current study provided a unique way of determining the appropriate number of sessions that need to be administered in order to reliably predict long-term outcome of the students' responses to the PPRI. The results support that the administration of seven PPRI sessions is sufficient to correlate a student's performance with his or her long-term outcome based on the slope of improvement. Extending the evaluation to an eighth session results in a correlation of

.84, suggesting powerful prediction with little advantage in collecting further data. The results are pivotal in the area and provide evidence of an appropriate length to continue PPRI prior to evaluating a student's RTI. Furthermore, the current study provided a methodology that allows for further investigations into the appropriate length of additional interventions. The study supports the frequent assessments to monitor student performance, which allows for an analysis of the trend of a student's performance as well as the availability of data to correlate short-term performance with long-term performance.

Limitations of the Study

The primary caution when interpreting the results of this study is a result of sequential assessment and intervention activities that increase in intensity as they decrease in the number of students participating in the activities. As a result of decisions based on students' CBM scores, not all students who were nominated as having a problem by the various methods analyzed were evaluated with the more intensive activities (i.e., motivational assessment and PPRI).

All problem nomination techniques have limitations; therefore, the concordance analyses conducted comparing problem nomination techniques did not provide an accurate or definitive criterion measure to which the techniques could be compared. The outcomes simply indicate that there is a relationship between imperfect methods of identifying students as having significant academic deficits. Specific limitations also occurred when analyzing the concordance between problem identification methods. When the teachers referred students to the school building level committee or nominated students for retention, the area of concern was not specified. For example, it was not

known if a teacher was primarily concerned about a student's reading skills, math skills, or other academic skills when evaluating the data. More specific analyses of the concordance of problem nomination procedures could have been conducted if this information had been available. Second, the problem identification method of nominating students for retention specifically may not be representative of the general population because of the high number of retention nominations (approximately 50% of students) that were identified as having a problem warranting retention according to teachers. Finally, the concordance of PPRI responders could not be analyzed through chi square analyses to other problem identification methods because the frequencies were too small in the cells.

In order to effectively address a student's response to intervention, many criteria must be met (Gresham, 2001; Vaughn & Fuchs, 2003). Perhaps one of the most critical criteria is the implementation of an **appropriate** and effective intervention. Although PPRI demonstrated overall effectiveness, it may not have been the most appropriate intervention to remediate their skills. It is important to remember that if a student was a nonresponder to PPRI, he or she may not be a nonresponder to a different intervention. Daly, Witt, Martens, and Dool (1997) recommend a functional assessment of academic concerns to formulate an appropriate intervention. "One size does not fit all" is applicable when deciding if one intervention is appropriate for all students. The specific deficits of students and their reason for occurrence need to be considered when designing an intervention for students. For example, an intervention that is based in oral reading fluency (Daly et al., 1997; Martens, Witt, Daly, & Vollmer, 1999) is not appropriate for

students who do not have the prerequisite phonics and decoding skills. A functional assessment of academic skills was not conducted in this study.

A final concern when analyzing the data is the lack of reading reliability checks. Although the university students and teachers were trained to 100% reliability, reading reliability checks were not conducted during the study. Experimenter performance was monitored during STEEP and intervention procedures, but it was not recorded formally.

Future Directions

The current study has served to supplement existing research (VanDerHeyden et al., 2003), provide new findings, and to lay the groundwork for future research. Further evaluations of the concordance of problem identification methods using chi-square analyses could be conducted. RTI could be included as a problem identification method for comparison if there are a greater number of participants included so sufficient expected cell frequency values can be established. Also, the concordance of problem identification methods based on local and criterion-referenced information to that of problem identification methods based on nationally standardized achievement tests need to be analyzed. The analyses of RTI and nationally standardized tests for comparison could also allow for a criterion reference to which other problem nomination techniques could be compared. The concordance between methods based on different comparison groups and reference points would be interesting to evaluate.

The growing emphasis on RTI as an important component of determining the eligibility of a student for placement in special education and determining the disability classification of students is emphasized in the Fuchs and Fuchs (1997, 1998) RTI model. This model involves evaluating a student's performance in response to accommodations,

modifications, and interventions of less intensity prior to recommending an individualized intervention. The current study only examined the students' responses to PPRI since evaluating the overall effectiveness of PPRI was an important research question in this study. However, additional research is needed to evaluate the students' response to less intensive services (i.e., classroom accommodations, classroom modifications, and small group interventions) that can be more easily continued in the general education setting as ongoing support for student learning. Empirical support for the implementation of RTI is essential to establish prior to recommending full implementation in the educational system.

The significant progress of some students who participated in PPRI is very promising. These effects need to be replicated in future research. The Regression Discontinuity Design was very useful in evaluating the effectiveness of PPRI. Future studies could evaluate a higher number of students allowing for cutoff scores specific to each grade level and separate analyses of the effectiveness of PPRI at each grade level.

Additionally, the applications of PPRI could be analyzed to determine the intensity level and components necessary to ensure student success. A successive application of the components of PPRI on an individual or group level of analysis could be conducted to answer this question. Currently, PPRI takes approximately 20 minutes to administer for each individual. If the necessary components could be identified, the intervention could potentially demonstrate similar strength with abbreviated procedures.

