Summer 8-2017

Differential Diagnosis of Specific Learning Disability Within a Response to Intervention Framework

Michael J. Boneshefski

Follow this and additional works at: https://knowledge.library.iup.edu/etd

🎨 Part of the Educational Psychology Commons, and the Special Education and Teaching Commons

Recommended Citation

DIFFERENTIAL DIAGNOSIS OF SPECIFIC LEARNING DISABILITY

WITHIN A RESPONSE TO INTERVENTION FRAMEWORK

A Dissertation

Submitted to the School of Graduate Studies and Research

in Partial Fulfillment of the

Requirements for the Degree

Doctor of Education

Michael J. Boneshefski

Indiana University of Pennsylvania

August 2017

Indiana University of Pennsylvania
School of Graduate Studies and Research
Department of Educational and School Psychology

We hereby approve the dissertation of

Michael J. Boneshefski

Candidate for the degree of Doctor of Education

_____   _____
                                 Joseph F. Kovaleski, D.Ed.
                                 Professor of Educational and School Psychology,
                                 Chair

_____   _____
                                 Timothy Runge, Ph.D.
                                 Associate Professor of Educational and School
                                 Psychology

_____   _____
                                 Daniel R. Wissinger, Ph.D.
                                 Assistant Professor of Professional Studies in
                                 Education

ACCEPTED

_____   _____
Randy L. Martin, Ph.D.
Dean
School of Graduate Studies and Research

ii

Title: Differential Diagnosis of Specific Learning Disability Within a Response to Intervention Framework

Author: Michael J. Boneshefski

Dissertation Chair: Dr. Joseph F. Kovaleski

Dissertation Committee Members: Dr. Timothy Runge
                                    Dr. Daniel R. Wissinger

The purpose of this study was to determine to what extent two major specific learning disability (SLD) criteria, including a student's level of academic achievement and rate of improvement (ROI), predict multidisciplinary evaluation teams' decision-making regarding referral for special education evaluation and special education eligibility. Reading curriculum-based measurement (CBM-R) and demographic data were obtained from 383 second and third grade students in a Midwestern state who were receiving general education intervention in reading but not referred for a special education evaluation or who were found eligible for special education. CBM-R data were analyzed to determine whether students found eligible for special education demonstrated dual discrepancies (Fuchs & Fuchs, 1998) and whether level of academic achievement and ROI predicted students' special education eligibility status. Results revealed students with SLD displayed lower levels of performance and ROIs than students receiving general education reading intervention who were not referred for a special education evaluation. Results also suggested that level of performance was predictive of SLD identification but that ROI did not significantly contribute to decisions about whether a student is identified as SLD. This result was found for both second and third grade students. Implications for implementing a multi-tiered system of supports (MTSS) and determining special education eligibility within a response to intervention framework are discussed.

TABLE OF CONTENTS

LIST OF TABLES

CHAPTER 1

INTRODUCTION

Although students with specific learning disability (SLD) make up the greatest proportion

of all students receiving special education services, determining exactly what SLD is has

historically been shrouded in controversy. Despite the passing of the Education for All

Handicapped Children Act (EHA) in 1975, which resulted in an operationalization of SLD in

public policy and guaranteed students with SLD access to a free and appropriate public

education, conceptual and practical issues arose regarding who is eligible for special education

services under the SLD label. In the years following the United States Department of

Education's (1977) regulations that operationally defined SLD, Ysseldyke et al. (1983) found

inconsistencies in special education team decisions regarding eligibility and placement, reliance

on assessments seeking within-child problems, and unreliable systems for identifying SLD.

As the legal definition of SLD remained largely unchanged over time, problems

associated with operationalizing SLD (Keogh, 1987; Mather & Roberts, 1994) for the purposes

of identifying SLD in the schools persisted (Kavale & Forness, 2000; Kavale & Reese, 1992,

Peterson & Shinn, 2002). In spite of these problems, Cortiella and Horowitz (2014) reported that

from 1976 to 2000 the number of students served under the SLD category grew by more than

300%. From 2002 to 2011, however, the number of students identified as having an SLD

decreased by 18%, although the total special education prevalence rate decreased by just three

percent over the same time period. Since the 1970s trends in research, policy, and practice,

including increased focus on curriculum-based measurement (CBM), educational accountability

movements (e.g., the No Child Left Behind Act of 2001, enacted in 2002), and inclusion of

response to intervention (RTI) practices in schools, have continued to alter the way SLD is

viewed. Reauthorizations of the Individuals with Disabilities Education Act[1] (IDEA) in 1997 and 2004 also influenced SLD identification practices.

Throughout history, practitioners have traditionally conceptualized ability as IQ (Kovaleski, VanDerHeyden, & Shapiro, 2013). Though the use of norm-referenced achievement and IQ tests for documenting a discrepancy between a student's ability and achievement was the regulatory standard until the most recent authorization of IDEA, researchers were developing assessment practices derived from students' curricular materials to evaluate the effects of instruction (Deno & Mirkin, 1977). This early work evolved into what is now widely-known as CBM (Deno, 1985, 2003). CBM, which measures performance on a general outcome (L. S. Fuchs & Deno, 1991, 1994), has evolved from its beginnings to guide instructional decision-making and program evaluation to provide myriad uses, including predicting performance on high-stakes assessments, identifying students at risk of academic failure, and replacing traditional special education evaluation practices (Deno, 2003).

With the increased importance of CBM in educational practices, the technical adequacy of CBM as an assessment tool has continued to be studied, as it must demonstrate reliability, validity, and sensitivity to be useful for higher-stakes decision-making (American Education Research Association, American Psychological Association, & National Council on Measurement in Education, 1999). Much research has demonstrated that CBM, and in particular CBM of oral reading fluency (CBM-R), meets these criteria (Marston, 1989; Wayman, Wallace, Wiley, Tichá, & Espin, 2007). As indicated by Ardoin, Christ, Morena, Cormier, and Klingbeil's (2013) literature review of CBM-R, however, much is still unknown about using individual students' CBM-R data for accurate decision-making about progress.

---

[1] All iterations of the Individuals with Disabilities Education Act will be referred to as IDEA

Following the passage of the NCLB act of 2001 (United States Department of Education, 2002) and its reauthorization as the Every Student Succeeds Act (United States Department of Education, 2015) and as educators are held increasingly accountable for student outcomes, formative assessments of students' progress have been emphasized to improve student outcomes on high-stakes assessments. As CBM may be used to predict performance on state assessments (Ferchalk, 2013; Silberglitt, 2008), schools have increasingly used CBM as a universal screener to identify students at risk throughout the school year (Kovaleski, et al., 2013; Kovaleski & Pedersen, 2008). Furthermore, schools are incorporating universal screening into their multi-tiered service delivery model for ensuring all students have access to robust core instruction, which was also recognized in the 2004 reauthorization of the IDEA.

With the 2004 reauthorization of IDEA, local education agencies (LEAs) could use students' RTI data to identify SLD (U.S. Department of Education, 2004). By allowing the use of RTI data, LEAs no longer had to rely on high-inference measures of ability to establish an ability-achievement discrepancy and could conduct frequent formative assessments using CBM to establish whether a child is demonstrating adequate achievement and progress to benefit from general education instruction.

Although as of 2006 all 50 states have provided guidance related to identifying SLD, more precise guidelines for identifying SLD have largely not been developed beyond federal regulations, which suggests that states are allowing LEAs leeway in how SLD is identified using RTI data (Hauerwas, Brown, & Scott, 2013). Additionally, review of the literature reveals varying methods for assessing a student's RTI, including final benchmark (Good, Simmons, & Kame'enui, 2001), slope discrepancy (Fuchs, Fuchs, & Compton, 2004), and dual discrepancy (Fuchs & Fuchs, 1998; Fuchs, Fuchs, & Speece., 2002; Speece & Case, 2001). The dual

discrepancy approach, which requires a student to demonstrate significant discrepancies in both level of performance and rate of improvement (ROI), appears to be the most consistently supported framework for assessing a student's RTI (Burns & Senesac, 2005; Fuchs, Compton, Fuchs, Bryant, & Davis, 2008; Fuchs et al., 2002; Fuchs et al., 2004; Speece & Case, 2001). Furthermore, the federal regulations stipulate that a student identified with SLD must demonstrate a dual discrepancy (United States Department of Education, 2006; Kovaleski et al., 2013).

Statement of the Problem

With the historical problems associated with SLD identification, the lack of specific guidelines for identifying SLD using RTI is concerning (Flinn, 2015; Hauerwas et al., 2013). Recently, Maki, Floyd, and Roberson (2015) reported varying identification practices across states. Additionally, when provided with identification criteria and evaluation data for SLD identification, Maki, Burns, and Sullivan (2016) found that practicing school psychologists did not consistently adhere to SLD criteria when making decisions about SLD eligibility, resulting in low consistency across SLD identification. Although in practice a dual discrepancy is required, when examining various models in the research literature for determining whether a student is a non-responder, there is limited agreement across methods used (Barth et al., 2008; Brown-Waesche, Schatschneider, Maner, Ahmed, & Wagner, 2011; Burns, Scholin, Kosciolek, & Livingston, 2010), which can have consequences when making a high-stakes decision, such as special education eligibility determinations.

Furthermore, L. S. Fuchs (2003) explained it must be determined which combination of factors used to assess responsiveness results in the best identification of students with persistent and pervasive reading problems (i.e., students with SLD). Given the limited understanding about

what RTI data truly are indicative of SLD, as different schools conduct evaluations using RTI data in different ways, it seems as though variation across contexts with regard to who is identified as having an SLD will emerge, challenging the concept of who has an SLD.

Additionally, researchers have demonstrated that, historically, the defining characteristic of SLD has been low achievement (Kavale & Reese, 1992; Peterson & Shinn, 2002; Ysseldyke, Algozzine, Shinn, & McGue, 1982). Although low achievement appears to be a defining characteristic of SLD, other studies have suggested seemingly extraneous variables may explain SLD categorization (Lester & Kelman, 1997; Singer Palfrey, Butler, & Walker, 1989; Ysseldyke et al., 1982; Ysseldyke & Algozzine, 1981). Lester and Kelman (1997) found that demographic and sociopolitical variables seemingly unrelated to organic disability prevalence more strongly influenced SLD diagnosis compared to "hard" disability diagnosis (e.g., physical disability). Moreover, Singer and colleagues (1989) suggested SLD is a catch-all or "category of last resort" (p. 277), potentially due to the heterogeneous nature of students identified with SLD. Earlier studies have demonstrated that, despite the necessity that students meet the four criteria outlined in IDEA regulations of SLD, school teams may not be relying on technically-sound assessment data when making educational decisions, including decisions about special education eligibility. Therefore, if RTI is to be established as a more viable option for accurately identifying students with SLD, further guidance based on technically adequate assessment practices is needed.

Research Question

The purpose of this study will be to determine to what extent two major SLD criteria, including a student's level of academic achievement and ROI, impacts multidisciplinary evaluation teams' decision-making regarding referral for special education evaluation and regarding special education eligibility. To provide insight into this issue, this study will focus on

the following question. Do level of academic achievement and ROI, as well as potentially

extraneous variables (i.e., student sex, race, and socioeconomic status), predict classification of

students into three groups: (a) students with oral reading fluency (ORF) skill deficits receiving

intensive reading intervention but not referred for special education evaluation, (b) students with

ORF deficits referred for a special education evaluation and found not eligible for special

education services, and (c) students with ORF deficits referred for special education evaluation

and found eligible for special education services?

1. It is hypothesized that level of performance will be the variable most related to students'
   group membership. Previous research demonstrated low achievement to be the defining
   characteristic of SLD (Brown-Waesche et al., 2011; Kavale & Reese, 1992; Peterson &
   Shinn, 2002; Ysseldyke et al., 1982), suggesting practitioners have historically
   determined need for more intensive support based on a student's performance level.

2. It is hypothesized that level of performance and ROI together will predict students' group
   membership to a greater extent than the extraneous variables. Although previous
   research demonstrated that extraneous variables influence SLD identification (Lester &
   Kelman, 1997; Singer et al., 1989; Ysseldyke et al., 1982; Ysseldyke & Algozzine,
   1981), more recent research suggested that decisions about student need based on level
   and ROI are less subject to influence from extraneous variables (Burns & Senesac, 2005;
   Marston, Muyskens, Lau, & Canter, 2003; Speece, Case, & Molloy, 2003). Additionally,
   federal regulations require students to be dually discrepant in both level and ROI (United
   States Department of Education, 2006).

Significance of the Study

The federal regulations suggest that a student identified as having an SLD based on RTI data needs to be dually discrepant in level of performance and ROI (United States Department of Education, 2006). Examining school teams' decisions about student need will provide insight regarding differential classification of students who demonstrate inadequate achievement. Examination of student characteristics, including performance level and ROI data, will be used to determine whether regulations and research that suggests SLD should be based on a dual discrepancy is borne out in practice.

This study will determine whether major SLD criteria identified by researchers and included in federal regulations explain team decisions about low achieving students' need for more intensive services, including special education as a student with an SLD. Although researchers have suggested that a dual discrepancy can be used to objectively identify students with the most persistent and pervasive learning needs (i.e., a student with an SLD; Fuchs & Fuchs, 1998; Speece & Case, 2001), decisions about student need have historically been influenced by extraneous student data (Lester & Kelman, 1997; Singer et al., 1989; Ysseldyke et al., 1982). Making decisions about student need using RTI data may be less prone to extraneous variables influencing teams' decisions (Burns & Senesac, 2005; Marston et al., 2003; Speece et al., 2003). This study will allow for a better understanding of whether the SLD construct based on a dual discrepancy is being implemented in actual practice, or whether extraneous factors continue to influence eligibility decision-making.

Assumptions

This study is based on a number of assumptions. First, it is assumed that participating school sites are implementing key features of an RTI framework, which includes providing

evidence-based core instruction, using regular universal screening and progress monitoring data for instructional decision-making, and providing targeted and intensive supports for students in need (United States Department of Education Office of Special Education and Rehabilitative Services, 2011). Given the frequency with which students' progress should be monitored, it is also assumed that the progress of struggling students is monitored at least weekly using CBM (Burns & Gibbons, 2008). It is assumed schools participating in the study are monitoring the progress of students who have performed below benchmark on two consecutive universal screening assessments approximately weekly. Lastly, it is assumed that student CBM data are accessible by school teams when making referrals for and completing evaluations and that special education evaluations are completed with integrity. It is assumed that educational practitioners making eligibility decisions are complying with special education regulations.

Limitations

Student data were obtained from school districts in a Midwestern state. Convenience sampling was used. Therefore, the demographics of the sample may not adequately represent the entire U. S. population, and generalizations may not be made to other students or settings. Data obtained reflect major SLD criteria identified by researchers and included in federal regulations to explain team decisions about low achieving students' need for more intensive services, including special education as a student with an SLD. Additionally, this study focuses only on students with reading skill deficits, and as such includes decisions made using CBM-R data as one source of data. Therefore, conclusions made do not necessarily generalize to CBM in math and written expression and oral language.

Although data were analyzed for the point at which a decision was made (e.g., decision point that a student is eligible for special education), students in the sample have different

numbers of data points comprising the slope depicting their ROI, which also means that students have received instruction for different amounts of time before a decision regarding the student's need (i.e., for special education services) is made. Therefore, at the point a decision about student need is made, students who are not referred for an evaluation until later in the school year may have additional opportunities to benefit from instruction compared to students referred for an evaluation in the beginning of the school year. This may impact student's performance and, thus, a team's decision about the student's need. Although this may be a limitation from a research design standpoint, the reality is that school teams make decisions about student need throughout the school year, which means, in practice, students have been provided with different amounts of instruction prior to a decision being made. Therefore, descriptive statistics about the dataset (e.g., number of progress monitoring data points available for students) will be reported in the results section.

Definition of Terms

Child Find: Child find is the process of identifying and evaluating children suspected of having a disability under the IDEA (U. S. Department of Education, 2004). Children who are thought to have a disability from birth to age 21 may be evaluated, identified as having a disability, and found to be in need of special education and related services.

Curriculum-Based Measurement: CBM is a brief (1 to 5 minutes) standardized assessment of academic skills used to determine students' performance in a particular skill area (Shinn, 2008). CBM may be used as a benchmark assessment and, when given over time, may serve as a measure of progress. CBM is a reliable and valid measure of general achievement (Deno, 1985; Deno, Fuchs, Marston, & Shin, 2001; Wayman et al., 2007) and may be used for

monitoring the effectiveness of instruction, making instructional changes, and determining

special education eligibility.

Diagnostic Assessment: A diagnostic assessment is administered to understand students'

strengths and needs for informing instruction. Diagnostic assessments provide information about

students' skill development to better individualize instruction.

Dual Discrepancy: Dual discrepancy is a discrepancy analysis that includes measuring a

student's level of performance and ROI to determine whether students are discrepant from peers

in both level and ROI (Fuchs, 2003). A dual discrepancy approach allows practitioners to

consider a low-performing student's growth in response to scientifically-based interventions.

Education for All Handicapped Children Act: The EHA is the first iteration of the

Individuals with Disabilities Education Improvement Act enacted by Congress in 1975. The

EHA guaranteed a free and appropriate public education for students with disabilities and

required local education agencies to engage in child find activities to identify students with

disabilities. The EHA and the 1977 Federal Regulations that accompanied it operationalized the

first definition of a specific learning disability in U. S. special education law.

General Outcome Measures: General outcome measures are assessments of proficiency

on global outcomes associated with a curriculum (Fuchs & Deno, 1991). General outcome

measures are efficient, reliable, and valid assessments used to improve instruction and evaluate

the effects of instruction on student growth. CBM is a type of general outcome measurement.

Individuals with Disabilities Educational (Improvement) Act: Previously known as the

EHA until 1990, the IDEA is a four-part legislative act that requires child find practices as well

as ensures a free and appropriate public education to students with disabilities. The 2004

reauthorization of IDEA and the 2006 Federal Regulations that followed allowed response to intervention data to be used for SLD eligibility determinations.

Level of Performance: Level of performance refers to a student's final benchmark status (Good, Simmons, & Kame'enui, 2001) on a measure of achievement (e.g., CBM-R) at the time an educational decision is made.

Local Education Agency: An agency with legal authority and control over public schools in a school district or other sector within a state (IDEA, 2004). Local education agency is sometimes synonymous with school district.

Multidisciplinary Evaluation Team: A multidisciplinary evaluation (MDE) team includes individuals who assess areas in which a disability is suspected and interpret data to determine eligibility. The MDE team works with the parents to make instructional recommendations based on the results of the evaluation and to set short- and long-term goals.

Multi-tiered System of Support: Multi-tiered system of support (MTSS) refers to a comprehensive service delivery model that incorporates evidence-based universal practices at the school- and class-wide levels, research-based supplemental intervention, and data-based decision-making across all levels of the system to improve academic and social outcomes for all students (Stoiber, 2014). An MTSS framework incorporates multiple tiers of support, including Tier 1 (i.e., universal instruction), which is the core instructional program; Tier 2 (i.e., targeted instruction), which includes research-based supplemental intervention for students at risk of failing; and Tier 3 (i.e., intensive intervention), which includes research-based individualized intervention for students with more intensive needs. In an MTSS framework, Tier 1 supports should sufficiently meet the needs of 80-90% of students; Tier 2 supports should meet the needs of 10-15% of students; and Tier 3 supports should be required for 1-5% of students (Tilly, 2008).

MTSS may integrate both academic and behavioral systems to address problems, but for the purposes of this study only academic systems will be considered.

Oral Reading Fluency: Oral reading fluency is the ability to read connected text accurately with prosody at an appropriate rate that allows for comprehension. Curriculum-based measures of oral reading fluency include reading a passage for a preset time (typically one minute) so that a student's rate (measured in words correct per minute [WCPM]) and accuracy (measured in percent of words read correctly out of the total words read or attempted) may be calculated.

Problem-Solving: Problem-solving is a decision-making framework that includes problem identification, problem analysis, intervention development, and intervention evaluation (Tilly, 2008). When using a problem-solving approach, school teams use objective problem definitions based on data that can be readily monitored through tools such as CBM. Rather than attempting to ameliorate within-child problems, problem-solving focuses on environmental, instructional, and curricular modifications to address the problem. Students' response to instruction and intervention data may serve as outcome data for evaluating the effectiveness of the intervention.

Progress Monitoring: Progress monitoring involves regular assessment of individual students who do not meet goals on benchmark assessments to monitor the effects of instruction on closing the achievement gap. Progress monitoring is used to determine the effectiveness of an intervention, inform decisions pertaining to instructional adjustments, and measure student progress toward instructional and grade-level goals (Fuchs & Fuchs, 2011). Progress monitoring of students receiving Tier 2 supports is recommended at least every other week, and progress monitoring of students receiving Tier 3 supports is recommended at least every week (Kovaleski

et al., 2013). Continual progress monitoring provides information regarding a student's response

to intervention (i.e., rate of improvement), which is required for documenting evidence of the

second criterion for SLD determination.

Rate of Improvement: Rate of improvement (ROI) refers to the progress a student is

making that allows educators to determine whether the achievement gap is closing. ROI, or

slope, can be visually analyzed by inspection of graphed progress monitoring data or quantified

through a calculation of the slope of the data, which provides the most accurate growth rate over

time (Christ, Zopluoglu, Long, & Monaghen, 2012; Deno et al., 2001; Shinn, Good, & Stein,

1989).

Race: In this study, race includes two categories, historically overrepresented and not

historically overrepresented groups. Groups historically overrepresented in the SLD category

include Native Americans, African Americans, and Hispanics (Skiba et al., 2008). Calculations

of risk ratios (Boneshefski & Runge, 2014) based on total enrollment data by race available from

the National Center for Education Statistics (n.d.) and the eligibility information by race and

disability status from the United States Department of Education (n.d.) indicated multi-racial

students are also overrepresented in the SLD designation. Groups not historically

overrepresented in the SLD category include Asians and Whites (Skiba et al., 2008).

Response to Intervention: RTI may be conceptualized as both a diagnostic approach and

an instructional model. Torgesen (2009) explained that the RTI diagnostic approach is a way of

"determining eligibility for special education services" (p. 38) under IDEA. Within the context

of IDEA (2004), RTI is an assessment approach for determining SLD eligibility, as teams "may

use a process that determines if the child responds to scientific, research-based intervention as a

part of the evaluation procedures" (Section 1414(b)(6)). Torgesen (2009) explained the RTI

instructional model is "a method for increasing the capacity of schools to respond effectively to the diverse learning and behavioral support needs of their students" (p. 38).

Specific Learning Disability: The federal definition of SLD has remained largely consistent since the EHA was passed in 1975. An SLD, as outlined in IDEA (2004) is as follows:

A disorder in one or more of the basic psychological processes involved in understanding or in using language, spoken or written, that may manifest itself in the imperfect ability to listen, think, speak, read, write, spell, or to do mathematical calculations, including conditions such as perceptual disabilities, brain injury, minimal brain dysfunction, dyslexia, and developmental aphasia. (34 CFR 300.8(c)(10))

State Education Agency: State education agency (SEA) refers to the state-level governing body that is responsible for disseminating information and allocating resources regarding education. SEAs are responsible for ensuring LEAs are in compliance with IDEA Part B requirements.

Universal Screening: Universal screening includes administering a benchmark assessment to all students approximately three times each year (fall, winter, and spring) to determine overall effects of instruction for groups of students and to identify students who are not proficient with a skill (Kovaleski & Pedersen, 2008).

Summary

This chapter provided a background for the SLD construct as well as its inclusion in IDEA. Traditional SLD identification practices (i.e., ability-achievement discrepancy methods) along with problems related to SLD identification practices were presented. Information related to the inclusion of RTI for SLD determinations within IDEA was briefly reviewed. Research

questions and hypotheses as well as assumptions and limitations of the study were described. Terms related to the study were also defined.

CHAPTER 2

REVIEW OF LITERATURE

This chapter reviews the historical context of specific learning disability (SLD), from the development of the construct and early identification practices through the paradigm shift that lead to the use of response to intervention (RTI) for identifying SLD. Furthermore, core features of RTI as a framework for improving student outcomes are discussed. Technically-sound assessment practices within an RTI framework for the purposes of identifying SLD are also discussed.

Early Definitions of SLD: Paving the Way for the IDEA

Research on learning disabilities can be traced back to the early 1800s (Hallahan & Mercer, 2001). Early European work focused on the connection between brain injury and impairment, especially related to speech and language, and researchers in the United States initially focused on the relationship between brain processes and behavior, including language and reading disabilities' relationship to perceptual, perceptual motor, and attention disabilities (Hallahan & Mercer, 2001). Orton (1937, as cited in Zumeta, Zirkel, & Danielson, 2014) theorized that reading disabilities were internal, brain-based problems that needed to be addressed by teaching phonics-based instruction, including phonological and phonemic awareness. Additionally, Monroe (1932, as cited in Zumeta et al., 2014), Orton's research associate, proposed that a discrepancy between potential and actual performance relates to the concept of learning disability. The term learning disability was invented by Samuel Kirk (1962) prior to the passage of the Education for all Handicapped Children Act (EHA), and parts of his definition are still used today. Kirk defined learning disability as

a retardation, disorder, or delayed development in one or more of the processes of speech, language, reading, writing, arithmetic, or other school subjects resulting from a psychological handicap caused by a possible cerebral dysfunction and/or emotional or behavioral disturbances. It is not the result of mental retardation, sensory deprivation, or cultural and instructional factors. (p. 263)

In the years following Kirk's first formal definition, which suggested that a disorder in psychological processing was the root of SLD, Bateman (1965), one of Kirk's students, included Monroe's (1932) idea of discrepancy in her definition. Bateman suggested that SLD could be identified by a significant discrepancy between ability and achievement. Bateman's new definition emphasized underachievement, and proposed that

Children who have learning disorders are those who manifest an educationally significant discrepancy between their estimated intellectual potential and actual level of performance related to basic disorders in the learning process, which may or may not be accompanied by demonstrable central nervous system dysfunction and which are not secondary to generalized mental retardation, educational or cultural deprivation, severe emotional disturbance, or sensory loss. (p. 220)

Bateman's definition laid the foundation for researchers to quantify unexpected underachievement as a discrepancy between ability and achievement.

As federal legislation authorizing supports to students with disabilities expanded in the 1950s and 1960s, increased lobbying resulted in federal legislation that supported programs aimed at educating students with SLD (Zumeta, Zirkel, & Danielson, 2014). In 1968, the National Advisory Committee on Handicapped Children, chaired by Samuel Kirk, developed a definition that served as the basis of the federal definition and further highlighted the specificity

of SLD, stipulating that it is not a generalized hindrance of skill development (Kavale & Forness, 2000). The National Advisory Committee's definition read:

> Children with special (specific) learning disabilities exhibit a disorder in one or more of
> the basic psychological processes involved in understanding or in using spoken and
> written language. These may be manifested in disorders of listening, thinking, talking,
> reading, writing, spelling, or arithmetic. They include conditions which have been
> referred to as perceptual handicaps, brain injury, minimal brain dysfunction, dyslexia,
> developmental aphasia, etc. They do not include learning problems that are primarily due
> to visual, hearing, or motor handicaps, to mental retardation, emotional disturbance, or to
> environmental disadvantage. (p. 34)

Following the first national definition of SLD, the Children with Specific Learning Disabilities Act of 1969 was passed (Zumeta et al., 2014), which made SLD a formal disability category and set the stage for its inclusion in the 1975 EHA. The definition posed by the National Advisory Committee on Handicapped Children in 1968 became the most popular definition among departments of education, leading to its adoption as the SLD definition in the EHA shortly thereafter (Hallahan & Mercer, 2001), and in 1975, the EHA defined SLD as follows:

> a disorder in one or more of the basic psychological processes involved in understanding
> or in using language, spoken or written, which disorder may manifest itself in imperfect
> ability to listen, think speak, read, write, spell or do mathematical calculations. Such
> disorders include such conditions as perceptual handicaps, brain injury, minimal brain
> dysfunction, dyslexia, and developmental aphasia. Such term does not include children
> who have learning problems which are primarily the result of visual, hearing, or motor

18

handicaps, of mental retardation, of emotional disturbance, or environmental, cultural, or

economic disadvantage. (§620[b][4][a])

Kavale and Forness (2000) explained that, despite SLD's inclusion in the EHA, it had not

been operationalized in a way that practitioners could apply, which lead to the operationalization

in the 1977 Federal Regulations, indicating

A team may determine that a child has a specific learning disability if: (1) The

child does not achieve commensurate with his or her age and ability levels in one

or more of the areas listed in paragraph (a)(2) of this section, when provided with

learning experiences appropriate for the child's age and ability levels; and (2) The

team finds the child has a severe discrepancy between achievement and ability in

one or more of the following areas: (1) Oral expression; (ii) Listening

comprehension; (iii) Written expression; (iv) Basic reading skill; (v) Reading

comprehension; (vi) Mathematics calculation; (vii) Mathematics reasoning. (US

Department of Education, 1977, p. 65083)

Since the passage of the 1975 EHA, LEAs were required to determine ways to document the

ambiguous criteria outlined in the 1977 regulations. Zumeta et al. (2014) explained that the

definition of SLD includes four main tenets: discrepancy, heterogeneity, exclusion, and internal

characteristics. Given the operationalization of SLD in 1977, Zumeta and colleagues argued that

discrepancy has historically been the single most important aspect in the identification of SLD.