Promising results were demonstrated for the predictability of long-term rates of improvement in a limited number of sessions (7) for PPRI. The significant correlation of the slope of improvement has only been applied to PPRI. The application of the same

methods that correlate the slope of CBM scores can be evaluated for different interventions to determine if the estimate is representative of typical response patterns. Future studies could also evaluate higher numbers of students and a standard that is based on a higher number of intervention sessions. Also, the correlation of scores obtained from progress monitoring sessions that include the administration of unrehearsed passages on a regular basis is also recommended.

Summary

The findings in the current study provide preliminary evidence supporting the use of STEEP and RTI as objective data to supplement and support ongoing activities conducted during classification procedures to increase the accuracy of students referred for special education services. Education is beginning to emphasize the importance of objective data when evaluating student performance. Thus, it is important that the objective data provided by STEEP were demonstrated to be related to other problem identification methods currently used in the school system. The objective data provided by STEEP assist in nominating students for prereferral intervention services. Effective reading interventions are a necessity in the school system.

Summarizing across the main research questions, class rankings based on CBM were stable suggesting stability in who would be identified as at-risk based on STEEP. STEEP was generally concordant with other methods of identifying problems, but identified problems at a lower rate than other methods in high poverty school that implemented these procedures. Responders and non-responders to intervention were readily identified using CBM progress monitoring data and the incidence of students responding favorably was higher in the lower grades. The PPRI remained effective in

this clinical replication. The number of sessions needed to predict outcome for the PPRI is seven sessions for a statistically significant degree of accuracy and eight sessions for an extremely precise prediction.

The methodological contributions of the current study include two relatively unique analyses for determining an appropriate intervention length to predict long-term outcomes, as well as evaluating responsiveness to intervention services are important to the field of education and school psychology. The study also demonstrated the use of Regression Discontinuity Design in the educational setting to allow assignment of services to students demonstrating the greatest need. The current study provides promising findings in the areas of classification activities for at-risk students based on ecologically proximal low inference assessments and RTI. STEEP and similar models may provide a means of identifying and responding to students needs that is flexible, practical, and can respond quickly to emerging student needs. However, substantial additional research is needed to clarify the conditions necessary for a successful application of STEEP or similar processes.

References

- Algozzine, B, Christenson, S., & Ysseldyke, J. E. (1982). Probabilities associated with the referral to placement process. Teacher Education and Special Education, *5*, 19-23.
- Algozzine, B., Ysseldyke, J. E., & Christenson, S. (1983). An analysis of the incidence of special class placement: The masses are burgeoning. The Journal of Special Education, *17*, 141-147.
- American Association of Colleges for Teacher Education (2004). Professional Development Schools. Retrieved March 20, 2004 from www.aacet.org
- Bahr, M. W., Fuchs, D., Stecker, P. M., & Fuchs, L. S. (1991). Are teachers' perceptions of difficult-to-teach students racially biased? School Psychology Review, *20*, 599-608.
- Bateman, B. (1965). Learning disabilities: An overview. Journal of School Psychology, *3*, 1-12.
- Berninger, V.W., & Abbott, R.D. (1994). Redefining learning disabilities: Moving beyond aptitude-achievement discrepancies to failure to respond to validated treatment protocols. In G. Reid Lyon (Ed.), Frames of reference for the assessment of learning disabilities (pp. 163-183). Baltimore, MD: Paul H. Brookes.
- Bocian, K., Beebe, M., MacMillan, D., & Gresham, F.M. (1999). Competing paradigms in learning disabilities classification by schools and variations in the meaning of discrepant achievement. Learning Disabilities Research & Practice, *14*, 1-14.
- Christenson, S., Ysseldyke, J. E., & Algozzine, B. (1982). Institutional constraints and external pressures influencing referral decisions. Psychology in the Schools, *19*, 341-345.
- Crawford, L., Tindal, F. & Stieber, S. (2001). Using oral reading rate to predict student performance on statewide achievement tests. Educational Assessment, *7*, 303-323.
- Daly III, E. J., Lentz, Jr., F. E. & Boyer, J. (1996). The instructional hierarchy: A conceptual model for understanding the effective components of reading interventions. School Psychology Quarterly, *11*, 369-386.
- Daly III, E. J., Witt, J. C., Martens, B. K., Dool, E. J. (1997). A model for conducting a functional analysis of performance problems. School Psychology Review, *26*, 554-574.

- Data Recognition Corporation & Pacific Metrics Corporation (2003). LEAP for the 21st Century: 2002 Operational Technical Report. Submitted to the Louisiana Department of Education, March 2003.
- Deno, S. L. (1985). Curriculum-based measurement: The emerging alternative. Exceptional Children, *52*, 219-232.
- Deno, S. L. (2003). Developments in curriculum-based measurement. Journal of Special Education, *37*, 184-189.
- Deno, S. L., Fuchs, L. S., Marston, D. B. & Shinn J. (2001). Using curriculum-based measurement to establish growth standards for student with learning disabilities. School Psychology Review, *30*, 507-524.
- Deno, S. L., Marston, D., & Mirkin, P. K. (1982). Valid measurement procedures for continuous evaluation of written expression. Exceptional Children, *48*, 368-371.
- Deno, S. L. & Mirkin, P. K. (1977). Data-based program modification: A manual. Reston, VA: Council for Exceptional Children.
- Elliott, S. N. & Fuchs, L. S. (1997). The utility of curriculum-based measurement and performance assessment as alternatives to traditional intelligence and achievement tests. School Psychology Review, *26*, 224-233.
- Fuchs, L., & Fuchs, D. (1997). Use of curriculum-based measurement in identifying students with disabilities. Focus on Exceptional Children, *30*, 1-16.
- 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., Fuchs, D., & Hamlett, C. (1989). Effects of alternative goal structures within curriculum-based measurement. Exceptional Children, *55*, 429-438.
- Fuchs, L. S., Fuchs, D. & Maxwell, L. (1988). The validity of informal reading comprehension measures. Remedial & Special Education, *9*, 20-28.
- Fuchs, L., Fuchs, D., & Speece, D. (2002). Treatment Validity as a unifying construct for identifying learning disabilities. Learning Disabilities Quarterly, *25*, 33-45.
- Gansle, K. A., Noell, G. H., & Freeland, J. T. (2002). Can't Jane read or won't Jane read? An analysis of pre-reading skills designed to differentiate skill deficits from performance deficits. Behavior Analyst Today, *3*, 161-165.
- Gerber, M. M., & Semmel, M. I. (1984). Teacher as imperfect test: Reconceptualizing the referral process. Educational Psychologist, *19*, 137-148.

- Good, R. H. & Kaminski, R. A. (1996). Assessment for instructional decisions: Toward a proactive/prevention model of decision-making for early literacy skills. School Psychology Quarterly, *11*, 326-336.
- Gravetter, F. J. & Wallnau, L. B. (1996). Statistics for the Behavioral Sciences (4th ed.). Minneapolis: West Publishing Company.
- Gresham, F.M. (1991). Conceptualizing behavior disorders in terms of resistance to intervention. School Psychology Review, *20*, 23-36.
- Gresham, F. M. (2001). Responsiveness to intervention: An alternative approach to the identification of learning disabilities. Paper prepared for the OSEP Learning Disabilities Initiative, Office of Special Education, Washington, DC. Lawrence Erlbaum, Inc., Mahwah, NJ.
- Gresham, F. M., MacMillan, D. L., & Bocian, K. M. (1997). Teachers as “tests”: Differential validity of teacher judgments in identifying students at-risk for learning difficulties. School Psychology Review, *26*, 47-60.
- Gresham, F. M., Reschly, D. J., & Carey, M. P. (1987). Teachers as “tests”: Classification accuracy and concurrent validation in the identification of learning disabled children. School Psychology Review, *16*, 543-553.
- Gresham, F. M., & Witt, J. C. (1997). Utility of intelligence tests for treatment planning, classification, and placement decisions: Recent empirical findings and future directions. School Psychology Quarterly, *12*, 249-267.
- Hartman, J.M., & Fuller, M.L. (1997). The development of curriculum-based measurement norms in literature-based classrooms. Journal of School Psychology, *35*, 377-389.
- Hayes, S. C. Nelson, R. O., & Jarrett, R. B. (1987). The treatment utility of assessment: A functional approach to evaluating assessment quality. American Psychologist, *42*, 963-974.
- Hoover, H. D., Hieronymus, A. N., Frisbie, D. A., & Dunbar, S. B. (1993). Iowa Test of Basic Skills, Form M. Riverside Publishing.
- Individuals with Disabilities in Education Act (IDEA). 20 U.S.C. Section 1400 (1990).
- Intervention Central. Retrieved March 20, 2004 from www.interventioncentral.org
- Kaminski, R. A. & Good, R. H. (1996). Toward a technology for assessing early literacy skills. School Psychology Review, *25*, 215-227.

- Kavale, K. A., Fuchs, D., & Scruggs, T. E. (1994). Setting the record straight on learning disability and low achievement: Implications for policymaking. Learning Disabilities Research & Practice, 9, 70-77.
- Keogh, B. (1994). A matrix of decision points in the measurement of learning disabilities. In G. R. Lyon (Ed.), Frames of reference for the assessment of learning disabilities (pp. 15-26). Baltimore: Paul H. Brookes.
- Kirk, S. (1962). Educating exceptional children. Boston: Houghton Mifflin.
- Lentz, F. E. (1998). Direct observation and measurement of academic skills: A conceptual review. In: Shapiro, E. S. & Kratochwill, T. R. (Eds.). Behavioral assessment in schools: Conceptual foundations and practical applications. New York, NY: Guilford Press. pp. 76-120.
- Louisiana State University School Psychology Reading Center. Accessed March 20, 2004, from <http://bitwww1.psyc.lsu.edu/Reading%20Center.htm>
- Lyon, G. R. (1996). Learning disabilities. The Future of Children, 6, 54-76.
- MacMillan, D.L., Gresham, F.M., & Bocian, K. (1998). Discrepancy between definitions of learning disabilities and what schools use: An empirical investigation. Journal of Learning Disabilities, 31, 314-326.
- MacMillan, D.L., & Speece, D. (1999). Utility of current diagnostic categories for research and practice. In R. Gallimore, L. Hernheimer, D. MacMillan, D. Speece, & S. Vaughn (Eds.), Developmental perspectives on children with high-incidence disabilities (pp. 111-133). Mahwah, NJ: Lawrence Erlbaum.
- Marston, D. (1989). Curriculum-based measurement: What is it and why do it? In M. R. Shinn (Ed.), Curriculum-based measurement: Assessing special children (pp. 18-78). New York: Guilford.
- Marston, D., & Magnusson, D. (1988). Curriculum-based assessment: District-level implementation. In J. Graden, J. Zins, & M. Curtis (Eds.), Alternative educational delivery systems: Enhancing instructional options for all children (pp. 137-172). Washington, DC: National Association of School Psychologists.
- Marston, D., Muyskens, P., Lau, M. & Canter, A. (2003). Problem-solving model for decision making with high-incidence disabilities: The Minneapolis Experience. Learning Disabilities Research & Practice, 18, 187-200.
- Martens, B. K., Witt, J. C., Daly, E. J., III, & Vollmer, T. R. (1999). Behavior analysis: Theory and practice in educational settings. In C. R. Reynolds & T. B. Gutkin (Eds.), The handbook of school psychology (3rd ed., pp. 638-663). New York: Wiley.