In fact, 2002 Federal Regulations indicated that a student must demonstrate a "severe

discrepancy between achievement and ability that is not correctable without special education

and related services" (34 CFR 300.543(a)(6), p. 76). Heterogeneity refers to the specific aspect

of SLD such that it is not a global deficiency of skills (i.e., as in an intellectual disability),

suggesting, for example, some students may have an SLD in basic reading skills whereas others may have an SLD in math computation. Exclusion suggests that other factors may not contribute to an individual's low achievement, such as lack of instruction, limited English proficiency, economic disadvantage, or other disabling conditions (e.g., intellectual disability, vision, hearing deficiencies). The fourth tenet, that SLD is internal to a student, is related to exclusion, as the low achievement cannot be contributed to an external factor, such as lack of instruction.

Despite SLD being the only IDEA disability category with documented regulations for determining eligibility above and beyond the federal definition (Lichtenstein, 2014), Kavale and Forness (2000) contended that the SLD operationalization has been too vague as it provides a general concept rather than describing specific criteria for meeting the condition of SLD. As such, assessing each aspect of SLD has been largely left up to LEAs. Given that the definition of SLD has remained largely unchanged over time, documentation of discrepancy as the "greatest single driver in identification and eligibility determinations for students," (Zumeta et al., 2014, p. 11) resulted in a prevalence of assessments to document a discrepancy.

Although much has changed regarding school psychological practice since SLD was included in the 1975 EHA, the definition of SLD has remained largely unchanged over time. The 2004 reauthorization of IDEA and the 2006 regulations, however, changed the ways in which LEAs could identify students with SLD, breaking away from the long-standing tradition of using an ability-achievement discrepancy model. As outlined in IDEA (2004), students must meet four criteria to be identified as having an SLD: 1) failure to meet age or grade-level standards (i.e., inadequate achievement); 2) discrepancy between ability and achievement or as indicated by an inadequate response to scientifically based instruction; 3) rule out of vision, hearing, moor problems, emotional disturbance, cultural or environmental factors, or limited

English proficiency; and 4) rule out of lack of instruction. Using RTI data for SLD

determination calls for reconsideration for applying the first two criteria of SLD identification as

outlined in IDEA, inadequate achievement and lack of progress despite receiving scientifically-

based instruction. Among other data sources, including state- and district-wide assessments,

curriculum-based measurement (CBM) is a key data source in driving educational decisions, and

can be used as evidence to meet Criteria 1 and 2 of the SLD regulation. Although CBM is a

useful tool for examining student achievement and responsiveness to instruction, much more is

known about the adequacy of CBM-R for making educational, and in some cases high-stakes,

decisions compared to CBM for math and written expression. Conclusions about the utility of

using CBM-R to make educational decisions do not necessarily generalize to CBM in math and

written expression (Hosp, Hosp, & Howell, 2006).

Furthermore, consideration of both Criteria 1 and 2 together suggests a student must

demonstrate a dual discrepancy (Fuchs & Fuchs, 1998), which has demonstrated utility for

identifying students with the most persistent and pervasive reading difficulties (Fuchs et al.,

2002; Vellutino et al., 1996). This suggests that students found to be dually discrepant likely

have an SLD when combining data about performance level and rate of improvement (ROI)

while ruling out external factors (i.e., Criterion 3), including lack of instruction (i.e., Criterion 4).

Ability-Achievement Discrepancy

One result of the 1977 Regulations was that practitioners were tasked with quantifying

the discrepancy between ability and achievement. Although intelligence is not mentioned in the

regulations, it was used to represent ability, leading to widespread use of tests of cognitive ability

when evaluating for SLD (Kovaleski et al., 2013). Psychometric approaches to SLD

identification using cognitive tests included identifying standard score discrepancies, using

21

mathematical approaches to identify discrepancies between scores, and by identifying discrepancies in a cognitive profile (Ahearn, 2003).

Tests of cognitive ability have historically been used to differentiate between slow learners (i.e., low cognitive abilities and low achievement scores) from students exhibiting unexpected low achievement (i.e., average cognitive abilities and low achievement scores). Additionally, myriad problems have been associated with the use of an ability-achievement discrepancy for SLD identification.

Difficulty Classifying SLD Using Ability-Achievement Discrepancy

Concerns related to identifying SLD based on the operationalization of an ability-achievement discrepancy is rampant in the SLD identification literature historically. Review of the literature reveals problems differentiating between low-achievers identified as SLD and not SLD based on their achievement discrepancies (Ysseldyke et al., 1982) and difficulties applying discrepancy criteria to correctly classify SLD students (Kavale & Reese, 1992; Peterson & Shinn, 2002). Additionally, Ysseldyke, Vanderwood, and Shriner (1997) reported that being referred for an evaluation is the best predictor of whether a student is found to have an SLD and be eligible for special education services.

An early study by Ysseldyke and colleagues (1982) raised concerns about which students were being identified as SLD in schools. By examining academic performance of students identified as SLD and students identified as low achievers (but not SLD), problems differentiating between students with SLD and students without SLD became evident. Results of this study revealed students demonstrating low achievement who were not receiving special education services were found with similar achievement discrepancies as students identified with an SLD. Despite the similarities between students determined to have or not have SLD, the

authors found students who were rated as misbehaving were more likely to be identified as SLD. The results suggested confusion about students who were SLD, that practitioners were potentially disregarding operational definitions of SLD, and that extraneous factors (e.g., student misbehavior) may be influencing eligibility decisions. Whatever the causes, two conclusions could be drawn. The first conclusion was practitioners were missing students demonstrating low achievement who were in fact students with SLD. The second conclusion one could postulate from Ysseldyke et al.'s results was many students who were underachievers were misidentified as SLD.

A classic study by Ysseldyke, Algozzine, and Epps (1983) echoed the idea that many students may be misidentified as SLD. Ysseldyke and colleagues examined the proportion of students who would be identified as SLD based on any one of 17 operationalizations of learning disability at the time. SLD definitions included ability-achievement discrepancies, low achievement, and scatter analyses. Out of 248 students receiving only general education support in their schools, 85% met criteria of at least one of the definitions of SLD used at the time.

Additionally, a study by Ysseldyke and Algozzine (1981) demonstrated team decisions about special education eligibility are strongly influenced by demographic (e.g., sex, socioeconomic status), physical (e.g., attractiveness based on photographs of previously judged attractive or unattractive children), and behavioral characteristics of students. Final decisions made about special education eligibility were based more on characteristics of students and the nature of the referral. The results suggested teams ignored assessment data, even if the data indicated a student demonstrated average performance, and more strongly utilized extraneous information to make eligibility decisions.

Kavale and Reese (1992) examined statewide data to obtain a profile of students identified with SLD in Iowa. Results of their analysis revealed that students with SLD generally had average IQs, were referred for evaluation due to poor academic performance, were identified as eligible for special education services by sixth grade (with approximately 60% of students identified by third grade), and demonstrated a significant academic discrepancy of performance approximately three years below grade level. When examining the ability-achievement discrepancy criterion, however, just 55% of the total sample of SLD students met this criterion (range 32% to 75% across all education agencies within the state), which suggests that low achievement is the primary factor driving SLD identification.

Peterson and Shinn (2002) further evaluated what type of low achievement constitutes SLD by examining three types of low achievement: an IQ-achievement discrepancy, a severe achievement discrepancy (i.e., low achievement compared to a national normative sample), and a relative achievement discrepancy (i.e., low achievement compared to a local standard). Since the federal operationalization of SLD included an ability-achievement discrepancy, it would be expected that a child identified as SLD would be identified across study sites using these criteria. Results revealed, however, the relative achievement discrepancy appeared to best explain school-based SLD identification practices.

Although low achievement appears to be a defining characteristic of SLD, other studies have suggested seemingly extraneous variables may explain SLD categorization. Lester and Kelman (1997) found that demographic and sociopolitical variables seemingly unrelated to organic disability prevalence were more strongly related to SLD diagnosis compared to disabilities with distinct physical indicators (e.g., physical disability). Results also indicated SLD prevalence rates across states were influenced by demographic and sociopolitical

information (e.g., statewide poverty level). Additionally, Singer et al. (1989) demonstrated wide variations in prevalence rates for students identified as SLD with much heterogeneity of students within this group. For students identified as SLD in their home district, 64% would retain their classification as having a learning disability if they moved. Furthermore, for all students that would be reclassified, students were most likely labeled as learning disabled, suggesting it is a catch-all or "category of last resort" (Singer et al., 1989, p. 277), potentially due to the heterogeneous nature of the group.

Historically, the concept of SLD has frequently been examined, with research demonstrating the questionable validity of the category as well as the methods used to identify it. Since the addition of SLD to IDEA in 1975, Ysseldyke and colleagues have revealed many problems with the way teams make decisions about students' with SLD, including inconsistent application of data to make eligibility decisions (Ysseldyke et al., 1982, 1983) and unrelated variables influencing team decisions more than student performance data (Ysseldyke & Algozzine, 1981). Singer et al.'s (1989) work echoes Ysseldyke et al.'s (1982) research demonstrating that the definition of SLD may be so broad that any number of students may be determined to meet criteria depending on which criteria the team deemed important. More recent research suggests not much has change as evidenced by Lester and Kelman (1997) and Peterson and Shinn (2002). Furthermore, contemporary research has demonstrated that an IQ-achievement discrepancy is an artificial classification (Francis et al., 2005) based on superficially objective criteria informed by federal regulation (Lichtenstein, 2014).

Wait to Fail

In addition to difficulties consistently applying ability-achievement discrepancy criteria, Lyon and colleagues (2001) explained that using an IQ-achievement discrepancy may do more

harm than good for students. An unanticipated outcome of the ability-achievement discrepancy approach is commonly referred to as *waiting to fail* (Brown-Chidsey, 2007; Lyon et al., 2001; Lyon & Fletcher, 2001). The ability-achievement discrepancy approach requires students to demonstrate a discrepancy between their IQ and achievement test scores, oftentimes resulting in students waiting without supports until their performance falls significantly below their predicted performance (based on IQ testing), essentially waiting until the discrepancy becomes large enough to be significant. This discrepancy, however cannot be reliably calculated until students are in approximately third grade (Lyon & Fletcher, 2001).

Using IQ as an indicator of SLD may not only delay identification, but more importantly, it also delays supports needed by students, until the point that problems are difficult to remediate (Coyne, Zipoli, & Ruby, 2006; Lyon et al., 2001). Lyon and Fletcher (2001) explained that adults' reading ability is related to their ability to read at age nine and younger. As evidence continued to build showing early identification and intervention practices could improve students' outcomes (Speece & Case, 2001; Torgesen, 2009; Vellutino et al., 1996; Vellutino, Scanlon, Zhang, & Schatscneider, 2008), it is clear why the United States Department of Education (2005) commented that relying on an IQ-achievement discrepancy may delay intervention.

Difficulty Informing Intervention

Lyon and colleagues (2001) explained that the detriments caused by delaying identification of SLD are further compounded by the emphasis of an evaluation on eligibility, not intervention. Initial educational evaluations have historically relied on identifying a discrepancy based on assessed IQ score instead of using assessments to identify students' individualized academic needs. Forness, Keogh, Macmillan, Kavale, and Gresham (1998) argued that

measuring IQ is irrelevant when attempting to ameliorate learning problems and measuring

improvements in student skills. They explained that defining outcomes of special education in

terms of only IQ and achievement is inappropriate. They also suggested that sophisticated

assessments are not needed to identify academic failure, and the search for within-child problems

has only perpetuated the search for measurement tools that provide little instructional relevance.

Furthermore, research has demonstrated the limitations of IQ testing when identifying

SLD. There are no meaningful differences between nondiscrepant and discrepant low achievers

on measures of literacy development (Gresham & Vellutino, 2010; Hoskyn & Swanson, 2000;

Stuebing, Fletcher, LeDoux, Shaywitz, & Shaywitz, 2002) or on their demonstrated response to

intervention (Gresham & Vellutino, 2010; Stuebing, Barth, Molfese, Weiss, & Fletcher, 2009;

Vellutino et al., 1996). Given the limited connections between IQ assessment and intervention, it

is no wonder that the effectiveness of special education services have not been realized to their

full potential (Kavale & Forness, 2000).

When questioning what is special about special education, Vaughn and Linan-Thompson

(2003) explained that as a result of cognitive testing, early specialized instruction focused on

remediating processing disorders, but many treatments failed to improve academic outcomes. As

explained by Ysseldyke and Reschly (2014), many treatments were based on aptitude-treatment

interactions (ATIs). An ATI approach involved examining aptitude and processing differences

of individuals, treatments for different problems, and interactions between them. ATI dates back

to Cronbach (1957), who, at the time, believed that individualized treatments could be selected

based on an individual's specific aptitudes and presenting problems. Although this idea was and

continues to be greatly attractive (Feifer, 2008), after many failed attempts to demonstrate the

effectiveness of ATI in the 1960s and 1970s, Cronbach himself described ATI as a "hall of mirrors that extends to infinity" (Cronbach, 1975, p. 119).

Although rare, statistically significant interactions were found, but when they were, the interactions were weak and often unable to be replicated (Ysseldyke & Reschly, 2014). Poor outcomes from an ATI approach were likely because underlying processing deficits were not reliably identified and treatments were not tailored to specific learning difficulties (Ysseldyke & Salvia, 1974). Hammill and Larson (1974) demonstrated early evidence that using treatments for students' assessed processing difficulties provided little guidance about what or how to teach. Ysseldyke and Salvia (1974) subsequently described two models for diagnostic-prescriptive teaching, including methods based on ATI and assessment of students' skills to determine what and how to teach. Problem solving and examining students' RTI is the modern application of Cronbach's (1975) recommendations after the ATI approach failed to improve student outcomes. The interest in ATI and its effects on learning, however, has persisted as it has "strong intuitive appeal" (Burns, 2016, p. 3).

Despite this appeal, research continues to demonstrate a lack of support for ATIs (Burns, 2016). Meta-analyses examining the impact of cognitive interventions on reading and math achievement found little impact when compared to academic interventions (Kearns & Fuchs, 2003), and working memory interventions for improving math and reading were negligible (Melby-Lervag & Hulme, 2013; Schwaighofer, Fischer, & Buhner, 2015). When applied to specially designed instruction, too much time had been spent attempting to ameliorate processing disorders rather than focusing on providing instruction to improve academic skill development (Vaughn & Linan-Thompson, 2003). Much of what was considered special about those teaching strategies has had little effect on student learning. Significant outcomes have been realized when

instruction provided to students with SLD is specific, explicit, intentional, and matched to student need. Although it is important to note that phonological processes are related to later reading skills, these processes may also be represented as phonemic awareness, which can be observed as a precursor reading skill (National Reading Panel, 2000) rather than an amorphous psychological process. Planning for and providing instruction that meets those criteria does not require assessment of IQ or underlying cognitive processes.

Education Reform: A Shift Towards Standards and Accountability

Concerned with the quality of public education in the United States, the National Commission on Excellence in Education (1983) assessed the status of the public school system. This report demonstrated concern that the United States' educational system was not competitive internationally and was the first step towards large-scale education reform calling for standards-based education and regular achievement testing to assess student progress and school quality. These principles were echoed in the 2002 reauthorization of the Elementary and Secondary Education Act (No Child Left Behind Act [NCLB], 2002).

One goal of NCLB, however, was guaranteeing progress towards attainment of basic skills for all students, including those with disabilities (NCLB, 2002). To promote this goal, NCLB required all students be assessed to document adequate yearly progress on state assessments. Furthermore, schools were required to employ highly qualified teachers (i.e., teachers with certification in a subject area that they teach) and use scientifically-based instructional practices to improve student achievement.

Around the time NCLB became enacted, two other influential events occurred, the release of the National Reading Panel's (2000) report and the Learning Disabilities Summit of 2001. The National Reading Panel reviewed available research to determine how students learn

to read, identified evidence-based teaching strategies for reading, and explained what should be taught to students. The Panel identified five big ideas in reading, including phonemic awareness, alphabetic principle, fluency, vocabulary, and comprehension; techniques for teaching these skills were summarized in the report as well. Additionally, NCLB (2002) incorporated conclusions of the Panel for improving outcomes.

The Learning Disabilities Summit was part of a national initiative by the United States Department of Education Office of Special Education Programs that examined federal regulations for identifying learning disabilities (Elksnin et al., 2001). The purpose of the Summit was to identify key concerns regarding SLD identification as well as to identify whether changes to SLD identification should occur based on review of research, input from experts in the field, and information gathered from application of SLD identification practices. A work group including researchers, practitioners, parents, and politicians was convened. White papers were prepared in the following areas: early childhood/early identification, classification of SLD, historical perspectives of SLD, decision-making approaches of SLD, discrepancy models, RTI approaches for SLD identification, processing weakness approaches, use of clinical judgment, and differentiating low achievement from SLD.

Themes emerged challenging federal SLD criteria at the time of the Summit. One such theme that emerged was the need to provide early intervention and prevention services, rather than relying on a discrepancy model that delays services to students in need (Jenkins & O'Connor, 2001). Additionally, early intervention to improve phonemic awareness and phonics skills is emphasized. Another theme that emerged was that it is impossible to reliably differentiate between students with SLD and regular low achievers using a discrepancy approach (Fletcher et al., 2001). Furthermore, given the questionable validity of discrepancy definitions of

SLD, dissatisfaction with the operationalization of SLD was noted (Kavale, 2001), and inability to meaningfully assess internal processes was explained (Torgesen, 2001). Given these themes, using an RTI model to conceptualize treatment resistant individuals as having an SLD was recommended (Gresham, 2001).

Enacting RTI Into Law

The culmination of each of these events was the inclusion of RTI as a way to identify SLD with the 2004 reauthorization of IDEA. The 2006 Federal Regulations stipulated that states

> (M)ust not require the use of a severe discrepancy between intellectual ability and achievement for determining whether a child has a specific learning disability, as defined in 34 CFR 300.8(c)(10); Must permit the use of a process based on the child's response to scientific, research-based intervention; and May permit the use of other alternative research-based procedures for determining whether a child has a specific learning disability, as defined in 34 CFR 300.8(c)(10). (U. S. Department of Education, 2006, p. 46786)

Furthermore, schools were now required to demonstrate that scientifically based instructional practices were used and had to document that students received the essential components of reading instruction, including the recommendations by the National Reading Panel (2000), in order to rule out the possibility that inadequate instructional practices were contributing to students' skill deficits. Although the federal regulations operationalized RTI for the purpose of special education identification practices, schools using RTI for SLD identification needed to implement system-wide RTI practices, including screening, intervening, and progress monitoring to adequately assess students' responsiveness.

Despite RTI's inclusion in federal law, the resulting regulations did not include specificity regarding RTI approaches or models to make eligibility decisions. Lack of clarity from the 2006 regulations, especially related to the needed magnitude of achievement and progress discrepancies, resulted in variability in identification practices, which may explain the continued reliance on IQ-achievement discrepancy approaches to SLD identification (Hauerwas et al., 2013; Zumeta et al., 2014). Additionally, how a student's RTI is assessed is left largely up to the states, as minimal guidance beyond federal language is often provided (Hauerwas et al., 2013). Flinn (2015) found that some states have provided specific guidelines. Additionally, Maki and colleagues (2016) reported variability across states' regulations and guidelines, and "outdated, unsupported, or vaguely described" (p. 466) SLD guidelines were noted. The inconsistency with which SLD is operationalized by states may result in inconsistent implementation of identification practices across states and practitioners. Even when given clear guidance, however, practitioners were found to inconsistently apply guidelines when making decisions about whether a student meets criteria for an SLD (Maki et al., 2016). Although RTI may be considered by IDEA regulations as a primary way for identifying SLD, few details have been addressed by the federal government, including how to determine appropriate instruction and inadequate progress. To date, the only professional contributions to this matter have been offered by Kovaleski et al. (2013) and the National Center on Learning Disabilities' (n.d.) RTI-Based SLD Identification Toolkit.

Moreover, using RTI for SLD identification is premised on well-functioning general education services (Zumeta et al., 2014), as RTI is also a school-improvement initiative that aims to improve outcomes for all students through successful general education supports. Without adequate core instruction it becomes increasingly difficult to determine whether a student's low

achievement is internal or resulting from lack of quality instruction. This problem is exacerbated when evaluating academic skill areas in which little evidence-based interventions exist, and SLD identification becomes increasingly challenging when system-wide RTI practices are not implemented with fidelity as poor implementation of RTI may result in students floundering in general education tiers of support before a referral for evaluation occurs.

RTI: A Multi-Tiered System of Support

Fuchs, Mock, Morgan, and Young (2003) explained that an RTI approach to instruction results in both a way for evaluating students for special education and a way to improve student outcomes. Torgesen (2009) further differentiated between an RTI instructional model and the RTI diagnostic approach. The RTI instructional model, hereon referred to as a multi-tiered system of support (MTSS), is "a method for increasing the capacity of schools to respond effectively to the diverse learning and behavioral support needs of their students" (p. 38). The RTI diagnostic approach is a way of "determining eligibility for special education services" (p. 38) under IDEA. For clarity throughout the rest of this manuscript, the tiered instructional model will be referred to as MTSS, and the diagnostic approach will be referred to as RTI.

An MTSS approach includes effective core instructional practices for all students, remediation and preventative supports for students demonstrating learning difficulties (rather than using a wait-to-fail approach), data-based decision making to improve instruction, and finally demonstrating that lack of appropriate instruction is not the primary reason for underachievement and eligibility for special education. Although limited guidance exists for using RTI for SLD identification (Hauerwas et al., 2013), MTSS has been and remains an important school-improvement framework. Despite disagreement regarding the use of RTI for special education eligibility decisions, researchers agree that an MTSS approach should be used

33

to improve instruction and students' outcomes (Elksnin et al., 2001; Kavale & Flanagan, 2007; Lichtenstein, 2014). As a framework for improving student outcomes, Tilly (2006) explained MTSS is a system that results in appropriate resource allocation to address all students' needs using research-based instructional practices by making data-based decisions.

Core Features of MTSS

A United States Department of Education Office of Special Education and Rehabilitative Services letter (2011) described core features of MTSS models. Included in their characteristics of MTSS models were high quality core instruction from general education, regular progress monitoring, universal screening, and tiered support systems with more intense services provided to students in need. This letter explained that timely identification of students in need of additional supports can be ensured when these features of an MTSS framework are in place. Additionally, the letter contends that LEAs implementing these core characteristics may be able to refer for evaluation and identify students with SLD based on their responsiveness to intervention.

A Three-Tiered Model

As explained by Tilly (2006), MTSS is an educational improvement system that benefits all students, such that general and special education supports are not distinct entities. Services are delivered along a continuum with students who have more intense needs receiving more intensive supports. Furthermore, as a result of the reauthorization of IDEA (2004), 15% of special education funds could be used to support general education initiatives for this type of prevention. Additionally, as Vaughn and Linan-Thompson (2003) pointed out, significant outcomes have been noted for students when instruction is specific, explicit, intentional, and matched to student need. Although these are common characteristics of specially designed

instruction, their application is not unique to special education as they are simply effective instructional features and may be applied across a continuum of supports. Although multi-tiered frameworks may differ in how many tiers of support they include, three-tiered models frequently represented as a triangle are most often implemented (Stoiber, 2014; Tilly, 2008).

Tier 1. Tier 1, also known as core or universal instruction, includes the standards and benchmarks taught to all students (Tilly, 2008). According to Kovaleski and colleagues (2013), beyond standards-based general education instruction, Tier 1 instruction includes three other essential components: scientifically-based instruction, universal screening, and data-based decision making.

According to Burns and Gibbons (2008), core instruction should also incorporate a brisk pace with many opportunities to respond. Feedback regarding student performance should be provided frequently, and student progress should guide instructional pacing. Class-wide intervention strategies such as peer tutoring may be incorporated as well. For literacy, core instruction includes components suggested by the National Reading Panel (2000), including the "big five ideas in reading:" phonological awareness, phonics, vocabulary, fluency, and comprehension. Additionally, many schools have implemented a literacy block of 90 minutes per day at the primary level for their core instructional block. All students in each grade level are provided with this core instruction. If effective, Tier 1 supports should allow for approximately 80% to 90% of students to be successful at meeting grade-level benchmarks (Tilly, 2008).

Tilly (2008) explained that student performance data are used to guide decision-making. Universal screening data and formative assessments allow teachers to determine whether instruction is meeting the majority of their students' needs. Ensuring there is an appropriate

match between core instruction and students' needs is critical, and ecological assessment may provide additional information about match (Burns & Gibbons, 2008). By collecting large-scale performance data at regular intervals throughout the year, schools can determine whether their instruction meets the requirement of being high quality, such that they are implementing instructional practices that have shown to be effective for bringing at least 80% of students to proficiency on measures of basic skills (Kovaleski et al., 2013).

Additionally, regular administration of benchmark assessments allows for the early identification of students in need of core plus supplemental supports. Teachers no longer have to rely on an unreliable referral process for identifying students in need of supports beyond core instruction (Tilly, 2008).

Tier 2. Tier 2, also known as supplemental or targeted instruction, is for the approximately 10% to 15% of students who are at risk for not meeting grade-level benchmarks with only core instruction (Tilly, 2008). As indicated by its name, supplemental instruction is just that – supplemental. Students receiving Tier 2 support receive supplemental support in addition to daily research-based core instruction; it is not a replacement for core instruction (Kovaleski et al., 2013; Stoiber, 2014; Tilly, 2008). Tilly (2008) explained that Tier 2 instruction can include more opportunities to access core instruction, including additional opportunities for practice and more time spent receiving core instruction. Additionally, Stoiber (2014) indicated supplemental instruction should be linked to the core but with more time, exposure, and opportunities to respond with teacher feedback. For some students, this additional access to core instruction is effective at bringing them to proficiency, and some students may not require supplemental instruction over the long term. Regardless, students receiving supplemental instruction are not suspected of having a disability at this point.

Although Tier 2 supports are more customized to the learners, they may be provided in a variety of ways. Fuchs et al. (2003) indicated there are two schools of thought regarding what supplemental instruction may look like. The first is based on a standard protocol approach. In this approach, students identified as being at risk receive more intensive instruction using programs that are largely effective at remediating a range of skills. Students with similar deficits may be identified and grouped by need based on information gathered from universal screeners (Burns & Gibbons, 2008; Kovaleski et al., 2013). Additionally, using a manualized intervention may allow for greater intervention integrity. The second approach is based on a problem-solving model. In this approach, additional assessments are administered to students identified as at risk to identify specific skills in need of remediation. It is believed that by targeting students' specific skills, more customized interventions may be provided to students, as students are not receiving additional support for skills that do not need remediated. This intermediate step before truly individualizing for students may allow for more efficient resource allocation (Burns & Gibbons, 2008).