- Messick, S. (1984). Assessment in context: Appraising student performance in relation to instructional quality. Educational Researcher, 13, 3-8.
- Messick, S. (1995). Validity of psychological assessment: Validation of inferences from persons' responses and performances as scientific inquiry into score meaning. American Psychologist, 50, 741-749.
- Nevin, J. A. (1988). Behavioral momentum and the partial reinforcement effect. Psychology Bulletin, 103, 44-56.
- Noell, G. H., Freeland, J. T., Witt, J. C., & Gansle, K. A. (2001). Using brief assessments to identify effective interventions for individual students. Journal of School Psychology, 39, 335-355.
- Noell, G. H., Gansle, K. A., Witt, J. C., Whitmarsh, E. L., Freeland, J. T., LaFleur, L. H., Gilbertson, D. N., & Northup, J. (1998). Effects of contingent reward and instruction on oral reading performance at differing levels of passage difficulty. Journal of Applied Behavior Analysis, 31, 659-663.
- Noell, G.H., & Witt, J.C. (1999). When does consultation lead to intervention implementation? The Journal of Special Education, 33, 29-35.
- PL 107-110 No Child Left Behind Act of 2001 (2001). Retrieved March 1, 2004 from <http://www.ed.gov/policy/elsec/leg/esea02/107-110.pdf>
- Pupil Appraisal Handbook (2000). Louisiana Department of Education.
- Rutter, M. & Yule, W. (1975) The concept of specific reading retardation. Journal of Child Psychology and Psychiatry, 16, 181-197.
- Shadish, W. R., Cook, T. D. & Campbell, D. T. (2002). Regression Discontinuity Designs. In Experimental and quasi-experimental designs for generalized causal inference. Boston, MA: Houghton Mifflin Co. pp. 207-245.
- Share, D. L., McGee, R., & Silva, P. (1989). IQ and reading progress: A test of the capacity notion of IQ. Journal of American Academy of Child and Adolescent Psychiatry, 28, 97-100.
- Shaywitz, B., Fletcher, J., Holahan, J., & Shaywitz, S. (1992). Discrepancy compared to low achievement definitions of reading disability: Results from the Connecticut Longitudinal Study. Journal of Learning Disabilities, 25, 639-648.
- Shaywitz, S., Shaywitz, B., Fletcher, J., & Escobar, M. (1990). Prevalence of reading disability in boys and girls: Results of the Connecticut longitudinal study. Journal of the American Medical Association, 264, 998-1002.

- Shepard, L., Smith, M. & Vojir, C. (1983). Characteristics of pupils identified as learning disabled. American Educational Research Journal, 20, 309-331.
- Shinn, M. R. (1989). Identifying and defining academic problems: CBM screening and eligibility procedures. In M. R. Shinn (Ed.), Curriculum-based measurement: Assessing special children (pp. 90-129). New York: Guilford.
- Shinn, M. (1995). Best practices in curriculum-based measurement and its use in a problem-solving model. In A. Thomas & J. Grimes (Eds.), Best practices in school psychology-III (pp. 547-567). Washington DC: National Association of School Psychologists.
- Shinn, M. R. & Marston, D. (1985). Differentiating mildly handicapped, low-achieving, and regular education students: A curriculum-based approach. Remedial and Special Education, 6, 31-45.
- Shinn, M. R., Tindal, G. A., & Spira, D. A. (1987). Special education referrals as an index of teacher tolerance: Are teachers imperfect tests? Exceptional Children, 54, 32-40.
- Siegel, L. S. (1989). IQ is irrelevant to the definition of learning disabilities. Journal of Learning Disabilities, 22, 469-486.
- Speece, D., Case, L., & Molloy, D. (2003). Responsiveness to general education instruction as the first gate to learning disabilities identification. Learning Disabilities Research and Practice, 18, 147-156.
- Stanovich, K. E., Siegel, L. S. (1994). Phenotypic performance profile of children with reading disabilities: A regression-based test of the phonological-core variable-difference model. Journal of Educational Psychology, 86, 24-53.
- U.S. Department of Education (1998). To assure the free appropriate public education of all children with disabilities: Twentieth annual report to Congress on the implementation of the Individuals with Disabilities Education Act. Washington, DC: U.S. Government Printing Office.
- U.S. Department of Education (2002). To assure the free appropriate public education of all children with disabilities: Twenty-fourth annual report to Congress on the implementation of the Individuals with Disabilities Education Act. Washington, DC: U.S. Government Printing Office.
- VanDerHeyden, A. M., Witt, J. C., & Naquin, G. (2003). Development and validation of a process for screening referrals to special education. School Psychology Review, 32, 204-227.

- Vaughn, S. & Fuchs, L. (2003). Redefining learning disabilities as inadequate response to instruction: The promise and potential problems. Learning Disabilities Research & Practice, 18, 137-146.
- Vaughn, S., Linan-Thompson, S., & Hickman, P. (2003). Response to Instruction as a means of identifying students with reading/learning disabilities. Exceptional Children, 69, 391-409.
- Velluntino, F., Scanlon, D., & Lyon, G.R. (2000). Differentiating between difficult-to-remediate and readily remediated poor readers: More evidence against the IQ-achievement discrepancy definition of reading disability. Journal of Learning Disabilities, 33, 223-238.
- Velluntino, F., Scanlon, D., & Tanzman, M. (1998). The case for early intervention in diagnosing reading disability. Journal of School Psychology, 36, 367-397.
- Velluntino, F., Scanlon, D., Sipay, E., Small, S., Pratt, A., Chen, R., & Denckla, M. (1996). Cognitive profiles of difficult-to-remediate and readily remediated poor readers: Early intervention as a vehicle for distinguishing between cognitive and experiential deficits as basic causes of specific reading disability. Journal of Educational Psychology, 88, 601-638.
- Witt, J. C. (1996). STEEP Administration Manual.
- Witt, J. C., Adler, Y., Thomas, P., & Naquin, G. (1999). Pupil Appraisal Model, Louisiana Department of Education, Grant # 324 50 4067.
- Witt, J. C., Daly, E., & Noell, G. (2000). Functional Assessments. Sopris West: Longmont, CO.
- Ysseldyke, J., Algozzine, B., Shinn, M., & McGue, M. (1982). Similarities and differences between low achievers and students classified as learning disabled. The Journal of Special Education, 16, 73-85.

Appendix A: University Course Readings

Psyc 491/691 Field Experience Bibliography

Daly III, E.J., Witt, J.C., Martens, B.K., & Dool, E.J. (1997). A model for conducting a functional analysis of academic performance problems. School Psychology Review, 26 (4), 554-574.

Deno, S.L., Fuchs, L.S., Marston, D., Shin, J. (2001) Using curriculum-based measurement to establish growth standards for students with learning disabilities. School Psychology Review, 30 (4), 507-524.

Fuchs, D. (2003) On responsiveness-to-intervention as a valid method of LD identification: Some things we need to know. Perspectives, 29 (2), 28-31.

Fuchs, L. S., & Fuchs, D. (1986). Effects of systematic formative evaluation: A meta-analysis. Exceptional Children, 53(3), 199-208.

Fuchs, L., Fuchs, D., & Bishop, N. (1992). Teacher planning for students with learning disabilities: Differences between general and special educators. Learning Disabilities Research and Practice, 7, 120-128.

Gansle, K.A., Noell, G. H., VanDerHayden, A.M., Naquin, G.M., Slider, N.J. (2002). Moving beyond total words written: The reliability, criterion validity, and time cost of alternative measures for curriculum-based measurement in writing. School Psychology Review, 31(4), 477-497.

Greenwood, C.R., Arreaga-Mayer, C., Utley, C.A., Gavin, K.M., Terry, B.J. (2001) Class wide peer tutoring learning management system: Applications with elementary-level English language learners. Remedial and Special Education, 22(1), 34-47.

Gresham, F. (1991). Conceptualizing behavior disorders in terms of resistance to intervention. School Psychology Review, 20, 23-36.

Learning Disabilities Research and Practice, 18 (3), 2003. Entire volume devoted to CBM and RTI.

Shinn, M.R. (1989). Volume Ed., Curriculum-Based Measurement: Assessing Special Children. S.N. Elliot & J.C. Witt (Series Eds.) New York, NY: The Guilford Press.

Stecker, P. & Fuchs, L. (2000). Effecting superior achievement using curriculum-based measurement: The importance of individual progress monitoring. Learning Disabilities Research and Practice, 15, 128-134.

VanDerHeyden, A. (2002). Problem Validation Screening. Presentation at National Association for School Psychologist Conference. Chicago, IL.

Witt, J. C. (2002). Nuclear Reading Intervention. Reading Center. Retrieved December 17, 2002, from <http://bitwww1.psyc.lsu.edu>

Witt, J. C. (2002). Over-representation of minority children in special education. Paper presented at the Annual Conference for Special Education Supervisors. Baton Rouge, LA.