Regardless of the way in which Tier 2 supports are provided, supplemental instruction needs to be more explicit (i.e., very clear, without assuming students will build skills on their own), more intensive (e.g., by increasing instructional time or decreasing student to teacher ratio), and more supportive (e.g., by providing emotional and cognitive support through scaffolded instruction; Torgesen, 2004). Likewise, Kovaleski et al. (2013) explained that although instruction may be intensified by adding more time, intensity of instruction should not be considered solely in number of minutes but rather in the type of instruction provided. Intensified instruction should include more opportunities to respond, include more immediate

37

feedback, opportunities for mastery of content before moving on to the next subject, and goal setting and progress monitoring.

Tier 2 instruction is typically provided for 30 minutes three to five times a week (Burns & Gibbons, 2008). Small groups of four to six students are formed based on need, which increase the intensity of the intervention. Intervention may be provided by a licensed teacher, a paraprofessional, a peer tutor, or a volunteer. Students' progress should be monitored at least every two weeks to evaluate the effects of supplemental instruction (Kovaleski et al., 2013). At regular intervals (e.g., every 10 weeks), student performance should be evaluated when determining whether to continue or discontinue supplemental supports for a student.

Tier 3. Tier 3, also known as intensive intervention, is required in addition to core instruction for approximately one to five percent of students (Burns & Gibbons, 2008; Tilly, 2008). Students who do not make adequate progress with supplemental intervention and who require more individualization may be considered for Tier 3 supports (Kovaleski et al., 2013). Although in some systems Tier 3 supports may be considered special education (Vaughn & Linan-Thompson, 2003), it does not necessarily equate with special education. Some argue that Tier 3 supports are a general education support that provides more individualized assessment and intervention (Kovaleski et al., 2013) or that may include special education services as a package of supports (Tilly, 2008). Tier 3 supports refer to intensive instructional supports, not a specific program or placement and may also include supports for students who are gifted and talented or English language learners.

Individualized Tier 3 supports include precise instructional targets, prerequisite skill instruction, and explicit and systematic instruction. Explicit and systematic instruction includes modeling, frequent performance feedback, and greater opportunities to respond (Kovaleski et al.,

2013). Data gathered from curriculum-based evaluation (Hosp, Hosp, Howell, & Allison, 2014)

and considered by problem-solving teams are used to design supports. Burns and Gibbons

(2008) explained the problem-solving process that occurs at Tier 3 includes defining the

problem, analyzing the problem, developing a hypothesis, developing a plan, implementing the

plan, and evaluating the plan. By defining a problem in terms of the performance gap between a

student's current level of performance and expected performance, school teams can better

understand what resources are required to accelerate students' learning to close the performance

gap (Torgesen, 2004).

At Tier 3 focused, individualized, and intensive interventions should be provided in

addition to core instruction for 60 to 120 minutes per day (Kovaleski et al., 2013) with some

students requiring up to 180 additional minutes of intensive intervention per day (Torgesen,

2004). Although Tier 3 supports are individualized, instruction can still be delivered in small

groups with students grouped based on need, not necessarily grade level. Furthermore, progress

monitoring needs to occur more frequently (i.e., at least weekly) so school teams can readily

determine whether the appropriate supports are in place (Kovaleski et al., 2013).

Special Education in an MTSS Framework

Vaughn and Linan-Thompson (2003) suggested that Tier 3 supports may be provided for

a length of time before determining whether a student is in need of special education. When

considering whether a student should be referred for a special education evaluation, it should be

noted that students eligible for special education are not given support based on a label but rather

their needs as determined by their performance on progress monitoring measures, with much

focus paid to level of performance and ROI. Special education should be thought of as supports

that are so intensive that they require resources beyond what can be provided solely by general

education alone. Students, regardless of their eligibility for special education, should continue to access core supports within an MTSS framework.

Additionally, Kovaleski and colleagues (2013) explained that special education supports are a "service (rather than a place) that can and should be delivered in the least restive environment" (p. 41). Supports may differ in how they are provided, as some students may receive specially designed instruction individually, in a small group, or within the general education classroom. The defining characteristics that guide special education supports are that they are provided by a special education teacher and governed by a student's individualized education program (IEP).

A benefit of using an MTSS framework for service delivery is that IEPs are driven by student need such that students eligible for special education demonstrate that they need supports that cannot be accessed solely through general education to make progress. In a sense, evidence of validity for special education supports is demonstrated through continued assessment, monitoring, and evaluation such that when successful growth is realized, these supports are written into a student's IEP (Fuchs et al., 2002). Furthermore, a treatment validity approach whereby nonresponsiveness to instruction is demonstrated as an indicator of needing more intensive supports is more defensible as special education's impact is evaluated in an ongoing manner through continued progress monitoring. If a student does not demonstrate successful growth, ongoing diagnostic assessment can occur to identify appropriate instructional supports.

RTI as an Assessment System

MTSS as a multi-tiered framework for supporting student growth includes components of assessment such as continuous screening and progress monitoring assessment that allows for an understanding of not only the effectiveness of the system but also the needs of individual

students. Therefore, the data collected as part of routine practices for improving student progress may be analyzed at the individual student level for determination of instructional need, including the potential need for special education services (Kovaleski et al., 2013). RTI has demonstrated benefits as a prereferral problem-solving process and as a tool for assessing students' needs as demonstrated by implementation at the local and state levels (Fuchs et al., 2003; Jimerson, Burns, & VanDerHeyden, 2007).

Technical Adequacy of RTI Decisions

By implementing an RTI framework, educators have data available that allow them to respond to students' needs, including determining when student needs cannot be supported using general education resources alone. VanDerHeyden (2011) explained that for RTI to demonstrate technical adequacy as an assessment tool, every data source must be reliable, valid, and able to identify at-risk students. Universal screening assessments, progress monitoring assessments, and diagnostic assessments must all demonstrate technical adequacy. Additionally, practitioners must appropriately follow decision-making procedures. All decisions must be made on sound decision rules, and all interventions should match student need and be implemented with fidelity. Problems related to classification may arise when using RTI for evaluations, as cut scores applied to decisions are often reflective of local resources or a criterion that is more representative of a research agenda rather than a meaningful outcome for a local sample (VanDerHeyden, 2011).

Although problems accurately identifying students who are academically at risk from a single measure may arise, a gated decision-making model such as RTI may actually improve accuracy of decisions made about student risk (VanDerHeyden, 2011). In fact, research has supported that decisions made about student risk within an RTI framework have both

discriminant and predictive validity (Speece et al., 2003; Vellutino et al., 1996). Furthermore,

VanDerHeyden, Witt, and Gilbertson (2007) demonstrated increased reliability for decisions

about student risk status when using an RTI assessment system.

Assessing Students' Responsiveness

Assessments in an RTI system not only include point-in-time measurements about a

student's current level of performance, but they also provide time series data that can be

analyzed to understand students' progress. As student data may be examined in a variety of

ways, practitioners must understand the best ways to examine students' responsiveness, such as

final benchmark (Good et al., 2001), slope discrepancy (Fuchs, Fuchs, & Compton, 2004), and

dual discrepancy (Fuchs & Fuchs, 1998; Fuchs et al., 2002; Speece & Case, 2001).

A final benchmark approach involves comparing a student's level of performance data to

a criterion suggestive of future success to determine how well a student responded to an

intervention (Good et al., 2001). Fuchs (2003) reported the final benchmark is feasible and

measures post-intervention level but does not consider rate of learning. Therefore, it does not

directly address student responsiveness and violates the tenet of measuring students' ROI

because a student may demonstrate exceptional growth but still not attain a benchmark if initial

level was significantly low. A slope discrepancy approach assesses growth and involves

measuring students' rate of progress to compare their slope to a normative group (Fuchs et al.,

2004). Growth methods base decisions on amount of learning but do not consider whether the

student attained some level of proficiency (Fuchs, 2003). Dual discrepancy models consider

both post-intervention status as well as growth and can be used at any time in the assessment

process, not just following intervention. To detect a dual discrepancy, students must demonstrate

performance significantly below a criterion as well as inadequate growth compared to classroom

peers when receiving effective classroom instruction. A dual discrepancy approach to assessing students' responsiveness to instruction has demonstrated consistent support from the literature (Fuchs et al., 2002; Fuchs & Fuchs, 1998; Speece & Case, 2001; Vellutino et al., 1996). The dual discrepancy's ability to reliably differentiate students most at risk for reading failure (i.e., students with learning disabilities) may be why it is included in federal regulations for identifying SLD (IDEA, 2004; Kovaleski et al., 2013).

Assessments Within an RTI System

Universal screening and progress monitoring are two types of assessment that frequently occur within an RTI system. CBM is frequently used for both types of assessments. For reading, commercially available assessments include the Dynamic Indicators of Basic Early Literacy Skills Next (DIBELS Next; Good et al., 2011), AIMSweb (Pearson Education, 2011), and Formative Assessment System for Teachers (FAST; FastBridge, 2015). Computer-adaptive testing (CAT) is also frequently used within an RTI/MTSS system. CAT selects and administers subsequent test items based on students' responses throughout the test in order to identify skills and problem areas within a particular progression of skills (Kovaleski et al., 2013). Commercially available CATs include STAR Math (Renaissance Learning, 2012a), STAR Reading (Renaissance Learning, 2012b), and Measures of Academic Progress (Northwest Evaluation Association, 2004). Purposes of assessments in RTI include using data to make instructional decisions at the individual (Deno & Mirkin, 1977) and group levels (Kovaleski & Pedersen, 2008) as well as ongoing monitoring of the effectiveness of instruction and intervention. Although both CBM and CAT provide relevant information within an RTI assessment system, this review will focus solely on CBM, particularly oral reading fluency due

to the scope of this study. Conclusions made do not necessarily generalize to CBM in math and written expression.

CBM

CBM was developed to test the effects of data-based program modification (Deno & Mirkin, 1997), which was based on the idea that formative evaluations could guide and improve instruction. Deno (2003) reported CBM was developed at the University of Minnesota Institute for Research on Learning Disabilities, and included generic progress monitoring assessments in reading, spelling, and written expression that were related to general outcomes. Since its development, CBM has been used for screening, assessment for guiding intervention, special education evaluation and placement decisions, formative assessment, evaluation of educational programs and initiatives, and prediction of performance on high-stakes assessments. CBM's use for educational decision making has also been found to improve student outcomes (Fuchs, 1986).

An assessment is curriculum-based when it draws directly from the instructional materials (Deno, 1985). However, Deno (2003) explained CBM refers to standardized assessment practices that include the following characteristics: technically adequate; standardized for reading, writing, and math; standardized stimulus materials that are equivalent and representative of skills; standardized administration practices; performance sampling; multiple equivalent assessments; time efficient; and easy to learn administration. CBM provides information regarding the number of correct and incorrect responses in a set period of time (Deno, 2003). CBM has been developed for reading (reading aloud, cloze passages), writing (word sequences when given a story starter or picture, letter sequences when given a spelling word orally), and solving math problems (digits correct, correct answers).

Standardized administration practices, including duration of measurement and scoring procedures, result in data that can be used for comparing between groups over time and for the development of local norms (Shinn, 1995). By using repeated measures of equivalent assessments, information can be gathered regarding student proficiency across time (e.g., an intervention phase, a school year, etc.). Although CBM was initially developed for teachers to evaluate the effects of their instruction, it has evolved to be used for a variety of other applications, including improving individual instructional programs (through setting goals, monitoring progress, making data-based instructional changes, and evaluating the effects of the changes), predicting performance on outcome assessments (including high-stakes statewide assessments), improving teachers' planning (identifying students' goals and making changes based on growth), developing norms (allowing for comparison of a student to expected performance), improving communication between educators and parents (due to the graphic depiction of progress), universal screening to determine at-risk students (comparing individual performance to group performance), determining the effectiveness of general education interventions and supports, reducing assessment bias, and evaluating special education practices within an RTI model (Deno, 2003).

Universal Screening

Universal screening typically occurs three times each year (i.e., fall, winter, and spring) and involves the administration of a benchmark assessment to all students (Burns & Gibbons, 2008). General outcome measurements using CBM are typically administered for each universal screener. Best practice for screening may be to use a median approach. Previous research suggests a standard set of three CBM-R probes should be used for universal screening (Ardoin & Christ, 2008). According to Kovaleski and colleagues (2013), universal screeners should

directly assess important skills for educational success, predict future performance on a future educational benchmark, be reliable, be administered efficiently with little time needed for administration, and be sensitive to changes in performance. Universal screening occurs for a variety of reasons, including early identification of students who might be at risk for not meeting grade-level expectations. Not meeting criteria on the benchmark assessment triggers the need to monitor students' progress on a more regular basis. Additionally, by understanding whether students are meeting benchmarks, teachers can more effectively differentiate their instruction to meet their students' diverse needs. In addition to helping teachers make instructional changes in their classroom, Burns and Gibbons (2008) explained that universal screening allows for schools to norm student performance for goal setting and evaluate the effectiveness of their instructional supports.

Data analysis teaming. Kovaleski and Pedersen (2014) reported that data analysis teams summarize and review data for decision-making. Data teams include teams of teachers, administrators, and specialist staff (e.g., school psychologists, reading consultants, etc.). Following fall, winter, and spring universal screening periods, data teams review group-level data to provide information about the current status and future outcomes for students and groups. For Tiers 2 and 3, data teams review small group and individual data to determine whether supplemental and intensive interventions are having the desired effect on student performance. By reviewing data, teams can understand the effectiveness instructional supports, identify students needing supplemental supports, identify instructional strategies and interventions to address areas of need, and manage logistical issues resulting from needed changes. Data analysis teaming can occur across all tiers of an MTSS framework with the general structure of a data

meeting including data review, goal setting, identification of strategies, and planning and support for implementation.

Progress Monitoring

Kovaleski and colleagues (2013) explained progress monitoring assessments should demonstrate the same characteristics as universal screeners such that they are associated with important educational outcomes, efficient to administer, provide reliable and valid scores, predict future outcomes, and be able to identify small changes in student performance. Despite these similarities with benchmark assessments, progress monitoring occurs with individual students for a reason beyond screening.

The progress of students not proficient on universal screeners is monitored. Kovaleski et al. (2013) suggested students receiving Tier 2 intervention should be monitored biweekly, and students receiving Tier 3 intervention should be monitored at least weekly. According to Burns and Gibbons (2008), the progress of students who are more discrepant from expectations may be monitored more often. Monitoring academic skills such as math and written expression should likely not be monitored more than biweekly due to the limited sensitivity to change for measurements of these skills. Additionally, monitoring reading at a frequency daily for shorter durations of time (e.g., six weeks) will likely not lead to improvements in growth estimates (Thornblad & Christ, 2014), but more frequent monitoring may allow for teachers to make better instructional decisions (Fuchs, 1986).

Burns and Gibbons (2008) suggested progress monitoring occurs for a variety of reasons. Collecting frequent measurements of student performance using CBM, which is a valid tool for monitoring progress (Shinn, 2008), allows for graphic depiction of students' improvements. Therefore, every student has a graphic depiction of progress towards a goal. Furthermore, the

47

visual depiction of improvement allows teachers, parents, and students to understand students' progress and make important adjustments to students' instruction (Kovaleski et al., 2013). Burns and Gibbons (2008) also explained that progress monitoring data allow school teams to make decisions about instructional materials and grouping of students to ensure that resources are used effectively. Additionally, collecting data results in teachers identifying ways to raise achievement of students by finding more powerful interventions.

Because interventions are implemented to close the achievement gap, it is recommended that progress of students be monitored using grade-level passages when possible (Burns & Gibbons, 2008). By monitoring on grade level, school teams can make decisions about student performance in comparison to peers or grade-level expectations. Shinn, Gleason, and Tindal (1989) suggested, however, that consistent difficulty of a passage set may be of greater importance when monitoring student progress than level of difficulty in passage sets. For students performing well below grade-level expectations, regular progress monitoring at a student's instructional level with occasional progress monitoring at grade-level may be more sensitive to changes in student improvement while still providing normative information about student performance. Additionally, collecting time series data allows for calculation of student ROI, which can be used as a tool to guide teams' decisions about student need. Furthermore, because Criterion 1 of the SLD criteria references failure to meet age or grade-level standards and because grade-level assessment is the only approach studied for SLD identification, assessment at grade level is the only approach sufficient for special education eligibility decision-making (Kovaleski et al., 2013).

Using CBM-R as Part of an Evaluation for Reading SLD Within an RTI System

Scores from CBM, which represent an individual's proficiency in some domain, can be used to identify performance discrepant from expectations to help inform decisions about need for special education (Deno et al., 2001). Given the nature of CBM, information regarding level of performance and rate of change can be obtained. Therefore, evidence of inter-individual differences (i.e., inadequate academic achievement as compared to peers) as well as intra-individual improvement (i.e., ROI over time in response to additional intervention) can be obtained.

Technical Adequacy of CBM-R

As described earlier, CBM data provide information about both current status as well as progress over time. Both are indicators of a student's RTI. Furthermore, as both performance level and ROI, two criteria indicative of SLD, can be obtained from CBM-R data, the technical adequacy of CBM-R for psychoeducational evaluations must be examined. Psychoeducational evaluations must be conducted using assessments that are be reliable, valid, and sensitive to change (American Educational Research Association, American Psychological Association, & National Council on Measurement in Education, 1999). To be useful for decision-making, CBM-R should demonstrate test-retest and alternate forms reliability. It should also be predictive of a future outcome. Research has demonstrated CBM-R meets these conditions (Marston, 1989; Wayman et al., 2007).

Wayman and colleagues (2007) reported that despite the limited face validity of CBM-R, research has supported the relationship between oral reading fluency (ORF) and overall reading proficiency. Comprehension was better predicted by ORF than other standard comprehension measures suggesting that ORF is not just an assessment of processing speed. However, ORF

49

may not be the most appropriate measure for younger and older students. ORF produced a floor effect when used with younger students, and word identification may be a better indicator of reading skill in younger grades in an RTI model. Cloze passages, which include passages where specific words (e.g., every seventh word) are deleted and students must provide the correct word, may be more appropriate for older students (i.e., secondary students) because initial research suggests that growth is not accurately reflected by ORF measures for middle school students.

CBM-R probes do not need to be developed directly from the material that is being instructed because the measures appear to function similarly when developed from different sources (Wayman et al., 2007). Additionally, CBM-R ORF measures can be used across a variety of curricula and pedagogies, and students can be assessed with materials that are at or above their instructional level with limited impact on the technical adequacy. Although CBM-R has a rich history of demonstrated reliability and validity for screening and benchmarking, much is still unknown about the decisions made based on CBM-R progress monitoring data (Ardoin et al., 2013).

Using CBM-R to Obtain Level of Performance Data

Shinn (2008) explained CBM-R may be used as a measure of current level of performance, which is an indicator from which to obtain information regarding a discrepancy in a student's skillset. Both universal screening and progress monitoring data obtained from CBM-R assessments can inform decisions about level of performance (Kovaleski et al., 2013). Descriptions about students' level of performance may be reported using a ratio (e.g., 2.0 times discrepant when dividing the expected score in words correct per minute [WCPM] by the student's score in WCPM) or a percentile rank. Shinn (2002) described a discrepancy of 2.0 times discrepant as significant. Germann and Tindal (1985) demonstrated that use of a 2.0

discrepancy approximated national placement rates. Marson, Tindal, and Dino (1984) however, suggested that using a single discrepancy cut score (e.g., 2.0 times discrepant) may result in different percentages of students identified as discrepant. Kovaleski and colleagues (2013) suggested performance below the 10th percentile rank may signify a discrepant level of performance. Additionally, in terms of level of performance, a wide range of performance levels have been used to signify non-responsiveness in the literature (Foorman et al., 1998; Torgesen et al., 1999; Vellutino et al., 1996).

When considering students' RTI, their performance level should be examined following the provision of an intervention for some period of time because it takes time for the effects of instruction to be realized before post-intervention performance is examined. Christ, Monaghen, Zopluoglu, and Van Norman (2013) explained it may take at least eight weeks for instructional effects to be apparent. Christ and colleagues also suggested that validity, reliability, and diagnostic precision may improve with increases in duration of instruction.

When examining performance following an intervention period, students' final status is examined as an indicator of performance level (Fuchs, 2003; Good et al., 2001). If using progress monitoring data to obtain data about students' level of performance, a series of CBM-R scores may be examined (Kovaleski et al., 2013). For students not responsive to instruction, low performance should be evident at the final performance on progress monitoring assessments.

Using CBM-R to Obtain ROI Data

A key feature of CBM is that it is sensitive to growth, which allows it to be used for assessing student progress over time (Deno, 1985). Collecting and graphing CBM-R time-series data allows for comparison of student progress in response to changes in instructional programs over time (Marston, 1989). ROI, an indicator of growth, provides a quantitative description of

51

student progress over time based on weekly CBM-R assessments. Obtaining information about students' ROI is invaluable for making instructional decisions, comparing instructional effects, and making a special education eligibility decision using RTI (Jenkins, Zumeta, & Dupree, 2005).

Due to the connection between ORF and overall reading ability, CBM-R has been a prime focus of measurement with regards to learning disabilities in reading. Administering CBM-R assessments over time provides an indicator of students' growth with their reading skills. When using CBM-R for monitoring progress, it must be assumed that changes in scores over time are representative in actual changes in reading ability. To model growth using CBM-R, scores must result in an interval scale without floor or ceiling effects; consistency of measured construct and difficulty level are required; and sufficient alternate forms must be available (Deno et al., 2001).

Calculating ROI

Many methods exist for calculating students' ROI. Last minus first, split-middle, Tukey, and ordinary least squares (OLS) regression all provide ways to quantify student progress with varying degrees of precision (Flinn & McCrea, 2013). Shinn et al. (1989) explained the best slope estimation technique must be easy to calculate and interpret, accurate (i.e., limit systematic over- and under-predictions), and precise (i.e., minimize individual prediction errors). Prior to the advent of computer-based methods for examining data trend lines, the methods besides OLS may have been more preferred due to the ease with which they could be calculated by hand. Research has shown, however, that OLS regression provides the most precise depiction of student ROI across time (Deno et al., 2001; Good & Shinn, 1990; Shinn et al., 1989). OLS trend line calculation was most frequently recommended in the literature (Ardoin et al., 2013).

Therefore, normative growth can be modeled using a linear function with slope depicting growth (Deno et al., 2001).

Ensuring a High Quality Data Set

Although CBM-R data collected over time can be used to calculate a slope for a student's ROI, error may impact interpretations of a student's data set and, therefore, decisions made about students. Christ (2006) demonstrated that the standard error of the slope estimate could be greater than actual student growth when error is introduced. Additionally, measurement error may actually influence students' ROI rather than actual growth (Hintze & Christ, 2004). Therefore certain factors must be controlled to have high quality data useful for decision-making at the individual student.

Error can be reduced by ensuring optimal testing conditions, including ensuring consistency across test administrator, location, and probe difficulty, testing in a quiet environment, and using standardized administration procedures (Christ, 2006). Optimal testing conditions that control for confounding variables result in a lower magnitude standard error of the estimate. Additionally, estimates of growth need to be derived from the same passage sets (Ardoin & Christ, 2009). Shinn and colleagues (1989) varied the difficulty of passage sets across groups of students to determine to what extent instructional level of CBM-R passages impacted students' ROI. They found that consistent difficulty of passages within a passage set may be of greater importance than the level of difficulty. More recent research also supports the argument that when estimating student growth using CBM-R, equivalent passages must be used (Ardoin & Christ, 2008; Christ & Ardoin, 2009; Jenkins et al., 2005). Given CBM-R's sensitivity to measurement conditions, higher control and standardization may lead to increased assessment reliability (Christ & Silberglitt, 2007), and greater confidence can be had for high-

stakes decisions by ensuring ideal testing conditions across longer progress monitoring durations (Christ, 2006).

Christ and colleagues (2013) explained validity, reliability, and diagnostic precision improve with increased duration of instruction, suggesting progress monitoring over longer periods of time improves the quality of the dataset. Although it is suggested that the length of time is an important factor to consider before making decisions about student progress, triannual universal screening outcome data likely do not yield satisfactory estimates of ROI (Ardoin & Christ, 2008). Christ et al. (2012) suggested collecting CBM-R progress monitoring data weekly for at least 14 weeks with high quality test conditions. Estimation of student ROI likely improves if a denser progress monitoring schedule is followed, but the quality of the slope estimate will likely not be improved if analyzed prior to 14 weeks of data collection. Thornblad and Christ (2014) substantiated those findings such that a longer duration of progress monitoring and more data points resulted in improved quality of slope estimates. Improved slope estimation was likely more related to longer duration of progress monitoring, however, because daily data collection for six weeks did not result in data that would be reliable or valid for making a high-stakes decision (e.g., special education eligibility).

Based on a review of 78 CBM-R studies, Ardoin et al. (2013) explained that students' progress should be monitored for at least 12 weeks with a dense progress monitoring schedule to make decisions about individual students. They found, however, that collecting seven CBM-R progress monitoring data points (with a range of 3 to 20) was most frequently reported in the literature as adequate for making decisions at the individual student level. Despite evidence that suggests CBM-R is a technically adequate tool that is sensitive for assessing group-level data, much less exists to support the claims made about analyzing individual student CBM-R data for

54

making high-stakes decisions that are much harder to reverse (e.g., eligibility for special education); however, there exist very few alternatives to CBM-R for monitoring progress and assessing the impact of instruction on students. It is important to note, however, that CBM has a rich history of being used at the individual level for making easily reversible decisions (e.g., teachers responding to student data for making instructional decisions; Deno, 2003; Deno & Mirkin, 1977). The important factor when considering the psychometric quality of CBM is that the quality of the data set must match the stakes of the decision being made. The focus of this study includes high-stakes decisions about special education eligibility, and therefore these decisions must be based on data with a higher level of psychometric quality.

Considerations for Decision-Making

Once optimal conditions are ensured for a high quality dataset, practitioners must consider a variety of factors when making decisions based on CBM-R data. ORF assessments using CBM-R have been the primary way educators have assessed reading fluency (Hasbrouck & Tindal, 2006). Additionally, normative ROI data have been developed (Deno et al., 2001; Hasbrouck & Tindal, 2006; Nese et al., 2013), and commercially available CBM-R assessments (e.g., FAST) provide normative growth information. Despite an understanding of normative ROI, much is still unknown about appropriately using this information for high-stakes decision-making.

Review of large-scale studies examining student growth over time revealed seasonal effects on growth across students receiving general education supports only and students provided with special education supports (Ardoin & Christ, 2008; Christ, Silberglitt, Yeo, & Cormier, 2010; Deno et al., 2001; Hasbrouck & Tindal, 2006; Nese et al., 2013). Seasonal differences in growth can be found across all levels of initial proficiency, including students

receiving gifted and talented supports (McGowan, Runge, & Pedersen, 2016). Each of these studies have demonstrated students' ROI is greater from fall to winter than winter to spring. Additionally, Nese et al. (2013) reported greater ROIs were found in the earlier grades such that growth slowed across Grades 1 to 4, suggesting these findings are consistent with the developmental trajectory of reading. Initially, dramatic increases are realized (e.g., in Grades 1 and 2) as students become automatic with decoding skills, and then growth continues to slow through third and fourth grade as students are more proficient at reading and are required to comprehend more texts.

Christ and colleagues (2010) found seasonal effects are less pronounced for special education students. Additionally, differences in growth rates emerged between general education and special education groups beginning in Grade 1 with the differences becoming smaller in the later grades until relatively identical growth is realized for both groups in Grades 5 and 6 (Deno et al., 2001). Students in special education and general education differed in their ROI even when initial level of performance was comparable. Furthermore, students in special education were found to typically grow at rates half as steep as their general education peers.