Witt, J.C. (1996) Screening to Enhance Equitable Placement in Special Education. Baton Rouge: Louisiana State University.

Appendix B: STEEP Instructions for Math

Scripted Instructions for Math Probe Administration

1. Pass out papers face-down instructing students not to turn them over until you tell them to do so.
2. **“Please write your first and last name on the back of your paper. Please write your teacher’s name next to your name.”** Pause briefly to allow students to write their names.
3. **“This is a math worksheet. All of the problems are _____ (addition, subtraction, multiplication, division, etc.). When I say ‘start,’ turn them over and begin answering the problems. Begin on the first problem on the left on the top row (point). Work across and then go to the next row. If you come to a problem that you do not know, cross it out. Are there any questions?”**
4. Set timer for two minutes. **“Start.”** Begin timer.
5. Monitor student performance to ensure that students work the problems in rows and do not skip around or answer only the easy problems.
6. When the timer rings, say, **“Stop. Put your pencils down and hold your papers up in the air so we can pick them up.”**
7. Collect math sheets and give to service provider/consultant.

Math Scoring Instructions

In scoring math questions, students get one point for each number they put in the correct location. This is called scoring for **digits correct**. Scoring for digits correct, rather than merely counting the number of correct problems, provides a way to give students “credit” for solving more complex problems which may involve several steps.

To score math problems, simply count number of digits correct in two minutes. Digits are correct if they are the correct digit located in the proper place value for that problem:

- $5 + 26 = \underline{31}$ (2 digits correct)
- $5 + 21 = 3\underline{6}$ (1 digit correct)
- $5 + 20 = \underline{2}3$ (1 digit correct)
- $5 + 16 = 30\underline{1}$ (1 digit correct)
- $5 + 10 = 23$ (0 digits correct)
- Total digits correct above = 5

The Math Instructions are from the STEEP Manual (Witt, 1996, pp. 8, 10).

Appendix C: Sample Math Probe

Curriculum-Based Assessment Mathematics
Single-Skill Computation Probe: Student Copy

$$\begin{array}{r} 27 \\ +16 \\ \hline \end{array}$$

$$\begin{array}{r} 77 \\ +16 \\ \hline \end{array}$$

$$\begin{array}{r} 73 \\ +27 \\ \hline \end{array}$$

$$\begin{array}{r} 38 \\ +23 \\ \hline \end{array}$$

$$\begin{array}{r} 19 \\ +61 \\ \hline \end{array}$$

$$\begin{array}{r} 24 \\ +67 \\ \hline \end{array}$$

$$\begin{array}{r} 29 \\ +61 \\ \hline \end{array}$$

$$\begin{array}{r} 55 \\ +39 \\ \hline \end{array}$$

$$\begin{array}{r} 88 \\ +13 \\ \hline \end{array}$$

$$\begin{array}{r} 45 \\ +36 \\ \hline \end{array}$$

$$\begin{array}{r} 58 \\ +24 \\ \hline \end{array}$$

$$\begin{array}{r} 49 \\ +39 \\ \hline \end{array}$$

The sample math probe was generated at Intervention Central (accessed March 20, 2004).

Appendix D: STEEP Instructions for Reading

Scripted Instructions for Reading Probe Administration

1. **“We’re reading with everyone (or several students) in your class/school today.”** Give the student the story. Write the student’s name, teacher, and grade on the probe sheet.
2. **“When I say ‘start,’ begin reading aloud at the top of the page. Read across the page (demonstrate by pointing). Try to read each word. If you come to a word that you do not know, I will tell it to you. Be sure to do your best reading. Do you have any questions?”**
3. Set the timer for one minute. **“Start.”** Allow the student to read for one minute. Follow along on your copy, marking the words that are read incorrectly. If the student pauses on a word, wait only three seconds, tell the student the word, and move on.
4. When the timer rings say, **“Stop reading.”** Draw a vertical line after the last word read. Thank the student for reading.

Note: Prior to beginning the reading assessment, it will save time if the assessor writes the name of each child on the reading probes sheets. By doing this, the assessor can use the time between children to score rather than to write the name of the next child. It also saves time if you have one child waiting at the door while testing a child. When finished testing, send that child to get the next child who will then be ready when you have finished. It saves time by having one child being tested, one child waiting, and one child in transit to be tested.

Reading Scoring Instructions

- ONE error is counted for the following
 - Mispronounced words
 - Skipped words
 - Transposition of word pairs (reads: “beautiful red tree” as red beautiful tree)
 - Substitutions (reads: “mother” as mom)
 - Word told to the student after 3-second hesitation.
- Words that are **NOT** counted as errors,
 - Words read correctly
 - Insertions
 - Repetitions
 - Self-corrections

Reading errors are noted as the child reads aloud. After the child has completed reading the passage, count the total number of words read and count the errors the child made. Subtract errors from words total words read for Correctly Read Words.

The reading instructions are from the STEEP manual (Witt, 1996, pp. 9,11).

Appendix E: Sample Reading Probe, Examiner Copy

CBM Procedures:

- Set your timer for **1 minute**.
- Give the student the Student Form of this assessment and instruct to **“Begin reading aloud now.”**
- Start the timer when the student reads the first word.
- Follow along (with pencil in hand) on the passage below as the student reads aloud.
- Only provide a word for the student if he/she hesitates on the word for 3 seconds.
- Put an **X** over words read incorrectly to mark reading errors.

<p><u>Mark (X) as incorrect if student:</u></p> <ul style="list-style-type: none"> ◆ Mispronounces a word. ◆ Skips a word. ◆ Transposes words (reads “big smile” as “smile big”). ◆ Substitutes words (reads “mom” for “mother”). ◆ Hesitates on words (longer than 3 seconds). 	<p><u>Leave as correct if student:</u></p> <ul style="list-style-type: none"> ◆ Inserts a word that is not in the text. ◆ Repeats a word that is in the text. ◆ Self-corrects reading errors.
---	---

- At the end of **1 minute**, put a **{bracket}** around the last word read and ask the student to stop reading.
- Complete the score box below.

Once upon a time there was a little girl whose home was	12
down a deserted road deep in the forest. The little girl	23
loved living in the forest among the animals. She was a	34
friend to all the animals of the forest. Yet, she was also	46
a lonely little girl because she had no family. One day	57
while skipping through the woods, she met a beautiful	66
fairy. The fairy questioned the girl about where she	75
lived. The girl told the fairy about her forest home. The	86
fairy asked the girl about her family. The girl told the	97
fairy that she did not have a family. The fairy could feel	109
that the girl was lonely, so she asked the girl if she	121
wanted to live with all the fairies. The girl happily	131
answered yes and went to live with the fairy. For the rest	143
of her life the girl had her friends, the animals, and a	155
family of her very own.	160

<p>Student: _____ Date: _____</p> <p style="text-align: center;">Number of words read <u>correctly</u> in 1 minute = _____</p> <p style="text-align: right; font-size: small;">Copy score onto the student's Room Placement Form.</p>
--

The probe was retrieved from the Louisiana State University
School Psychology Reading Center (2004)

Appendix F: Instructions for Can't Do/Won't Do Assessment in Math

Scripted Instructions for a Can't Do/Won't Do Assessment in Math

1. Greet student. **“We’re going to do some math today.”**
2. **“The last time you did a math sheet like this one you scored _____ (score from class wide assessment) correct.”**
3. **“I’m going to give you an opportunity to do this worksheet again. If you can beat your score, then you can pick anything you like from the treasure chest.”** Show student the treasure chest. Allow student to briefly sample items in the treasure chest.
4. Ask the student **“Do you see anything in there that you would like to earn?”** If the student does not seem excited about any of the items in the treasure chest, offer free time, outside time, visit with favorite teacher, or ask the student to nominate something reasonable.
5. **“This is a math worksheet. All of the problems are _____ (addition, subtraction, multiplication, division, etc.). When I say ‘start,’ you may begin answering the problems. Begin on the first problem on the left on the top row (point). Work across and then go to the next row. If you come to a problem that you do not know, cross it out. Do you have any questions?”**
6. Set timer for two minutes. **“Start.”** Wait 2 minutes.
7. Monitor student performance to ensure that the student works the problems in rows and does not skip around or answer only the easy problems.
8. When timer rings say **“Stop.”**
9. Count the number of digits correct. If the student increased his/her score by one digit or more, allow student to select something from the treasure chest. If the student did not increase his/her score by one digit or more, do not allow the student to make a selection from the treasure chest.

The instructions are from the STEEP manual (Witt, 1996, p. 13).

Appendix G: Instructions for Can't Do/Won't Do Assessment in Reading

Scripted Instructions for a Can't Do/Won't Do Assessment in Reading

1. Greet student. **“We’re going to do some reading today.”**
2. **“The last time you read a story, you read _____ words correctly. I’m going to give you a chance to read another story. If you can read more words correctly this time than you did last time, then you can pick anything you like from the treasure chest.”** Show student the treasure chest. Allow student to briefly sample items in the treasure chest.
3. Ask the student **“Do you see anything in there that you would like to earn?”** If the student does not seem excited about any of the items in the treasure chest, you may offer free time, outside time, visit with favorite teacher, or get the student to nominate something reasonable.
4. **“When I say ‘start,’ begin reading aloud at the top of the page. Read across the page (demonstrate by pointing). Try to read each word. If you come to a word that you do not know, I will tell it to you. Be sure to do your best reading. Do you have any questions?”**
5. **“Start.”** Allow the student to read for one minute. Follow along on your copy, marking the words that are read incorrectly. If the student pauses on a word, wait only three seconds, tell the student the word, and move on.
6. At the end of one minute, **“Stop reading.”** Draw a vertical line after the last word read. Thank the student for reading.
7. Count number of words read correctly and number of errors. If student exceeded goal, deliver reward.

The instructions are from the STEEP manual (Witt, 1996, p. 14).

Appendix H: Integrity Checklist for Math Probe Administration

Integrity Checklist for Math Probe Administration

Circle each item below as it is completed correctly.