Although normative data are available, the norm group used for comparison must be considered. Shapiro and Guard (2014) explained risks and benefits of using local and national norms. Local comparison groups may result in students being eligible for special education in one district but not another. Although this has been a criticism of traditional evaluation models, Burns and Gibbons (2008) indicated this is appropriate in an RTI framework because it allows for appropriate resource allocation. When considering the spirit of IDEA, however, special education eligibility should be based on indicators of a "hard" disability rather than simple identification of the most deficient students in a district. Shapiro and Guard (2014) suggested

that, although nationally normed ROIs provide information similar to well-established norm-referenced assessments, over- or underrepresentation of students in special education may occur depending on the overall performance level of a school's student population.

One benefit to using CBM-R data for evaluating ROI is that it can be graphically displayed. Comparison of the trend line of students' progress monitoring data to an aim line allows for analyzing actual student progress compared to desired progress. Visual analysis of student progress, however, may be problematic due to differences in initial performance level between students (VanDerHeyden, Witt, & Barnett, 2005). Students with a higher initial level of performance need to grow at a slower rate than students with more significant discrepancies in initial performance level. So, students with much lower baseline performance levels may be judged as inadequate responders despite making progress that would be greater than typical students. Different decisions about students' RTI may be made depending on whether an aim line is used to guide decisions compared to a dual discrepancy (Burns et al., 2010).

Given that no empirically-validated indicator of what constitutes a deficient ROI exists, Kovaleski et al. (2013) provided recommendations for making decisions about inadequate ROI. They explained teams may examine ROI using a gap analysis or by using ROI trajectories to determine how long it would take a student to reach some level of proficiency. When using a gap analysis, teams compare a student's attained ROI to typical peers' ROI to calculate a discrepancy. Teams may also compare a student's attained ROI to the ROI needed to close the gap. Just conducting an ROI gap analysis, however, provides little information about how much time or what resources will be required to help the student close the gap. Furthermore, there is no published guidance as to what constitutes a deficient ROI. Therefore, teams may consider analyzing a student's growth trajectory based on students' ROIs to determine how much time it

57

would take a student to reach an acceptable level of performance (e.g., the 25th or 40th percentile). By considering what an acceptable level of proficiency is, teams can use ROI trajectories to determine how much time and what resources are required to allow students to sustain their growth to reach that level. For some, it may be reasonable to expect general education interventions over some time will allow students to attain some level of proficiency. For others, the team may determine that the time and resources required for the student to attain an acceptable level of performance exceeds the capacity of general education resources alone, suggesting special education services would better support these students.

Associated Problems With Using RTI for SLD Identification

Although support within the field exists for using RTI as a framework for improving instruction, using RTI for SLD identification is not without its critics. Reynolds and Shaywitz (2009) contended that many unresolved issues associated with RTI, especially related to diagnosing SLD, remain. They argued that RTI is another discrepancy-based model with vague guidelines regarding what constitutes an appropriate response. These concerns are apparent when examining operationalization of inadequate RTI in state guidance documents (Hauerwas et al., 2013). Zumeta et al. (2014) suggested the regulations specifically allowed for flexibility for implementation across states, allowing for frameworks to meet the needs of different environments and populations.

Additionally, Reynolds and Shaywitz (2009) contended that vagaries with regard to inadequate achievement in comparison to peers may result in unreliable diagnoses of SLD. It is suggested that federal guidelines have not adequately defined the peer group to which a comparison should be made. Furthermore, unclear guidelines as to what constitutes an appropriate comparative group may result in contextual diagnoses of SLD whereby a student is

identified in one classroom or school district but not another. Burns and Gibbons (2008) have suggested that this is to be expected; however, this changes the idea that an SLD is related to an internal or biological deficit rather than a combination of environmental factors interacting with a student resulting in the presentation of SLD.

Additionally, as there is no empirically validated method for determining inadequate response, practitioners may not make reliable decisions about students' RTI. Although inclusion of RTI approaches for SLD identification has allowed for better evaluation of the exclusionary and student-centered aspects of SLD due to assessment of environmental characteristics, the lack of clear guidance has led to problems identifying SLD (Zumeta et al., 2014). Zumeta and colleagues also suggested that poor implementation of MTSS and remaining questions about inadequate response may result in students floundering in general education tiers of support before a referral for evaluation occurs, suggesting that the "wait to fail" problem is not addressed. Additionally, Burns and colleagues (2010) explained that decisions about students' RTI would vary depending on whether an aim line or dual discrepancy was used to aid decision-making. Given these concerns, RTI does not have a consistent way for determining lack of response, and various methods of applying RTI criteria may likely result in different students being identified as unresponsive (Reynolds & Shaywitz, 2009). Although much is known about effective practices to support RTI, questions remain about the utility of RTI for validly and reliably identifying SLD (Fuchs & Deshler, 2007).

Summary

As discussed throughout this chapter, the definition of SLD has been surrounded by questions since its invention by Kirk (1962). The lack of clarity with regard to who has an SLD along with the problems associated with the ability-achievement discrepancy, which has

historically been the primary means for identifying students with SLD, has called into question the validity of the SLD construct. There exists a considerable amount of research that suggests extraneous variables influence team decisions about special education eligibility to a greater extent than supposed indicators of SLD (Lester & Kelman, 1997; Singer et al., 1989; Ysseldyke et al., 1982; Ysseldyke & Algozzine, 1981). Also, as many operationalizations of SLD have been used in practice, up to 85% of students may meet criteria for at least one definition of SLD (Ysseldyke et al., 1983).

Based on both local and large scale implementation, RTI has demonstrated effectiveness as a prereferral problem-solving process and as a means for assessing students' needs for more intensive services, including special education (Fuchs et al., 2003; Jimerson et al., 2007). Thus, with the 2004 reauthorization of IDEA, RTI could be used as a means for evaluating SLD. The 2006 regulations attempted to operationalize SLD within an RTI system, but further clarity beyond federal language was warranted. Despite the lack of clarity within federal guidelines, only some states have provided more specific guidelines (Flinn, 2015).

A dual discrepancy, which includes demonstrating achievement significantly below age or grade-level peers (i.e., Criterion 1) and an inadequate ROI (i.e., Criterion 2), is specified in the federal operationalization of SLD. The dual discrepancy appears to have the most support for identifying students with the most significant and persistent reading problems (Fuchs et al., 2002; Fuchs & Fuchs, 1998; Speece & Case, 2001; Vellutino et al., 1996) suggesting it is a valid means for identifying SLD. Despite the dual discrepancy's apparent utility for determining which students are likely SLD and the federal regulations' inclusion of a dual discrepancy, much is still unknown about the extent to which these guidelines are borne out in practice.

Decisions made about students' need using RTI data may be less prone to influence from extraneous variables (Burns & Senesac, 2005; Marston et al., 2003; Speece et al., 2003), suggesting using RTI data improves team decision-making compared to decisions made using ability-achievement discrepancy data. This study will attempt to address the issues associated with identifying SLD by examining the extent that performance level and ROI, two major SLD criteria in an RTI system, predict team decisions about students' need for special education services beyond extraneous student characteristics (e.g., demographic information). This study will provide insight as to whether the SLD construct as indicated by a dual discrepancy is supported by practice.

CHAPTER 3

METHODOLOGY

Federal special education regulations stipulate that a student identified as having a specific learning disability (SLD) based on response to intervention (RTI) data needs to be dually discrepant in level of academic achievement and rate of improvement (ROI; United States Department of Education, 2006). Furthermore, several researchers have suggested that a dual discrepancy can be used to objectively identify students with SLD (Fuchs & Fuchs, 1998; Speece & Case, 2001) and that decisions based on RTI data (i.e., a dual discrepancy) are less prone to influence from extraneous variables (Marston et al., 2003; Speece et al., 2003). Therefore, using a dual discrepancy model for SLD identification appears to overcome many of the limitations to historical SLD identification practices, including the influence of extraneous demographic, physical, and behavioral characteristics of students (Ysseldyke & Algozzine, 1981). As Flinn (2015) and Maki and colleagues (2015) suggested, practitioners are still faced with applying vague guidelines for identifying SLD. Additionally, recent research by Maki and colleagues (2016) suggested that even when objective criteria for making decisions about SLD are given, practitioners may still come to different conclusions about student eligibility for special education.

The purpose of this study was to determine to what extent two major SLD criteria, including a student's level of academic achievement and ROI, predict multidisciplinary evaluation teams' decision-making regarding referral for special education evaluation and regarding special education eligibility. When using RTI data to inform special education eligibility decisions, little is known about how much value is added to a team's decision by ROI

data. By examining school teams' decisions about student need, differential classification of students demonstrating inadequate achievement may be better understood.

Design

This study was intended to be a predictive study (Creswell, 2012) with one outcome variable (student classification) and five predictor variables (level of performance, ROI, sex, race, free or reduced lunch status). Student classification (i.e., the outcome variable) was planned to have three categories: students with oral reading fluency (ORF) skill deficits receiving intensive reading intervention but not referred for special education evaluation, students with ORF deficits referred for a special education evaluation and found not eligible for special education services, and students with ORF deficits referred for special education evaluation and found eligible for special education services. Continuous predictor variables included level of performance and ROI. Categorical predictor variables included sex, race, and free or reduced lunch status. The variable sex has two levels, male and female. Race was collapsed into two levels, historically disproportionately overrepresented in the SLD category and not historically overrepresented in the SLD category. As demonstrated by Skiba and colleagues (2008), Native Americans, African Americans, and Hispanics have historically been overrepresented in the SLD category. Calculations of risk ratios (Boneshefski & Runge, 2014) based on total enrollment data by race available from the National Center for Education Statistics (n.d.) and the eligibility information by race and disability status from the United States Department of Education (n.d.) indicated multi-racial students are also overrepresented in the SLD designation. Asians and Whites have not been historically overrepresented in the SLD category (Skiba et al., 2008). Free or reduced lunch status included two levels, receiving free or reduced lunch and not receiving free or reduced lunch. Level of performance and rate of

improvement were the predictor variables of primary interest. Ancillary variables were to include sex, race, and free and reduced lunch status. These data were planned to be analyzed to predict group membership and determine whether there are meaningful differences between the three groups of students examined in this study.

Population

Anonymous student data obtained from routine progress monitoring practices were collected from the 2015-2016 school year and from August to December of the 2016-2017 school year. Data were collected from second and third grade students who were below benchmark on two or more consecutive universal screening periods using the FAST CBM-R assessment and from second and third grade students with ORF deficits referred for an initial educational evaluation. In the schools involved in this study, students scoring below benchmark on two or more consecutive universal screeners received daily general education intervention in addition to 90-minutes of universal instruction, and their progress was expected to be monitored weekly. Data were collected from elementary schools in districts located in a Midwestern state that have been identified as having at least 75% of students reach proficiency on a state-approved universal screener. Although it is desirable to have at least 80% of students proficient with universal supports (Tilly, 2008), universal screening data were not available for every student in each participating site to account for any potential error in the universal screening scores. Therefore, having 75% of students proficient on the universal screener was selected to include schools that would round up to 80% proficient. Additionally, school sites identified as having 75% of students proficient were selected as a *de facto* measure of fidelity of universal supports within a multi-tiered system of support (MTSS) framework (Kovaleski, Marco-Fies, &

Boneshefski, 2013). Seventy-five public school districts in the Midwestern state met the

inclusion criteria. Participating schools' names were withheld to protect confidentiality.

Sample

Inclusion Criteria

Consent for participation was sought for school districts identified as having at least 75%

of students reach proficiency on a state-approved universal screener during the 2015-2016 school

year. Although in an aspirational three-tiered model, it is desired to have at least 80% of

students identified as proficient with universal supports alone (Tilly, 2008), all measurements

include error. Christ and Silberglitt (2007) reported the median standard error of measurement

for curriculum-based measurement in reading (CBM-R) across grade levels was 10 words correct

per minute (WCPM), which is a generic estimate of SEM. Given the inherent error within any

measurement tool, and the principal investigator's inability to access universal screening data

from every student within each site, it was decided to include districts with scores that round up

to 80% proficient (i.e., 75% proficient and above). District consent was sought for access to

schools' data management systems that housed progress monitoring data, special education

eligibility information, and demographic information. During December 2016 and January 2017,

consent letters were sent to 24 public school districts served by an intermediate educational

service agency that met the inclusion criteria, which was 32% of all public districts meeting

inclusion criteria. Once school districts provided consent for participation, intermediate

educational agency staff were recruited for participation to voluntarily collect data. Districts that

returned consent letters and that had intermediate educational agency staff who agreed to

volunteer for data collection were selected for participation in the study. Consent letters were

received from six school districts (25% of districts recruited), which included 20 schools. Intermediate educational agency staff who agreed to participate provided data for 19 schools.

Data were collected from second and third grade students identified as at risk with ORF deficits receiving intensive general education reading intervention (as indicated by two consecutive scores below benchmark on a universal screener) and second and third grade students referred for an initial educational evaluation with ORF deficits during the 2015-2016 school year and during the first semester of the 2016-2017 school year. Data were collected by intermediate educational agency staff for every second and third grade student with ORF skill deficits referred for an initial educational evaluation during the 2015-2016 school year and during the first semester of the 2016-2017 school year as well as every second and third grade student at risk for reading failure who had not yet been referred during the 2015-2016 school year and during the first semester of the 2016-2017 school year. Students referred for an initial educational evaluation who were also simultaneously evaluated in another performance domain (e.g., math, behavior, etc.) were also included in the sample.

To ensure equivalence of CBM-R assessments, only data obtained from the FAST CBM-R assessment (FastBridge, 2015) was used to determine level and rate of progress. One school district that was recruited for the study reported using AIMSweb; therefore they were excluded from the study. Ardoin and colleagues (2013) indicated that the most frequent recommendation for number of progress monitoring data points in the literature was seven. Therefore, students with at least seven progress monitoring data points were included in the sample.

The expected age range of students was approximately 7 to 9 years-old because this age range is expected to include the majority of second and third grade students. Data were collected for both male and female students of any race. Data were inclusive of students with varying

access to resources as indicated by their free and reduced lunch status. Data were collected for students who may be considered to be English language learners, because limited English proficiency must be ruled out when identifying SLD.

Exclusion Criteria

School districts that did not return a consent letter indicating consent for all data systems listed on the consent form were excluded. Schools who did not have intermediate educational agency staff that volunteered to collect data were excluded. Students enrolled in schools served by the intermediate educational agency who were not identified as persistently at risk based on FAST CBM-R data and also not referred for a special education evaluation during the 2015-2016 school year or first semester of the 2016-2017 school year were excluded. Students who did not receive intensive reading intervention to address ORF deficits during the 2015-2016 school year or during the first semester of the 2016-2017 school year were excluded. Data from incomplete records was excluded. Incomplete records included: missing level of performance data, missing rate of improvement data, missing sex, missing race, and missing eligibility status information.

Because data were obtained across multiple years, it was possible that students would be included in multiple groups (e.g., not referred one year and referred and eligible for special education services the next). During database generation, each student was randomly assigned a code number by a third-party employee of the intermediate educational agency to ensure the data were aligned for analysis. Students were assigned a code number based on unique names listed in the database. Duplicate students were excluded from the dataset to meet the independence of observations assumption of multinomial logistic regression based on the following criteria. Student data were included based on these priorities: (a) keep data aligning with the year a student was referred and determined to be not eligible to increase the size of this group; (b) keep

67

data for students referred and eligible for special education services; (c) for students not referred, and therefore not eligible for special education services, keep data from the year that includes the most progress monitoring data points.

Assignment

Convenience sampling was used because districts meeting the criteria of having at least 75% of students being proficient on universal screeners were recruited. Only school districts providing consent for participation were included in the sample. Only data that occurred as a result of typical educational practices (i.e., naturally occurring educational data) were analyzed. Students were designated as belonging to one of the three groups (i.e., students with ORF skill deficits receiving intensive reading intervention but not referred for special education evaluation, students with ORF deficits referred for a special education evaluation and found not eligible for special education services, and students with ORF deficits referred for special education evaluation and found eligible for special education services) as part of regular educational practice. Student data from each of the three groups were analyzed to determine whether their data could be used to predict classification of students into the three naturally-occurring groups.

<div align="center">

Measurement

</div>

Dependent Variable

The latent dependent variable was planned to be students' referral and special education eligibility status. The observed dependent variable was planned to be the school team's decision regarding whether they belong to one of three groups: students with ORF deficits receiving intensive reading intervention but not referred for an initial educational evaluation, students with ORF deficits referred for an initial educational evaluation and found not eligible to receive

special education services, and students with ORF deficits referred for an initial educational evaluation and found eligible to receive special education services.

These groups were selected because no clear guidelines exist indicating at what point a referral for an initial educational evaluation should be made. There are also no guidelines about what constitutes a deficient level and rate of improvement. The dual discrepancy approach, which requires a student to demonstrate significant discrepancies in both level of performance and rate of improvement, appears to be the most consistently supported framework for assessing a student's RTI (Burns & Senesac, 2005; Fuchs et al., 2008; Fuchs et al., 2004; Fuchs et al., 2002; Speece & Case, 2001). Additionally, the federal regulations stipulate that a student identified with SLD should demonstrate a dual discrepancy when using an RTI approach (Kovaleski et al., 2013; United States Department of Education, 2006). Despite the research support and inclusion in federal regulations, much is still unknown about whether the SLD construct based on a dual discrepancy is being implemented in actual practice, or whether extraneous factors continue to influence eligibility decision-making. Conceptualization of a dual discrepancy suggests that both level of performance and rate of progress are equally important and necessary when determining whether a student has an SLD. Assignment to one of the three groups based on school teams' decision-making was selected because clinical judgment continues to be frequently used when making special education eligibility decisions (Lichtenstein, 2014).

Predictor Variables

The latent predictor variables were level of academic achievement and student responsiveness to intervention. These variables were selected because they are operationalizations of Criteria 1 and 2 of an SLD (United States Department of Education,

69

2004): (a) failure to meet age or grade-level standards (i.e., inadequate achievement); and (b) an inadequate response to scientifically based instruction. Consideration of data for Criteria 1 and 2 are required for establishing a dual discrepancy (Fuchs & Fuchs, 1998).

Observed level of achievement was operationalized as level of performance, which was obtained from the median of the students' last three progress monitoring data points prior to the time an educational decision was made. Level of performance data were obtained from FAST CBM-R progress monitoring data. The alternate form correlation between individual passages is a median of $r = 0.75$ (National Center on Intensive Intervention, n.d.). The test-retest reliability coefficient for third grade students from fall to spring is $r = 0.89$ for Grade 2 and $r = 0.90$ for Grade 3. The interrater reliability for passage sets used with second and third grade students is $r = 0.97$. When examining validity compared to AIMSweb (Pearson Education, 2011) an established CBM-R assessment, the concurrent validity coefficient is 0.97 for second grade students and 0.95 for third grade students. The predictive validity coefficient at approximately 12 weeks is 0.92 for second grade students and 0.90 for third grade students.

Observed student RTI was obtained from the ROI calculated from the slope of the students' progress monitoring data during the 2015-2016 school year or first semester of the 2016-2017 school year. ROI data were obtained from FAST CBM-R progress monitoring data. The National Center on Intensive Intervention (n.d.) reported the reliability of the slope for second grade students with up to 20 progress monitoring data points was $r = 0.83$, and the reliability of the slope for third grade students with up to 20 progress monitoring data points was $r = 0.70$. No validity coefficients are reported for the slope. Based on previous research, however, using ordinary least squares to calculate slope appears to be a valid way to document rate of improvement (Christ, 2006; Deno, Fuchs, Marston, & Shin, 2001; Shinn, Good, & Stein,

1989). Recent research, however, has provided statistical criticisms of using OLS to model

growth including the vulnerability of OLS to extreme outliers (Haupt, Lösel, & Stemmler, 2013)

and that OLS assumes a linear ROI, which may not be the case (Haupt et al., 2013; Nese et al.,

2013).

Based on the recommendation by Runge, Bennyhoff, Ferchalk, and McCrea (in review),

ROI were calculated using the formula $y = mx + b$, where m is interpreted as ROI. ROI should

be calculated using a consistent approach with the same measurement interval, and Runge et al.

suggested that using the actual date of the progress monitoring assessment to calculate the OLS

regression line may be the most valid interval for measurement. Therefore, data provided to the

principal investigator included actual dates of progress monitoring assessments so they could be

used when calculating ROI. To calculate each students' ROI, an OLS regression line was fit to

the data such that the Y-values were the students' score on the CBM-R assessment in WCPM

and the X-values were the actual date of the progress monitoring assessment, which resulted in a

daily ROI. Each students' daily ROI was multiplied by seven to obtain a weekly ROI (i.e.,

WCPM gained per week).

CBM-R was selected as the tool to measure the continuous predictor variables because

CBM-R is a reliable and valid indicator of student achievement and progress (Deno, 1985; Deno

et al., 2001). Guidelines released by the American Educational Research Association, American

Psychological Association, and National Council on Measurement in Education (1999) indicate

psychoeducational evaluations should involve reliable, valid, and sensitive measurements.

CBM-R meets these recommendations (Marston, 1989; Wayman, Wallace, Wiley, Ticha, &

Espin, 2007).

The latent ancillary variables were planned to be student sex, race, and free and reduced lunch status. The blocks for sex were male and female and were obtained from students' school records. The blocks for race were historically overrepresented and not historically overrepresented groups. Race data were obtained from the student records. The blocks for free and reduced lunch status were planned to be eligible for free or reduced lunch and not eligible for free or reduced lunch. Free and reduced lunch status were planned to be obtained by reviewing students' school records.

Previous research demonstrated demographic information impacts special education eligibility determinations (Lester & Kelman, 1997; Singer et al, 1989; Ysseldyke & Algozzine, 1981). Additionally, disproportionality has existed for special education placements of males, ethnic minority students, and students from disadvantaged backgrounds (Hosp & Reschly, 2004; Skiba, Poloni-Staudinger, Gallini, Simmons, & Feggins-Azziz, 2006). Although disproportionality has historically existed with special education rates, recent research has suggested using RTI data to make educational decisions may ameliorate problems related to disproportionate representation of these groups in special education (Burns & Senesac, 2005; Marston et al., 2003; Speece et al., 2003). Therefore, the ancillary predictor variables were planned to be included to determine to what extent they impact teams' decisions about students' special education status using RTI data. See Table 1 for an overview of the variables related to the research question, the instruments, and their reliability and validity.

Procedures

Intermediate educational agency staff working in the school sites were given data collection spreadsheets (see Appendix A for a sample data spreadsheet) to record the following information: student grade, sex, race, free and reduced lunch status, FAST CBM-R progress

monitoring data from the 2015-2016 school year or the first semester of the 2016-2017 school year, referral date for initial educational evaluation if applicable, and special education eligibility decision if applicable. Intermediate educational agency staff were instructed to record that information for each second and third grade student evaluated with ORF deficits during the 2015-2016 school year and during the first semester of the 2016-2017 school year. Intermediate educational agency staff were instructed to record that information for students not evaluated by the end of the 2015-2016 school year or at the end of the first semester of the 2016-2017 school year for second and third grade identified as persistently at risk for reading difficulties.

Data sheets were provided to a third party employee of the intermediate educational agency for compilation into a database. Data sheets were combined using Filemaker Pro to create a single spreadsheet, which was exported to Microsoft Excel. During database generation, each student was randomly assigned a code number to ensure the data were aligned for analysis. A random number was assigned to every unique name in the database. These data were supplied in redacted form in Excel spreadsheets to the primary investigator for analysis.

Sample Size

The sample size for this study was based on the recommendation that the minimum sample size for multinomial logistic regression is 10 cases per independent variable (Garson, 2016; Petrucci, 2009). Peduzzi, Concato, Kemper, Holford, and Feinstein (1996) recommended at least 10 events per level of each independent variable.

Using Peduzzi's and colleagues' formula $N = 10k/p$, where p is the smallest proportion of cases and k is the total number of levels for the predictor variables, Long (1997) made the following recommendation. If N is greater than or equal to 100, the sample size is adequate. Vittinghoff and McCulloch (2006), however, reported sample sizes resulting in five to nine

73

events per variable is sufficient. Therefore, the sample size will likely need to include at least

100 students using the recommendations of Garson (2016), Peduzzi et al. (1996), and Petrucci

(2009).

Table 1

Research Questions and Measurement Tools

Latent Variable	Observed Variables	Instrument/ Source	Validity	Reliability
Student Group Membership	Eligible, Not Eligible, Not Referred	School Record	Excellent	Excellent
Level of Academic Achievement	Median of 3 most recent progress monitoring data points prior to a decision about student group membership	FAST CBM-R	Referenced	Referenced
Rate of Improvement	Slope (based on OLS regression) of progress monitoring data	FAST CBM-R	Referenced	Referenced
Sex	Male/Female	School Records	Excellent	Excellent
Race	Historically overrepresented in special education and historically underrepresented in special education	School Records	Excellent	Excellent
Access to Resources	Free and Reduced Lunch Status	School Records	Excellent	Excellent

Research Question

The purpose of this study was to determine to what extent two major SLD criteria, including a student's level of academic achievement and ROI, impacts multidisciplinary evaluation teams' decision-making regarding referral for special education evaluation and regarding special education eligibility. To provide insight into this issue, this study was planned to focus on the following question. Do level of academic achievement and ROI, as well as potentially extraneous variables (e.g., student sex, race, and socioeconomic status), predict classification of students into three groups: (a) students with ORF skill deficits receiving intensive reading intervention but not referred for special education evaluation, (b) students with ORF deficits referred for a special education evaluation and found not eligible for special education services, and (c) students with ORF deficits referred for special education evaluation and found eligible for special education services?

1. It was hypothesized that level of performance would be the variable most related to students' group membership. Previous research demonstrated low achievement to be the defining characteristic of SLD (Brown-Waesche et al., 2011; Kavale & Reese, 1992; Peterson & Shinn, 2002; Ysseldyke et al., 1982), suggesting practitioners have historically determined need for more intensive support based on a student's performance level.

2. It was also hypothesized that level of performance and ROI together would predict students' group membership to a greater extent than the extraneous variables. Although previous research demonstrated that extraneous variables influence SLD identification (Lester & Kelman, 1997; Singer et al., 1989; Ysseldyke et al., 1982; Ysseldyke & Algozzine, 1981), more recent research suggested that decisions about student need based

on level and ROI are less subject to influence from extraneous variables (Burns & Senesac, 2005; Marston, Muyskens, Lau, & Canter, 2003; Speece, Case, & Molloy, 2003). Additionally, federal regulations require students to be dually discrepant in both level and ROI (United States Department of Education, 2006).

Statistical Analyses

Results were analyzed using the Statistical Package for the Social Sciences (SPSS), version 24 for Windows 64-bit version. The research question was planned to be examined using the following statistical analyses. Multinomial logistic regression (MLR) was planned to be used to predict group membership based on data considered by the school MDE teams.

MLR is an extension of binary logistic regression (Garson, 2016). Binary logistic regression is a type of regression that is used when the dependent variable is dichotomous and the predictor variables are continuous or categorical. Binary logistic regression, MLR, and ordinal logistic regression are all related types of logistic regression. The purpose of MLR is to predict membership to a categorical dependent variable with three or more categories. MLR is generally used when the dependent variable is nominal, but it may also be used for ordinal dependent variables if the data do not meet the assumptions of ordinal regression.

As an extension of binary logistic regression, MLR shares many similarities with it. MLR uses maximum likelihood estimation to calculate the probability of group membership (Starkweather & Moske, n.d.). Additionally, like binary logistic regression, MLR has many uses including: predicting categorical dependent variables; identifying the importance of predictor variables in a model; assessing interaction effects; and determining covariate control variables' impact (Garson, 2016).

Because MLR is an extension of binary logistic regression, it is similar to linear regression (Garson, 2016). Whereas linear regression involves continuous outcome variables, MLR has a categorical outcome variable with three or more categories. To develop a model, linear regression fits a line to the data, which is not desirable for categorical outcome variables. Logistic regression uses the logistic curve to model the probability of an outcome based on individual characteristics (i.e., the predictor variables). Garson (2016) explained the logistic curve "comes closer to hugging the $y = 0$ and $y = 1$ points on the y-axis" (p. 14). This concept is extended to MLR creating a series of binary comparisons, comparing one level of the outcome variable to a reference group of the outcome variable. This creates "k-1 binary model equations" (p. 14) to which the logistic curve can be fitted. When using statistical software to conduct an MLR, the reference category is usually the highest-coded value entered in the statistical software package, but it may also be selected by the researcher (UCLA Statistical Group, n.d.).

Garson (2016) further described the concepts on which MLR is built, including odds, odds ratios, log odds, and logit. Odds refers to the ratio formed by dividing the probability that an event occurs by the probability that it does not occur. An odds ratio is a ratio developed by comparing two odds for an outcome (Petrucci, 2009). Odds ratios in MLR reflect the difference a predictor variable has on the dependent variable (Garson, 2016). Odds ratios of 1 suggest a variable has no effect. The further an odds ratio is from 1 in either direction, the greater the effect the predictor variable has on the dependent variable (Garson, 2016; Petrucci, 2009). An odds ratio greater than one indicates a greater risk for belonging to the comparison group with each increase in the predictor variable of interest (UCLA Statistical Group, n.d.). Odds ratios greater than one suggest belonging to the comparison outcome is more likely. An odds ratio less than one suggests a greater risk for the outcome falling in the reference group as the predictor

variable increases. An odds ratio less than one suggests belonging to the reference group is more likely. Petrucci (2009) suggested an odds ratio may be interpreted as whether a predictor variable is more likely (i.e., odds ratio greater than 1) or less likely (i.e., odds ratio less than 1) to affect the outcome variable.

When conducting an MLR, each level of the outcome variable is compared against a reference group (Petrucci, 2009). Examining the frequency distribution of the outcome variable may inform reference group selection (UCLA Statistical Group, n.d.). Maximum likelihood estimation is used to produce odds ratios for each independent variable using paired comparisons of two levels of the outcome variable (Garson, 2016; Petrucci, 2009). Additionally, pseudo R^2 values were developed as an estimate of effect size (Petrucci, 2009). Because pseudo R^2 values are not equivalent to R^2 values in linear regression, they may be interpreted with caution (UCLA Statistical Group, n.d.). Likelihood ratio tests provide information about the contribution of each predictor variable in the model, and the Wald statistic tests for significance of each predictor variable (Garson, 2016; Petrucci, 2009).

Although a strength of MLR is to identify different characteristics of groups, Petrucci (2009) noted various limitations. Limitations of MLR include the need for larger sample sizes across all levels of the predictor and dependent variables, limited tools for assessing model fit compared to linear regression, and difficulty interpreting the models when more than four groups are included in the dependent variable.

Assumptions of MLR

Starkweather and Moske (n.d.) indicated MLR is appealing to researchers because it does not need to meet many assumptions required for similar multivariate analyses, including multivariate analysis of variance and discriminant function analysis. Normality, linearity, and

homoscedasticity do not need to be assumed by MLR. If those assumptions are met, however, discriminant function analysis may be the more appropriate tool because it is a more powerful alternative to MLR. Although MLR has less associated assumptions than similar analyses, it still requires attention to the sample size. A minimum of 10 cases for every predictor variable is required (Garson, 2016; Petrucci, 2009). Sample sizes resulting in five to nine events per variable, however, may also be adequate (Vittinghoff & McCulloch, 2006).

MLR has six assumptions (Multinomial Logistic Regression using SPSS, n.d.). The dependent variable should be nominal. Ordinal dependent variables may also be used if the assumptions for ordinal regression are not met. The model includes one or more predictor variables that are continuous, ordinal, or nominal. The model should have independence of observations, and it should be properly specified so all relevant variables are included and all irrelevant variables are excluded from the regression model. There should be no multicollinearity, meaning the independent variables should not be highly correlated with each other. There should also be a linear relationship between the continuous predictor variables and the logit transformation of the dependent variable. Finally, there should be no outliers or highly influential data points.

<div align="center">**Summary**</div>

The planned design and methods for predicting student referral and special education eligibility group membership based on school MDE teams' decision making were discussed in this chapter. A review of the population and sample was provided to discuss to whom the results may be applied. The planned predictor and outcome variables were reviewed. The statistical analysis MLR, as well as its assumptions, to answer the research question was described.

CHAPTER 4

RESULTS

The purpose of this study was to determine to what extent two major specific learning

disability (SLD) criteria, including a student's level of academic achievement and rate of

improvement (ROI), predict multidisciplinary evaluation teams' decision-making regarding

referral for special education evaluation and regarding special education eligibility. The original

research question was, "Do level of academic achievement and ROI, as well as potentially

extraneous variables (i.e., student sex, race, and socioeconomic status), predict classification of

students into three groups: (a) students with oral reading fluency (ORF) skill deficits receiving

intensive reading intervention but not referred for special education evaluation, (b) students with

ORF deficits referred for a special education evaluation and found not eligible for special

education services, and (c) students with ORF deficits referred for special education evaluation

and found eligible for special education services?"

It was hypothesized that level of performance would be the variable most related to

students' group membership. Previous research demonstrated low achievement to be the

defining characteristic of SLD (Brown-Waesche et al., 2011; Kavale & Reese, 1993; Peterson &

Shinn, 2002; Ysseldyke et al., 1982), suggesting practitioners have historically determined need

for more intensive support based on a student's performance level. It was also hypothesized that

level of performance and ROI together would predict students' group membership to a greater

extent than the extraneous variables. Although previous research demonstrated that extraneous

variables influence SLD identification (Lester & Kelman, 1997; Singer et al., 1989; Ysseldyke et

al., 1982; Ysseldyke & Algozzine, 1981), more recent research suggested that decisions about

student need based on level and ROI are less subject to influence from extraneous variables

(Burns & Senesac, 2005; Marston, Muyskens, Lau, & Canter, 2003; Speece, Case, & Molloy, 2003). Additionally, federal regulations require students to be dually discrepant in both level and ROI (United States Department of Education, 2006).

Complications related to the sample, however, resulted in modifications to the research question, study design, and subsequent analyses. An overview of the sample is offered to provide insight about why the originally planned analyses would have been flawed, and changes to the study are then reviewed prior to reporting the results.

Complications Related to Sample

Data Obtained

Data were obtained from 19 elementary schools in a Midwestern state. Included in the original dataset were 575 cases. Cases with missing data were excluded from the dataset. Cases that were found eligible for special education prior to the 2015-2016 school year were excluded from the dataset. Cases that included scores from alternate progress monitoring assessments (e.g., monitoring FAST CBM-R off grade-level, monitoring letter sounds rather than CBM-R) were excluded. Cases with suspect progress monitoring data points (e.g., 7788 words correct per minute [WCPM]) were excluded.

Since it was possible that a student's data from Grade 2 (i.e., the 2015-2016 school year) and Grade 3 (i.e., the 2016-2017 school year) could be included in the dataset, a selection prioritization was used to exclude data from duplicate cases. The selection process included the following prioritizations: (a) include data from the year a student was referred for an evaluation but found not eligible; (b) include data from the year a student was referred and found eligible; (c) if not referred for an evaluation during either school year, include data with more progress monitoring points; (d) if a student had the same number of progress monitoring data points for

both Grade 2 and 3, flip a coin to determine which data to keep. For all duplicate cases, both

Prioritizations A and B were never considered for the same student. Following the processing of

the original dataset, the final dataset included 383 cases.

Revised Research Questions and Hypotheses

Due to the small number of cases in the group referred for a special education evaluation

but found not eligible for special education services, the sample size required for a multinomial

logistic regression was not obtained. Therefore, the study's research questions, design, and

subsequent analyses were modified. The revised research questions are as follows. Table 2

provides a side-by-side comparison of the original and revised research questions.

1. Do level of performance and ROI differ as a function of grade level (Grade 2, Grade 3)

   and special education status (not eligible, eligible)?

   a. Hypotheses: It was hypothesized that students eligible for special education

      would display lower levels of performance and ROIs than students receiving

      general education reading intervention who were not referred for a special

      education evaluation. It was also hypothesized that students in Grade 2 would

      demonstrate lower levels of performance than students in Grade 3 but that

      students in Grades 2 and 3 would demonstrate similar ROIs.

2. Do level of performance and ROI predict whether a persistently at-risk student continues

   to receive only general education reading intervention or is found eligible for special

   education?

   a. Hypotheses: It was hypothesized that level of performance would be the variable

      most predictive of special education eligibility status. It was hypothesized that

level of performance and ROI together would predict students' group membership

to a greater extent than level of performance alone.

Table 2

Comparison of Original Research Question to Revised Research Questions

Original Question	Revised Questions
Do level of academic achievement and ROI, as well as potentially extraneous variables (i.e., student sex, race, and socioeconomic status), predict classification of students into three groups: (a) students with ORF skill deficits receiving intensive reading intervention but not referred for special education evaluation, (b) students with ORF deficits referred for a special education evaluation and found not eligible for special education services, and (c) students with ORF deficits referred for special education evaluation and found eligible for special education services?	1. Do level of performance and ROI differ as a function of grade level (Grade 2, Grade 3) and special education status (not eligible, eligible)? 2. If significant differences in level of performance and ROI are found, do level of performance and ROI predict whether a persistently at-risk student continues to receive only general education reading intervention or is found eligible for special education?

Note. ROI = rate of improvement, ORF = oral reading fluency

Modified Design and Analyses

Research Question 1

To answer Question 1, a 2 (Grade 2, Grade 3) x 2 (not eligible, eligible) factorial design

with two dependent variables, level of performance and ROI, was used. The independent

variables are grade level and special education status. Grade level has two groups, Grade 2 and Grade 3. Special education status has two groups, not eligible and eligible. The continuous dependent variables are level of performance and ROI. A two-way multivariate analysis of variance (MANOVA) was used to determine whether there was an interaction between grade and eligibility status on the dependent variables, level of performance and ROI, whether there was a main effect of grade, and whether there was a main effect of special education eligibility status.

The assumptions for a two-way MANOVA include: (a) two or more continuous dependent variables; (b) two independent variables with each consisting of at least two categorical groups; (c) independence of observations; (d) linearity of variables; (e) no multicollinearity; (f) no significant outliers; (g) normal distribution of dependent variables; (h) adequate sample size; (i) homogeneity of variance-covariance matrices; and (j) homogeneity of variances (Lund & Lund, 2013c).

The assumption for two or more continuous dependent variables was met, as the dependent variables include level of performance measured in WCPM and ROI measured in WCPM per week. Both level of performance and ROI are ratio data. The assumption of two categorical independent variables was met because grade level is dichotomous with two groups (Grade 2 and Grade 3) and special education eligibility status is dichotomous with two groups (not referred and referred and eligible). The independence of observations assumption was met because each student in the sample provides scores for only one case. The sample size assumption was met, as there were more cases within each group than the number of dependent variables (see Table 3; Lund & Lund, 2013c).

A linear relationship between each pair of dependent variables was examined using a scatterplot of the dependent variables for each group in the design (Lund & Lund, 2013c).

Visual inspection of the scatterplots suggested a linear relationship between the dependent variables.

Multicollinearity of the dependent variables was tested by obtaining Pearson correlation coefficients using SPSS for the dependent variables of level of performance and ROI. Correlation coefficients greater than 0.9 indicate multicollinearity of the dependent variables (Lund & Lund, 2013c). As indicated in Table 3, there was no evidence of multicollinearity, as assessed by Pearson correlation ($|r| < 0.9$).

Table 3

Correlations Between Level of Performance and ROI to Test for Multicollinearity of the

Dependent Variables

Grade	Eligibility Status	N	r	p
Grade 2	Not Referred	233	.258	<.0005
	Referred and Eligible	15	.402	.138
Grade 3	Not Referred	117	.066	.482
	Referred and Eligible	11	.627	.039

Visual inspection of boxplots were used to examine univariate outliers. Lund and Lund (2013c) suggested parameters for examining outliers using boxplots. Data were considered to be outliers if they were 1.5 box-lengths from the edge of the box in a boxplot, and data were considered to be extreme outliers if they were greater than 3 box-lengths from the edge of the box in a boxplot. For Grade 2 not referred students, there were two univariate outliers for level of performance (144 WCPM and 137 WCPM) and four univariate outliers for ROI (2.5 WCPM per week, 2.6 WCPM per week, 2.6 WCPM per week, and 2.6 WCPM per week); none were extreme outliers, however. For Grade 3 not referred students, there were two univariate outliers for level of performance (24 WCPM and 34 WCPM) and five univariate outliers for ROI (2.1

WCPM per week, 2.4 WCPM per week, 2.5 WCPM per week, 2.7 WCPM per week, and 3.1 WCPM per week). The student improving at 3.1 WCPM per week was identified as an extreme outlier. Johnson and Wichern (2007) explained that outliers that include accurate data may "justifiably be part of the group and may lead to a better understanding of the phenomena being studied" (p. 187), suggesting that outliers may be included in the analysis if it is believed they are genuine. Furthermore, Lund and Lund (2013b) referred to "genuinely unusual values," (para. 4) which are values not resulting from data entry or measurement error that are most likely genuine data points belonging to the group being studied. They explained that, although including outliers may not be "ideal from a statistical perspective (i.e., they violate one of the assumptions of the two-way MANOVA), there is no good reason to reject them as invalid." (para. 4). For referred and eligible Grade 2 students, there were no univariate outliers for level of performance or ROI. For referred and eligible Grade 3 students, there was one univariate outlier for level of performance (129 WCPM), and there were no univariate outliers for ROI. Outliers were double-checked for accuracy of data entry and calculation of ROI. The data for each outlier were graphed to detect whether there were any extreme scores among the individual data points that would have skewed the calculated levels of performance and ROIs. There were no such individual data point outliers, so each datum was maintained in the calculation of ROI and determination of final level of performance. Examination of the outliers' individual progress monitoring scores does not provide any reason to reject them as invalid (Johnson & Wichern, 2007; Lund & Lund, 2013b); therefore, they were kept in the analysis.

Maholanobis distance values were generated using SPSS to test for multivariate outliers (Lund & Lund, 2013c). There was one multivariate outlier, as assessed by Mahalanobis distance ($p > .001$). The multivariate outlier was a not referred Grade 3 student with a level of

performance of 94 WCPM and an ROI of 3.1 WCPM per week. The data for the multivariate

outlier were graphed to detect whether there were any extreme scores among the individual data

points that would have skewed the calculated ROI. There were no such individual data point

outliers, so each datum was maintained in the calculation of ROI and determination of final level

of performance. Examination of the outliers' individual progress monitoring scores does not

provide any reason to reject them as invalid (Johnson & Wichern, 2007; Lund & Lund, 2013b);

therefore, they were kept in the analysis.

Table 4

Skewness and Kurtosis Values for Dependent Variables Level of Performance and ROI Across

Combinations of Grades and Eligibility Groups

Cell	Dependent Variable	Skewness	Kurtosis
Grade 2, Not Referred	Level of Performance	.063	.355
	ROI	.455	.249
Grade 2, Referred and Eligible	Level of Performance	.253	-.749
	ROI	-.074	-.161
Grade 3, Not Referred	Level of Performance	-.515	.565
	ROI	1.116	2.126
Grade 3, Referred and Eligible	Level of Performance	.556	1.785
	ROI	-.071	-.560

The normality assumption was assessed using skewness and kurtosis values because the

sample size is greater than 50 (Lund & Lund, 2013c). Cameron (2011) described values between

-2 and +2 as acceptable for skewness (which describes symmetry of a distribution) and kurtosis

(which describes peakness of a distribution). Table 4 provides skewness and kurtosis values for

each cell included in the analysis. Results showed that skewness and kurtosis values were within

acceptable limits for the dependent variables across all cells except for skewness and kurtosis values for ROI for not referred students in Grade 3. Although the normality assumption was not met, MANOVA is robust to violations of the normality assumption with respect to Type 1 error (Bray & Maxwell, 1985; Pituch & Stevens, 2016). Bray and Maxwell (1985) reported that MANOVA may still be used if the normality assumption is violated. Therefore, it was decided to continue with the MANOVA.

Box's test of equality of covariance matrices was used to test the assumption of homogeneity of covariance matrices (Lund & Lund, 2013c). This assumption is met if $p > .001$. As displayed in Table 5, this assumption was met as assessed by Box's M test, $p = .048$.

Table 5

Box's Test of Equality of Covariance Matrices

Box's M Statistic	F	$df1$	$df2$	p
17.871	1.896	9	7522.737	.048

Levene's test of equality of error variances was used to test the assumption of homogeneity of variances (Lund & Lund, 2013c). As depicted in Table 6, this assumption was met as assessed by Levene's test, as results for each variable were not significant ($p > .05$), indicating relative equality of error variances.

Table 6

Levene's Test of Equality of Error Variances

Dependent Variable	F	$df1$	$df2$	p
Level of Performance	1.314	3	372	.269
ROI	.253	3	372	.859

Research Question 2

To answer Research Question 2 a predictive design (Creswell, 2012) with one outcome variable (special education eligibility status) and two predictor variables (level of performance and ROI) was used. The outcome variable has two groups (not eligible and eligible). The two predictor variables, level of performance and ROI, are both measured at the continuous level. Binary logistic regression (logistic regression) was used to examine whether level of performance and ROI are predictive of students' special education eligibility status (i.e., not referred for an evaluation vs. eligible for special education). It was expected that students in Grades 2 and 3 demonstrated different levels of performance (FastBridge, 2015) making it inappropriate to compare performance levels across grade levels. Therefore, the logistic regression was completed for Grade 2 students and then repeated with Grade 3 students to examine the predictiveness of level of performance and ROI at each grade level. Logistic regression is a type of regression that is used when the dependent variable is dichotomous and the predictor variables are continuous or categorical. The purpose of logistic regression is to predict membership to a categorical dependent variable with two dichotomous categories.

Garson (2016) described the assumptions of logistic regression, which include: (a) one dichotomous dependent variable; (b) one or more predictor variables that are continuous or nominal; (c) independence of observations; (d) no multicollinearity; (e) linear relationship between the continuous predictor variables and the logit transformation of the dependent variable; and (f) no outliers or highly influential data points. Although logistic regression has less associated assumptions than similar analyses, it still requires attention to the sample size. Peduzzi et al. (1996) recommended at least 10 events per independent variable, referring to positive cases (i.e., students identified with SLD in the sample). Vittinghoff and McCulloch

(2006) reported that sample sizes resulting in five to nine events per variable, however, may also be adequate.

The assumption of one dichotomous dependent variable was satisfied as the outcome variable (special education eligibility status) has two groups (not referred for an evaluation and eligible for special education). The assumption of one or more continuous or nominal predictor variables was met, because two predictor variables (level of performance and ROI) were included in the model. Both predictor variables are continuous variables measured at the ratio level. The independence of observations assumption was met because each student in the sample provides scores for only one case (Garson, 2016).

When considering the sample size, 20 positive cases (i.e., students eligible for special education) are required with the inclusion of two predictor variables based on the recommendations of Peduzzi and colleagues (1996). Using Vitinghoff and McCulloch's (2006) study about the minimum events per variable required for logistic regression, it may be appropriate to have as few as 10 positive cases in the sample based on their findings about including five to nine events per variable in a logistic regression model. For Grade 2 students, there were 15 positive cases, and for Grade 3 students there were 11 positive cases. Although the sample size is smaller than generally recommended (Peduzzi et al., 1996), it may still be sufficient (Vittinghoff & McCullogh, 2006).

Multicollinearity was analyzed using SPSS to obtain variance inflation factor values (Salkind, 2011). Variance inflation factor values less than 10 indicate the absence of multicollinearity. As demonstrated in Table 7, which shows the results of tests for multicollinearity, no multicollinearity exists in the dataset.

Table 7

Multicollinearity Statistics

	VIF
Grade 2	
Level of Performance	1.108
ROI	1.108
Grade 3	
Level of Performance	1.104
ROI	1.104

Note. VIF = Variance inflation factor value

The Box-Tidwell (1962) procedure was used to test the assumption that there is a linear relationship between the continuous predictor variables with respect to the logit of the dependent variable. A Bonferroni correction was applied using all four terms in the model resulting in statistical significance being accepted when $p < .0125$ (Tabachnick & Fidell, 2007). The linearity assumption is met if $p > .0125$. Table 8 displays the results of the Box-Tidwell procedure indicating that across Grades 2 and 3, the linearity assumption was met for both continuous predictor variables because $p > .0125$ (Tabachnick & Fidell, 2007).

To assess for outliers, cases were reviewed for studentized residuals greater than 2.5 standard deviations (Lund & Lund, 2013a). For Grade 2 data, there were four cases with studentized residual values greater than 2.5 standard deviations. For Grade 3 data, there were five cases with studentized residual values greater than 2.5 standard deviations. The data for each outlier were graphed to detect whether there were any extreme scores among the individual data points that would have skewed the calculated levels of performance and ROIs. There were no such individual data point outliers, so each datum was maintained in the calculation of ROI

and determination of final level of performance. Data from these cases were examined and appear to be genuine cases. To exclude them would result in a loss of information (Johnson & Wichern, 2007; Lund & Lund, 2013b). Therefore, they were included in the analyses.

Table 8

Results of Box-Tidwell Procedure to Test Linear Relationship between Continuous Predictors and Logit of the Dependent Variable

	p
Grade 2	
Level of Performance x Natural Log Transformation of Level of Performance	.552
ROI x Natural Log Transformation of Level of Performance	.407
Grade 3	
Level of Performance x Natural Log Transformation of Level of Performance	.398
ROI x Natural Log Transformation of Level of Performance	.413

Note. ROI = rate of improvement

Descriptive Statistics

Table 9 includes data for number of students in the three special education eligibility status categories disaggregated by grade, sex, and race. Mean level of performance and ROI data for second and third grade students disaggregated by sex and race are displayed in Table 10. As indicated in Table 9, the total number of students in the referred but not eligible category was seven students. For that reason, the original analyses were unable to be conducted; additionally, because the referred but not eligible group is too small to meet the sample size requirements of the planned analyses, this group was excluded from further analyses.

The mean number of progress monitoring data points for all second grade students was 22.6. For Grade 2 students, a mean of 22.9, 19.1, and 15.0 progress monitoring data points were

available for students in the not referred, referred and eligible, and referred but not eligible

groups, respectively. The mean number of progress monitoring data points for all third grade

students was 21.8. For Grade 3 students, a mean of 22.4, 19.1, and 13.3 progress monitoring

data points were available for students in the not referred, referred and eligible, and referred but

not eligible groups, respectively. The average number of data points in the sample exceeded 14,

which was recommended by Christ and colleagues (2012). All students included in the sample

had at least seven progress monitoring data points, which has been frequently recommended in

the research (Ardoin et al., 2013), and 90.3% of students in the sample had at least 14 data points

as recommended by Christ et al. (2012).

<div align="center">

Tests of Hypotheses

</div>

Research Question 1

Table 11 displays mean values for level of performance in WCPM and ROI in WCPM

per week for students across eligibility groups and grade levels. A two-way (2x2) MANOVA

with two independent variables (level of performance and ROI) and two dependent variables

(grade and special education eligibility status) was conducted to determine whether level of

performance and ROI differ as a function of grade level and special education eligibility status.

The data were analyzed to determine whether there are main effects of eligibility status and

grade as well as interactions between these independent variables on the dependent variables

level of performance and ROI.

As indicated in Table 12, the interaction effect between grade and eligibility decision on

the combined dependent variables was not statistically significant, $F(2, 371) = 1.06$, $p = .348$.

Therefore, the main effects were interpreted. There was a statistically significant effect of

<div align="center">

93

</div>

eligibility decision on the combined dependent variables $F(2, 371) = 23.30$, $p < .001$ and of

grade on the combined dependent variables $F(2, 371) = 29.57$, $p < .001$.

Table 9

Descriptive Information About the Sample

	Sex		Race		Total
Group	F	M	Over	Not	
Grade 2					
Not Referred	88	145	53	180	233
Referred and Eligible	4	11	4	11	15
Referred not Eligible	1	2	3	0	3
Grade 3					
Not Referred	47	70	29	80	117
Referred and Eligible	6	5	4	7	11
Referred not Eligible	1	3	0	4	4
Total	147	236	93	290	383

Note. F = Female, M = Male, Over = Overrepresented racial groups, Not = Not overrepresented racial groups

Follow-up univariate tests were conducted. First, a follow-up two-way (2x2) ANOVA

was conducted on the dependent variable of level of performance. Table 11 displays mean

scores for students across the two eligibility groups and two grades.

As indicated in Table 13, there was a statistically significant main effect of eligibility

decision $F(1, 372) = 45.35$, $p < .001$ and grade $F(1, 372) = 50.64$, $p < .001$. For eligibility status,

results indicated students in the referred and eligible group displayed significantly lower levels

of performance than students in the not referred group. For grade, results revealed students in

Grade 2 displayed significantly lower levels of performance than students in Grade 3. This was

to be expected due to the developmental progression in level of performance for ORF. As

expected, no significant interaction was found between the two independent variables on level of

performance. This supports the hypotheses that students eligible for special education display

lower levels of performance than students not referred for an evaluation and that students in

Grade 2 demonstrate lower levels of performance than students in Grade 3.

Table 10

Mean Level of Performance and ROI Data for Second and Third Grade Students Disaggregated

by Sex and Race

	Level		ROI	
Group	*M*	*SD*	*M*	*SD*
Grade 2				
Female	73.16	21.25	1.22	.52
Male	71.01	22.11	1.21	.53
Overrepresented racial groups	63.63	18.77	1.06	.54
Not overrepresented racial groups	74.38	22.06	1.26	.52
Grade 3				
Female	99.02	23.77	.92	.53
Male	99.51	24.72	.93	.58
Overrepresented racial groups	98.48	24.08	.82	.48
Not overrepresented racial groups	99.59	24.41	.97	.58

Table 11

Descriptive Statistics for Level of Performance and ROI by Eligibility Status and Grade

	Level				ROI		
	M	*SD*	*n*		*M*	*SD*	*n*
Not Referred							
Total	83.1	25.3	350		1.1	.6	350
Grade 2	73.7	20.6	233		1.2	.5	233
Grade 3	101.9	23.3	117		.9	.6	117
Referred and Eligible							
Total	55.3	26.2	26		.8	.4	26
Grade 2	40.6	15.4	15		.8	.4	15
Grade 3	75.4	24.9	11		.8	.4	11
Grade 2							
Total	71.7	21.8	248		1.2	.5	248
Grade 3							
Total	99.6	24.5	128		.9	.6	128

Note. ROI = rate of improvement

Table 12

Results of MANOVA for the Effect of Grade and Eligibility Decision on Level of Performance

and ROI

Effect	Wilks' λ	*F*	Hypothesis *df*	Error *df*	*p*	η^2
Grade	.863	29.57	2	371	< .001	.137
Eligibility Decision	.888	23.30	2	371	< .001	.112
Grade x Eligibility Decision	.994	1.06	2	371	.348	.006

A follow-up two-way (2x2) ANOVA was also conducted on the dependent variable ROI. As reported in Table 13, a statistically significant main effect was found for eligibility decision $F(1, 372) = 6.21$ but not for grade $F(1, 372) = 2.13$, $p = .145$. Mean ROIs for both eligibility groups and grade levels are presented in Table 11. Results revealed that students in the referred and eligible group demonstrated significantly lower ROIs than students in the not referred group. With no significant main effect for grade, it can be inferred that students in Grades 2 and 3 do not demonstrate significantly different ROIs. As expected, there was no significant interaction between grade and eligibility status on ROI. This supports the hypotheses that students eligible for special education display lower ROIs than students not referred for an evaluation and that students demonstrating low achievement in Grades 2 and 3 demonstrate similar ROIs.