1. Papers were passed out face down.
2. Instructions were given to not to turn paper over until told to do so.
3. The following phrase was read word by word: **“Please write your first and last name on the back of your paper. Please write your teacher’s name next to your name.”**
4. The following passage was read work for word **“This is a math worksheet. All of the problems are _____ (addition, subtraction, multiplication, division, etc.). When I say ‘start,’ turn them over and begin answering the problems. Begin on the first problem on the left on the top row (point). Work across and then go to the next row. If you come to a problem that you do not know, cross it out. Are there any questions?”**
5. Teacher said start and then immediately started the timer.
6. Timing was for exactly 2 minutes.
7. Teacher monitored student performance to ensure that students worked the problems in rows and did not skip around.
8. Teacher upon hearing timer, immediately said: **“Stop. Put your pencils down and hold your papers up in the air so we can pick them up.”**
9. Math sheets were collected.
10. Scoring was completed accurately.

The instructions are from the STEEP manual (Witt, 1996, p. 17).

Appendix I: Integrity Checklist for Reading Probe Administration

Integrity Checklist for Reading Probe Administration

Circle each item below as it is completed correctly.

1. Assessor read the following phrase work for word: **“We’re reading with everyone (or several students) in your class/school today.”**
2. Assessor gave the student the story.
3. Assessor wrote the student’s name, teacher, and grade on the assessors copy.
4. Assessor read the following phrase word for word: **“When I say ‘start,’ begin reading aloud at the top of the page. Read across the page (demonstrated by pointing). Try to read each word. If you come to a word that you do not know, I will tell it to you. Be sure to do your best reading. Do you have any questions?”**
5. Assessor started timer and then said, **“Start.”**
6. If the student paused on a word, assessor waited three seconds, and told the student the word, and move on.
7. Timing was for exactly one minute.
8. When the timer rang, the assessor said **“Stop reading.”**
9. A vertical line was drawn after the last word read..

The instructions are from the STEEP manual (Witt, 1996, p. 18).

Appendix J: PPRI Implementation Procedures

Instructions for Implementing the Intervention

1. Provide the student with Lesson A from the appropriate Room.
 - a. If it is the first day of intervention then Lesson 1A should be used.
 - b. The intervention lesson is determined by the student's reading of story *C* during the previous intervention session.
2. Read the story to the student slowly.
3. Have the student read the story back to you twice.
 - a. Provide immediate error correction.
4. Have the student identify words in the Word Recognition Quiz.
 - a. Ask the student to make an "X" near the word that you say.
 - b. Say the word that has an "X" next to it on the teacher's copy.
 - c. If the student makes an error, say "No that's not right, try finding the word _ again."
 - d. If the student fails to choose the correct word again, show the student the correct word.
 - e. Perform these steps for the remaining word identification problems
5. Have the student answer the comprehension questions for the passage.
 - a. Read the question to the student and have the student mark an "x" near the correct answer.
 - b. If the student does not get the answer correct, help the student find the answer in the passage.
 - c. Complete all three comprehension questions for this passage.
6. Repeat steps 1 through 5 for stories B and C.
7. After the student has read all three stories (a, b, & C) in a level and has answered all word identification and comprehension questions, use CBM procedures to assess students reading fluency on story C. (This will be the third time that the student reads story C.)
 - a. CBM involves giving the student one minute to read as many words as he/she can and recording the student's errors.
 - b. Record the number of words the student read in a minute for story C on the Level Determination Form.
 - c. The number of words read correctly in one minute of story C determines the intervention set that will be used during the next intervention session.
8. If a student is reading stories out of the Blue Circle or Red Star rooms the student must read 60 words correct in a minute in order to move to the next level in that room. If the student reads less than 60 words in a minute, implement the intervention at the same level during the next session.
9. If a student is reading stories out of the Green Square, Yellow Triangle, or Purple Diamond rooms the student must read 100 words correct in a minute in order to move to the next level in that room. If the student reads less than 100 words in a minute, implement the intervention at the same level during the next session.

The instructions were retrieved from the Louisiana State University School Psychology Reading Center (2004)

Appendix K: Treatment Integrity Checks for PPRI

Treatment Integrity Checks for PPRI

Student Code: _____ Ix Level: _____ Lesson #: _____

- ___ 1. Provided correct lesson based on performance on C from previous passage
- ___ 2. Passage A: Student identified words on Recognition Quiz as evidenced by “X” or “√” marks.
- ___ 3. Passage A: Student answered comprehension questions as evidenced by “X” or “√” marks.
- ___ 4. Passage B: Student identified words on Recognition Quiz as evidenced by “X” or “√” marks.
- ___ 5. Passage B: Student answered comprehension questions as evidenced by “X” or “√” marks.
- ___ 6. Passage C: Student identified words on Recognition Quiz as evidenced by “X” or “√” marks.
- ___ 7. Passage C: Student answered comprehension questions as evidenced by “X” or “√” marks.
- ___ 8. Number of words read per minute on passage C correctly recorded on Level Determination Form.
- ___ 9. Correctly checked to repeat lesson or proceed to next session based on student fluency score recorded on form.

Vita

Susan Lockhart Gatti is currently a graduate student in the School Psychology Program at Louisiana State University under the direction of Dr. George Noell. She received both her Bachelor of Arts degree (1997) and her Master of Arts degree (2000) from Louisiana State University in the major area of psychology. Susan Gatti is a candidate for the degree of Doctor of Philosophy to be awarded in May of 2004.