Research Question 2

A binomial logistic regression was performed to determine the effects of level of performance and ROI on the likelihood that students are identified as SLD in Grade 2. The logistic regression model was statistically significant, $\chi^2(2) = 37.303$, $p < .001$, and Hosmer Lemeshow goodness of fit tests resulted in non-significant χ^2 values ($p > .05$), indicating the model was a good fit to the data. The model explained 38.1% (Nagelkerke R^2) of the variance in SLD identification and correctly classified 93.1% of cases. Sensitivity was 13.3% and specificity was 98.3%. Of the two predictor variables only level of performance was statistically significant, $p < .001$, and ROI was not, $p = .277$. Increasing level of performance was associated with a decreased likelihood of being identified as SLD. For every one unit increase in level of performance, the odds of being found eligible for special education decreased by 7.9%. The results of the logistic regression for Grade 2 are displayed in Table 14.

Table 13

Post Hoc ANOVA Results for Grade and Eligibility Decision on Level of Performance and ROI

	Mean Square	df	F	p	η²
Grade					
Level of Performance	23255.56	1	50.64	< .001	.120
ROI	.59	1	2.13	.145	.006
Eligibility Decision					
Level of Performance	20825.64	1	45.35	< .001	.109
ROI	1.73	1	6.21	.013	.016
Grade x Eligibility Decision					
Level of Performance	253.62	1	.55	.458	.001
ROI	.53	1	1.90	.169	.005
Error					
Level of Performance	459.20	372			
ROI	.28	372			

Note. ROI = rate of improvement

Table 14

Results of Logistic Regression for Grade 2

	B	SE	Wald	df	p	Exp(B)	95% CI for EXP(B)	
							Lower	Upper
Level of Performance	-.08	.02	17.84	1	< .001	.921	.887	.957
ROI	-.76	.70	1.18	1	.227	.469	.120	1.835
Constant	2.65	.97	7.48	1	.006	14.175		

Note. CI = Confidence Interval.

A binomial logistic regression was performed to determine the effects of level of performance and ROI on the likelihood that students are identified as SLD in Grade 3. The logistic regression model was statistically significant, $\chi^2(2) = 11.203$, $p = .004$, and Hosmer Lemeshow goodness of fit tests resulted in non-significant χ^2 values ($p > .05$), indicating the model was a good fit to the data. The model explained 18.9% (Nagelkerke R^2) of the variance in SLD identification and correctly classified 91.4% of cases. Sensitivity was 9.1% and specificity was 99.1%. Of the two predictor variables only level of performance was statistically significant, $p = .002$, and ROI was not $p = .853$. Increasing level of performance was associated with a decreased likelihood of being identified as SLD. For every one unit increase in level of performance, the odds of being found eligible for special education decreased by 4.1%. The results of the logistic regression for Grade 3 are displayed in Table 15.

To analyze differences in the explained variance using level of performance alone and level of performance plus ROI, hierarchical logistic regression was used to examine Nagelkerke R^2 values at the two stages of the analysis. Stage one included only level of performance as a predictor variable. Stage two included level of performance and ROI as predictor variables. This process was completed for Grade 2 and then for Grade 3.

Table 15

Results of Logistic Regression for Grade 3

	B	*SE*	Wald	*df*	*p*	Exp(B)	95% CI for EXP(B) Lower	Upper
Level of Performance	-.04	.01	9.35	1	.002	.959	.933	.985
ROI	-.11	.59	.03	1	.853	.897	.282	2.846
Constant	1.50	1.20	1.57	1	.210	4.491		

Note. CI = Confidence Interval.

For Grade 2, level of performance contributed significantly to the model, $p < .001$ at stage one, and the variance explained at stage one was 36.9% (Nagelkerke R^2). At stage two, which included level of performance and ROI as predictors, the variance explained only increased by 1.3%, resulting in 38.1% (Nagelkerke R^2) of the variance explained in stage two.

For Grade 3, level of performance contributed significantly to the model at stage one, $p = .002$, and the variance explained at this stage was 18.8% (Nagelkerke R^2). At stage two, which included level of performance and ROI as predictors, the explained variance only increased by .1% to 18.9% (Nagelkerke R^2).

It was hypothesized that level of performance would be the variable most predictive of special education eligibility status and that level of performance and ROI together would predict students' group membership to a greater extent than level of performance alone. Inclusion of level of performance and ROI as predictor variables resulted in a model that explained 38.1% (Nagelkerke R^2) of the variance in SLD identification at Grade 2 and 18.9% (Nagelkerke R^2) of the variance in SLD identification at Grade 3. Examination of the significance of both level of performance and ROI at Grade 2 revealed that only level of performance was a statistically significant predictor, $p < .001$, and ROI was not, $p = .277$. The same was found at Grade 3. Only level of performance was a statistically significant predictor, $p = .002$, and ROI was not $p = .853$. Across Grades 2 and 3, adding ROI as a predictor variable only slightly increased the percent of variance explained as indicated by Nagelkerke R^2 values. The data support the hypothesis that level of performance would be most predictive of special education eligibility status. The data do not, however, support the hypothesis that ROI is predictive of SLD identification or that, when considered with level of performance, ROI adds more to a decision about special education eligibility.

Summary

This chapter reviewed the modifications made to the research question and design as a result of complications with the sample. A two-by-two factorial design with two dependent variables, level of performance and ROI, was used for Research Question 1, which examined whether level of performance and ROI differ as a function of grade level and special education eligibility status. Results of a two-way MANOVA and subsequent post-hoc tests supported the hypothesis that students with SLD displayed lower levels of performance and ROIs than students receiving general education reading intervention who were not referred for a special education evaluation. The hypotheses that Grade 3 students demonstrate higher levels of performance than Grade 2 students but that ROIs are similar across grade levels was also supported.

A predictive design was used to answer Research Question 2, which examined whether level of performance and ROI predict whether a student was determined to be SLD by an MDE team or continued to receive intensive general education reading intervention without being referred for a special education evaluation. Results of a logistic regression supported the hypothesis that level of performance would be predictive of SLD identification. This result was found for both second and third grade students. Analyses did not support the hypothesis that ROI significantly contributes to decisions about whether a student is identified as SLD in Grades 2 and 3.

CHAPTER 5

DISCUSSION

The primary goal of this study was to determine to what extent two major specific

learning disability (SLD) criteria, a student's level of academic achievement and rate of

improvement (ROI), predict multidisciplinary evaluation teams' decisions about special

education eligibility. The original aim of the study was to also examine the impact of potentially

extraneous variables (i.e., student sex, race, and socioeconomic status) on eligibility decisions in

relation to the two major SLD criteria. The planned research question was, "Do level of

academic achievement and ROI, as well as potentially extraneous variables (i.e., student sex,

race, and socioeconomic status), predict classification of students into three groups: (a) students

with oral reading fluency (ORF) skill deficits receiving intensive reading intervention but not

referred for special education evaluation, (b) students with ORF deficits referred for a special

education evaluation and found not eligible for special education services, and (c) students with

ORF deficits referred for special education evaluation and found eligible for special education

services?" Complications related to the sample limited the scope of the study and resulted in

changes to the research question, study design, and subsequent analyses. The primary goal of the

study, however, was maintained.

As previously indicated, complications related to the sample impacted the scope of the

study. Data were obtained from 14 elementary schools in a Midwestern state, resulting in a final

dataset that included 383 cases. Socioeconomic status information for each student in the sample

was not available, so it was excluded as a variable. After processing the dataset, data were

available for only seven students (3 students in Grade 2 and 4 students in Grade 3) from the

group that was referred for a special education evaluation but found not eligible. Additionally,

because 15 students remained in the referred and eligible group in the Grade 2 sample and 11 students remained in the referred and eligible group in the Grade 3 sample, extraneous variables (sex, race) had to be excluded from the regression analysis due to sample size constraints. Therefore, the sample size required for the planned analysis was not obtained, and two new research questions were developed. The questions were, "Do level of performance and ROI differ as a function of grade level (Grade 2, Grade 3) and special education status (not eligible, eligible)?" and "Do level of performance and ROI predict whether a persistently at-risk student continues to receive only general education reading intervention or is found eligible for special education?"

Research Question 1 was addressed using a 2 (grade) by 2 (special education eligibility status) factorial design with two dependent variables (level of performance and ROI). Data for Research Question 1 were analyzed using a two-way MANOVA to examine main effects grade level and special education eligibility status, as well as an interaction between grade level and special education eligibility status on the combined dependent variables, level of performance and ROI. Research Question 2 was addressed using a predictive design with one dichotomous outcome variable (special education eligibility status) and two continuous predictor variables (level of performance and ROI). Data for Research Question 2 were analyzed using logistic regression to examine whether level of performance and ROI are predictive of eligibility for special education.

Summary of Findings

Discussion of Descriptive Data

Review of descriptive information about the sample suggests consistencies with published opinions regarding best practices in assessment using response to intervention (RTI)

data, including curriculum-based measurement in reading (CBM-R), were found. Examination of the number of progress monitoring data points across eligibility status groups and grade levels in the sample revealed a mean of 22.6 progress monitoring data points for all Grade 2 students and a mean of 21.8 progress monitoring data points for all Grade 3 students included in the sample. For Grade 2 students, a mean of 22.9, 19.1, and 15.0 progress monitoring data points were available for students in the not referred, referred and eligible, and referred but not eligible groups, respectively. For Grade 3 students, a mean of 22.4, 19.1, and 13.3 progress monitoring data points were available for students in the not referred, referred and eligible, and referred but not eligible groups, respectively. Additionally, all students in the final sample had at least seven progress monitoring data points, which was reportedly most often recommended in the literature (Ardoin et al., 2013).

Descriptive information about progress monitoring data points in the sample is largely consistent with recommendations from research. After reviewing 78 studies, Ardoin and colleagues (2013) concluded that students' progress should be monitored for at least 12 weeks with a dense progress monitoring schedule to make decisions about growth for individual students. They found, however, that the most consistently recommended number of progress monitoring data points in the literature was seven. Using simulated models of progress monitoring data with different durations, Christ and colleagues (2012) recommended that at least 14 progress monitoring data points are required to obtain a reliable slope for decisions about students' growth. Furthermore, findings by Thornblad and Christ (2014) substantiated previous research that longer progress monitoring durations resulted in improved slope estimation.

Comparisons of the sample data for this study to recommendations from previous studies indicate that across all second and third grade students in the sample, the mean number of

progress monitoring data points across Grade 2 and 3 students in the sample exceeds 14 as recommended by Christ et al. (2012). Inclusive of all 383 students in the dataset, 90.3% of students had at least 14 progress monitoring data points available from which ROI could be calculated. Descriptive information about progress monitoring data points in the sample suggests that, overall, practitioners have adequate datasets on hand to make high-stakes educational decisions such as the determination of eligibility for special education. Additionally, because practitioners are monitoring students' progress over durations consistent with recommended practices, it appears that practitioners are allowing for sufficient durations to realize the effects of instruction, which may take at least eight weeks for instructional effects to become apparent (Christ et al., 2013).

When examining referral rates for special education evaluations, 33 out of 383 (8.6%) students receiving intensive general education reading intervention were referred for a special education evaluation. The low overall rate of referrals was not surprising, because school districts included in this study all had at least 75% of students proficient on state-approved benchmark assessments. In theory, by including schools that met this *de facto* measure of fidelity of universal supports within a multi-tiered system of support (MTSS; Kovaleski, Marco-Fies, & Boneshefski, 2013), the majority of students' needs are met with the provision of universal instruction, potentially allowing for more efficient allocation of general education resources. By not overwhelming systems with large numbers of students in need of supplemental or intensive supports, general education resources may be used to meet the needs of most struggling students (Tilly, 2008).

Of the 33 students referred for an evaluation, 26 (79%) were found eligible for special education, revealing a high "hit rate" of students referred for an evaluation who were

subsequently found eligible for special education. Once again, this high hit rate was not surprising, because it is consistent with previous findings that high percentages (greater than 70%) of students referred for special education evaluations were subsequently found eligible for special education (Algozzine, Christenson, & Ysseldyke; 1982; Ysseldyke et al., 1997). As demonstrated by the percentage of students found eligible for special education, once a referral to evaluate was made, it appears that a referral for a special education evaluation is the most important predictor of whether a student is eligible for special education.

Previous research has demonstrated that when using traditional SLD identification practices (e.g., ability-achievement discrepancy), extraneous variables, such as sex and race, influence eligibility decisions (Lester & Kelman, 1997; Singer et al., 1989; Ysseldyke et al., 1982; Ysseldyke & Algozzine, 1981). Within RTI assessment systems, however, research has suggested that using RTI data to make educational decisions may ameliorate problems related to disproportionate representation of these groups in special education (Burns & Senesac, 2005; Marston et al., 2003; Speece & Case, 2001; Speece et al., 2003). Within the sample, 6.8% of males and 6.8% of females were identified as having SLD, and 8.6% of the students in the overrepresented racial group and 6.2% of students in the not overrepresented racial group were identified as having SLD. Demographic data were not available across each tier of support from the schools included in this study.

Unfortunately, due to sample size constraints, the impact of race and sex on eligibility status could not be analyzed in this sample. Although recent research has suggested that disproportionate identification of historically overrepresented students in the SLD category may be addressed using RTI data for SLD identification (Burns & Senesac, 2005; Speece & Case, 2001), those findings were based on decision-making criteria that were specific to their

respective studies. What is promising from those studies, however, is that effective instruction combined with clear decision rules about student need appeared to improve equitable practices. More information is needed about the effects of RTI assessment systems on disproportionate identification of historically overrepresented students in practice.

Discussion of Research Question 1

Research Question 1 had two hypotheses. It was hypothesized that students eligible for special education would display lower levels of performance and ROIs than students not referred for special education but who were receiving general education reading intervention. It was also hypothesized that students in Grade 2 would demonstrate lower levels of performance than students in Grade 3 but that students in Grades 2 and 3 would demonstrate similar ROIs. No significant interactions were found between grade and eligibility status on level of performance and ROI based on the results of a 2x2 MANOVA. There were significant main effects of eligibility status and grade on the combined dependent variables. Interpretation of follow-up analysis on the dependent variable of level of performance supported the hypotheses that students eligible for special education would display lower levels of performance than students not referred for an evaluation and that students in in Grade 2 would demonstrate lower levels of performance than students in Grade 3. A follow-up analysis was also conducted on the dependent variable ROI. Results of the follow-up analysis supported the hypotheses that students eligible for special education would display lower ROIs than students not referred for an evaluation and that students in Grades 2 and 3 would demonstrate similar ROIs.

Grade differences. As expected given the developmental progression of reading development, students in Grade 2 demonstrated lower levels of performance than students in Grade 3. Normative information exists that demonstrates students have higher ORF rates as they

progress through grade school (Fast Bridge, 2015; Hasbrouck & Tindal, 2006). Students in

Grades 2 and 3 demonstrated similar ROIs as hypothesized. When considering growth rates of

students, Deno et al. (2001) reported second and third grade students who were eligible for

special education demonstrated very similar ROIs. Additionally, FastBridge (2015) normative

information for ROIs based on students' CBM-R level indicates similar growth rates for second

and third grade students performing below the 30th percentile, growing at an average of 1.38

words correct per minute (WCPM) per week and 1.27 WCPM per week from fall to spring,

respectively. However, when considered as a whole, growth appears to quickly increase at the

early elementary grades and then slows as students age (Hasbrouck & Tindal, 2006; Nese et al.,

2014). Readers with low initial levels of performance have been shown to demonstrate slower

ROIs (Silberglitt & Hintze, 2007).

Eligibility status differences. As anticipated, students found eligible for special

education demonstrated lower levels of performance than students not referred for a special

education evaluation. This finding is consistent with previous research that suggests low

achievement is the defining characteristic of SLD (Kavale & Reese, 1992; Peterson & Shinn,

2002; Ysseldyke et al., 1982). In the current study's sample, second grade students who were

found eligible for special education performed approximately 33 WCPM lower than students not

referred for a special education evaluation. Third grade students who were found eligible for

special education performed approximately 26 WCPM lower than students not referred for a

special education evaluation. Review of the dates that eligibility decisions were made suggests

most eligibility decisions were made during winter and spring. So, to examine at what percentile

rank teams may consider to indicate a discrepant level, mean levels of performance for Grade 2

(40.6 WCPM) and for Grade 3 (75.4 WCPM) were compared to FastBridge (2015, 2016) literacy

norms for winter and spring. These data suggest the mean score of second grade students with

SLD in the sample would fall between the fifth and 10th percentile ranks in winter and below the

fifth percentile rank in spring. Additionally, the mean score of third grade students with SLD in

the sample would fall between the fifth and 10th percentile ranks in winter and below the fifth

percentile rank in spring. These data are consistent with the recommendation of Kovaleski and

colleagues (2013) that a discrepant performance level may be considered to be performance

below the 10th percentile. Therefore, when comparing differences in level of performance

between eligible and not eligible students, the data suggest that Criterion 1 of the SLD criteria

(failure to meet age or grade-level standards as evidenced by achievement significantly below

grade-level peers) was met for this sample of students.

Also, as hypothesized, students found eligible for special education demonstrated lower

ROIs than students who were receiving general education reading intervention. Previous

research has found that growth rates differ as a function of SLD status (Deno et al., 2001). Deno

and colleagues found that, up to and including Grade 4, students eligible for special education

demonstrated lower ROIs than students receiving only general education supports. Growth rates

of students eligible for special education reported by Deno and colleagues revealed mean ROIs

of .57 WCPM per week and .58 WCPM per week for students in Grade 2 and 3, respectively.

In the present study's sample, students found eligible for special education demonstrated

significantly lower ROIs than students not referred for an evaluation. Second grade students

who were found eligible for special education demonstrated a mean ROI of .8 WCPM per week

compared to the mean ROI of 1.2 WCPM per week for students not referred for an evaluation.

Third grade students who were found eligible for special education demonstrated a mean ROI .8

WCPM per week compared to the mean ROI of .9 WCPM per week for students not referred for

an evaluation. When considering differences in ROI across eligibility status groups, students found eligible for special education demonstrated significantly lower ROIs than students not eligible for special education, suggesting that Criterion 2 of the SLD criteria (inadequate response to scientifically based instruction) was met for this sample of students.

Taken together, the data are suggestive that students found eligible for special education in this study demonstrated a dual discrepancy, in that they demonstrated inadequate achievement, based on calculated final levels of performance, and insufficient growth, based on quantified ROIs, compared to students in the sample who were not referred for an evaluation. This is consistent with research proposing that a dual discrepancy appears to be the most consistently-supported framework for assessing students' RTI (Burns & Senesac, 2005; Fuchs et al., 2008; Fuchs et al., 2002; Fuchs et al., 2004; Speece & Case, 2001), which is likely why federal regulations require that a student identified with an SLD using RTI data must demonstrate a dual discrepancy (United States Department of Education, 2006; Kovaleski et al., 2013).

Discussion of Research Question 2

Research Question 2 had two hypotheses. It was hypothesized that level of performance would be the variable most predictive of special education eligibility status. It was also hypothesized that level of performance and ROI together would predict students' eligibility status to a greater extent than level of performance alone. These hypotheses were analyzed for Grade 2 and Grade 3 data separately by conducting two logistic regressions. Results of the logistic regression revealed a statistically significant model with level of performance and ROI as predictor variables, indicating the model was a good fit to the data. When the model was built for both Grade 2 and Grade 3 data, level of performance was found to be a significant predictor of eligibility status, and ROI was not found to be a significant predictor of eligibility status.

Based on the results of the analysis, for every one unit increase in level of performance, the odds of being found eligible for special education decreased by 7.9% in Grade 2 and by 4.1% in Grade 3. Across Grades 2 and 3, the hypothesis that level of performance was most predictive of special education eligibility status was supported. The hypothesis that level of performance and ROI together would predict students' eligibility status to a greater extent than level of performance alone was not supported, as only a minimal increase in explained variance was found when ROI was added to the regression model. This is not to say that assessing students' progress is not important, however. Obtaining a quantified ROI may allow practitioners to more readily assess the needs of students. By monitoring students' progress at regular intervals, school teams can understand whether the supports provided to students are allowing them to make adequate progress toward a meaningful benchmark. Although two students may have a similar post-intervention discrepancy, they may have received very different levels of support to allow them to reach that level of performance. For students whose progress has been accelerated with very intensive supports, those supports may be documented in an educational plan so that they can continue to be provided. For students who have not demonstrated an accelerated ROI, this may trigger teams to further intensify intervention supports until progress is accelerated to close the performance gap, which may include special education.

The results of the regression imply that ROI does not appear to add significantly to team decisions about special education eligibility. This finding is likely influenced based on how Criterion 2 (inadequate response to instruction) is interpreted. By primarily considering level of performance, students would seemingly be identified as SLD based on a point-in-time piece of datum, suggesting teams considered how far behind a student is compared to expectations at the time a decision is made. Although measuring a student's post-intervention status may provide

some indication of responsiveness to instruction, it provides little meaningful information about a student's rate of learning in the absence of a student's initial level of performance or a quantified ROI (Fuchs, 2003). It also provides little guidance to teams as they consider growth in relation to the student's history and the resources required to attain the post-intervention level. Additionally, when testing the assumption of multicollinearity for the MANOVA, moderate correlations were found between level of performance and ROI for referred and eligible students in Grades 2 and 3 in the sample (see Table 3). This may mean that students who were determined to be SLD had shown a pattern of low pre-intervention level of performance and ROI and slow growth during intervention, resulting in a persistently low level of performance that was still apparent at the time an eligibility decision was made. It appeared that students' final benchmark was strongly considered by multidisciplinary evaluation teams, which is consistent with previous research that shows low-achieving students have historically been identified as SLD, making the argument that low achievement is the defining characteristic of SLD (Brown-Waesche et al., 2011; Kavale & Reese, 1992; Peterson & Shinn, 2002; Ysseldyke et al., 1982).

That the results of the regression are consistent with previous research that suggests low achievement is the primary characteristic of SLD is not surprising. Level of performance is likely much easier to quantify and understand in relation to grade-level expectations and standards. As a point-in-time measurement, teams can definitively make judgements about the discrepancy between a student's performance compared to expectations. When judging a student's ROI, however, multiple factors may need to be considered by teams, including selecting an appropriate method for quantifying ROI, determining adequate intra-individual improvement, and understanding a student's ROI in relation to normative growth rates.

For this study, all ROIs were calculated by the primary investigator using OLS and recommendations by Runge et al. (in review) to ensure ROIs could be accurately calculated and compared. Although OLS is generally recommended in the research literature as a way to quantify growth (Deno et al., 2001; Good & Shinn, 1990; Shinn et al., 1989), many other methods exist for calculating ROI, which include last minus first, split middle, and Tukey, among others (Flinn & McCrea, 2013). Additionally, recent research has provided statistical criticisms of using OLS to model growth including the vulnerability of OLS to extreme outliers (Haupt et al., 2013) and that OLS assumes a linear ROI, which may not be the case (Haupt et al., 2013; Nese et al., 2013). Flinn and McCrea demonstrated that different growth rates may be obtained depending on which method practitioners use for calculating a student's ROI. As has been further revealed by Runge and colleagues, even if OLS is used by practitioners as the method for calculating ROI, the measurement interval selected by the practitioner (e.g., actual date of progress monitoring assessment, instructional day on which progress monitoring occurred, school week during which progress monitoring occurred) can greatly impact the calculated ROI. Runge and colleagues recommended that the OLS regression line be calculated using a consistent measurement interval, and ideally the actual date of progress monitoring assessment. Although all ROIs were quantified in a consistent manner for this study, practitioners in the field may use different methods for calculating ROI as part of special education evaluations, which impacts their ability to interpret a student's growth. Until all practitioners quantify ROI in a consistent manner, continued difficulty interpreting students' responsiveness to instruction will likely occur.

Although one benefit of using CBM-R is that it provides information about both inter-individual and intra-individual performance, another difficulty that lends itself to determining

acceptable growth is evaluating intra-individual performance. Multidisciplinary evaluation teams are faced with the task of determining acceptable growth, which has been shown to be impacted by a student's initial level of performance, with students with low performance levels demonstrating much slower growth rates (Silberglitt & Hintze, 2007). Students with high performance levels, however, have also demonstrated slower growth rates, which complicates interpretation of ROI even more. Furthermore, although normative ROI data have been obtained (Deno et al., 2001; Hasbrouck & Tindal, 2006; Nese et al., 2013), and commercially available CBM-R assessments (e.g., FAST CBM-R) provide normative growth information, much is still unknown about appropriately using this information for high-stakes decision-making. Teams need to consider a variety of variables to understand whether intra-individual growth is adequate, including a student's initial level of performance, which impacts ROI, as well as reasonable or ambitious growth rates (Fuchs & Fuchs, 1993) that would allow a student to close the achievement gap.

FastBridge (2015) provided weekly growth rates for three performance groups based on high achieving (>85th percentile rank on CBM-R), typically achieving (30th to 84th percentile rank on CBM-R), and low achieving (less than the 30th percentile on CBM-R). They also provided ranked growth rates from the lowest to highest percentile in the norm group, which are growth rates independent of CBM-R level. To illustrate the difference between these two growth scores, a second grade student with an ROI of 1.70 WCPM per week from fall to spring would have a growth rate at the 50th percentile; however, a typically achieving student (based on CBM-R level) would be expected to grow at a rate of 1.36 WCPM per week. If a team compares a student's ROI to normative data, it may appear that a student is responding similarly to peers, but if the growth is not accelerated, the student may not close the achievement gap. Therefore,

114

simply examining a student's growth in relationship to a normative sample may provide information about whether a student's progress is typical, but it may not provide information about need.

When considering the concerns about calculating and interpreting ROIs noted above, it is not surprising that ROI was not a significant predictor of eligibility status in this study's sample. Furthermore, without clear and consistent direction across federal and state regulations (Flinn, 2015; Maki et al., 2015), practitioners are left to make decisions about whether a student's responsiveness to instruction is inadequate with little guidance. That is not to say, however, that simply providing guidelines about what constitutes a discrepant ROI (e.g., growth below a certain percentile) would solve the problem. Maki and colleagues (2016) reported that even with objective decision-making criteria for identifying SLD, practitioners may come to different conclusions about a student's eligibility for special education. In the present study, students in the referred and eligible group demonstrated significantly lower levels of performance and ROIs, however, which suggests MDE teams in this study did appear to identify students who displayed persistently poor responses to intervention as envisioned by the model's developers and institutionalized in the IDEA regulations.

Given that limited empirical guidance about insufficient ROIs exist, Kovaleski and colleagues (2013) suggested that teams can use ROI trajectories, projecting growth into the future to examine the amount of time it would take for a student to attain a meaningful benchmark as well as the resources required to sustain that growth. This means that students would inevitably demonstrate different ROIs that would be deemed acceptable by multidisciplinary evaluation teams. Thus, ROI may be best understood in relation to initial and final level as well as the resources that were required to get the student to the final level of

performance. A strong ROI for a student with a significantly discrepant initial level may still result in a student being unacceptably behind, whereas a student with a minimally discrepant initial level and a slower ROI may have a post-intervention level that is comparable to peers. Therefore, teams may likely consider the student's final level more strongly when making a decision about special education eligibility, which is suggested by the results of this study. It appears that ROI may nevertheless be interpreted within the context of a student's initial level (Silberglitt & Hintze, 2007) and within the context of a student's history, such that teams need to decide whether general education resources can sustain adequate growth towards a meaningful target (Kovaleski et al., 2013).

Limitations

There were multiple limitations associated with this study. The first set of limitations involved the study's sample, which ultimately affected the scope of the study. Complications related to the sample arose, which included difficulty obtaining free and reduced lunch data for students in the sample. Because free and reduced lunch data were not available, the variable was excluded from analyses. Complications related to the sample resulted in changes to the research question, study design, and subsequent analyses. Sample size complications impacted the ability to include potentially extraneous variables (e.g., student sex and race) in the regression equation.

A second limitation, which may have impacted the internal validity of this study is related to the interventions provided to students across each school site. Although guidance from the state indicates that students included in the study's sample must receive intensive intervention and provides recommendations about acceptable interventions, it is unknown what intervention each student in the sample was receiving. Students' levels of performance and ROIs are related to the quality of the instruction and intervention that was provided. In the absence of

intervention information, the inclusion criteria for schools, which required sites to have at least 75% of students proficient on state-approved benchmark assessment, may serve as a proxy for the quality of instruction at the school sites and suggest that the sites included in the sample meet a minimum standard of instructional quality. Future studies may control for the effects of intervention quality as it relates to level, ROI, and thus eligibility status.

A third limitation that may have impacted the internal validity of this study is related to the fidelity of MTSS implementation in the sites included in this study. As previously reported, having 75% of students proficient on a state-approved universal screener was used as a *de facto* measure of fidelity for universal supports (Kovaleski, Marco-Fies, & Boneshefski, 2013). In practice, however, much more is required for MTSS implementation than having most students proficient in basic skills. To implement MTSS, schools must also provide evidence-based core instruction, use regular universal screening and progress monitoring data for instructional decision-making, and provide targeted and intensive supports for students in need (United States Department of Education Office of Special Education and Rehabilitative Services, 2011). Therefore, in practice, schools that may bring at least 75% of students to proficiency may not be implementing MTSS. No fidelity assessment information (e.g., Self-Assessment of MTSS; Stockslager, Castillo, Brundage, Childs, & Romer, 2016) was available to gain further insight about the differential effectiveness of practices in schools that are implementing MTSS with fidelity.

A limitation that impacts the external validity is the characteristics of the sample and location of the study. The sample was obtained from schools that were deemed to be implementing MTSS with fidelity based on having at least 75% of students proficient on a state-approved benchmark assessment. In the absence of more specific fidelity data related to MTSS

117

implementation (e.g., as assessed by the Self-Assessment of MTSS; Stockslager et al., 2016), it cannot be determined to what extent the results of this study would generalize to schools implementing MTSS with a similar level of fidelity. Additionally, because the sites included in this study demonstrated effective universal supports based on proficiency levels from the sites, the results may not be applicable to school systems that have not been successful at bringing most students up to proficiency because of the inherent differences related to resource allocation that arise between schools with effective and ineffective universal supports.

One delimitation that may increase internal validity for the study relates to the assessment materials that were used to obtain level of performance and ROI data. Ardoin and Christ (2009) suggested that growth estimates should be derived from the same passage sets, and Shinn and colleagues (1989) found that consistent difficulty of passages within a passage set may be of greater importance than the level of difficulty of the passages for estimating growth. All CBM-R data were obtained from FAST CBM-R passages (FastBridge, 2015), which ensures consistency of assessment materials across all students included in the sample.

Implications for Practice

Implementing MTSS

The descriptive data obtained as part of this study shed some light on the implications of having effective universal supports for students. The referral rate of 8.6% is suggestive that schools that are able to meet most students' needs (i.e., at least 75% of the students) with universal supports may have fewer referrals for special education. It should be noted that the referral rate of 8.6% was out of a sample of second and third grade students who were identified as having reading concerns through universal screeners. This indicates that the percentage of students referred for evaluations would be much lower when considering it in the context of all

second and third grade students enrolled in these schools. By limiting the time school

psychologists are required to spend on evaluations for special education eligibility due to

effective instructional supports, increased time may be spent in other roles, such as consultation

and coaching to improve educational practices and engaging in problem-solving across the

school system.

Additionally, although a cause-effect conclusion may not be made about the quality of

the dataset and its impact on team decisions about student need, previous research suggests

LEAs should be encouraged to have effective systems in place for screening and monitoring

students' progress to allow practitioners to have an adequate dataset for making high-stakes

decisions. By setting standards for monitoring students' progress and providing guidance about

the requirements of an adequate dataset, the effects of instruction and intervention may be

allowed to be realized (Christ et al., 2013).

A limited sample size prevented statistical analyses of level of performance and ROI by

sex and race. Although in the literature the effects of MTSS implementation and assessment

using RTI data are promising at ameliorating disproportionate identification of historically

disadvantaged students (Burns & Senesac, 2005; Marston et al., 2003; Speece & Case, 2001;

Speece et al., 2003), the effects of MTSS on these students' performance should continue to be

explored.

Eligibility Decision-Making Using RTI Data

Consistent with previous findings by Algozzine et al. (1982) and Ysseldyke et al. (1997),

most students who were referred for a special education evaluation were found eligible. When

considering this information along with the minimal changes in classification accuracy in the

regression model once level and ROI were included, the results may be suggestive that the single

most important predictor of special education eligibility is whether or not a student was referred for an evaluation. Theoretically, within an MTSS framework, data about student level and progress are available to practitioners prior to making a referral for a special education evaluation. At the point that consent to evaluate is received, the practitioner may simply be allowed to examine the available data in a different light (i.e., for special education eligibility rather than just instructional decision-making). Thus, by implementing an MTSS framework and using RTI data to drive decision-making, special education evaluations may simply involve summarizing the available data that are indicative of SLD. More importantly, however, it may be argued that the original conceptualizations of an MTSS framework and an RTI assessment process (Batsche et al., 2005; Tilly, 2008) where a longitudinal gating procedure for identifying which students are most likely to be in need of increasingly intensive intervention, and thus likely to be eligible for special education, can indeed be realized. Key features of this process were noted in this study, as demonstrated by the low number of referrals for special education evaluations, significant differences in level of performance and ROI between students receiving intensive reading intervention and students determined to be eligible for special education, and a high hit rate of students referred for an evaluation who were subsequently found eligible for special education.

When considering level of performance, it appears that practitioners differentiate between students with SLD and those without based on level of performance. When critiquing identification of SLD using RTI data, Reynolds and Shaywitz (2009) argued that practitioners continued to rely on a discrepancy-based approach to SLD identification. It may be inferred from the data that eligible students demonstrate lower levels of performance, and a discrepant performance level may be more readily interpreted by practitioners as indicative of SLD. The

critique of Reynolds and Shaywitz may be pertinent, because practitioners are often left with insufficient guidelines about interpreting growth and using ROI data along with information about students' history to make determinations about student need.

When considering level of performance and ROI together, the data are suggestive that students with SLD demonstrate a dual discrepancy. The lack of significance for ROI as a predictor in the regression model, however, suggests that it may not be used by practitioners to reliably differentiate between students with SLD and those without SLD in practice. This may be because ROIs invariably need to be interpreted in relation to initial level. Therefore, simply providing interpretation guidelines for ROI based on normative data (e.g., an ROI that is below the 10th percentile) or a discrepancy criterion (e.g., an ROI that is 2.0 times discrepant from expectations) may not provide meaningful information about students' progress.

Assigning criteria for interpreting ROI as is done for interpreting level (e.g., 2.0 times discrepant), may lead to a large number of students for whom practitioners would deem it to be acceptable to not close the achievement gap. Because the results suggested that a quantified ROI in and of itself does not predict eligibility, practitioners may examine students' progress as allowing them to attain a meaningful benchmark (Kovaleski et al., 2013). Therefore, even if a student's progress is similar to a typically-progressing student (e.g., growth at the 50th percentile), a student's ROI may still be determined to be inadequate if it has not allowed the student to reach a meaningful benchmark.

The suggestion to not interpret ROI based solely on normative growth rates does not mean that students' growth may be evaluated in the absence of guidance, however. As with any measurement tool, CBM-R includes error, and thus calculations of students' growth include error. Even though OLS may be the most reliable way to estimate students' growth (Deno et al.,

2001; Good & Shinn, 1990; Shinn et al., 1989), Runge and colleagues (in press) have demonstrated that different ROIs may be obtained depending on how the OLS regression line is calculated. Therefore, practitioners should use a consistent metric, such as the exact date the progress monitoring assessment was administered, to further reduce error associated with measuring growth when making special education eligibility decisions.

Lastly, because students' ROIs were not predictive of special education eligibility status, the results may suggest that ROI is difficult to interpret in isolation. The recommendation of Kovaleski and colleagues (2013) to interpret growth using ROI trajectories may ameliorate the problem of evaluating ROI in isolation. By considering a student's history in relation to the amount of growth needed to reach a meaningful benchmark and the amount of time and resources required to reach that benchmark, teams may more readily assess whether general education or special education resources are required to sustain the growth.

Implications for Future Research

The results of this study may be used to advance future research. First of all, the study may be replicated with a larger sample size, allowing for the model to be built with one set of data and cross validated with another. This may provide insight as to how the model may generalize to another dataset. Furthermore, by replicating this study with a larger sample size, the extent that potentially extraneous variables (e.g., sex, race, access to resources) predict special education eligibility within an RTI system may be examined.

Next, because a proxy for fidelity of universal supports within an MTSS framework was based on the sites' proficiency levels, the results of the current study may not generalize to locations that either do not implement MTSS with fidelity or that recognize fidelity based on a more comprehensive measure of MTSS fidelity (e.g., Self-Assessment of MTSS; Stockslager et

al., 2016). Future studies may explore how predictive level of performance and ROI are of SLD in systems that are implementing MTSS based on a more comprehensive fidelity measure. Additionally, the level of implementation and type of MTSS implementation that would be sufficient to produce similar results to this study may be explored. This study should also be conducted in locations with different demographics to determine whether similar results are obtained.

Given the low referral rate and high hit rate found in the sample, future studies may explore the differential referral rates based on proficiency levels, level of MTSS implementation (e.g., using a more formal MTSS fidelity assessment), and based on type of SLD identification method used. Additionally, differences in the quality of progress monitoring datasets available when making educational decisions, such as the frequency of progress monitoring and the number of data points available to practitioners, for students eligible and not eligible for special education may be explored.

Although FAST CBM-R was the assessment used to obtain progress-monitoring data for this study, additional commercially-available CBM assessments are frequently used in practice. DIBELS Next (Good et al., 2011) and AIMSweb (Pearson Education, 2011) are two such assessments. Additional research should explore whether the results of this study can be replicated using other progress monitoring assessments. Likewise, computer-adaptive tests, such as STAR Reading (Renaissance Learning, 2012b) and Measures of Academic Progress (Northwest Evaluation Association, 2004), which report scores as standard scores, are also frequently used within MTSS frameworks. Future research should examine whether levels of performance and ROIs calculated using standard scores predict special education eligibility.

Lastly, for this study, final level of performance was examined as a predictor of eligibility along with ROI. Because ROI may be best understood in relation to initial and final level, initial level of performance may be explored as a covariate. By examining to what extent a student's initial level of performance influences eligibility decisions, additional insight may be obtained about interpreting growth over time.

Summary

The aim of this study was to examine to what extent level of performance and ROI, which are two major SLD criteria, predict special education eligibility decisions. Previous research has supported the use of a dual discrepancy, which requires students to be discrepant in both level of performance and ROI (Fuchs et al., 2002; Fuchs & Fuchs, 1998; Speece & Case, 2001; Vellutino et al., 1996). So, the extent to which this is borne out in practice was investigated.

Although the characteristics of the sample limited the extent to which statistical analyses could occur, some interesting, but tentative, trends emerged in the data. Examination of mean number of data points in the sample suggested, in general, practitioners had adequate data sets available to make high-stakes decisions based on recommendations by Ardoin et al. (2013) and Christ et al. (2012). Review of the progress monitoring data points across the sample revealed 90.3% of cases had at least 14 progress monitoring data points in the dataset.

Additionally, low rates of referrals for evaluation emerged at Grades 2 and 3, with 8.6% of students in the sample being referred for a special education evaluation. These data suggest that, in systems where most students are able to reach proficiency with universal supports, general education resources may be allocated efficaciously across the system to meet most students' needs. This may allow students to access general education curricula without

124

demonstrating a need for special education. Of the students referred for an evaluation in the sample, 79% were found eligible for special education. This finding is consistent with previous research indicating there was a high probability that students would be found eligible for special education if they were referred for an evaluation (Algozzine et al., 1982; Ysseldyke et al., 1997). Having historical data about students' performance over time may impact the way evaluations are conducted using RTI data. By having discrepancy and progress data available at the time a consent for evaluation is obtained, evaluators may be able to summarize data that are indicative of SLD. Additionally, using RTI assessment data may be an efficient and effective way to identify SLD, as demonstrated by low rates of referrals for special education evaluations, significant differences in levels of performance and ROI between eligible and not eligible students, and high hit rates of students referred for an evaluation who were subsequently found eligible for special education.

No statistical analyses could be conducted on potentially extraneous variables (e.g., sex, race) due to the characteristics of the sample. Previous research has suggested that disproportionate identification of overrepresented students may be ameliorated by using RTI data (Burns & Senesac, 2005; Speece & Case, 2001). Further research is needed to understand how the use of RTI data in practice impacts the identification of historically overrepresented groups in special education.

When analyzing performance levels and ROIs across grade levels, the results supported all hypotheses. As expected given the developmental nature of learning to read, students in Grade 2 demonstrated lower levels of performance than students in Grade 3. Much normative information exists that reveals ORF rates generally increase as students progress through grade school (Fast Bridge, 2015; Hasbrouck & Tindal, 2006). In this sample, students in Grades 2 and

3 demonstrated similar ROIs. Normative information about growth has generally suggested that growth quickly accelerates at the early elementary grades and slows as students age (Hasbrouck & Tindal, 2006; Nese et al., 2014). Deno and colleagues (2001) reported that growth rates for students eligible for special education were lower than that of their general education peers and were similar across grade levels in the elementary grades. Although the majority of the students in the sample were not eligible for special education, they were students with reading skill deficits, so it may be that students with low initial levels of performance may demonstrate similar growth rates.

When analyzing level of performance and ROI across eligibility status groups, the results supported all hypotheses. As expected, students who were found eligible for special education demonstrated lower levels of performance than students receiving general education reading intervention but not referred for an evaluation, which is consistent with previous research demonstrating performance discrepancies between students with SLD and those without (Kavale & Reese, 1992; Peterson & Shinn, 2002; Ysseldyke et al., 1982). Also, as hypothesized, students who were found eligible for special education demonstrated lower ROIs than students who were not referred for a special education evaluation. Previous research has demonstrated that struggling readers demonstrate lower growth rates than typical readers (Deno et al., 2001; Fuchs et al., 2004). Taken together, the data are suggestive that students found eligible for special education have dually discrepant levels of performance and ROIs (Fuchs & Fuchs, 1998). The data imply that a dual discrepancy may be borne out in practice and are consistent with federal regulations (United States Department of Education, 2006).

Results of the regression analyses revealed level of performance was a significant predictor of eligibility status, but ROI was not. Additionally, adding ROI to the model resulted

in minimal changes in explained variance. Although there was some evidence that students who were found eligible for special education demonstrated dual discrepancies, it seems as though students' post-intervention level influences eligibility decisions to a greater extent than quantified growth rates. Although federal SLD regulations indicate that students must demonstrate inadequate achievement and an inadequate response to scientifically-based instruction, the results of this study may suggest that a student's response is quantified not as ROI but perhaps as post-intervention status because a numerical ROI did not appear to add significantly to team decisions about special education eligibility. Obtaining a quantified ROI, however, may allow practitioners to more readily assess the needs of students. By monitoring students' progress at regular intervals, school teams can understand whether the supports provided to students are allowing them to make adequate progress toward a meaningful benchmark and adjust supports as necessary. By examining the resources required to sustain adequate growth, school teams may make more informed decisions about the need for special education services.

It appears that multidisciplinary evaluation teams used student's final benchmark (Good et al, 2001) for determining responsiveness to intervention, which also provides information about discrepancy from grade-level standards and expectations. This finding is consistent with previous research that indicated low achievement is the defining characteristic of SLD (Brown-Waesche et al., 2011; Kavale & Reese, 1992; Peterson & Shinn, 2002; Ysseldyke et al., 1982). Final benchmark does not provide information about rate of learning, however. So, ROI may be best understood in relation to initial and final level as well as the resources that were required to get the student to the final level of performance. Therefore, when making decisions about student need, teams may interpret ROI in relation to students' initial level and final level, which

may allow teams to better understand whether general education resources can bring students to

a meaningful benchmark in a reasonable amount of time (Kovaleski et al., 2013).

References

Ahearn, E. M. (2003). *Specific learning disability: Current approaches to identification and proposals for change*. Alexandria, VA: National Association of State Directors of Special Education. Retrieved from http://nasdse.org/DesktopModules/DNNspot-Store/ProductFiles/191_b35826a4-9b94-4029-9b19-82a2983071ed.pdf

Algozzine, B., Christenson, S. L., & Ysseldyke, J. E. (1982). Probabilities associated with the referral to placement process. *Teacher Education and Special Education*, *5*(3), 19-23.

American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (1999). *The standards for educational and psychological testing*. Washington, DC: American Educational Research Association.

Ardoin, S. P., & Christ, T. J. (2008). Evaluating curriculum-based measurement slope estimates using data from triannual universal screenings. *School Psychology Review*, *37*, 109-125.

Ardoin, S. P., & Christ, T. J. (2009). Curriculum-based measurement of oral reading: Standard errors associated with progress monitoring outcomes from DIBELS, AIMSweb, and an experimental passage set. *School Psychology Review, 38,* 266-283.

Ardoin, S. P., Christ, T. J., Morena, L. S., Cormier, D. C., & Klingbeil, D. A. (2013). A systematic review and summarization of the recommendations and research surrounding curriculum-based measurement of oral reading fluency (CBM-R) decision rules. *Journal of School Psychology*, *51*, 1-18. doi:10.1016/j.jsp.2012.09.004

Barth, A. E., Stuebing, K. K., Anthony, J. L., Denton, C. A., Mathes, P. G., Fletcher, J. M., & Francis, D. J. (2008). Agreement among response to intervention criteria for identifying responder status. *Learning and Individual Differences*, *18*, 296-307.

Bateman, B. (1965). An educational view of a diagnostic approach to learning disorders. In J. Hellmuth (Ed.), *Learning disorders: Vol. 1* (pp. 219-239). Seattle, WA: Special Child Publications.

Boneshefski, M. J., & Runge, T. J. (2014). Addressing disproportionate discipline practices within a school-wide positive behavioral interventions and supports framework: A practical guide for calculating and using disproportionality rates. *Journal of Positive Behavior Interventions*, *16*, 149-158. doi:10.1177/1098300713484064

Bray, J. H., & Maxwell, S. E. (1985). *Multivariate analysis of variance*. Thousand Oaks, CA: Sage.

Brown-Chidsey, R. (2007). No more "waiting to fail." *Educational Leadership*, *65*, 40-46.

Brown-Waesche, J. S., Schatschneider, C. Maner, J. K., Ahmed, Y., & Wagner, R. K. (2011). Examining agreement and longitudinal stability among traditional and RTI-based definitions of reading disability using affected-status agreement statistic. *Journal of Learning Disabilities, 44,* 296-307.

Burns, M. K. (2016). Effect of cognitive processing assessments and interventions on academic outcomes: Can 200 studies be wrong? *Communiqué*, *44*(5) 1, 26-29.

Burns, M. K., & Gibbons, K. A. (2008). *Implementing response-to-intervention in elementary and secondary schools: Procedures to assure scientific-based practices*. New York, NY: Routledge.

Burns, M. K., Scholin, S. E., Kosciolek, S., & Livingson, J. (2010). Reliability of decision-making frameworks for response to intervention for reading. *Journal of Psychoeducational Assessment*, *28*, 102-114. doi:10.1177/0734282909342374

Burns, M. K., & Senesac, B. V. (2005). Comparison of dual discrepancy criteria to assess response to intervention. *Journal of School Psychology, 23,* 393-406.

Cameron, A. C. (2011). Kurtosis. In M. S. Lewis-Beck, A. Bryman, & T. Futing Liao (Eds.) *The SAGE Encyclopedia of Social Science Research Methods*. Thousand Oaks, CA: Sage Publications, Inc.

Christ, T. J. (2006). Short-term estimates of growth using curriculum-based measurement of oral reading fluency: Estimating standard error of the slope to construct confidence intervals. *School Psychology Review*, *35*, 128-133.

Christ, T. J., & Ardoin, S. P. (2009). Curriculum-based measurement of oral reading: Passage equivalence and probe-set development. *Journal of School Psychology*, *47*, 55-75. doi:10.1016/j.jsp.2008.09.004

Christ, T. J., Monaghen, B. D, Zopluoglu, C., & Van Norman, E. R. (2013). Curriculum-based measurement of oral reading: Evaluation of growth estimates derived with pre-post assessment methods. *Assessment for Effective Instruction*, *38*, 139-153.

Christ, T. J., & Silberglitt, B. (2007). Estimates of standard error of measurement for curriculum-based measures of oral reading fluency. *School Psychology Review, 1,* 130-146.

Christ, T. J., Silberglitt, B., Yeo, S., & Cormier, D. (2010). Curriculum-based measurement of oral reading: An evaluation of growth rates and seasonal effects among students served in general and special education. *School Psychology Review*, *39*, 447-462.

Christ, T. J., Zopluoglu, C., Long, J. D., & Monaghen, B. D. (2012). Curriculum-based measurement of oral reading: Quality of progress monitoring outcomes. *Exceptional Children*, *78*, 356-373.

Cortiella, C., & Horowitz, S. H. (2014). *The state of learning disabilities: Facts, trends, and emerging issues*. New York, NY: National Center for Learning Disabilities.

Coyne, M. D., Zipoli, R. P., & Ruby, M. F. (2006). Beginning reading instruction for students at risk for reading disabilities: What, how, and when. *Intervention in School and Clinic*, *41*, 161-168.

Cronbach, L. J. (1957). The two disciplines of scientific psychology. *American Psychologist*, *12*, 671-684.

Cronbach, L. J. (1975). Beyond the two disciplines of scientific psychology. *American Psychologist*, *30*, 116-127.

Deno, S. L. (1985). Curriculum-based measurement: The emerging alternative. *Exceptional Children*, *52*, 219-232.

Deno, S. L. (2003). Developments in curriculum-based measurement. *The Journal of Special Education*, *37*, 184-192.

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,* 507-524.

Deno, S. L., & Mirkin, P. K. (1977). *Data-based program modification: A manual*. Reston, VA: Council for Exceptional Children.

Education for All Handicapped Children Act of 1975, Pub. L. No. 94-142, 20 USC 1401 *et seq*. (1975).

Elksnin, L. K., Pedrotty Bryant, D., Gartland, D., King-Sears, M., Rosenberg, M. S., Scanlon, D., … Wilson, R. (2001). LD summit: Important issues for the field of learning disabilities. *Learning Disability Quarterly*, *24*, 297-305.

FastBridge Learning (2015). *Formative Assessment System for Teachers: Benchmarks and norms for 2015-16*. Minneapolis, MN: Author.

FastBridge Learning (2016). *Formative Assessment System for Teachers: Benchmarks and norms for 2016-17*

Feifer, S. G. (2008). Integrating response to intervention (RTI) with neuropsychology: A scientific approach to reading. *Psychology in the Schools*, *45*, 812-825.

Ferchalk, M. R. (2013). Test nationally, benchmark locally: Using local DIBELS benchmarks to predict performance on the PSSA. Retrieved from ProQuest Digital Dissertations. (AAT 3589383).

Fletcher, J. M., Lyon, G. R., Barnes, M., Stuebing, K. K., Francis, D. J., Olson, R. K., … Shaywitz, B. A. (2001, October 1). *Classification of learning disabilities: An evidence-based evaluation.* Paper presented at the 2001 Learning Disabilities Summit – Building a Foundation for the Future. Executive summary retrieved from http://www.air.org/ldsummit

Flinn, C. (2015). Recommended practices for determining rate of improvement: A review of state-level guidance. Retrieved from ProQuest Digital Dissertations. (AAT 3665889).

Flinn, C. & McCrea, A. E. (2013, June). *Student growth: Graphing, calculating, and interpreting rate of improvement data*. Session presented at the Pennsylvania RTII Implementers' Forum, Hershey, PA.

Foorman, B. R., Francis, D. J., Fletcher, J. M., Schatschneider, C., & Mehta, P. (1998). The role of instruction in learning to read: Preventing reading failure in at-risk children. *Journal of Educational Psychology*, *90*, 37-55.

Forness, S. R., Keogh, B. K., Macmillan, D. L., Kavale, K. A., & Gresham, F. M. (1998). What is so special about IQ? The limited explanatory power of cognitive abilities in the real world of special education. *Remedial and special education*, *19*, 315-322. doi:10.1177/074193259801900602

Francis, D. J., Fletcher, J. M., Stuebing, K. K., Lyon, G. R., Shaywitz, B. A., & Shaywitz, S. E. (2005). Psychometric approaches to the identification of LD: IA and achievement scores are not sufficient. *Journal of Learning Disabilities*, *38*, 98-108.

Fuchs, D., Compton, D. L., Fuchs, L. S., Bryant, J., & Davis, G. N. (2008). Making "secondary intervention" work in a three-tier responsiveness-to-intervention model: Findings from the first grade longitudinal reading study of the National Research Center on Learning Disabilities. *Reading and Writing: An Interdisciplinary Journal*, *21*, 413-436.

Fuchs, D., & Deshler, D. D. (2007). What we need to know about responsiveness to intervention (and shouldn't be afraid to ask). *Learning Disabilities Research & Practice*, *22*, 129-136.

Fuchs, D., Fuchs, L. S., & Compton, D. (2004). Identifying reading disabilities by responsiveness-to-instruction: Specifying measures and criteria. *Learning Disability Quarterly*, *27*, 216-227. doi:10.2307/1593674

Fuchs, D., Mock, D., Morgan, P. L., & Young, C. L. (2003). Responsiveness-to-intervention: Definitions, evidence, and implications for the learning disabilities construct. *Learning Disabilities Research and Practice*, *18*, 157-171.

Fuchs, L. S., & Deno, S. L. (1991). Paradigmatic distinctions between instructionally relevant measurement models. *Exceptional Children*, *57*, 488-500.

Fuchs, L. S., & Deno, S. L. (1994). Must instructionally useful performance assessment be based in the curriculum? *Exceptional Children*, *64*, 15-24.

Fuchs, L. S. (1986). Monitoring progress among mildly handicapped pupils: Review of current

    practice and research. *Remedial and Special Education*, *7*, 5-12.

Fuchs, L. S. (2003). Assessing intervention responsiveness: Conceptual and technical issues.

    Learning Disabilities Research & Practice, 18, 172-186.

Fuchs, L. S., & Fuchs, D. (1998). Treatment validity: A unifying concept for reconceptualizing

    the identification of learning disabilities. *Learning Disabilities Research & Practice*, *13*,

    204-219.

Fuchs, L. S., & Fuchs, D. (2011). *Using CBM for progress monitoring in reading*. Washington,

    DC: National Center on Student Progress Monitoring. Retrieved from ERIC Database.

    (ED519252).

Fuchs, L. S., Fuchs, D., & Speece, D. L. (2002). Treatment validity as a unifying construct for

    identifying learning disabilities. *Learning Disability Quarterly*, *25*, 33-45.

    doi:10.2307/1511189

Garson, G. D. (2016). *Logistic regression: Binary & multinomial*. Asheboro, NC: Statistical

    Publishing Associates.

Good, R. H., Kaminski, R. A., Cummings, K. D., Dufour-Martel, C., Petersen, K., Powell-Smith,

    K., … Wallin, J. (2011). DIBELS Next. Retrieved from http://dibels.org

Good, R. H., & Shinn, M. R. (1990). Forecasting the accuracy of slope estimates for reading

    curriculum-based measurement: Empirical evidence. *Behavioral Assessment*, *12*, 179-

    193.

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

Gresham, F. (2001, October 1). *Responsiveness to intervention: An alternative approach to the identification of learning disabilities*. Paper presented at the 2001 Learning Disabilities Summit – Building a Foundation for the Future. Executive summary retrieved from http://www.air.org/ldsummit

Gresham, F. M., & Vellutino, F. R. (2010). What is the role of intelligence in the identification of specific learning disabilities? Issues and clarifications. *Learning Disabilities Research & Practice*, *25*, 194-206.

Hallahan, D. P., & Mercer, C. D. (2001). *Learning disabilities: Historical perspectives*. Executive Summary from ERIC

Hammill, D. D., & Larsen, S. C. (1974). The effectiveness of psycholinguistic training. *Exceptional Children*, *41*, 5-14.

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

Hauerwas, L. B., Brown, R., & Scott, A. M. (2013). Specific learning disability and response to intervention: State-level guidance. *Exceptional Children*, *80*, 101-120.

Haupt, H., Lösel, F., & Stemmler, M. (2013). Quantile regression analysis and other alternatives to ordinary least squares regression: A methodological comparison on corporal punishment. *Methodology, 10*, 81-91. doi:10.1027/1614-2241/a000077

Hintze, J. M., & Christ, T. J. (2004). An examination of variability as a function of passage variance in CBM progress monitoring. *School Psychology Review*, *33*, 204-217.

Hoskyn, M., & Swanson, H. L. (2000). Cognitive processing of low achievers and children with reading disabilities: A selective meta-analytic review of the published literature. *School Psychology Review*, *29*, 102-119.

Hosp, J. L., Hosp, M. K., Howell, K. W., & Allison, R. (2014). *The ABCs of curriculum-based evaluation: A practical guide to effective decision making*. New York, NY: The Guilford Press.

Hosp, J. L., & Reschly, D. J. (2004). Disproportionate representation of minority students in special education: Academic, demographic, and economic predictors. *Exceptional Children*, *70*, 185-199.

Individuals with Disabilities Education Improvement Act of 2004 (IDEA), Pub. L. No. 108-446. (2004).

Jenkins, J., & O'Connor, R. (2001, October 1). *Early identification and intervention for young children with reading/learning disabilities*. Paper presented at the 2001 Learning Disabilities Summit – Building a Foundation for the Future. Executive summary retrieved from http://www.air.org/ldsummit

Jenkins, J. R., Zumeta, R., & Dupree, O. (2005). Measuring gains in reading ability with passage reading fluency. *Learning Disabilities Research & Practice, 20,* 245-253.

Jimerson, S. R., Burns, M. K., & VanDerHeyden, A. M. (Eds.). (2007). *Handbook of response to intervention: The science and practice of assessment and intervention*. New York, NY: Springer.

Johnson, R. A., & Wichern, D. W. (2007). *Applied multivariate statistical analysis* (6th ed.). Upper Saddle River, NJ: Pearson.

Kavale, K. A. (2001, October 1). *Discrepancy models in the identification of learning disability*. Paper presented at the 2001 Learning Disabilities Summit – Building a Foundation for the Future. Executive summary retrieved from http://www.air.org/ldsummit

Kavale, K. A., & Flanagan, D. P. (2007). Ability–achievement discrepancy, response to intervention, and assessment of cognitive abilities/processes in specific learning disability Identification: Toward a contemporary operational definition. In S. R. Jimerson, M. K. Burns & A. VanDerHeyden (Eds.), *Handbook of response to intervention: The science and practice of assessment and intervention.* (pp. 130-147). New York, NY: Springer.

Kavale, K. A., & Forness, S. R. (2000). What definitions of learning disability say and don't say: A critical analysis. *Journal of Learning Disabilities*, *33*, 239-256.

Kavale, K. A., & Reese, J. H. (1992). The character of learning disabilities: An Iowa profile. *Learning Disability Quarterly*, *15*, 74-94.

Kearns, D. M., & Fuchs, D. (2013). Does cognitively focused intervention improve the academic performance of low-achieving students? *Exceptional Children*, *79*, 263-290.

Keogh, B. K. (1987). Learning disabilities: In defense of a construct. *Learning Disabilities Research*, *3*, 4-9.

Kirk, S. A. (1962). *Educating exceptional children*. Boston, MA: Houghton Mifflin.

Kovaleski, J. F., Marco-Fies, C. M., & Boneshefski, M. J. (2013). *Treatment integrity: Ensuring the "I" in RtI*. RtI Action Network. Available at: http://rtinetwork.org/getstarted/evaluate/treatment-integrity-ensuring-the-i-in-rti

Kovaleski, J. F., & Pedersen, J. A. (2008). Best practices in data analysis teaming. In A. Thomas
& J. Grimes (Eds.), *Best practices in school psychology, V* (Vol. 2, pp. 115-129).
Bethesda, MD: National Association of School Psychologists.

Kovaleski, J. F., & Pedersen, J. A. (2014). Best practices in data-analysis teaming. In P. L.
Harrison & A. Thomas (Eds.), *Best practices in school psychology: Data-based decision
and collaborative decision making*. Bethesda, MD: National Association of School
Psychologists.

Kovaleski, J. F., VanDerHeyden, A. M., & Shapiro, E. S. (2013). *The RTI approach to
evaluating learning disabilities*. New York, NY: Guilford.

Lester, G., & Kelman, M. (1997). State disparities in the diagnosis and placement of pupils with
learning disabilities. *Journal of Learning Disabilities*, *30*, 599-607.

Lichtenstein, R. (2014). Best practices in identification of learning disabilities. In P. Harrison &
A. Thomas (Eds.), *Best practices in school psychology: Data-based and collaborative
decision making* (pp. 331-354). Bethesda, MD: National Association of School
Psychologists.

Long, J. S. (1997). *Regression models for categorical and limited dependent variables*.
Thousand Oaks, CA: Sage Publications.

Lund, A., & Lund, M. (2013a). Binomial logistic regression. Retrieved from
https://statistics.laerd.com/premium/spss/blr/binomial-logistic-regression-in-spss.php

Lund, A., & Lund, M. (2013b). Identifying the type of univariate outlier. Retrieved from
https://statistics.laerd.com/premium/spss/twm/two-way-manova-in-spss-11.php#dealing-
outliers

Lund, A., & Lund, M. (2013c). Two-way MANOVA in SPSS statistics. Retrieved from

    https://statistics.laerd.com/spss-tutorials/two-way-manova-using-spss-statistics.php

Lyon, G. R., & Fletcher, J. M. (2001). Early warning system. *Education Matters*, *1*, 23-29.

    Retrieved from http://educationnext.org/files/ednext20012_22.pdf

Lyon, G. R., Fletcher, J. M., Shaywitz, S. E., Shaywitz, B. A., Torgesen, J. K., Wood, F. B,. …

    Olson, R. (2001). Rethinking learning disabilities. In C. E. Finn Jr., A. J., Rotherham, &

    C. R. Hokanson Jr. (Eds.) *Rethinking special education for a new century* (pp. 259-287).

    Washington, DC: Thomas B. Fordham Foundation. Retrieved from https://edex.s3-us-

    west-2.amazonaws.com/publication/pdfs/special_ed_final_7.pdf

Maki, K. E., Burns, M. K., & Sullivan, A. (2016). Learning disability identification consistency:

    The impact of methodology and student evaluation data. *School Psychology Quarterly*, 1-

    14. doi:10.1037/spq0000165

Maki, K. E., Floyd, R. G., & Roberson, T. (2015). State learning disability criteria: A

    comprehensive review. *School Psychology Quarterly*, *30*, 457-469.

    doi:10.1037/spq0000109

Marston, D. (1989). A curriculum-based measurement approach to assessing academic

    performance: What it is and why do it. In M. Shinn (Ed.), *Curriculum-based*

    measurement: Assessing special children (pp.18–78). New York: Guilford Press.

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 and Practice, *18*, 187-200.

Mather, N., & Roberts, R. (1994). Learning disabilities: A field in danger of extinction?

    Learning Disabilities Research & Practice, *9*, ,49-58.

McGowan, M. R., Runge, T. J., & Pedersen, J. A. (2016). Using curriculum-based measures for identifying gifted learners. *Roeper Review, 38*, 93-106. doi:10.1080/02783193.2016.1150376

Melby-Lervag, M., & Hulme, C. (2013). Is working memory training effective? A meta-analytic review. *Developmental Psychology*, *49*, 270-291. doi:10.1037/a0028228

Salkind, N. J. (Ed.). (2011). *Encolopedia of Measurement and Statistics*. Thousand Oaks, CA: Sage Publications, Inc. doi:http://dx.doi.org/10.4135/9781412952644

National Center for Education Statistics. (n. d.). *Enrollment and percentage distribution of enrollment in public elementary and secondary schools, by race/ethnicity and region: Selected years, fall 1995 through fall 2025*. Retrieved from https://nces.ed.gov/programs/digest/d15/tables/dt15_203.50.asp

National Center on Intensive Intervention. (n.d.). *Academic progress monitoring GOM*. Retrieved from http://www.intensiveintervention.org/chart/progress-monitoring

National Center on Learning Disabilities. (n.d.). *RTI-based SLD identification toolkit*. Retrieved from: http://www.rtinetwork.org/toolkit

National Commission on Excellence in Education. (1983). *A nation at risk: The imperative for educational reform*. Retrieved from http://www2.ed.gov/pubs/NatAtRisk/index.html

National Reading Panel. (2000). *Teaching students to read: An evidence-based assessment of the scientific research literature on reading and its implications for reading instruction*. Washington, DC: National Institute of Child Health and Human Development, National Institutes of Health, U. S. Government Printing Office. Retrieved from: https://www.nichd.nih.gov/publications/pubs/nrp/documents/report.pdf

Nese, J. F. T, Biancarosa, G., Cummings, K., Kennedy, P., Alonzo, J., & Tindal, G. (2013). In search of average growth: Describing within-year oral reading fluency growth across grades 1-8. *Journal of School Psychology*, *51*, 625-642.

Northwest Evaluation Association. (2004). *Measures of academic progress*. Lake Oswego, OR: Author.

Office of Special Education and Rehabilitative Services. (2011). *Letter to State Directors of Special Education*. Retrieved from https://www2.ed.gov/policy/speced/guid/idea/memosdcltrs/osep11-07rtimemo.pdf

Pearson Education, Inc. (2011). *AIMSWeb*. San Antonio, TX: Author.

Peduzzi, P., Concato, J., Kemper, E., Holford, T. R., & Feinstein, A. R. (1996). A simulation study of the number of events per variable in logistic regression analysis. *Journal of Clinical Epidemiology*, *49*, 1373-1379.

Peterson, K. M. H., & Shinn, M. R. (2002). Severe discrepancy models: Which best explains school identification practices for learning disabilities? *School Psychology Review*, *31*, 459-476.

Petrucci, C. J. (2009). A primer for social worker researchers on how to conduct a multinomial logistic regression. *Journal of Social Service Research*, *35*, 193-205.

Pituch, K. A., & Stevens, J. P. (2016). *Applied multivariate statistics for the social sciences* (6th ed.). New York, NY: Routledge.

Renaissance Learning Inc. (2012a). *STAR Math: Computer-adaptive math test*. Wisconsin Rapids, WI: Author.

Renaissance Learning Inc. (2012b). *STAR Reading: Computer-adaptive reading test*. Wisconsin Rapids, WI: Author.

Reynolds, C. R., & Shaywitz, S. E. (2009). Response to intervention: Ready or Not? Or, from wait-to-fail to watch-them-fail. *School Psychology Quarterly*, *24*, 130-145. doi:10.1037/a0016158

Runge, T. J., Bennyhoff, C. F., Ferchalk, M. R., & McCrea (in review). Data analyzer beware: The role of measurement interval in rate of improvement calculation.

Schwaighofer, M., Fische, F., & Buhner, M. (2015). Does working memory training transfer? A meta-analysis including training conditions as moderators. *Educational Psychologist*, 50, 138-166.

Shapiro, E. S., & Guard, K. B. (2014). Best practices in setting progress monitoring goals for academic skill improvement. In P. Harrison & A. Thomas (Eds.), *Best practices in school psychology: Student-level services* (pp. 51-66). Bethesda, MD: National Association of School Psychologists.

Shinn, M. (1995). Best practices in curriculum-based measurement and its use in a problem-solving model. In J. Grimes & A. Thomas (Eds.), *Best practices in school psychology* (Vol. 3., pp. 547-568). Silver Spring, MD: National Association of School Psychologists.

Shinn, M. R. (2002). Best practices in curriculum-based measurement in a problem-solving model. In A. Thomas & J. Grimes (Eds.), *Best practices in school psychology IV* (Vol. 2, pp. 671-679). Washington, DC: National Association of School Psychologists.

Shinn, M. R. (2008). Best practices in using curriculum-based measurement in a problem-solving model. In A. Thomas & J. Grimes (Eds.), *Best practices in school psychology V* (pp. 243-262). Bethesda, MD: National Association of School Psychologists.

Shinn, M. R., Gleason, M. M., & Tindal, G. (1989). Varying the difficulty of testing materials: Implications for curriculum-based measurement. *Journal of Special Education, 89*, 223-233.

Shinn, M. R., Good, R. H., & Stein, S. (1989). Summarizing trend in student achievement: A comparison of methods. *School Psychology Review*, *18*, 356-370.

Silberglitt, B. (2008). Best practices in using technology for data-based decision making. In A. Thomas & J. Grimes (Eds.), *Best practices in school psychology V* (pp. 1869-1884). Bethesda, MD: National Association of School Psychologists.

Silberglitt, B., & Hintze, J. M. (2007). How much growth can we expect? A conditional analysis of R-CBM growth rates by level of performance. *Exceptional Children*, *74*, 71-84. doi:https://doi.org/10.1177/001440290707400104

Singer, J. D., Palfrey, J. S., Butler, J. A., & Walker, D. K. (1989). Variation in special education classification across school districts: How does where you live affect what you are labeled? *American Educational Research Journal*, *26*, 261-281.

Skiba, R. J., Poloni-Staudinger, L. Gallini, S., Simmons, A. B., & Feggins-Azziz, R. (2006). Disparate access: The disproportionality of African American students with disabilities across educational environments. *Exceptional Children*, *72*, 411-424.

Skiba, R. J., Simmona, A. B., Ritter, S., Gibb, A. C., Rausch, M. K., Cuadrado, J., & Chung, C. (2008). Achieving equity in special education: History, status, and current challenges. *Exceptional Children*, *74*, 264-288.

Speece, D. L., & Case, L. P. (2001). Classification in context: An alternative approach to identifying early reading disability. *Journal of Educational Psychology*, *93*, 735-749. doi:10.11003377//A/O022-O663.93.4.735

Speece, D. L., Case, L. P., & Molloy, D. E. (2003). Responsiveness to general education instruction as the first gate to learning disabilities identification. *Learning Disabilities Research and Practice*, *18*, 147-156.

Starkweather, J., & Moske, A. K. (n.d.). Multinomial logistic regression. Retrieved from http://www.unt.edu/rss/class/Jon/Benchmarks/MLR_JDS_Aug2011.pdf

Stockslager, K., Castillo, J., Brundage, A., Childs, K., & Romer, N. (2016). *Self-assessment of MTSS (SAM)*. Florida's Problem Solving/Response to Intervention Project and Florida's Positive Behavior Intervention and Support Project, University of South Florida. Retrieved from http://floridarti.usf.edu/resources/program_evaluation/sam/sam.pdf

Stoiber, K. C. (2014). A comprehensive framework for multitiered systems of support in school psychology. In P. L. Harrison & A. Thomas (Eds.), *Best practices in school psychology: Data-based and collaborative decision making* (pp. 41-70). Bethesda, MD: National Association of School Psychologists.

Stuebing, K., Barth, A., Molfese, P., Weiss, B., & Fletcher, J. (2009). IQ is not strongly related to response to reading instruction: A meta-analytic interpretation. *Exceptional Children*, *76*, 31-51.

Stuebing, K., Fletcher, J., LeDoux, J., Lyon, G. R., Shaywitz, D., & Shaywitz, B. (2002). Validity of IQ discrepancy classifications of reading disabilities: A meta-analysis. *American Educational Research Journal*, *39*, 469-518.

Thornblad, S. C., & Christ, T. J. (2014). Curriculum-based measurement of reading: Is 6 weeks of daily progress monitoring enough? *School Psychology Review*, *43*, 19-29.

Tilly, W. D. (2006). Response to intervention: An overview. What is it? Why do it? Is it worth it? *The Special Edge, 19*(2), 1, 4, 5, 10.

Tilly, W. D. (2008). The evolution of school psychology to science-based practice: Problem solving and the three-tiered model. In A. Thomas & J. Grimes (Eds.), *Best practices in school psychology V* (pp. 17-36). Bethesda, MD: National Association of School Psychologists.

Torgesen, J. K. (2001, October 1). *Empirical and theoretical support for direct diagnosis of learning disabilities by assessment of intrinsic processing weaknesses*. Paper presented at the 2001 Learning Disabilities Summit – Building a Foundation for the Future. Executive summary retrieved from http://www.air.org/ldsummit

Torgesen, J. K. (2004). Avoiding the downward spiral: The evidence that early intervention prevents reading failure. *American Educator*, *28*, 6-19.

Torgesen, J. K. (2009). The response to intervention instructional model: Some outcomes from a large-scale implementation in Reading First schools. *Child Development Perspectives*, *3*, 38-40.

Torgesen, J. K., Wanger, R. K., Rashotte, C. A., Rose, E., Lindamood, P., Conway, T., & Garvan, C. (1999). Preventing reading failure in young children with phonological processing disabilities: Group and individual responses to instruction. *Journal of Educational Psychology*, *91*, 579-593.

UCLA Statistical Group. (n.d.). Annotated SPSS output multinomial logistic regression. Retrieved from http://www.ats.ucla.edu/stat/spss/output/mlogit.htm

United States Department of Education. (n.d.). *Number and percent of children in race/ethnicity category ages 6 through 21 with disabilities served under IDEA, Part B, by disability category and state*. Retrieved from https://www2.ed.gov/programs/osepidea/618-data/static-tables/2015-2016/part-b/child-count-and-educational-environment/1516-bchildcountandedenvironment-11.xlsx

United States Department of Education. (1977). Definition and criteria for defining students as learning disabled. *Federal Register, 42*, 250.

United States Department of Education. (2002). No Child Left Behind Act of 2001. Pub. L. No. 107-110, 115 Stat. 1425.

United States Department of Education. (2005). Commentary and explanation about proposed regulations for IDEA 2004. Retrieved from http://www.wrightslaw.com/idea/law/idea.regs.explain.pdf

United States Department of Education. (2006). Assistance to states for the education of children with disabilities. Final Rule, 71 Fed. Reg. 46539-46845. (August 14, 2006).

United States Department of Education. (2015). Every Student Succeeds Act. Pub. L. No 114-95, 129 Stat. 1802.

VanDerHeyden, A. M. (2011). Technical adequacy of response to intervention decisions. *Exceptional Children*, *77*, 335-350.

VanDerHeyden, A. M., Witt, J. C., & Barnett, D. W. (2005). The emergence and possible futures of response to intervention. *Journal of Psychoeducational Assessment*, *23*, 339-361. doi:10.1177/073428290502300404

VanDerHeyden, A. M., Witt, J. C., & Gilbertson, D. (2007). A multi-year evaluation of the effects of a response to intervention (RTI) model on identification of children for special education. *Journal of School Psychology*, *45*, 225-256.

Vaughn, S., & Linan-Thompson, S. (2003). What is so special about special education for students with learning disabilities? *The Journal of Special Education*, *37*, 140-247.

Vellutino, F. R., Scanlon, D. M., Sipay, E. R., Small, S. G., Pratt, A., Chen, R., Denckla, M. B. (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.

Vellutino, F. R., Scanlon, D. M., Zhang, H., & Schatschneider, C. (2008). Using response to kindergarten and first grade intervention to identify children at risk for long-term reading difficulties. *Reading and Writing: An Interdisciplinary Journal*, *21*, 437-480.

Vittinghoff, E., & McCulloch, C. E. (2006). Relaxing the rule of ten events per variable in logistic and cox regression. *American Journal of Epidemiology*, *165*, 710-718. doi:10.1093/aje/kwk052

Wayman, M. M., Wallace, T., Ives Wiley, H., Ticha, R., & Espin, C. A. (2007). Literature synthesis on curriculum-based measurement in reading. *The Journal of Special Education*, *41*, 85-120.

Ysseldyke, J. E., & Algozzine, B. (1981). Diagnostic classification decisions as a function of referral information. *The Journal of Special Education*, *15*, 429-435.

Ysseldyke, J. E., Algozzine, B., & Epps, S. (1983). A logical and empirical analysis of current practice in classifying students as handicapped. *Exceptional Children*, *50*, 160-166.

Ysseldyke, J. E., Algozzine, B., Shinn, M. R., & McGue, M. (1982). Similarities and differences between low achievers and students classified learning disabled. *The Journal of Special Education*, *16*, 73-85.

Ysseyldyke, J. E., & Reschly, D. J. (2014). The evolution of school psychology: Origins, contemporary status, and future directions. In P. L. Harrison & A. Thomas (Eds.), *Best practices in school psychology: Data-based and collaborative decision making* (pp. 71-84). Bethesda, MD: National Association of School Psychologists.

Ysseldyke, J. E., & Salvia, J. (1974). Diagnostic prescriptive teaching: Two models. *Exceptional Children*, *41*, 181-185.

Ysseldyke, J. E., Thurlow, M., Graden, J., Wesson, C., Algozzine, B., & Deno, S. (1983). Generalizations from five years pf research on assessment and decision making. *Exceptional Education Quarterly, 4*, 75-93.

Ysseldyke, J. E., Vanderwood, M. L., & Shriner, J. (1997). Changes over the past decade in special education referral to placement probability: An incredibly reliable practice. *Diagnostique*, *23*, 193-201.

Zumeta, R. O., Zirkel, P. A., & Danielson, L. (2014). Identifying specific learning disabilities: Legislation, regulation, and court decisions. *Top Language Disorders*, *34*, 8-24.

Appendix A

Sample Data Collection Spreadsheet

	A	B	C	D	E	F	G	H	I	J	K	L	AQ	AR	AS	AT	AU	AV	CA	CB	CC
	Name	Referred for S	Eligible for Sp	Attending Building	Grade	Gender	Ethnicity	Date of Special Education Eligibility Decision	Receiving Free/Reduced Lunch	Date 1	Date 2	Date 3	Date 34	Date 35	Date 36	PM 1	PM 2	PM 3	PM 34	PM 35	PM 36
		No	No					N/A													
		No	No					N/A													
		No	No					N/A													
		No	No					N/A													
		No	No					N/A													
		No	No					N/A													
		No	No					N/A													
		No	No					N/A													
		No	No					N/A													
		No	No					N/A													
		No	No					N/A													
		No	No					N/A													
		No	No					N/A													
		No	No					N/A													
		No	No					N/A													
		No	No					N/A													
		No	No					N/A													
		No	No					N/A													
		No	No					N/A													
		No	No					N/A													
		No	No					N/A													
		No	No					N/A													
		No	No					N/A													
		No	No					N/A													
		No	No					N/A													
		No	No					N/A													
		No	No					N/A													
		No	No					N/A													
		No	No					N/A													
		No	No					N/A													
		No	No					N/A													
		No	No					N/A													
		No	No					N/A													
		No	No					N/A													
		No	No					N/A													

Indiana University of Pennsylvania
www.iup.edu

Institutional Review Board for the
Protection of Human Subjects
School of Graduate Studies and Research
Stright Hall, Room 113
210 South Tenth Street
Indiana, Pennsylvania 15705-1048

P 724-357-7730
F 724-357-2715
irb-research@iup.edu
www.iup.edu/irb

January 18, 2017

Michael Boneshefski
Dept of Educational and School Psychology
Stouffer Hall

Dear Mr. Boneshefski:

Your proposed research project, "Differential Diagnosis of Specific Learning Disabilities within a Response to Intervention Framework," (Log No. 16-319) has been reviewed by the IRB and is approved. In accordance with 45CFR46.101 and IUP Policy, your project is exempt from continuing review. This approval does not supersede or obviate compliance with any other University requirements, including, but not limited to, enrollment, degree completion deadlines, topic approval, and conduct of university-affiliated activities.

You should read all of this letter, as it contains important information about conducting your study.

Now that your project has been approved by the IRB, there are elements of the Federal Regulations to which you must attend. IUP adheres to these regulations strictly:

1. You must conduct your study exactly as it was approved by the IRB.

2. Any additions or changes in procedures must be approved by the IRB before they are implemented.

3. You must notify the IRB promptly of any events that affect the safety or well-being of subjects.

4. You must notify the IRB promptly of any modifications of your study or other responses that are necessitated by any events reported in items 2 or 3.

The IRB may review or audit your project at random *or* for cause. In accordance with IUP Policy and Federal Regulation (45CFR46.113), the Board may suspend or terminate your project if your project has not been conducted as approved or if other difficulties are detected

Although your human subjects review process is complete, the School of Graduate Studies and Research requires submission and approval of a Research Topic Approval Form (RTAF) before you can begin your research. If you have not

yet submitted your RTAF, the form can be found at
http://www.iup.edu/page.aspx?id=91683 .

While not under the purview of the IRB, researchers are responsible for adhering
to US copyright law when using existing scales, survey items, or other works in
the conduct of research. Information regarding copyright law and compliance at
IUP, including links to sample permission request letters, can be found at
http://www.iup.edu/page.aspx?id=165526.

I wish you success as you pursue this important endeavor.

Sincerely,

Jennifer Roberts, Ph.D.
Chairperson, Institutional Review Board for the Protection of Human Subjects
Professor of Criminology

JLR:jeb

Cc: Dr. Joseph Kovaleski, Dissertation Advisor
      Ms. Brenda Boal, Secretary

Appendix C

Institutional Review Board Approval of Modifications to Project

Indiana University of Pennsylvania
www.iup.edu

Institutional Review Board for the
Protection of Human Subjects
School of Graduate Studies and Research
Stright Hall, Room 113
210 South Tenth Street
Indiana, Pennsylvania 15705-1048

P 724-357-7730
F 724-357-2715
irb-research@iup.edu
www.iup.edu/irb

March 01, 2017

Mr. Michael Boneshefski

Dear Mr. Boneshefski:

Your proposed modifications to your previously approved research project, "Differential Diagnosis of Specific Learning Disabilities within a Response to Intervention Framework," (Log No. 16-319) been reviewed by the IRB and are approved. In accordance with 45CFR46.101 and IUP Policy, your project is exempt from continuing review in addition to the approval of your request for changes. This approval does not supersede or obviate compliance with any other University requirements, including, but not limited to, enrollment, degree completion deadlines, topic approval, and conduct of university-affiliated activities.

You should read all of this letter, as it contains important information about conducting your study.

Now that your project has been approved by the IRB, there are elements of the Federal Regulations to which you must attend. IUP adheres to these regulations strictly:

1. You must conduct your study exactly as it was approved by the IRB.

2. Any additions or changes in procedures must be approved by the IRB before they are implemented.

3. You must notify the IRB promptly of any events that affect the safety or well-being of subjects.

4. You must notify the IRB promptly of any modifications of your study or other responses that are necessitated by any events reported in items 2 or 3.

The IRB may review or audit your project at random *or* for cause. In accordance with IUP Policy and Federal Regulation (45CFR46.113), the Board may suspend or terminate your project if your project has not been conducted as approved or if other difficulties are detected.

While not under the purview of the IRB, researchers are responsible for adhering to US copyright law when using existing scales, survey items, or other works in the conduct of research. Information regarding copyright law and compliance at IUP, including links to sample permission request letters, can be found at http://www.iup.edu/page.aspx?id=165526.

I wish you success as you pursue this important endeavor.

Sincerely,

Jennifer Roberts, Ph.D.
Chairperson, Institutional Review Board for the Protection of Human Subjects
Professor of Criminology

JLR:jeb

Cc: Dr. Joseph Kovaleski, Faculty Advisor